

DIFFUSION APPROXIMATIONS FOR LOAD BALANCING MECHANISMS IN CLOUD STORAGE SYSTEMS

AMARJIT BUDHIRAJA* ** AND

ERIC FRIEDLANDER, * *** *University of North Carolina at Chapel Hill*

Abstract

In large storage systems, files are often coded across several servers to improve reliability and retrieval speed. We study load balancing under the batch sampling routing scheme for a network of n servers storing a set of files using the maximum distance separable (MDS) code (cf. Li (2016)). Specifically, each file is stored in equally sized pieces across L servers such that any k pieces can reconstruct the original file. When a request for a file is received, the dispatcher routes the job into the k -shortest queues among the L for which the corresponding server contains a piece of the file being requested. We establish a law of large numbers and a central limit theorem as the system becomes large (i.e. $n \rightarrow \infty$), for the setting where all interarrival and service times are exponentially distributed. For the central limit theorem, the limit process takes values in ℓ_2 , the space of square summable sequences. Due to the large size of such systems, a direct analysis of the n -server system is frequently intractable. The law of large numbers and diffusion approximations established in this work provide practical tools with which to perform such analysis. The power-of- d routing scheme, also known as the supermarket model, is a special case of the model considered here.

Keywords: Mean-field approximation; diffusion approximation; stochastic network; cylindrical Brownian motion; propagation of chaos; cloud storage system; supermarket model; MDS coding; power-of- d

2010 Mathematics Subject Classification: Primary 60K25; 60K35; 60F05; 60H30
Secondary 90B22; 60J28; 60J70; 60B12

1. Introduction

In the world of cloud-based computing, large data centers are often used for file storage. In these data centers, large networks of servers are used to store even larger sets of files. In order to improve reliability and retrieval speed, these files are often ‘coded’. By coded, we mean that the file is broken down into smaller pieces which are stored on multiple servers. Consider the situation in which there are four servers and one file. One can store the entire file on one server, but in such a configuration the file would be inaccessible if that server were to fail. In order to improve reliability, one can replicate the file across all four servers, but such a method would require much more memory. Suppose instead that we split the file into halves, A and B , and then store A , B , $A + B$, $A - B$ in each of the four servers, respectively. Then the original file can be constructed from any two pieces. One can extend this idea to the case where equally sized

Received 5 December 2017; revision received 16 September 2018.

* Postal address: Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, NC 27599, USA.

** Email address: budhiraj@email.unc.edu

*** Email address: ebf2@live.unc.edu

pieces of a file are stored across L servers and any k pieces can reconstruct the original file. This can be accomplished using the maximum distance separable (MDS) code with parameters (L, k) [29]. The MDS code greatly improves reliability since $L - k + 1$ servers must fail before the file becomes irretrievable, while only requiring enough total memory to store L/k files. Given a coding scheme, one can consider load balancing mechanisms to improve file retrieval speed. In [28], two routing schemes, called batch sampling (BS) and redundant request with killing (RRK), were considered. In BS routing, incoming jobs are routed to the k -shortest queues containing the file being requested, while in RRK routing jobs are routed to all servers containing the requested file and then removed from the queue (killed) once k pieces of the file have been returned. Li [28] formally calculated the steady state ($T \rightarrow \infty$) queue length distribution in the large system limit ($n \rightarrow \infty$) and gave simulation results for different values of L and k in both routing schemes.

In this work we are interested in developing a rigorous limit theory for such load balancing schemes for systems with MDS coding as n becomes large. Specifically, we establish law of large numbers and diffusion approximations for such systems under an appropriate scaling, as $n \rightarrow \infty$. Such limit theorems provide useful model simplifications that can then be employed for approximate simulation of the large and complex n -server systems (see Section 6 for some numerical results). These limit theorems are also the first steps towards making rigorous the program initiated in [28] of developing steady state approximations for such systems, with provable convergence properties as n becomes large.

We will focus in this work only on BS routing and leave analysis of the RRK scheme for future work. Specifically, we consider a system with n servers on which $I(n)$ files are stored using MDS coding with parameters (L, k) . A key assumption to our analysis is that the files are stored such that each combination of L servers has exactly c files. We further assume that jobs arrive in the system according to a Poisson process with rate $n\lambda$ and request a file uniformly at random. This is another simplifying assumption on our model that roughly says that all files are in equal demand. These structural assumptions imply a convenient exchangeability property of the system which allows for the use of certain mean-field approximation techniques. A single file request spawns k jobs which are then routed into the k -shortest queues within the set of L servers containing the file being requested. Each server processes the jobs in their queue with exponential service rate k according to the first-in–first-out (FIFO) discipline and processing times are mutually independent. Regarding each server as a ‘particle’, the above formulation describes an interacting particle system with *simultaneous jumps*. Note that the symmetry structure introduced above implies that every time a file request arrives, it leads to a selection of L servers uniformly at random (from which the k servers with shortest queues are chosen). In particular, this says that the well-studied ‘power-of- d ’ routing scheme (also known as the ‘supermarket model’) is a special case of the scheme considered here on taking $L = d$ and $k = 1$. Direct analysis of such large and complex n -server systems is challenging even by simulation methods as frequently the servers in networks of interest number in the hundreds of thousands with arrival rates of file requests of similar order. The goal of this work is to develop suitable approximate approaches to such systems.

The starting point of our analysis is to consider, as the state descriptor, the empirical measure of the n queue lengths rather than the individual values of the queue lengths. Thus, the state space for our system will be the space $\mathcal{P}_n(\mathbb{N}_0)$ of probability measures on \mathbb{N}_0 that assign weights in \mathbb{N}_0/n to sets in \mathbb{N}_0 rather than the space \mathbb{R}_+^n . With this formulation, the state processes for all n -server systems can be regarded as taking values in a common space $\mathcal{S} := \mathcal{P}(\mathbb{N}_0)$ (the space of probability measures on \mathbb{N}_0). It follows from our symmetry

assumptions that the state evolution of the n -server system describes a pure-jump Markov process with values in $\mathcal{P}(\mathbb{N}_0)$ and thus one can bring to bear the theory of weak convergence of Markov processes to study the scaling limits as n becomes large. In particular, in Theorem 1 we prove a law of large numbers for the empirical measure process $(\pi^n(t))_{0 \leq t \leq T}$ as $n \rightarrow \infty$. We show that π^n converges to a deterministic function π in $\mathbb{D}([0, T]: \mathcal{S})$, where $\mathbb{D}([0, T]: \mathcal{S})$ is the space of functions from $[0, T]$ to \mathcal{S} that are right continuous and have left limits, equipped with the usual Skorokhod topology. Next we consider the fluctuation process $X^n := \sqrt{n}(\pi^n - \pi)$. This process can be regarded as taking values in the space of signed measures on \mathbb{N}_0 ; however, for an asymptotics analysis, it is convenient to view it as taking values in the Hilbert space of square summable real sequences, ℓ_2 . The study of the asymptotics of these fluctuations is the subject of Theorem 2, which shows that $X^n := \sqrt{n}(\pi^n - \pi)$ converges in $\mathbb{D}([0, T]: \ell_2)$ to an ℓ_2 -valued diffusion process.

Limit theorems of the form studied in this work can be used for model simplification and for computing approximations for performance measures, e.g. through simulation methods. Direct simulation of the underlying n -server system would in general be prohibitively expensive for large n since the jumps in the system occur at a rate proportional to n . The asymptotic approximations given in this work allow a system manager to simulate performance metrics for the system at a coarser scale via numerical ordinary differential equation (ODE) solvers or Euler discretizations for stochastic differential equations (SDEs) (see Section 6 for an example). Although the systems considered here are required to satisfy certain symmetry conditions (all files are equally sized and all jobs are in equal demand), the simplified models given by the limiting ODE and SDE give useful qualitative insights into the behavior of large storage networks employing these types of coding schemes. The results obtained here are useful in analyzing the long-time behavior of such systems as well, e.g. providing information on the rate at which the queue lengths decay in steady state and how such a decay is impacted by different values of L and k [15]. Furthermore, techniques developed in this work can also be used for models satisfying weaker symmetry conditions (e.g. for multitype populations with a fixed finite number of types). The poisson representations used in the proofs of Theorems 1 and 2 crucially rely on the fact that interarrival and service times are exponentially distributed. Different methods will need to be used in order to treat the case of general distributions. We refer the reader to [6] and Section 6 of [35] for work in this direction.

Load balancing mechanisms similar to the type considered here have been studied in many works. Specifically, the join-the-shortest-queue (JSQ), join-the-idle-queue (JIQ), and power-of- d routing schemes have garnered quite a bit of attention (see [6], [13], [16], [31], [32], [35], [37], and the references therein). Mitzenmacher [31] and Vvedenskaya *et al.* [37] first analyzed the power-of- d routing scheme, showing that in steady state the fraction of queues with lengths exceeding m decay super exponentially in m , a large improvement over the exponential rate for the setting where jobs are routed to servers uniformly at random. Graham [16] established a functional law of large numbers for π^n on $\mathbb{D}([0, T]: \mathcal{S})$ in the power-of- d routing scheme using characterization results for nonlinear martingale problems. In [13], the authors derived a diffusion approximation for the JSQ routing policy in the large-system limit under heavy-traffic scaling. It was shown that the limit can be characterized through a two-dimensional diffusion. In [32], it was shown that the JIQ routing scheme yields the same diffusion approximation as the JSQ routing scheme. In both these works, the diffusion approximations were derived under the same scaling regime as considered here. However, unlike for the JSQ and JIQ routing regimes where the diffusion approximations can be described through two-dimensional processes, in the current work the limit is an

infinite-dimensional diffusion described as an ℓ_2 -valued process driven by a cylindrical Brownian motion. We refer the reader to [1], [12], [23], and [33] for other queueing network systems where infinite-dimensional diffusions arise as approximate models. As noted earlier, the power-of- d scheme is a special case of our results, and, thus, Theorems 1 and 2 also provide law of large numbers and diffusion approximations for this classical load balancing scheme (see Corollaries 1 and 2). In particular, Corollary 1 recovers the law of large numbers established in [16]. Limit theorems giving fluctuation results for power-of- d schemes have not been studied previously.

Diffusion approximation methods have been used extensively in stochastic network theory. In particular, they have been very useful in the study of critically loaded stochastic processing networks (see [4], [5], [8], [9], [11], [17], [27], [38], and the references therein). For such systems, the key mathematical tool is the functional central limit theorem for renewal processes which provides Brownian motion approximations for a finite collection of centered renewal processes with rates approaching ∞ . The scaling regime and mathematical tools that are relevant for the analysis in the current context are quite different from those used in the above works. In particular, here the number of nodes approach ∞ and the tools for proving convergence come from martingale problems and Markov process theory. A similar scaling regime was considered in [7] for certain systems motivated by ad-hoc wireless network models introduced in [3]. A key simplifying feature there is that the state space of an individual queue is a finite set. Consequently, the limit diffusion in [7] is finite dimensional and, thus, for diffusion approximations, classical convergence theorems from [20] and [24] can be invoked. In contrast, the queue length processes in this work are unbounded, taking values in \mathbb{N}_0 , and thus one needs to study diffusion approximations in an infinite-dimensional state space, namely the Hilbert space ℓ_2 . The proofs employ appropriate criteria for tightness and characterization results for Hilbert space-valued stochastic processes.

A basic assumption in our analysis of the fluctuations around the law of large number limit (see statement of Theorem 2) is a uniform (in n) bound on the second moment of the empirical measure at time 0. This condition is not very stringent and allows for many types of initial configurations, e.g. one where no queue contains more than k jobs (where k is independent of n). We argue that these integrability properties at time 0 propagate through to any finite future time T . Tightness of the scaled fluctuation processes X^n which is shown by establishing, uniform in n , second moment bounds (on X^n) and by employing criteria for tightness of Hilbert space-valued semimartingales (cf. [20] and [30]), relies on these integrability properties. Another ingredient in the proof of tightness is a suitable Lipschitz property of the map F introduced in (5) that enables the use of a Gronwall argument. For this argument, one needs a Lipschitz estimate in the ℓ_2 norm; however, it is not clear that F , as a map from ℓ_2 to ℓ_2 , is Lipschitz. We instead restrict attention to a smaller space

$$\mathcal{V}_M := \left\{ r \in \ell_2 : r_i \geq 0, \sum_{i=0}^{\infty} r_i = 1, \sum_{i=0}^{\infty} ir_i \leq M \right\},$$

and argue that, for each M , the map F is Lipschitz from \mathcal{V}_M to ℓ_2 . This ‘local’ Lipschitz property plays an important role in the proof of Proposition 4.

For characterization of limit points in the proof of the central limit theorem, one needs to argue that the associated SDE in ℓ_2 (see (11)) has a unique weak solution in an appropriate class of processes. It turns out that arguing this uniqueness among adapted processes with paths in $\mathbb{C}([0, T]: \ell_2)$ (the space of continuous functions from $[0, T]$ to ℓ_2) is not straightforward due

to a lack of suitable regularity of the function G introduced in (16). In particular, once more, the Lipschitz property of the map $x \mapsto G(x, \pi)$ (for a fixed $\pi \in \mathcal{P}(\mathbb{N}_0)$) from ℓ_2 to itself is not immediate. The key observation here is that this map is Lipschitz when restricted to the space

$$\tilde{\ell}_2 := \left\{ x \in \ell_2 : \sum_{j=0}^{\infty} j^2 x_j^2 < \infty, \sum_{j=0}^{\infty} x_j = 0 \right\}.$$

This observation, together with the property that the limit points X of $X^n = \sqrt{n}(\pi^n - \pi)$ satisfy $X(t) \in \tilde{\ell}_2$ for all $t \geq 0$ almost surely, is key to the characterization of the limit points as the unique solution of SDE (11) in a suitable class (see Proposition 2).

The paper is organized as follows. In Section 2 we give a precise mathematical formulation of our model and a statement of our main results. Specifically, Theorem 1 provides the convergence in probability of the empirical measure process in $\mathbb{D}([0, T]: \mathcal{S})$ to the unique solution of the ODE defined in (7). In Theorem 2 we present the main diffusion approximation result. This result says that the sequence of centered and scaled processes X^n , defined in (10), converges to the unique solution (in a suitable class) of the ℓ_2 -valued SDE, driven by a cylindrical Brownian motion, given in (11). In Section 2.1 we record the corollaries of these results for the special setting of power-of- d schemes. We then proceed to the proofs of Theorems 1 and 2. In Section 3 we give a convenient representation of the state processes through a countable number of time-changed unit-rate Poisson processes. Such Poisson representations have been used extensively (cf. [2], [22], and [25]) in the study of diffusion approximations for pure-jump processes. Using this, we obtain a semimartingale decomposition (see (20)) that is central to our analysis. Section 4 is devoted to the proof of Theorem 1. In Section 4.1 we prove tightness of the sequence of state processes $\{\pi^n\}_{n \in \mathbb{N}}$ (see Proposition 3) and the proof of Theorem 1 is completed in Section 4.2. In Section 5 we prove Theorem 2. In Section 5.1 we prove the propagation of integrability properties that was discussed earlier, and in Section 5.2 (see Proposition 4) we prove the key tightness property for the sequence of processes $\{X^n\}_{n \in \mathbb{N}}$ which relies on the Lipschitz property of F , in the ℓ_2 norm, on \mathcal{V}_M (Lemma 4). Theorem 2 is then proved in Section 5.3. Finally, in Section 6 we present some numerical results. In particular, we use our results to give numerical confidence intervals for several performance measures of interest and compare the results to those obtained from a direct simulation of the corresponding n -server systems.

1.1. Notation

The following notation will be used. Fix $T < \infty$. All stochastic processes will be considered over the time horizon $[0, T]$. We will use the notation $(X_t)_{0 \leq t \leq T}$ and $(X(t))_{0 \leq t \leq T}$ interchangeably for stochastic processes. The space of probability measures on a Polish space \mathbb{S} , equipped with the topology of weak convergence, will be denoted by $\mathcal{P}(\mathbb{S})$. When $\mathbb{S} = \mathbb{N}_0$, we will metrize $\mathcal{P}(\mathbb{S})$ with the metric d_0 defined as

$$d_0(\mu, \nu) := \sum_{j=0}^{\infty} \frac{|\mu(j) - \nu(j)|}{2^j}, \quad \mu, \nu \in \mathcal{P}(\mathbb{N}_0).$$

For \mathbb{S} -valued random variables $X, X_n, n \geq 1$, convergence in distribution of X_n to X as $n \rightarrow \infty$ will be denoted as $X_n \Rightarrow X$. The space of functions that are right continuous with left limits (RCLL) from $[0, T]$ to \mathbb{S} will be denoted as $\mathbb{D}([0, T]: \mathbb{S})$ and equipped with the usual

Skorokhod topology. Similarly $\mathbb{C}([0, T]: \mathbb{S})$ will be the space of continuous functions from $[0, T]$ to \mathbb{S} , equipped with the uniform topology. We will usually denote by $\kappa, \kappa_1, \kappa_2, \dots$ the constants that appear in various estimates within a proof. The values of these constants may change from one proof to another and, unless stated otherwise, will take values in the set $(0, \infty)$. Let $\ell_2 = \{(a_j)_{j=0}^\infty \mid \sum_{j=0}^\infty a_j^2 < \infty\}$ be the space of square summable real sequences. This space is a Hilbert space with inner product

$$\langle x, y \rangle_2 = \sum_{j=0}^\infty x_j y_j.$$

We denote the corresponding norm as $\|\cdot\|_2$. Similarly, $\ell_1 = \{(a_j)_{j=0}^\infty \mid \sum_{j=0}^\infty |a_j| < \infty\}$ and $\|\cdot\|_1$ is the norm on this Banach space. The Hilbert–Schmidt norm of a Hilbert–Schmidt operator A on ℓ_2 will be denoted by $\|A\|_{\text{HS}}$ (cf. Appendix A.3). We denote by I the identity operator. For a Hilbert Space \mathbb{H} , $\mathcal{M}_T^2(\mathbb{H})$ will denote the space of all \mathbb{H} -valued continuous, square-integrable martingales M , such that $M(0) = 0$. For a real number a , $(a)_+$ will denote the positive part of a .

2. Model description and main result

We consider a system with n servers each with its own infinite capacity queue. In all, there are $I(n)$ equally sized files stored over the n servers. Each file is stored in equally sized pieces at L servers such that any k pieces can reconstruct the original file. The files are distributed such that each combination of L servers has exactly c files. This, in particular, implies that $I(n) = c \binom{n}{L}$. Jobs arrive from outside according to a Poisson process with rate $n\lambda$ and request one of the $I(n)$ files uniformly at random. Such a request corresponds to selection of one of the $\binom{n}{L}$ sets of L servers, uniformly at random, which is the set of servers containing the pieces of the requested file. The job is then routed to the k -shortest queues among this set of L servers. Each server processes queued jobs according to the FIFO discipline. Processing times at each server are mutually independent and exponentially distributed with mean k^{-1} .

Let $Q^n(t) = \{Q_i^n(t)\}_{i=1}^n$, where $Q_i^n(t)$ represents the length of the i th queue at time t , and let $\pi^n(t) = \{\pi_i^n(t)\}_{i \in \mathbb{N}_0}$, where $\pi_i^n(t)$ represents the proportion of queues with length exactly i at time t . This can explicitly be written as

$$\pi_i^n(t) = \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{\{Q_j^n(t)=i\}}. \tag{1}$$

We will assume for simplicity that $Q^n(0) = q^n$ is nonrandom and, thus, $\pi^n(0) = (1/n) \sum_{j=1}^n \mathbf{1}_{\{q_j^n=i\}}$ is nonrandom as well. We identify $\mathcal{P}(\mathbb{N}_0)$ with the infinite-dimensional simplex $\mathcal{S} = \{s \in \mathbb{R}_+^\infty \mid \sum_{i=0}^\infty s_i = 1\}$, and let $\mathcal{S}_n = \mathbb{N}_0^\infty / n \cap \mathcal{S}$. It follows that $\pi^n(t) \in \mathcal{S}_n$ for all $t \in [0, T]$. Let $\Sigma = \{\ell = (\ell_i)_{i=1}^L \in \mathbb{N}_0^L \mid \ell_1 \leq \ell_2 \leq \dots \leq \ell_L\}$ and, for $\ell \in \Sigma$, define $\rho_i(\ell) := \sum_{j=1}^L \mathbf{1}_{\{\ell_j=i\}}$, $i \in \mathbb{N}_0$. Roughly speaking, Σ will represent the set of possible states for L selected queues arranged by nondecreasing queue length. Note that each file will be stored at L servers and that at any given time t the queue lengths of these L servers (up to a reordering) will correspond to an element in Σ . We will refer to the elements of Σ as ‘queue length configurations’. Given a configuration $\ell \in \Sigma$, $\rho_i(\ell)$ gives the number of queues of length i (among the L selected). From the above description of the system, it follows that the

empirical measure process, $\pi^n(t)$, is a continuous-time Markov chain with state space \mathcal{S}_n and generator

$$\begin{aligned} \mathcal{L}^n f(r) = & \frac{n\lambda}{\binom{n}{L}} \sum_{\ell \in \Sigma} \left(\prod_{i=0}^{\infty} \binom{nr_i}{\rho_i(\ell)} \right) \left[f\left(r + \frac{1}{n} \Delta_\ell\right) - f(r) \right] \\ & + k \sum_{i=1}^{\infty} nr_i \left[f\left(r + \frac{1}{n}(e_{i-1} - e_i)\right) - f(r) \right] \end{aligned} \tag{2}$$

for $f: \mathcal{S}_n \rightarrow \mathbb{R}$, where

$$\Delta_\ell := \sum_{i=1}^k e_{\ell_i+1} - \sum_{i=1}^k e_{\ell_i} \tag{3}$$

and, for $y \in \mathbb{N}_0$, $e_y \in \ell_2$, is a vector with 1 at the y th coordinate and 0 elsewhere. Here we use the standard conventions that $0^0 = \binom{0}{0} = 0! = 1$, and $\binom{a}{b} = 0$ when $a < b$. The above generator can be understood as follows. A typical term in the second expression corresponds to a jump as a result of a server, with exactly i jobs queued, completing a job. The term in the square brackets gives the change in value of f as a result of such a jump and the prefactor knr_i corresponds to the fact that servers process jobs at rate k and there are in all nr_i queues (prior to the jump) with exactly i jobs. The first expression in (2) corresponds to a jump resulting from an arrival of a job to the system. Typically, such an arrival makes a request for L servers with queue length configuration $\ell_1 \leq \ell_2 \leq \dots \leq \ell_L$ and results in the jump Δ_ℓ/n . The sum in (3) only goes up to k (instead of L) since only the smallest k queues are affected by such a jump. Since, prior to the jump, there are nr_i queues with exactly i jobs, the overall rate associated with the configuration $\ell = \{\ell_1 \leq \ell_2 \leq \dots \leq \ell_L\} \in \Sigma$ equals

$$\frac{n\lambda}{\binom{n}{L}} \left(\prod_{i=0}^{\infty} \binom{nr_i}{\rho_i(\ell)} \right).$$

In our setting the first entry in an element of ℓ_2 will typically correspond to the number of empty queues, and, thus, we refer to it as the ‘0th’ coordinate and any $r \in \ell_2$ will correspond to a vector of the form (r_0, r_1, \dots) . For notational convenience, for $r \in \ell_2$, we set $r_{-1} := 0$.

The main results in this work provide scaling limits for π^n . We first present the law of large numbers which describes the nominal state of the system for large n . Define, for $r \in \ell_1$,

$$\bar{\zeta}^\delta(j, r) := \bar{\zeta}(j - 1, r) - \bar{\zeta}(j, r),$$

where, adopting the convention that $\sum_{i=b}^a x_i = 0$ for $a < b$,

$$L! \bar{\zeta}(j, r) := \sum_{i_1=0}^{k-1} \sum_{i_2=1}^{L-i_1} \frac{L!}{i_1! i_2! (L - i_1 - i_2)!} \left(\sum_{m=0}^{j-1} r_m \right)^{i_1} (r_j)^{i_2} \left(\sum_{m=j+1}^{\infty} r_m \right)^{L-i_1-i_2} [i_2 \wedge (k - i_1)]. \tag{4}$$

This allows us to define, for $r \in \ell_1$,

$$F(r) := \lambda L! \sum_{j=0}^{\infty} \bar{\zeta}^\delta(j, r) e_j + k \sum_{j=0}^{\infty} [r_{j+1} - r_j] e_j + kr_0 e_0. \tag{5}$$

For $j \geq 0$, the quantities $k[r_{j+1} - r_j]$ in (5) roughly represent the rate at which the j th coordinate of the state changes (in the limit) as a result of job completions. The final term in the display

cancels the extra term in the second summation for the boundary case $j = 0$. The quantity $\lambda L! (\bar{\zeta}(j - 1, r) - \bar{\zeta}(j, r))$ represents a similar quantity as a result of job arrivals. The various terms in (4) can be interpreted as follows. An arrival to a queue with j jobs implies that a queue length configuration vector $\ell = \{\ell_1 \leq \ell_2 \leq \dots \leq \ell_L\}$ was selected which has the property that at least one of the k smallest ℓ_i equals j , or, equivalently, exactly i_1 ($i_1 = 0, 1, \dots, k - 1$) of the smallest L selected are less than j , i_2 ($i_2 = 1, \dots, L - i_1$) of these are equal to j , and $L - i_1 - i_2$ are greater than j . The three ratios in (4) are contributions from these three types of queues. The term $[i_2 \wedge (k - i_1)]$ arises from the fact that only the smallest k of the L queues are affected.

Also observe that, for some $c_\zeta \in (0, \infty)$,

$$\bar{\zeta}(j, r) \leq c_\zeta r_j \quad \text{for all } j \in \mathbb{N}_0 \quad \text{and} \quad r = (r_j)_{j=0}^\infty \in \mathcal{S}. \tag{6}$$

Thus, the infinite sum in (5) is well defined since $\sum_{j=0}^\infty r_j = 1$, and, consequently, F is a well-defined map from \mathcal{S} to ℓ_1 . A similar estimate shows that F is a well-defined map from ℓ_1 to ℓ_1 and $\sum_{j=0}^\infty F_j(r) = 0$ for all $r \in \ell_1$.

Consider the system of ODEs

$$\dot{\pi}(t) = F(\pi(t)), \quad \pi(0) = \pi_0, \tag{7}$$

where F is defined in (5) and $\pi_0 \in \mathcal{S}$. The solution of the equation is a continuous map $\pi : [0, T] \rightarrow \mathcal{S}$ such that

$$\pi(t) = \pi_0 + \int_0^t F(\pi(s)) \, ds, \quad t \in [0, T], \tag{8}$$

where the integral on the right-hand side is the classical Bochner integral which is well defined since, from (5) and (6),

$$\sup_{0 \leq s \leq T} \|F(\pi(s))\|_1 \leq \sup_{r \in \mathcal{S}} \|F(r)\|_1 < \infty. \tag{9}$$

Equation (7) will characterize the law of large number limit of π^n .

The following result on the well posedness of (7) will be shown in Section 4.2.

Proposition 1. *Let $\pi_0 \in \mathcal{S}$. Then there exists a $\pi \in \mathbb{C}([0, T]: \mathcal{S})$ that solves (7). Furthermore, if π and $\tilde{\pi}$ are two elements of $\mathbb{C}([0, T]: \mathcal{S})$ solving (7) with $\pi(0) = \tilde{\pi}(0) = \pi_0$, then $\pi = \tilde{\pi}$.*

The next theorem gives a law of large numbers for the sequence $\{\pi^n\}_{n \in \mathbb{N}}$. Recall that we take $\pi^n(0)$ to be nonrandom.

Theorem 1. *Suppose that $\pi^n(0) \rightarrow \pi_0$ in \mathcal{S} as $n \rightarrow \infty$. Then $\pi^n \rightarrow \pi$ in probability in $\mathbb{D}([0, T]: \mathcal{S})$, where π is the unique solution of (7) in $\mathbb{C}([0, T]: \mathcal{S})$.*

The proof of Theorem 1 will be given in Section 4.2.

Remark 1. The occupancy process π^n satisfies a natural monotonicity property. Let $\gamma_i^n(t) := \sum_{j=i}^\infty \pi_j^n(t)$, starting from some initial state $\gamma^{n,0}$ and let $\tilde{\gamma}^n$ be the corresponding process starting from an initial state $\tilde{\gamma}^{n,0}$. Suppose that $\tilde{\gamma}_i^{n,0} \geq \gamma_i^{n,0}$ for every i . Then $\gamma^n(t)$ is componentwise stochastically dominated by $\tilde{\gamma}^n(t)$ for every $t \geq 0$. A deterministic analogue of this property will hold for the limiting trajectories π . Such a monotonicity can be used to deduce various qualitative properties of π^n and the limit path π . Indeed, Friedlander [15] used this monotonicity behavior crucially in order to study the long-time behaviors of π^n and π .

Our second main result studies the fluctuations of π^n from its law of large number limit. Consider

$$X^n(t) = \sqrt{n}[\pi^n(t) - \pi(t)], \quad t \in [0, T], \tag{10}$$

where π^n is the state process introduced in (1) and π is the unique solution of (7) in $\mathbb{C}([0, T]: \mathcal{S})$.

We will show that, under conditions, X^n converges in distribution in $\mathbb{D}([0, T]: \ell_2)$ to a stochastic process that can be characterized as the solution of an SDE of the form

$$dX(t) = G(X(t), \pi(t)) dt + a(t) dW(t), \quad X(0) = x_0. \tag{11}$$

The equation is again interpreted in the integrated form

$$X(t) = x_0 + \int_0^t G(X(s), \pi(s)) ds + \int_0^t a(s) dW(s), \quad t \in [0, T]. \tag{12}$$

In the above equations, a is a measurable map from $[0, T]$ to the space of Hilbert–Schmidt operators from ℓ_2 to ℓ_2 such that $\int_0^T \|a(t)\|_{\text{HS}}^2 dt < \infty$, where $\|\cdot\|_{\text{HS}}$ denotes the Hilbert–Schmidt norm (see Appendix A.3), and W is an ℓ_2 -cylindrical Brownian motion. Precise definitions are given in Appendix A.4, but roughly speaking, W can be identified with an independent and identically distributed sequence $\{\beta_i\}_{i \in \mathbb{N}_0}$ of standard real Brownian motions over $[0, T]$ and the stochastic integral $\int_0^t a(s) dW(s)$ represents an ℓ_2 -valued Gaussian martingale $M(t)$ given as

$$M_i(t) = \sum_{j=0}^{\infty} \int_0^t A_{ij}(s) d\beta_j(s), \quad t \in [0, T], \quad i \in \mathbb{N}_0, \tag{13}$$

where $A_{ij}(s) = \langle e_i, a(s)e_j \rangle_2$, $s \in [0, T]$, $i, j \in \mathbb{N}_0$. We refer the reader to Chapter 4 of [10] for construction and properties of the stochastic integral in (12). The Hilbert–Schmidt norm and integrability property of a ensure that the infinite sum in (13) converges. The operator $a(t)$ is determined from the system parameters and the law of large number limit π in Theorem 1 as the symmetric square root of the following nonnegative trace class operator:

$$\Phi(t) := \lambda L! \sum_{\ell \in \Sigma} \Delta_\ell \Delta_\ell^T \prod_{i=0}^{\infty} \frac{\pi_i(t)^{\rho_i(\ell)}}{\rho_i(\ell)!} + k \sum_{i=1}^{\infty} (e_{i-1} - e_i)(e_{i-1} - e_i)^T \pi_i(t). \tag{14}$$

The trace class property of $\Phi(t)$ and the integrability of the squared Hilbert–Schmidt norm of $a(t)$ are shown in Lemma 7. Define the space $\tilde{\ell}_2 \subset \ell_2$ as

$$\tilde{\ell}_2 := \left\{ x \in \ell_2 : \sum_{j=0}^{\infty} j^2 x_j^2 < \infty, \sum_{j=0}^{\infty} x_j = 0 \right\}. \tag{15}$$

In (11) G is a map from $\tilde{\ell}_2 \times \mathcal{S}$ to ℓ_2 defined as

$$G_i(x, r) := \left. \frac{\partial}{\partial u} F_i(r + ux) \right|_{u=0}, \quad i \in \mathbb{N}_0, \quad u \in \mathbb{R}. \tag{16}$$

One of the difficulties in the analysis is that G as a map from $\ell_2 \times \mathcal{S}$ to ℓ_2 is not well behaved and we need to restrict attention to the smaller space $\tilde{\ell}_2 \times \mathcal{S}$ in order to get unique solvability of

(11). Note that, under the condition that $\sum_{j=0}^{\infty} j^2 x_j^2 < \infty$, the series $\sum_{j=0}^{\infty} |x_j| < \infty$, and thus the series $\sum_{j=0}^{\infty} x_j$, is convergent. Additionally, the right-hand side of (16) is well defined for every $x \in \tilde{\ell}_2$ and $r \in \mathcal{S}$, since, for each $j \in \mathbb{N}_0$ and $r \in \ell_1$ with $\sum_{i=0}^{\infty} r_i = 1$, $r \mapsto F_j(r)$ is a polynomial in $(r_0, r_1, \dots, r_{j+1})$ given as

$$F_j(r) = \lambda L! [\bar{\zeta}(j - 1, r) - \bar{\zeta}(j, r)] + k(r_{j+1} - r_j),$$

where

$$\bar{\zeta}(j, r) = \sum_{i_1=0}^{k-1} \frac{(\sum_{m=0}^{j-1} r_m)^{i_1}}{i_1!} \sum_{i_2=1}^{L-i_1} [i_2 \wedge (k - i_1)] \frac{(r_j)^{i_2}}{i_2!} \frac{(1 - \sum_{m=0}^j r_m)^{L-i_1-i_2}}{(L - i_1 - i_2)!}.$$

Also, from (4) and (5), it is easily checked that there is a $c \in (0, \infty)$ such that, for all $x \in \tilde{\ell}_2$ and $r \in \mathcal{S}$,

$$|G_i(x, r)| \leq c \left[|x_{i-1}| + |x_i| + |x_{i+1}| + (r_{i-1} + r_i) \sum_{m=0}^{\infty} |x_m| \right].$$

This in particular implies that $G(x, r) := (G_i(x, r))_{i \in \mathbb{N}_0} \in \ell_1 \subset \ell_2$ for all $(x, r) \in \tilde{\ell}_2 \times \mathcal{S}$.

The following result shows the well posedness of (12). The definition of an ℓ_2 -cylindrical Brownian motion is given in Section A.4.

Proposition 2. *There exists a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\})$ on which is given an ℓ_2 -cylindrical Brownian motion W and a continuous $\{\mathcal{F}_t\}$ -adapted process $(X(t))_{0 \leq t \leq T}$ with sample paths in $\mathbb{C}([0, T]: \ell_2)$ that satisfies the integral equation (12) and is such that $X(t) \in \tilde{\ell}_2 \subset \ell_2$ for all $t \in [0, T]$ almost surely. Furthermore, if $\{\tilde{X}_t\}_{0 \leq t \leq T}$ is another such process then $\tilde{X}_t = X_t$ for all $t \in [0, T]$ almost surely.*

The above result establishes weak existence and pathwise uniqueness of (12). By a standard argument (cf. [18, Section IV.1]), it follows that (12) has a unique weak solution. We can now present our main result on the fluctuations of π^n . Recall that $X^n(0) = \sqrt{n}(\pi^n(0) - \pi_0)$ is deterministic.

Theorem 2. *Suppose that $\sup_{n \in \mathbb{N}} \sum_{j=0}^{\infty} j^2 \pi_j^n(0) < \infty$ and $\pi^n(0) \rightarrow \pi_0$ in \mathcal{S} as $n \rightarrow \infty$. Let π be the unique solution of (7) and, with X^n defined as in (10), $X^n(0) \rightarrow x_0$ in ℓ_2 . In addition, suppose that*

$$\sup_{n \in \mathbb{N}} \sum_{j=0}^{\infty} j^2 (X_j^n(0))^2 < \infty. \tag{17}$$

Then $X^n \Rightarrow X$ in $\mathbb{D}([0, T]: \ell_2)$, where X is the unique weak solution to (11) given by Proposition 2.

Proposition 2 and Theorem 2 will be proved in Section 5. In Section 6 we will describe how Theorems 1 and 2 can be used for numerical computation of various performance measures using simulation of diffusion processes.

2.1. Supermarket model

Consider a system of n servers, each with its own queue. Jobs arrive in the system according to a Poisson process with rate $n\lambda$. When a job enters the system, d servers are chosen uniformly at random and the job is routed to the shortest of the d selected queues. All servers process jobs

according to the FIFO discipline. Service times are mutually independent and exponentially distributed with mean 1. This model has been well studied and is known as power-of- d routing or the ‘supermarket model’ (see [16], [31], and [37]). The model is a special case of the system considered in the current work, corresponding to $L = d$ and $k = 1$. Theorems 1 and 2 then provide, as corollaries, the following law of large numbers and central limit theorem for the power-of- d routing scheme.

Define by π_d^n the empirical measure process of queue lengths in the power-of- d system. For $r \in \ell_1$, define

$$F_d(r) := \lambda \left[\sum_{i=1}^d \binom{d}{i} r_{j-1}^i \left(\sum_{m=j}^{\infty} r_m \right)^{d-i} - \sum_{i=1}^d \binom{d}{i} r_j^i \left(\sum_{m=j+1}^{\infty} r_m \right)^{d-i} \right] e_j + \sum_{j=0}^{\infty} [r_{j+1} - r_j] e_j.$$

The following is a direct corollary of Theorem 1.

Corollary 1. *Suppose that $\pi_d^n(0) \rightarrow \pi_d(0)$ in \mathcal{S} as $n \rightarrow \infty$. Then $\pi_d^n \rightarrow \pi_d$ in probability in $\mathbb{D}([0, T]: \mathcal{S})$, where π_d is the unique solution in $\mathbb{C}([0, T]: \mathcal{S})$ to the ODE*

$$\dot{\pi}_d(t) = F_d(\pi_d(t)), \quad \pi_d(0) = \pi_0.$$

Remark 2. This result has been established in [16] (see Theorem 3.4 therein). In particular, it is easy to verify that $v_m(t) := \sum_{j=m}^{\infty} (\pi_d(t))_j$ is the same function as in Equation (3.9) of [16] (see also [37]).

Our second corollary studies the fluctuations of π_d^n from its law of large number limit. Consider

$$X_d^n(t) = \sqrt{n}[\pi_d^n(t) - \pi_d(t)], \quad t \in [0, T].$$

Analogous to $a(t)$ introduced in (11), let $a_d(t)$ be the symmetric square root of the nonnegative operator

$$\begin{aligned} \Phi_d(t) := & \lambda \sum_{j=0}^{\infty} (e_{j+1} - e_j)(e_{j+1} - e_j)^T \left(\sum_{i=1}^d \binom{d}{i} [(\pi_d)_j(t)]^i \left(\sum_{m=j+1}^{\infty} (\pi_d)_m(t) \right)^{d-i} \right) \\ & + \sum_{j=1}^{\infty} (e_{j-1} - e_j)(e_{j-1} - e_j)^T (\pi_d)_j(t). \end{aligned} \tag{18}$$

Analogous to G in (16), let G_d be a map from $\tilde{\ell}_2 \times \mathcal{S}$ to ℓ_2 , where $\tilde{\ell}_2$ is as in (15), defined as

$$(G_d)_i(x, r) := \frac{\partial}{\partial u} (F_d)_i(r + ux) \Big|_{u=0}, \quad i \in \mathbb{N}_0, u \in \mathbb{R}. \tag{19}$$

In the special case that $d = 2$, this function simply reduces to

$$(G_2)_i(x, r) = 2\lambda \sum_{m=i}^{\infty} [x_{i-1}r_m + r_{i-1}x_m - x_i r_{m+1} - r_i x_{m+1}] + (x_{i+1} - x_i).$$

The following result is immediate from Theorem 2.

Corollary 2. *Suppose that $\sup_{n \in \mathbb{N}} \sum_{j=0}^{\infty} j^2 (\pi_d^n)_j(0) < \infty$ and $\pi_d^n(0) \rightarrow \pi_0$ in \mathcal{S} as $n \rightarrow \infty$. Also, suppose that $X_d^n(0) = \sqrt{n}[\pi_d^n(0) - \pi_0] \rightarrow x_0$ in probability in ℓ_2 and that*

$$\sup_{n \in \mathbb{N}} \sum_{j=0}^{\infty} j^2 ((X_d^n)_j(0))^2 < \infty.$$

Then $X_d^n \Rightarrow X_d$ in $\mathbb{D}([0, T]: \ell_2)$, where X_d is the unique weak solution to (11) with values in $\tilde{\ell}_2$, with G replaced by G_d defined by (19) and $a(t)$ replaced by $a_d(t)$ which is given as the symmetric square root of the operator $\Phi_d(t)$ in (18).

3. Semimartingale representation

In this section we write the state processes using compensated time-changed Poisson processes to give a semimartingale representation for the system. Let $\{N_\ell, \ell \in \Sigma\}$ and $\{D_i, i \in \mathbb{N}_0\}$ be collections of mutually independent unit-rate Poisson processes. The process N_ℓ will be used to represent the stream of jobs requesting files which are stored at servers with queue length configuration (immediately before the time of arrival of the request) $\ell = (\ell_1, \dots, \ell_L)$. Similarly, D_i will represent the stream of jobs completed by servers whose queue length (immediately before the time of completion) is equal to i . From the form of the generator in (2) we see that the state process π^n can be expressed as

$$\begin{aligned} \pi^n(t) = \pi^n(0) &+ \frac{1}{n} \sum_{\ell \in \Sigma} \Delta_\ell N_\ell \left(\int_0^t \frac{n\lambda}{\binom{n}{L}} \prod_{i=0}^{\infty} \binom{n\pi_i^n(s)}{\rho_i(\ell)} ds \right) \\ &+ \frac{1}{n} \sum_{i=1}^{\infty} (e_{i-1} - e_i) D_i \left(k \int_0^t n\pi_i^n(s) ds \right). \end{aligned}$$

By adding and subtracting the compensators of the Poisson processes we can write the state process as a semimartingale. Namely,

$$\pi^n(t) = \pi^n(0) + A^n(t) + M^n(t), \tag{20}$$

where

$$A^n(t) := \sum_{\ell \in \Sigma} \Delta_\ell \int_0^t \frac{\lambda}{\binom{n}{L}} \prod_{i=0}^{\infty} \binom{n\pi_i^n(s)}{\rho_i(\ell)} ds + k \sum_{i=1}^{\infty} (e_{i-1} - e_i) \int_0^t \pi_i^n(s) ds \tag{21}$$

and

$$\begin{aligned} M^n(t) := &\sum_{\ell \in \Sigma} \frac{1}{n} \Delta_\ell N_\ell \left(\frac{n\lambda}{\binom{n}{L}} \int_0^t \prod_{i=0}^{\infty} \binom{n\pi_i^n(s)}{\rho_i(\ell)} ds \right) - \sum_{\ell \in \Sigma} \Delta_\ell \frac{\lambda}{\binom{n}{L}} \int_0^t \prod_{i=0}^{\infty} \binom{n\pi_i^n(s)}{\rho_i(\ell)} ds \\ &+ \sum_{i=1}^{\infty} \frac{1}{n} (e_{i-1} - e_i) D_i \left(k \int_0^t n\pi_i^n(s) ds \right) - k \sum_{i=1}^{\infty} (e_{i-1} - e_i) \int_0^t \pi_i^n(s) ds. \end{aligned} \tag{22}$$

It will follow from (45) and (54) that both $M^n(t)$ and $A^n(t)$ take values in ℓ_2 . Specifically, (45) and (54) imply that, for some $c_\zeta \in (0, \infty)$,

$$A_j^n(t) \leq \int_0^t \left(\frac{\lambda}{\binom{n}{L}} c_\zeta n^L [\pi_{j-1}^n(s) + \pi_j^n(s)] + k[\pi_{j+1}^n(s) + \pi_j^n(s)] \right) ds$$

for all $t \in [0, T]$, $n \in \mathbb{N}$, and $j \in \mathbb{N}_0$. Thus, there exists a $\kappa \in (0, \infty)$ such that

$$\sum_{j=0}^{\infty} A_j^n(t)^2 \leq \kappa \sum_{j=0}^{\infty} \int_0^t [\pi_{j-1}^n(s)^2 + \pi_{j+1}^n(s)^2 + \pi_j^n(s)^2] ds \leq 3\kappa t$$

for all $t \in [0, T]$. A similar argument shows that $A^n(t)$ in fact takes values in ℓ_1 . In the next section we establish tightness of $\{M^n\}$ as a sequence of ℓ_2 -valued processes.

Similarly, using (20) and (8) for $\pi(t)$, we can express X^n as a semimartingale through the equation

$$X^n(t) = X^n(0) + A^{n,c}(t) + M^{n,c}(t), \tag{23}$$

where

$$A^{n,c}(t) = \sqrt{n} \left[A^n(t) - \int_0^t F(\pi(s)) ds \right] \tag{24}$$

and $M^{n,c}(t) = \sqrt{n}M^n(t)$. We note that there is a natural filtration $\{\mathcal{F}_t^n\}_{0 \leq t \leq T}$ on the probability space where the processes N_ℓ , D_i , and π^n are defined such that A^n , M^n , π^n , X^n , $M^{n,c}$, and $A^{n,c}$ are RCLL processes adapted to the filtration, and M^n and $M^{n,c}$ are $\{\mathcal{F}_t^n\}$ -local martingales.

4. Law of large numbers

In this section we present the proof of Theorem 1. First, in Section 4.1 we use the semimartingale representation from Section 3 to prove a key tightness property (see Proposition 3). Then, in Section 4.2 we prove the unique solvability of (7) and complete the proof of Theorem 1 by proving convergence of π^n to the unique solution of (7) in $\mathbb{C}([0, T]: \mathcal{S})$.

4.1. Tightness

In this section we prove tightness of $\{(\pi^n, M^n)\}_{n \in \mathbb{N}}$. We first recall the notion of \mathbb{C} -tightness. From [14, Theorem 3.10.2], it can be seen that the definition presented below is equivalent to the more standard definition of \mathbb{C} -tightness (see [19, Definition VI.3.25]).

Definition 1. Let $(\mathcal{Z}, d_{\mathcal{Z}})$ be a Polish space. For $z \in \mathbb{D}([0, T]: \mathcal{Z})$, let

$$j_T(z) := \sup_{0 \leq t \leq T} d_{\mathcal{Z}}(z(t), z(t-)).$$

A sequence $\{Z_n\}_{n \in \mathbb{N}}$ of $\mathbb{D}([0, T]: \mathcal{Z})$ -valued random variables is said to be \mathbb{C} -tight if it is tight in $\mathbb{D}([0, T]: \mathcal{Z})$ and $j_T(Z_n) \Rightarrow 0$.

If Z_n and Z are $\mathbb{D}([0, T]: \mathcal{Z})$ -valued random variables and $Z_n \Rightarrow Z$ then $\mathbb{P}(Z \in \mathbb{C}([0, T]: \mathcal{Z})) = 1$ if and only if $\{Z_n\}_{n \in \mathbb{N}}$ is \mathbb{C} -tight [14]. The following proposition proves the \mathbb{C} -tightness of $\{\pi^n\}_{n \in \mathbb{N}}$ and convergence of M^n to the zero process.

Proposition 3. *Suppose that $\pi^n(0) \rightarrow \pi_0$ in \mathcal{S} as $n \rightarrow \infty$. Then $\{(\pi^n, M^n)\}_{n \in \mathbb{N}}$ is a \mathbb{C} -tight sequence of $\mathbb{D}([0, T]: \mathcal{S} \times \ell_2)$ -valued random variables. Furthermore, $M^n \Rightarrow 0$ in $\mathbb{D}([0, T]: \ell_2)$.*

Proof. We first prove the second statement by arguing that $\mathbb{E} \sup_{0 \leq s \leq T} \|M^n(s)\|_2^2 \rightarrow 0$ as $n \rightarrow \infty$. For this, from Doob’s inequality, it suffices to show that $\mathbb{E}|\langle M^n \rangle(T)| \rightarrow 0$ as $n \rightarrow \infty$, where

$$\langle M^n \rangle(s) := \sum_{j=0}^{\infty} \langle M_j^n \rangle(s), \quad s \in [0, T].$$

Note that $\langle e_i, e_k e_j^T e_m \rangle_2 = \delta_{ik} \delta_{jm}$ for all $i, j, k, l \in \mathbb{N}_0$, where δ_{xy} is 1 if $x = y$ and 0 otherwise. It then follows that

$$\begin{aligned} \sum_{i=1}^{\infty} \langle e_j, (e_{i-1} - e_i)(e_{i-1} - e_i)^T e_j \rangle_2 \pi_i^n(s) &= \sum_{i=1}^{\infty} \langle e_j, (e_{i-1} e_{i-1}^T + e_i e_i^T) e_j \rangle_2 \pi_i^n(s) \\ &= \pi_{j+1}^n(s) + \pi_j^n(s). \end{aligned} \tag{25}$$

Since $\{N_\ell, D_i \mid \ell \in \Sigma, i \in \mathbb{N}_0\}$ are mutually independent Poisson processes, we now have, from (22),

$$(M_j^n)(t) = \frac{\lambda}{n \binom{n}{L}} \int_0^t Z(j, n\pi^n(s)) ds + \frac{k}{n} \int_0^t [\pi_{j+1}^n(s) + \pi_j^n(s)] ds, \tag{26}$$

where

$$Z(j, n\pi^n(s)) = \sum_{\ell \in \Sigma} \langle e_j, \Delta_\ell \Delta_\ell^T e_j \rangle_2 \prod_{i=0}^{\infty} \binom{n\pi_i^n(s)}{\rho_i(\ell)}. \tag{27}$$

The ℓ th term in the sum on the right-hand side of (27) is the contribution from jobs that request servers with queue length configuration ℓ . A fixed $\ell \in \Sigma$ will make a nonzero contribution to $\langle e_j, \Delta_\ell \Delta_\ell^T e_j \rangle_2$ if j or $j - 1$ is one of the k -smallest coordinates in ℓ . Thus, for a fixed $\ell \in \Sigma$, the ℓ th term in (27) is nonzero only if j or $j - 1$ is a member of the set (ℓ_1, \dots, ℓ_k) . The contribution from all such ℓ in the sum (27) can be counted as follows. Suppose that $0 \leq i_1 \leq k - 1$ servers are selected among those with queue length less than $j - 1$. This corresponds to $\binom{n \sum_{m=0}^{j-2} \pi_m^n(s)}{i_1}$ different choices of servers. In addition, suppose that $i_2 \leq L - i_1$ and $i_3 \leq L - i_1 - i_2$ servers are selected among those with queue length equal to $j - 1$ and j , respectively. This corresponds to $\binom{n\pi_{j-1}^n(s)}{i_2}$ and $\binom{n\pi_j^n(s)}{i_3}$ choices, respectively. It follows that $L - i_1 - i_2 - i_3$ servers must be selected which have queue length larger than j , which corresponds to $\binom{n \sum_{m=j+1}^{\infty} \pi_m^n(s)}{L - i_1 - i_2 - i_3}$ possible choices. Since jobs are only routed to the k -shortest servers, we have, with

$$\Sigma_j(i_1, i_2, i_3) := \left\{ \ell \in \Sigma \mid \sum_{i=1}^{j-2} \rho_i(\ell) = i_1, \rho_{j-1}(\ell) = i_2, \rho_j(\ell) = i_3 \right\}, \tag{28}$$

for $\ell \in \Sigma_j(i_1, i_2, i_3)$,

$$\langle e_j, \Delta_\ell \Delta_\ell^T e_j \rangle_2 = [i_2 \wedge (k - i_1) - i_3 \wedge (k - i_1 - i_2)]_+^2, \tag{29}$$

and, thus, for $x \in n\mathcal{S}_n$,

$$\begin{aligned} Z(j, x) &= \sum_{i_1=0}^{k-1} \binom{\sum_{m=0}^{j-2} x_m}{i_1} \sum_{i_2=0}^{L-i_1} \binom{x_{j-1}}{i_2} \\ &\quad \times \sum_{i_3=0}^{L-i_1-i_2} [i_2 \wedge (k - i_1) - i_3 \wedge (k - i_1 - i_2)]_+^2 \binom{x_j}{i_3} \binom{\sum_{m=j+1}^{\infty} x_m}{L - i_1 - i_2 - i_3}, \end{aligned} \tag{30}$$

recalling that we adopt the convention that, for $a < b$, $\sum_{i=b}^a x_i = 0$.

Note that, for nonnegative integers $a, b, a \geq b$,

$$\binom{a}{b} \leq \frac{a^b}{b!}. \tag{31}$$

This fact, combined with (30) and recalling the fact that $\pi^n(s) \in \mathcal{S}$ for $s \in [0, T]$, gives the following bound on $Z(j, n\pi^n(s))$:

$$\begin{aligned} Z(j, n\pi^n(s)) &\leq \sum_{i_1=0}^{k-1} \frac{(n \sum_{m=0}^{j-2} \pi_m^n(s))^{i_1}}{i_1!} \sum_{i_2=0}^{L-i_1} \frac{(n\pi_{j-1}^n(s))^{i_2}}{i_2!} \\ &\quad \times \sum_{i_3=0}^{L-i_1-i_2} k^2 \mathbf{1}_{\{i_2 \vee i_3 > 0\}} \frac{(n\pi_j^n(s))^{i_3}}{i_3!} \frac{(n \sum_{m=j+1}^{\infty} \pi_m^n(s))^{L-i_1-i_2-i_3}}{(L-i_1-i_2-i_3)!} \\ &\leq n^L \sum_{i_1=0}^{k-1} \sum_{i_2=0}^{L-i_1} \sum_{i_3=0}^{L-i_1-i_2} k^2 \mathbf{1}_{\{i_2 \vee i_3 > 0\}} (\pi_{j-1}^n(s))^{i_2} (\pi_j^n(s))^{i_3} \\ &\leq c_Z n^L (\pi_{j-1}^n(s) + \pi_j^n(s)) \end{aligned} \tag{32}$$

for some $c_Z \in (0, \infty)$. Using (32) in (26) gives

$$\begin{aligned} \mathbb{E}| \langle M^n \rangle(t) | &\leq \mathbb{E} \left| \frac{2\lambda(n-L)! L! c_Z n^L}{n \times n!} \int_0^t \sum_{j=0}^{\infty} \pi_j^n(s) ds \right| + \mathbb{E} \left| \frac{2k}{n} \int_0^t \sum_{j=0}^{\infty} \pi_j^n(s) ds \right| \\ &\leq \left| \frac{2\lambda(n-L)! L! c_Z n^L}{n \times n!} t \right| + \left| \frac{2k}{n} t \right|. \end{aligned} \tag{33}$$

Thus, $\mathbb{E}| \langle M^n \rangle_T | \rightarrow 0$ and, consequently, $\mathbb{E} \sup_{0 \leq s \leq T} \|M^n(s)\|_2^2 \rightarrow 0$ as $n \rightarrow \infty$. It follows that $M^n \Rightarrow 0$ in $\mathbb{D}([0, T]: \ell_2)$, which completes the proof of the second statement.

Tightness of $\{\pi^n\}_{n \in \mathbb{N}}$ in $\mathbb{D}([0, T]: \mathcal{S})$ follows as in the proof of Theorem 3.4 of [16]. Namely, it suffices to show tightness of $\{Q_1^n\}_{n \in \mathbb{N}}$ in $\mathbb{D}([0, T]: \mathbb{N})$ (cf. [36]). However, this tightness is an immediate consequence of the fact that the jumps of Q_1^n can be embedded in a Poisson process with rate $\lambda L + k$.

Finally, in order to show that $\{\pi^n\}_{n \in \mathbb{N}}$ is \mathbb{C} -tight, it suffices to show that

$$j_T(\pi^n) := \sup_{0 \leq t \leq T} d_0(\pi^n(t), \pi^n(t-)) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Note that the arrivals to the n th system occur according to a Poisson process with rate $n\lambda$. When a job arrives in the system, the dispatcher assigns it to k different servers, causing the queue length of each of the k chosen servers to increase by one. The n servers in the system process jobs according to Poisson processes with rate $1/k$. Any completion of job processing results in the corresponding queue length dropping by 1. These $n + 1$ Poisson processes are mutually independent from the construction in Section 3, which ensures that the compensated processes,

$$N_\ell \left(\int_0^t \frac{n\lambda}{\binom{n}{L}} \prod_{i=0}^{\infty} \binom{n\pi_i^n(s)}{\rho_i(\ell)} ds \right) - \int_0^t \frac{n\lambda}{\binom{n}{L}} \prod_{i=0}^{\infty} \binom{n\pi_i^n(s)}{\rho_i(\ell)} ds$$

and

$$D_i \left(k \int_0^t n \pi_i^n(s) ds \right) - k \int_0^t n \pi_i^n(s) ds,$$

are martingales with respect to a common filtration. Therefore, these Poisson processes do not have simultaneous jumps. Each arrival creates a jump for the vector $\pi^n(t)$ with

$$d_0(\pi^n(t), \pi^n(t-)) \leq \frac{2k}{n},$$

and each service completion event produces a jump for the vector $\pi^n(t)$ with

$$d_0(\pi^n(t), \pi^n(t-)) \leq \frac{2}{n}.$$

Thus, almost surely, at any instant the jump size $d_0(\pi^n(t), \pi^n(t-))$ is at most $2k/n$. Therefore,

$$j_T(\pi^n) \leq \frac{2k}{n} \rightarrow 0,$$

which completes the proof. □

4.2. Convergence

In this section we provide the proof of Theorem 1. Since we have already proved tightness of $\{\pi^n\}_{n \in \mathbb{N}}$ in Section 4.1, all that remains is to prove uniqueness of the solutions of (7) in an appropriate class and to characterize the limit of any weakly convergent subsequence as the unique solution to (7). We first present the following Lipschitz property for the map $F: \mathcal{S} \rightarrow \ell_1$, defined in (5), that will give uniqueness of the solutions to (7). We note that in the proof of Theorem 2 we will need a stronger Lipschitz property of F in the ℓ_2 norm. This Lipschitz property is not immediate on the space \mathcal{S} , but, as shown in Lemma 4, is satisfied on a smaller class \mathcal{V}_M .

Lemma 1. *The map F is a Lipschitz function from \mathcal{S} to ℓ_1 . Namely, there exists a $C_1 \in (0, \infty)$ such that, for any $r, \tilde{r} \in \mathcal{S}$,*

$$\|F(r) - F(\tilde{r})\|_1 \leq C_1 \|r - \tilde{r}\|_1. \tag{34}$$

Proof. Let $r, \tilde{r} \in \mathcal{S}$ and, for $i_1 \in \mathbb{N}_0$ and $j, i_2 \in \mathbb{N}$, define $R_{j,i_1,i_2}(r, \tilde{r})$ as

$$R_{j,i_1,i_2}(r, \tilde{r}) := \left(\sum_{m=0}^{j-1} r_m \right)^{i_1} \left(\sum_{m=j+1}^{\infty} r_m \right)^{L-i_1-i_2} r_j^{i_2} - \left(\sum_{m=0}^{j-1} \tilde{r}_m \right)^{i_1} \left(\sum_{m=j+1}^{\infty} \tilde{r}_m \right)^{L-i_1-i_2} \tilde{r}_j^{i_2}. \tag{35}$$

Note that, for any $a, b, c, \tilde{a}, \tilde{b}, \tilde{c} \in \mathbb{R}_+$,

$$abc - \tilde{a}\tilde{b}\tilde{c} = ab(c - \tilde{c}) + a(b - \tilde{b})\tilde{c} + (a - \tilde{a})\tilde{b}\tilde{c}. \tag{36}$$

Combining (35), (36), and the fact that $r, \tilde{r} \in \mathcal{S}$, we have

$$\begin{aligned} |R_{j,i_1,i_2}(r, \tilde{r})| &\leq |r_j^{i_2} - \tilde{r}_j^{i_2}| + \tilde{r}_j^{i_2} \left| \left(\sum_{m=j+1}^{\infty} r_m \right)^{L-i_1-i_2} - \left(\sum_{m=j+1}^{\infty} \tilde{r}_m \right)^{L-i_1-i_2} \right| \\ &\quad + \tilde{r}_j^{i_2} \left| \left(\sum_{m=0}^{j-1} r_m \right)^{i_1} - \left(\sum_{m=0}^{j-1} \tilde{r}_m \right)^{i_1} \right|. \end{aligned} \tag{37}$$

For any $a, b \in \mathbb{R}$ and $i \in \mathbb{N}$, $(a^i - b^i) = (a - b) \sum_{j=1}^i a^{i-j} b^{j-1}$. Thus, if $a, b \in [0, 1]$ and $i \leq L$, $|a^i - b^i| \leq |a - b|L$. This inequality along with (37) implies that there exist $\kappa_1, \kappa'_1 > 0$ such that, for all $i_1, i_2 \leq L$, $i_2 > 0$,

$$\begin{aligned} |R_{j,i_1,i_2}(r, \tilde{r})| &\leq \kappa'_1 \left(|r_j - \tilde{r}_j| + \tilde{r}_j^{i_2} \sum_{m=j+1}^{\infty} |r_m - \tilde{r}_m| + \tilde{r}_j^{i_2} \sum_{m=0}^{j-1} |r_m - \tilde{r}_m| \right) \\ &\leq \kappa_1 (|r_j - \tilde{r}_j| + \tilde{r}_j \|r - \tilde{r}\|_1). \end{aligned} \tag{38}$$

The definition of F (see (5)) and the triangle inequality imply that

$$\|F(r) - F(\tilde{r})\|_1 \leq \lambda L! \sum_{j=0}^{\infty} |\bar{\zeta}^\delta(j, r) - \bar{\zeta}^\delta(j, \tilde{r})| + k \sum_{j=0}^{\infty} |(r - \tilde{r})_{j+1} - (r - \tilde{r})_j|. \tag{39}$$

Noting that

$$\bar{\zeta}^\delta(j, r) - \bar{\zeta}^\delta(j, \tilde{r}) = [\bar{\zeta}(j - 1, r) - \bar{\zeta}(j - 1, \tilde{r})] - [\bar{\zeta}(j, r) - \bar{\zeta}(j, \tilde{r})],$$

it follows that

$$\sum_{j=0}^{\infty} |\bar{\zeta}^\delta(j, r) - \bar{\zeta}^\delta(j, \tilde{r})| \leq 2 \sum_{j=0}^{\infty} |\bar{\zeta}(j, r) - \bar{\zeta}(j, \tilde{r})| \leq \kappa_2 \sum_{j=0}^{\infty} \sum_{i_1=0}^{k-1} \sum_{i_2=1}^{L-i_1} |R_{j,i_1,i_2}(r, \tilde{r})|, \tag{40}$$

where the second inequality follows from the the definitions of $\bar{\zeta}$ and R . Combining (40) with (39) and applying (38) yields, for some $\kappa_3 > 0$,

$$\begin{aligned} \|F(r) - F(\tilde{r})\|_1 &\leq \kappa_2 \lambda L! \sum_{j=0}^{\infty} \sum_{i_1=0}^{k-1} \sum_{i_2=1}^{L-i_1} |R_{j,i_1,i_2}(r, \tilde{r})| + 2k \sum_{j=0}^{\infty} |r_j - \tilde{r}_j| \\ &\leq \kappa_3 \sum_{j=0}^{\infty} [|r_j - \tilde{r}_j| + \tilde{r}_j \|r - \tilde{r}\|_1] + 2k \|r - \tilde{r}\|_1, \end{aligned}$$

and, thus, with $C_1 := 2(\kappa_3 + k)$, (34) is satisfied for all $r, \tilde{r} \in \mathcal{S}$, which proves the result. □

Using the above Lipschitz property of F , we can now complete the proof of Proposition 1.

Proof of Proposition 1. Existence of a $\pi \in \mathbb{C}([0, T]: \mathcal{S})$ that solves (7) will be shown below in the proof of Theorem 1. We now argue uniqueness. Suppose that π and $\tilde{\pi}$ are two elements of $\mathbb{C}([0, T]: \mathcal{S})$ satisfying (7) with $\pi(0) = \tilde{\pi}(0) = \pi_0$. The Lipschitz property of F proved in Lemma 1 implies that, for all $t \in [0, T]$,

$$\begin{aligned} \|\pi(t) - \tilde{\pi}(t)\|_1 &= \left\| \int_0^t [F(\pi(s)) - F(\tilde{\pi}(s))] ds \right\|_1 \\ &\leq \int_0^t \|F(\pi(s)) - F(\tilde{\pi}(s))\|_1 ds \\ &\leq C_1 \int_0^t \|\pi(s) - \tilde{\pi}(s)\|_1 ds. \end{aligned}$$

The result follows. □

We now proceed to the proof of Theorem 1.

Proof of Theorem 1. From Proposition 3, $\{\pi^n\}_{n \in \mathbb{N}}$ is a \mathbb{C} -tight sequence of $\mathbb{D}([0, T]: S)$ -valued random variables.

Note from (20) that, for all $j \in \mathbb{N}_0$,

$$\pi^n(t) = \pi^n(0) + V^n(t) + M^n(t) + \int_0^t F(\pi^n(s)) ds, \tag{41}$$

where

$$V^n(t) := A^n(t) - \int_0^t F(\pi^n(s)) ds.$$

From the definition of A^n in (21) we see that

$$A_j^n(t) = \int_0^t \left(\sum_{\ell \in \Sigma} \langle \Delta_\ell, e_j \rangle_2 \frac{\lambda}{\binom{n}{L}} \prod_{i=0}^{\infty} \binom{n\pi_i^n(s)}{\rho_i(\ell)} + k[\pi_{j+1}^n(s) - \pi_j^n(s)] \right) ds. \tag{42}$$

By a similar argument (see the comments given below (44)) used to obtain the representation in (30),

$$\sum_{\ell \in \Sigma} \langle \Delta_\ell, e_j \rangle_2 \prod_{i=0}^{\infty} \binom{n\pi_i^n(s)}{\rho_i(\ell)} = [\zeta(j-1, n\pi^n(s)) - \zeta(j, n\pi^n(s))], \tag{43}$$

where, for $x \in n\mathcal{S}_n$,

$$\zeta(j, x) := \sum_{i_1=0}^{k-1} \binom{\sum_{m=0}^{j-1} x_m}{i_1} \sum_{i_2=1}^{L-i_1} [i_2 \wedge (k-i_1)] \binom{x_j}{i_2} \binom{\sum_{m=j+1}^{\infty} x_m}{L-i_1-i_2}. \tag{44}$$

One can interpret $\zeta(j, x)$ as the rate at which jobs are being routed into queues of length j when the system is in state x . Recall that any incoming job corresponds to the selection of L queues. The term on the right-hand side of (44) then sums over all possible queue length configurations of this selection. In particular, i_1 represents the number of queues with lengths less than j , i_2 corresponds to the queues of length equal to j , and $L - i_1 - i_2$ are the queues of length greater than j . Since we are routing jobs to the k -shortest queues, the rate must be multiplied by the factor $[i_2 \wedge (k - i_1)]$ rather than i_2 . From our convention that $x_{-1} = 0$, we see that $\zeta(-1, x) = 0$. In addition, recalling the conventions that, for $a < b$, $\sum_{i=b}^a x_i = 0$ and $\binom{0}{0} = 1$, we see that $\zeta(0, x)$ is well defined. Combining (42), (43), and (44) gives the following representation for A_j^n :

$$A_j^n(t) = \frac{\lambda}{\binom{n}{L}} \int_0^t [\zeta(j-1, n\pi^n(s)) - \zeta(j, n\pi^n(s))] ds + k \int_0^t [\pi_{j+1}^n(s) - \pi_j^n(s)] ds. \tag{45}$$

For each fixed $j, i_1 \in \mathbb{N}_0$ and $i_2 \in \mathbb{N}$ with $i_1, i_2 \leq L$, we have

$$\begin{aligned} & \binom{n \sum_{m=0}^{j-1} \pi_m^n(s)}{i_1} [i_2 \wedge (k - i_1)] \binom{n\pi_j^n(s)}{i_2} \binom{n \sum_{m=j+1}^{\infty} \pi_m^n(s)}{L - i_1 - i_2} \\ &= n^L \frac{(\sum_{m=0}^{j-1} \pi_m^n(s))^{i_1}}{i_1!} [i_2 \wedge (k - i_1)] \frac{(\pi_j^n(s))^{i_2}}{i_2!} \frac{(\sum_{m=j+1}^{\infty} \pi_m^n(s))^{L-i_1-i_2}}{(L - i_1 - i_2)!} \\ &+ \hat{R}_n(j, i_1, i_2, s), \end{aligned} \tag{46}$$

where

$$\sup_{i_1, i_2 \leq L} |\hat{R}_n(j, i_1, i_2, s)| \leq \kappa_1 n^{L-1} \pi_j^n(s),$$

and, thus, from the definitions of ζ and $\bar{\zeta}$ in (44) and (4),

$$\left| \zeta(j, n\pi^n(s)) - \frac{n!}{(n-L)!} \bar{\zeta}(j, \pi^n(s)) \right| \leq \kappa_2 n^{L-1} \pi_j^n(s) \quad \text{for all } s \in [0, T]. \tag{47}$$

Furthermore, using the definitions of A^n in (45) and F in (5), (47) implies that

$$\sup_{0 \leq t \leq T} \|V^n(t)\|_2 = \sup_{0 \leq t \leq T} \left\| A^n(t) - \int_0^t F(\pi^n(s)) \, ds \right\|_2 \leq \frac{\kappa_3}{n}. \tag{48}$$

Also, from Proposition 3, $M^n \Rightarrow 0$ in $\mathbb{D}([0, T]: \ell_2)$. Combining these observations with tightness of π^n , we have subsequential convergence of (π^n, M^n, V^n) to $(\pi, 0, 0)$, in distribution, in $\mathbb{D}([0, T]: \mathcal{S} \times \ell_2 \times \ell_2)$ for some $\mathbb{C}([0, T]: \mathcal{S})$ -valued π . By appealing to the Skorokhod representation theorem we can assume that this convergence holds almost surely. Noting that $r \mapsto F_j(r)$ is a continuous map from \mathcal{S} to \mathbb{R} for each $j \in \mathbb{N}_0$, we have $F_j(\pi^n(s)) \rightarrow F_j(\pi(s))$ as $n \rightarrow \infty$ for all $j \in \mathbb{N}_0$ and $s \in [0, T]$. Thus, upon sending $n \rightarrow \infty$ in (41), (9) and the dominated convergence theorem imply that, almost surely,

$$\pi_j(t) = (\pi_0)_j + \int_0^t F_j(\pi(s)) \, ds \quad \text{for all } t \in [0, T], j \in \mathbb{N}_0.$$

This shows that π satisfies (7). The result now follows from the uniqueness property shown in Proposition 1. □

5. Diffusion approximation

In this section we prove Theorem 2. Section 5.1 presents some moment estimates on π^n which will be used in the proof of Theorem 2. Section 5.2 then proves tightness of the sequence of centered and scaled state processes $\{X^n\}_{n \in \mathbb{N}}$. Section 5.3 completes the proof of Theorem 2 by proving unique solvability of the SDE (11) (Theorem 2) and characterizing limit points of X^n as this unique solution.

5.1. Moment bounds

The following elementary lemma will be useful in the proof of Lemma 3.

Lemma 2. *For all $t \geq 0$, $k \in \mathbb{N}$, and $n \in \mathbb{N}$, $\lim_{m \rightarrow \infty} \mathbb{E} m^k \sup_{0 \leq s \leq t} \pi_m^n(s) = 0$.*

Proof. Fix $n \in \mathbb{N}$. Note that file requests arrive at rate $n\lambda$. Let N be a Poisson process representing the total flow of such file requests. Also, let $m^* = \sup\{m: \pi_m^n(0) > 0\}$ be the length of the largest queue at time 0. Note that, since the system consists of n queues, m^* must be finite for any fixed n . Then, for $m > m^*$,

$$\begin{aligned} \mathbb{E} m^k \sup_{0 \leq s \leq t} \pi_m^n(s) &= \mathbb{E} \sup_{0 \leq s \leq t} \mathbf{1}_{\{N(t) \geq m - m^*\}} m^k \pi_m^n(s) + \mathbb{E} \sup_{0 \leq s \leq t} \mathbf{1}_{\{N(t) < m - m^*\}} m^k \pi_m^n(s) \\ &\leq m^k \mathbb{P}(N(t) \geq m - m^*). \end{aligned}$$

Thus, from Markov's inequality, for $m > m^*$,

$$\mathbb{E} m^k \sup_{0 \leq s \leq t} \pi_m^n(s) \leq m^k e^{-(m - m^*)} e^{n\lambda t(e-1)}.$$

The result follows. □

In the next lemma we will establish two key moment bounds that will be needed in the tightness proof (see proof of Proposition 4).

Lemma 3. *Suppose that $\sup_{n \in \mathbb{N}} \sum_{j=0}^{\infty} j^2 \pi_j^n(0) =: c_{\pi(0)} < \infty$. Then*

$$\sup_{n \in \mathbb{N}} \mathbb{E} \sup_{0 \leq t \leq T} \left(\sum_{j=0}^{\infty} j \pi_j^n(t) \right)^2 < \infty \tag{49}$$

and

$$\sup_{n \in \mathbb{N}} \mathbb{E} \int_0^T \sum_{j=0}^{\infty} j^2 \pi_j^n(t) dt < \infty. \tag{50}$$

Proof. Since $\pi^n(t) = \pi^n(0) + A^n(t) + M^n(t)$, we can write, for fixed $K \in \mathbb{N}$,

$$\mathbb{E} \sup_{0 \leq t \leq T} \left| \sum_{j=0}^K j \pi_j^n(t) \right|^2 \leq 3 \left| \sum_{j=0}^K j \pi_j^n(0) \right|^2 + 3 \mathbb{E} \sup_{0 \leq t \leq T} \left| \sum_{j=0}^K j A_j^n(t) \right|^2 + 3 \mathbb{E} \sup_{0 \leq t \leq T} \left| \sum_{j=0}^K j M_j^n(t) \right|^2. \tag{51}$$

Using (43), for $K \in \mathbb{N}$, we can write

$$\begin{aligned} \sum_{j=0}^K j \sum_{\ell \in \Sigma} \langle \Delta_{\ell}, e_j \rangle_2 \prod_{i=0}^{\infty} \binom{n \pi_i^n(s)}{\rho_i(\ell)} &= \sum_{j=1}^K j [\zeta(j-1, n\pi^n(s)) - \zeta(j, n\pi^n(s))] \\ &= \sum_{j=0}^{K-1} \zeta(j, n\pi^n(s)) - K \zeta(K, n\pi^n(s)) \end{aligned} \tag{52}$$

and

$$k \sum_{j=0}^K j [\pi_{j+1}^n(s) - \pi_j^n(s)] = -k \left(\sum_{j=1}^K \pi_j^n(s) - K \pi_{K+1}^n(s) \right). \tag{53}$$

Using similar bounds as in (32), for some $c_{\zeta} \in (0, \infty)$,

$$\zeta(j, n\pi^n(s)) \leq c_{\zeta} n^L \pi_j^n(s). \tag{54}$$

The above bound implies that, for some $\kappa_1 \in (0, \infty)$ and all $n, K \in \mathbb{N}$,

$$\begin{aligned} \mathbb{E} \sup_{0 \leq t \leq T} \left[\frac{\lambda}{\binom{n}{L}} \int_0^t \sum_{j=1}^K \zeta(j-1, n\pi^n(s)) + k \int_0^t \sum_{j=0}^K \pi_j^n(s) ds \right]^2 &\leq \mathbb{E} \left[\left(c_{\zeta} n^L \frac{\lambda}{\binom{n}{L}} + k \right) T \right]^2 \\ &\leq \kappa_1. \end{aligned}$$

Combined with (45), (52), and (53), the above estimate gives, for all $n, K \in \mathbb{N}$,

$$\mathbb{E} \sup_{0 \leq t \leq T} \left| \sum_{j=0}^K j A_j^n(t) \right|^2 \leq \kappa_2 \left(1 + K \mathbb{E} \left[\sup_{0 \leq t \leq T} (\pi_K^n(t) + \pi_{K+1}^n(t)) \right] \right). \tag{55}$$

We now consider $\mathbb{E} \sup_{0 \leq t \leq T} |\sum_{j=0}^K jM_j^n(t)|^2$. Since $\sum_{j=0}^K jM_j^n(t)$ is a martingale, Doob's inequality implies that

$$\mathbb{E} \sup_{0 \leq t \leq T} \left| \sum_{j=0}^K jM_j^n(t) \right|^2 \leq 4\mathbb{E} \left\langle \sum_{j=0}^K jM_j^n \right\rangle(T) = 4\mathbb{E} \sum_{j_1=0}^K \sum_{j_2=0}^K j_1 j_2 \langle M_{j_1}^n, M_{j_2}^n \rangle(T). \tag{56}$$

The diagonal terms ($j_1 = j_2$) in the above sum are given by (26). We now consider the off-diagonal terms. Fix $0 \leq j_1 < j_2 \leq K$, and note that in order to compute $\langle M_{j_1}^n, M_{j_2}^n \rangle(T)$, we must expand

$$Z(j_1, j_2, n\pi^n(s)) := \sum_{\ell \in \Sigma} \langle e_{j_1}, \Delta_\ell \Delta_\ell^T e_{j_2} \rangle_2 \prod_{i=0}^{\infty} \binom{n\pi_i^n(s)}{\rho_i(\ell)}. \tag{57}$$

Similar to (27), the ℓ th term in (57) is the contribution from jobs that request servers with queue length configuration ℓ . A fixed $\ell \in \Sigma$ will make a nonzero contribution to $\langle e_{j_1}, \Delta_\ell \Delta_\ell^T e_{j_2} \rangle_2$ if (j_1 or $j_1 - 1$) and (j_2 or $j_2 - 1$) are among the k -smallest coordinates in ℓ . That is, for a fixed $\ell \in \Sigma$, the ℓ th term is nonzero only if (j_1 or $j_1 - 1$) is a member of the set (ℓ_1, \dots, ℓ_k) and (j_2 or $j_2 - 1$) is also a member. The contribution from all such ℓ in the sum (57) can be counted in a method analogous to that used to obtain (30). Namely, we count the numbers of choices of servers with queue length less than $j_1 - 1$, equal to $j_1 - 1$, equal to j_1 , between j_1 and $j_2 - 1$, equal to $j_2 - 1$, equal to j_2 , and larger than j_2 . One must be careful in the cases $j_2 - 1 = j_1$ and $j_2 - 1 = j_1 + 1$. In both cases there are no servers with length between j_1 and $j_2 - 1$. In the first case above ($j_2 - 1 = j_1$), we must also be careful not to double count. To ensure this, we include an indicator function $\mathbf{1}_{\{j_2 > j_1 + 1\}}$ in the upper index of the binomial coefficient corresponding to the selection of servers with queue length equal to $j_2 - 1$. Combining these observations we see that, for $x \in n\mathcal{S}_n$,

$$\begin{aligned} Z(j_1, j_2, x) &= \sum_{\ell \in \Sigma} \langle e_{j_1}, \Delta_\ell \Delta_\ell^T e_{j_2} \rangle_2 \prod_{i=0}^{\infty} \binom{x_i}{\rho_i(\ell)} \\ &= \sum_{i_1=0}^{k-2} \binom{\sum_{m=0}^{j_1-2} x_m}{i_1} \sum_{i_2=0}^{k-i_1-1} \binom{x_{j_1-1}}{i_2} \sum_{i_3=0}^{k-i_1-i_2-1} [i_2 - i_3] \binom{x_{j_1}}{i_3} \\ &\quad \times \sum_{i_4=0}^{k-i_1-i_2-i_3-1} \binom{\sum_{m=j_1+1}^{j_2-2} x_m}{i_4} \sum_{i_5=0}^{L-\sum_{n=1}^4 i_n} \binom{x_{j_2-1} \mathbf{1}_{\{j_2 > j_1 + 1\}}}{i_5} \\ &\quad \times \sum_{i_6=0}^{L-\sum_{n=1}^5 i_n} \left[(\mathbf{1}_{\{j_2 = j_1 + 1\}}(i_3 - i_5) + i_5) \wedge \left(k - \sum_{n=1}^4 i_n \right)_+ - i_6 \wedge \left(k - \sum_{n=1}^5 i_n \right)_+ \right] \\ &\quad \times \binom{x_{j_2}}{i_6} \binom{\sum_{m=j_2+1}^{\infty} x_m}{L - \sum_{n=1}^6 i_n}. \end{aligned} \tag{58}$$

For $j_1 > j_2$, we define $Z(j_1, j_2, x) := Z(j_2, j_1, x)$. The contribution to $\langle M_{j_1}^n, M_{j_2}^n \rangle(T)$ for $j_1 \neq j_2$ from completed jobs is given by the term

$$\sum_{i=1}^{\infty} \langle e_{j_1}, (e_{i-1} - e_i)(e_{i-1} - e_i)^T e_{j_2} \rangle_2 2\pi_i^n(s) = -\mathbf{1}_{\{j_1 = j_2 - 1\}} \pi_{j_2}^n(s) - \mathbf{1}_{\{j_1 - 1 = j_2\}} \pi_{j_1}^n(s). \tag{59}$$

This follows on noting that if a job is completed from a queue of length j then its queue length becomes $j - 1$. This implies that the contribution is zero unless $j_1 = j_2 - 1$ or $j_1 - 1 = j_2$, which results in the above expression. Combining (58) and (59) gives, for $j_1, j_2 \in \mathbb{N}_0$,

$$\begin{aligned} \langle M_{j_1}^n, M_{j_2}^n \rangle(T) &= \frac{\lambda}{n \binom{n}{L}} \int_0^T Z(j_1, j_2, n\pi^n(s)) \, ds \\ &\quad + \frac{k}{n} \int_0^T [\mathbf{1}_{\{j_1=j_2\}}[\pi_{j_1}^n(s) + \pi_{j_1+1}^n(s)] - \mathbf{1}_{\{j_1=j_2-1\}}\pi_{j_2}^n(s) \\ &\quad - \mathbf{1}_{\{j_1-1=j_2\}}\pi_{j_1}^n(s)] \, ds, \end{aligned} \tag{60}$$

where, by convention, $Z(j, j, x) := Z(j, x)$. Referring to the definition of Z in (58), note that, for $j_2 > j_1 + 1$, $Z(j_1, j_2, x) = 0$ unless $(i_2$ or $i_3)$ is greater than 0 and $(i_5$ or $i_6)$ is greater than 0. In the case that $j_2 = j_1 + 1$, $Z(j_1, j_2, x) = 0$ unless $(i_2$ or $i_3)$ is greater than 0 and $(i_3$ or $i_6)$ is greater than 0. Therefore, (31) implies there exists a $\tilde{c}_Z \in (0, \infty)$ such that, for $r \in \mathcal{S}_n$ and $j_1 < j_2$,

$$Z(j_1, j_2, nr) \leq \tilde{c}_Z n^L [r_{j_1} r_{j_2} + r_{j_1-1} r_{j_2} + r_{j_1} r_{j_2-1} + r_{j_1-1} r_{j_2-1} + \mathbf{1}_{\{j_2=j_1+1\}} r_{j_1}]. \tag{61}$$

Combining this with (32) and (60), we have, for some $\kappa'_3, \kappa_3 \in (0, \infty)$ and all $n, K \in \mathbb{N}$,

$$\begin{aligned} &\sum_{j_1=0}^K \sum_{j_2=0}^K j_1 j_2 \langle M_{j_1}^n, M_{j_2}^n \rangle(T) \\ &\leq \frac{\kappa'_3}{n} \left[\int_0^T \sum_{j_1=0}^K \sum_{j_2=0}^K (j_1 + 1)(j_2 + 1) \pi_{j_1}^n(t) \pi_{j_2}^n(t) \, dt + \int_0^T \sum_{j=1}^{K+1} j(j + 1) \pi_j^n(t) \, dt \right] \\ &\leq \frac{\kappa_3}{n} \left[\int_0^T \left(\sum_{j=0}^K j^2 \pi_j^n(t) + (K + 1)^2 \pi_{K+1}^n(t) \right) dt + 1 \right]. \end{aligned} \tag{62}$$

Recalling that $\pi^n(t) = \pi^n(0) + A^n(t) + M^n(t)$, we have, for all $K, n \in \mathbb{N}$,

$$\begin{aligned} \mathbb{E} \int_0^T \sum_{j=0}^K j^2 \pi_j^n(t) \, dt &= \int_0^T \sum_{j=0}^K j^2 \pi_j^n(0) \, dt + \mathbb{E} \int_0^T \sum_{j=0}^K j^2 A_j^n(t) \, dt + \int_0^T \mathbb{E} \sum_{j=0}^K j^2 M_j^n(t) \, dt \\ &\leq \mathbb{E} \int_0^T \sum_{j=0}^K j^2 A_j^n(t) \, dt + \kappa_4, \end{aligned}$$

where $\kappa_4 = c_\pi(0)T$ and the last inequality follows on using the fact that $M_j^n(t)$ is a martingale. Thus, from (45), for some $\kappa_5 \in (0, \infty)$ and all $K, n \in \mathbb{N}$,

$$\begin{aligned} \mathbb{E} \int_0^T \sum_{j=0}^K j^2 \pi_j^n(t) \, dt &\leq \frac{\kappa_5}{n^L} \mathbb{E} \int_0^T \sum_{j=1}^K j^2 \int_0^t [\zeta(j - 1, n\pi^n(s)) - \zeta(j, n\pi^n(s))] \, ds \, dt \\ &\quad + \kappa_5 \mathbb{E} \int_0^T \sum_{j=1}^K j^2 \int_0^t [\pi_{j+1}^n(s) - \pi_j^n(s)] \, ds \, dt + \kappa_5. \end{aligned} \tag{63}$$

Using the facts that, for any $a_0, \dots, a_K \in \mathbb{R}$,

$$\begin{aligned} \sum_{j=1}^K j^2 [a_{j-1} - a_j] &= \sum_{j=1}^K [(j-1)^2 a_{j-1} - j^2 a_j + (2j-1)a_{j-1}] \\ &= -K^2 a_K + \sum_{j=1}^K (2j-1)a_{j-1} \end{aligned}$$

and

$$\begin{aligned} \sum_{j=0}^K j^2 [a_{j+1} - a_j] &= \sum_{j=0}^K [(j+1)^2 a_{j+1} - j^2 a_j - (2j+1)a_{j+1}] \\ &= (K+1)^2 a_{K+1} - \sum_{j=0}^K (2j+1)a_{j+1} \end{aligned}$$

in (63) we have, for some $\kappa_6 \in (0, \infty)$ and all $K, n \in \mathbb{N}$,

$$\begin{aligned} \mathbb{E} \int_0^T \sum_{j=0}^K j^2 \pi_j^n(t) dt &\leq \frac{\kappa_5}{nL} \mathbb{E} \int_0^T \int_0^t \sum_{j=0}^K (2j-1)\zeta(j-1, n\pi^n(s)) ds dt \\ &\quad + \kappa_5 \mathbb{E} \int_0^T \int_0^t (K+1)^2 \pi_{K+1}^n(s) ds dt + \kappa_5 \\ &\leq \kappa_6 \mathbb{E} \int_0^T \left[K^2 \sup_{0 \leq s \leq t} \pi_{K+1}^n(s) + \sup_{0 \leq s \leq t} \sum_{j=0}^K j \pi_j^n(s) \right] dt + \kappa_6, \end{aligned} \tag{64}$$

where the second inequality follows from (54). Thus, it follows from (56) and (62) that, for some $\kappa_7 \in (0, \infty)$,

$$\begin{aligned} \mathbb{E} \sup_{0 \leq t \leq T} \left| \sum_{j=0}^K j M^n(t) \right|^2 &\leq \frac{\kappa_3}{n} \left[\int_0^T \mathbb{E} \sum_{j=0}^K j^2 \pi_j^n(t) dt + \gamma_K^n T + 1 \right] \\ &\leq \frac{\kappa_7}{n} \left[1 + \gamma_K^n + \int_0^T \mathbb{E} \sup_{0 \leq u \leq s} \left| \sum_{j=0}^K j \pi_j^n(u) \right|^2 ds \right], \end{aligned} \tag{65}$$

where $\gamma_K^n = \mathbb{E}(K^2 \sup_{0 \leq s \leq T} \pi_{K+1}^n(s))$. Combining (51), (55), and (65), and using the fact that $\left| \sum_{j=0}^\infty j \pi_j^n(0) \right| \leq c_{\pi(0)}$,

$$\begin{aligned} \mathbb{E} \sup_{0 \leq t \leq T} \left| \sum_{j=0}^K j \pi_j^n(t) \right|^2 &\leq \kappa_8 \left(1 + \mathbb{E} \sup_{0 \leq t \leq T} \left| \sum_{j=0}^K j A_j^n(t) \right|^2 + \mathbb{E} \sup_{0 \leq t \leq T} \left| \sum_{j=0}^K j M_j^n(t) \right|^2 \right) \\ &\leq \kappa_9 \left(1 + \gamma_K^n + \frac{1}{n} \int_0^T \mathbb{E} \sup_{0 \leq s \leq t} \left| \sum_{j=0}^K j \pi_j^n(s) \right|^2 ds \right). \end{aligned}$$

By Gronwall’s lemma (since the above inequality also holds for all $T_1 \leq T$), there is a $\kappa_{10} \in (0, \infty)$ such that, for all $n, K \in \mathbb{N}$,

$$\mathbb{E} \sup_{0 \leq t \leq T} \left| \sum_{j=0}^K j \pi_j^n(t) \right|^2 \leq \kappa_{10}(1 + \gamma_K^n).$$

Sending $K \rightarrow \infty$ and recalling from Lemma 2 that, for each fixed n , as $K \rightarrow \infty$, $\gamma_K^n \rightarrow 0$ we have, for all n ,

$$\mathbb{E} \sup_{0 \leq t \leq T} \left| \sum_{j=0}^{\infty} j \pi_j^n(t) \right|^2 \leq \kappa_{10},$$

where κ_{10} is independent of n . This proves (49). Finally, (50) follows from (49) upon sending $K \rightarrow \infty$ in (64). □

5.2. Tightness

We now proceed with the proof of the tightness of $\{(X^n, M^{n,c})\}_{n \in \mathbb{N}}$. Let, for $M \in \mathbb{R}_+$,

$$\mathcal{V}_M := \left\{ r \in \mathcal{S} \mid \sum_{i=0}^{\infty} i r_i \leq M \right\},$$

where \mathcal{V}_M is equipped with the topology inherited from ℓ_2 . We begin by establishing the following Lipschitz property for F on \mathcal{V}_M .

Lemma 4. *The map F is a Lipschitz function from \mathcal{V}_M to ℓ_2 for each $M \in \mathbb{R}_+$. Namely, there exists an $C(M) \in (0, \infty)$ such that, for any $r, \tilde{r} \in \mathcal{V}_M$,*

$$\|F(r) - F(\tilde{r})\|_2 \leq C(M) \|r - \tilde{r}\|_2. \tag{66}$$

Proof. Fix $M \in \mathbb{R}_+$. Let $r, \tilde{r} \in \mathcal{V}_M$ and, for $i_1 \in \mathbb{N}_0$ and $j, i_2 \in \mathbb{N}$, recall $R_{j,i_1,i_2}(r, \tilde{r})$ from (35). Using (36) and the fact that $r, \tilde{r} \in \mathcal{S}$, we have

$$\begin{aligned} (R_{j,i_1,i_2}(r, \tilde{r}))^2 &\leq 3[r_j^{i_2} - \tilde{r}_j^{i_2}]^2 + 3\tilde{r}_j^{2i_2} \left[\left(\sum_{m=j+1}^{\infty} r_m \right)^{L-i_1-i_2} - \left(\sum_{m=j+1}^{\infty} \tilde{r}_m \right)^{L-i_1-i_2} \right]^2 \\ &\quad + 3\tilde{r}_j^{2i_2} \left[\left(\sum_{m=0}^{j-1} r_m \right)^{i_1} - \left(\sum_{m=0}^{j-1} \tilde{r}_m \right)^{i_1} \right]^2. \end{aligned}$$

By an argument similar to that used to derive (38) and an application of the Cauchy–Schwarz inequality we have the following inequality for all $i_1, i_2 \leq L, i_2 > 0$:

$$(R_{j,i_1,i_2}(r, \tilde{r}))^2 \leq \kappa_1 ([r_j - \tilde{r}_j]^2 + (j+1)\tilde{r}_j \|r - \tilde{r}\|_2^2). \tag{67}$$

Using arguments analogous to those used in the proof of Lemma 1, we have

$$\begin{aligned} \|F(r) - F(\tilde{r})\|_2 &\leq \kappa_2 \lambda L! \left(\sum_{j=0}^{\infty} \sum_{i_1=0}^{k-1} \sum_{i_2=1}^{L-i_1} [R_{j,i_1,i_2}(r, \tilde{r})]^2 \right)^{1/2} + 2k \left(\sum_{j=0}^{\infty} (r - \tilde{r})_j^2 \right)^{1/2} \\ &\leq \kappa_3 \left(\sum_{j=0}^{\infty} [[r_j - \tilde{r}_j]^2 + (j+1)\tilde{r}_j \|r - \tilde{r}\|_2^2] \right)^{1/2} + 2k \|r - \tilde{r}\|_2 \\ &\leq \kappa_4 \|r - \tilde{r}\|_2 \left(1 + \sum_{j=0}^{\infty} j \tilde{r}_j \right)^{1/2} + 2k \|r - \tilde{r}\|_2. \end{aligned} \tag{68}$$

Since $r, \tilde{r} \in \mathcal{V}_M$, (68) gives

$$\|F(r) - F(\tilde{r})\|_2 \leq \kappa_4 (M + 1)^{1/2} \|r - \tilde{r}\|_2 + 2k \|r - \tilde{r}\|_2,$$

and, thus, with $C(M) := \kappa_4 (M + 1)^{1/2} + 2k$, (66) is satisfied for all $r, \tilde{r} \in \mathcal{V}_M$, which proves the result. \square

Recall the process X^n introduced in (10) and $M^{n,c}$ defined below (24). The following proposition gives tightness of $\{(X^n, M^{n,c})\}_{n \in \mathbb{N}}$.

Proposition 4. *Suppose that $\{\pi^n\}_{n \in \mathbb{N}}$ is as in the statement of Theorem 1 with $\sup_{n \in \mathbb{N}} \sum_{j=0}^{\infty} j^2 \pi_j^n(0) < \infty$. Let $X^n(0) = \sqrt{n}(\pi^n(0) - \pi_0)$ and suppose that (17) is satisfied. Then $\{(X^n, M^{n,c})\}_{n \in \mathbb{N}}$ is a \mathbb{C} -tight sequence of $\mathbb{D}([0, T]: (\ell_2)^2)$ -valued random variables.*

Proof. We will make use of Theorem 4 in Appendix A.2. We first prove that $\{M^{n,c}\}_{n \in \mathbb{N}}$ is tight. In order to show that condition (A) in Theorem 4 is satisfied for $\{M^{n,c}\}_{n \in \mathbb{N}}$ it suffices (cf. Theorem 2.3.2 of [20]) to show that the condition is satisfied for the real-valued process $\langle M^{n,c} \rangle(t) := \sum_{j=0}^{\infty} \langle M_j^{n,c} \rangle(t)$. Fix $\varepsilon \in (0, T]$ and $T_0 \in [0, T - \varepsilon]$. Let $\tau_n \leq T_0$ be a sequence of $\{\mathcal{F}_t^n\}$ -stopping times. Then, (43) and (54) imply that, for $\theta \in [0, \varepsilon]$,

$$\begin{aligned} &|\langle M^{n,c}(\tau_n + \theta) \rangle - \langle M^{n,c}(\tau_n) \rangle| \\ &= \left| \sum_{j=0}^{\infty} \left[\int_{\tau_n}^{\tau_n + \theta} \sum_{\ell \in \Sigma} \langle \Delta_\ell, e_j \rangle_2 \frac{\lambda}{I(n)} \prod_{i=0}^{\infty} \binom{n\pi_i^n(s)}{\rho_i(\ell)} + k \int_{\tau_n}^{\tau_n + \theta} [\pi_{j+1}^n(s) - \pi_j^n(s)] ds \right] \right| \\ &\leq \kappa_1 \sum_{j=0}^{\infty} \int_{\tau_n}^{\tau_n + \theta} [\pi_j^n(s) + \pi_{j-1}^n(s) + \pi_{j+1}^n(s)] ds \\ &\leq \kappa_1 \varepsilon. \end{aligned}$$

The proof of (A) is now immediate.

We next show that $\{M^{n,c}\}_{n \in \mathbb{N}}$ satisfies condition (T₁) of Theorem 4. For this, we will apply Theorem 3. We first verify that $\{M^{n,c}(t)\}_{n \in \mathbb{N}}$ satisfies Theorem 3(a) for all $t \in [0, T]$. It follows from (33) that

$$\sup_{n \in \mathbb{N}} \mathbb{E} \langle M^{n,c} \rangle(T) = \sup_{n \in \mathbb{N}} n \mathbb{E} \langle M^n \rangle(T) \leq \kappa_2. \tag{69}$$

This, combined with Doob’s inequality, implies that, for each $n_0 \in \mathbb{N}$,

$$\sup_{n \in \mathbb{N}} \sum_{i=0}^{n_0} \mathbb{E} \sup_{0 \leq t \leq T} |M_i^{n,c}(t)| \leq n_0 + \sup_{n \in \mathbb{N}} \sum_{i=0}^{n_0} \mathbb{E} \left(\sup_{0 \leq t \leq T} M_i^{n,c}(t) \right)^2 \leq n_0 + \kappa_3.$$

Using Markov’s inequality, Theorem 3(a) follows.

We now verify Theorem 3(b) for $\{M_j^{n,c}(t)\}_{n \in \mathbb{N}}$ for each fixed $t \in [0, T]$. Note that $\langle M_j^{n,c} \rangle(t) = n \langle M_j^n \rangle(t)$ and, thus, from (26) and (32),

$$\langle M_j^{n,c} \rangle(t) \leq \kappa_4 \int_0^t (\pi_{j-1}^n(s) + \pi_j^n(s) + \pi_{j+1}^n(s)) \, ds. \tag{70}$$

It follows from (70) and the Cauchy–Schwarz inequality that

$$\begin{aligned} \sum_{j=n_0}^{\infty} \mathbb{E} (M_j^{n,c}(t))^2 &= \sum_{j=n_0}^{\infty} \mathbb{E} \langle M_j^{n,c} \rangle(t) \\ &\leq \kappa_5 \mathbb{E} \int_0^t \sum_{j=n_0-1}^{\infty} \pi_j^n(s) \, ds \\ &\leq \kappa_5 \left(\sum_{j=n_0-1}^{\infty} \frac{1}{j^2} \right)^{1/2} \int_0^t \mathbb{E} \left(\sum_{j=n_0-1}^{\infty} j^2 \pi_j^n(s) \right)^{1/2} \, ds. \end{aligned} \tag{71}$$

From (50),

$$\sup_{n \in \mathbb{N}} \mathbb{E} \int_0^T \sum_{j=0}^{\infty} j^2 (\pi_j^n(s))^2 \, ds \leq \sup_{n \in \mathbb{N}} \mathbb{E} \int_0^T \sum_{j=0}^{\infty} j^2 \pi_j^n(s) \, ds =: \kappa_6 < \infty. \tag{72}$$

Using this observation in (71), we have

$$\begin{aligned} \sum_{j=n_0}^{\infty} \mathbb{E} (M_j^{n,c}(t))^2 &\leq \kappa_7 \left(\sum_{j=n_0-1}^{\infty} \frac{1}{j^2} \right)^{1/2} \int_0^t \mathbb{E} \left(\sum_{j=n_0-1}^{\infty} j^2 \pi_j^n(s) \right)^{1/2} \, ds \\ &\leq \kappa_8 \left(\sum_{j=n_0-1}^{\infty} \frac{1}{j^2} \right)^{1/2}. \end{aligned}$$

From Markov’s inequality we now see that, for any $\delta > 0$,

$$\lim_{n_0 \rightarrow \infty} \sup_{n \in \mathbb{N}} \mathbb{P} \left(\sum_{j=n_0}^{\infty} (M_j^{n,c}(t))^2 > \delta \right) = 0,$$

which verifies Theorem 3(b). Thus, we have shown that $\{M_j^{n,c}(t)\}_{n \in \mathbb{N}}$ is a tight sequence of ℓ_2 -valued random variables for all $t \in [0, T]$. From Theorem 4, it now follows that $\{M_j^{n,c}\}_{n \in \mathbb{N}}$ is a tight sequence of $\mathbb{D}([0, T]; \ell_2)$ -valued random variables.

We will now argue that $\{X^n\}_{n \in \mathbb{N}}$ is a tight sequence of $\mathbb{D}([0, T]; \ell_2)$ -valued random variables. Again, via Theorem 4, it suffices to show that $\{X^n(t)\}_{n \in \mathbb{N}}$ is tight for every $t \in [0, T]$ (which will follow from verifying conditions (a) and (b) of Theorem 3) and that $\{X^n\}_{n \in \mathbb{N}}$

satisfies condition (A) of Theorem 4. We first show that, for all $t \in [0, T]$, Theorem 3(a) holds for $\{X^n(t)\}_{n \in \mathbb{N}}$. Namely, we show that, for each $n_0 \in \mathbb{N}$ and $t \in [0, T]$,

$$\lim_{A \rightarrow \infty} \sup_{n \in \mathbb{N}} \mathbb{P} \left(\sum_{j=0}^{n_0} |X_j^n(t)| > A \right) = 0. \tag{73}$$

Fix $\varepsilon > 0$. From Lemma 3, there is a $M \in (0, \infty)$ such that

$$\sup_{n \in \mathbb{N}} \mathbb{E} \left(\sup_{0 \leq t \leq T} \sum_{j=0}^{\infty} j \pi_j^n(t) \right) \leq \frac{M\varepsilon}{2}. \tag{74}$$

Let $B_M^n := \{\sup_{0 \leq t \leq T} \sum_{j=0}^{\infty} j \pi_j^n(t) \leq M\}$. Then, for $t \in [0, T]$ and $n_0 \in \mathbb{N}$,

$$\begin{aligned} \mathbb{P} \left(\sum_{j=0}^{n_0} |X_j^n(t)| > A \right) &\leq \mathbb{P} \left(\sup_{0 \leq t \leq T} \sum_{j=0}^{\infty} j \pi_j^n(t) > M \right) + \mathbb{P} \left(\sum_{j=0}^{n_0} |X_j^n(t)| > A, B_M^n \right) \\ &\leq \frac{\varepsilon}{2} + \mathbb{P} \left(\sum_{j=0}^{n_0} |X_j^n(t)| > A, B_M^n \right). \end{aligned} \tag{75}$$

The Cauchy–Schwarz inequality yields

$$\sum_{i=0}^{n_0} |X_i^n(t)| \leq \sqrt{n_0} \left(\sum_{j=0}^{n_0} |X_j^n(t)|^2 \right)^{1/2} \leq \sqrt{n_0} \|X^n(t)\|_2. \tag{76}$$

Furthermore, from (20) and the triangle inequality,

$$\|X^n(t)\|_2 \leq \|X^n(0)\|_2 + \|A^{n,c}(t)\|_2 + \|M^{n,c}(t)\|_2. \tag{77}$$

The definition of $A^{n,c}$ in (24) gives

$$\|A^{n,c}(t)\|_2 = \sqrt{n} \left\| A^n(t) - \int_0^t F(\pi(s)) \, ds \right\|_2.$$

The moment bound (49) proved in Lemma 3 implies that

$$\sup_{n \in \mathbb{N}} \mathbb{E} \sup_{0 \leq t \leq T} \left| \sum_{j=0}^{\infty} j \pi_j^n(t) \right|^2 =: \kappa_7 < \infty, \tag{78}$$

and, thus, for some $\kappa_8 \in (0, \infty)$,

$$\sup_{0 \leq t \leq T} \left| \sum_{j=0}^{\infty} j \pi_j(t) \right|^2 \leq \kappa_8 \tag{79}$$

as well. From (48) and the Lipschitz property proved in Lemma 4, with $M \geq \kappa_7 \vee \kappa_8$ on the set B_M^n ,

$$\|A^{n,c}(t)\|_2 \leq \sqrt{n} \int_0^t \|F(\pi^n(s)) - F(\pi(s))\|_2 ds + \frac{\kappa_9}{\sqrt{n}} \leq C(M) \int_0^t \|X^n(s)\|_2 ds + \frac{\kappa_9}{\sqrt{n}}.$$

Thus, from (77) and Gronwall’s lemma, on the set B_M^n , for all $n \geq 1$,

$$\sup_{0 \leq t \leq T} \|X^n(t)\|_2 \leq \kappa_{10} \left(\frac{1}{\sqrt{n}} + \|X^n(0)\|_2 + \sup_{0 \leq t \leq T} \|M^{n,c}(t)\|_2 \right) e^{C(M)T}. \tag{80}$$

From (69) and Doob’s inequality,

$$\sup_{n \in \mathbb{N}} \mathbb{E} \sup_{0 \leq t \leq T} \|M^{n,c}(t)\|_2^2 < \infty. \tag{81}$$

Also, by assumption, $X^n(0) \rightarrow x_0$ in ℓ_2 . Thus, for the given $\varepsilon > 0$, we can find α_0 such that, for all $\alpha \geq \alpha_0$,

$$\mathbb{P} \left(\sup_{0 \leq t \leq T} \|X^n(t)\|_2 \geq \frac{\alpha}{\sqrt{n_0}}, B_M^n \right) \leq \frac{\varepsilon}{2}.$$

Therefore, from (75) and (76) we have, for all $A \geq \alpha_0/\sqrt{n_0}$,

$$\sup_{n \in \mathbb{N}} \mathbb{P} \left(\sum_{j=0}^{n_0} |X_j^n(t)| > A \right) \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary we get (73). Thus, we have verified Theorem 3(a) for $\{X^n(t)\}_{n \in \mathbb{N}}$ for each $t \in [0, T]$.

We now consider Theorem 3(b). Namely, we show that, for every $\delta > 0$ and $t \in [0, T]$,

$$\lim_{n_0 \rightarrow \infty} \sup_{n \in \mathbb{N}} \mathbb{P} \left(\sum_{j=n_0}^{\infty} (X_j^n(t))^2 > \delta \right) = 0.$$

For this, it suffices to show that

$$\sup_{n \in \mathbb{N}} \mathbb{E} \sup_{0 \leq t \leq T} \sum_{j=0}^{\infty} j^2 (X_j^n(t))^2 < \infty. \tag{82}$$

Recalling that $X_j^n(t) = X_j^n(0) + A_j^{n,c}(t) + M_j^{n,c}(t)$ for each $j \in \mathbb{N}$, it follows that

$$\begin{aligned} \mathbb{E} \sup_{0 \leq t \leq T} \sum_{j=0}^{\infty} j^2 (X_j^n(t))^2 &\leq 3 \sup_{n \in \mathbb{N}} \sum_{j=0}^{\infty} j^2 (X_j^n(0))^2 + 3 \sup_{n \in \mathbb{N}} \mathbb{E} \sup_{0 \leq t \leq T} \sum_{j=0}^{\infty} j^2 (A_j^{n,c}(t))^2 \\ &\quad + 3 \sup_{n \in \mathbb{N}} \mathbb{E} \sup_{0 \leq t \leq T} \sum_{j=0}^{\infty} j^2 (M_j^{n,c}(t))^2. \end{aligned} \tag{83}$$

Using the definitions of $A^{n,c}$, A^n , and F in (24), (45), and (5), respectively, we can write

$$\begin{aligned} (A_j^{n,c}(t))^2 &\leq \kappa_{11} \left\{ \int_0^t n \left[\frac{(n-L)!}{n!} \zeta(j, n\pi^n(s)) - \bar{\zeta}(j, \pi(s)) \right]^2 ds \right. \\ &\quad + \int_0^t n \left[\frac{(n-L)!}{n!} \zeta(j-1, n\pi^n(s)) - \bar{\zeta}(j-1, \pi(s)) \right]^2 ds \\ &\quad \left. + \int_0^t n [\pi_j^n(s) - \pi_j(s)]^2 ds + \int_0^t n [\pi_{j+1}^n(s) - \pi_{j+1}(s)]^2 ds \right\}. \end{aligned} \tag{84}$$

From (47) and in a similar manner as in (40) we have

$$\begin{aligned} n \left[\frac{(n-L)!}{n!} \zeta(j, n\pi^n(s)) - \bar{\zeta}(j, \pi(s)) \right]^2 &\leq \kappa_{12} \{ (\pi_j^n(s))^2 + n [\bar{\zeta}(j, \pi^n(s)) - \bar{\zeta}(j, \pi(s))]^2 \} \\ &\leq \kappa_{13} \left\{ (\pi_j^n(s))^2 + n \sum_{i_1=0}^{k-1} \sum_{i_2=1}^{L-i_1} R_{j,i_1,i_2} (\pi^n(s), \pi(s))^2 \right\}, \end{aligned}$$

where R_{j,i_1,i_2} is as in (35). By (67) and the Cauchy–Schwarz inequality we now have

$$\begin{aligned} nR_{j,i_1,i_2} (\pi^n(s), \pi(s))^2 &\leq \kappa_{14} \left[(X_j^n(s))^2 + \pi_j(s) \left(\sum_{m=0}^{\infty} |X_m^n(s)| \right)^2 \right] \\ &\leq \kappa_{14} \left[(X_j^n(s))^2 + \pi_j(s) \left(\sum_{m=0}^{\infty} \frac{1}{m^2} \right) \sum_{m=0}^{\infty} m^2 (X_m^n(s))^2 \right]. \end{aligned}$$

Therefore,

$$n \left[\frac{(n-L)!}{n!} \zeta(j, n\pi^n(s)) - \bar{\zeta}(j, \pi(s)) \right]^2 \leq \kappa_{15} \left\{ (\pi_j^n(s))^2 + (X_j^n(s))^2 + \pi_j(s) \sum_{m=0}^{\infty} m^2 (X_m^n(s))^2 \right\}.$$

Combining this estimate with (72) and (84) yields

$$\begin{aligned} \mathbb{E} \sum_{j=0}^{\infty} j^2 (A_j^{n,c}(t))^2 &\leq \kappa_{16} \mathbb{E} \left\{ \int_0^t \sum_{j=1}^{\infty} j^2 \left[(X_{j-1}^n(s))^2 + (X_j^n(s))^2 + (X_{j+1}^n(s))^2 \right. \right. \\ &\quad \left. \left. + (\pi_j(s) + \pi_{j-1}(s)) \sum_{m=0}^{\infty} m^2 (X_m^n(s))^2 \right] ds \right\} + \kappa_{16} \\ &\leq \kappa_{17} \mathbb{E} \int_0^t \left(1 + \sum_{j=1}^{\infty} j^2 \pi_j(s) \right) \left(\sum_{j=1}^{\infty} j^2 (X_j^n(s))^2 \right) ds + \kappa_{17}. \end{aligned} \tag{85}$$

Additionally, it follows from (70) and (50) that

$$\mathbb{E} \sup_{0 \leq t \leq T} \sum_{j=0}^{\infty} j^2 M_j^{n,c}(t)^2 \leq \kappa'_{18} \mathbb{E} \sum_{j=0}^{\infty} j^2 \langle M_j^{n,c} \rangle_T \leq \kappa_{18} \int_0^T \left[1 + \mathbb{E} \sum_{j=0}^{\infty} j^2 \pi_j^n(s) \right] ds \leq \kappa_{19}.$$

Therefore, from (17), (83), and (85), for all $t \in [0, T]$,

$$\mathbb{E} \sup_{0 \leq t \leq T} \sum_{j=0}^{\infty} j^2 (X_j^n(t))^2 \leq \kappa_{20} + \kappa_{20} \int_0^T \left(1 + \sum_{j=1}^{\infty} j^2 \pi_j(t) \right) \mathbb{E} \sup_{0 \leq s \leq t} \left(\sum_{j=1}^{\infty} j^2 (X_j^n(s))^2 \right) dt.$$

From (50) and Fatou’s lemma, $\int_0^T \sum_{j=1}^{\infty} j^2 \pi_j(s) ds < \infty$, and, thus, by Gronwall’s lemma,

$$\sup_{n \in \mathbb{N}} \mathbb{E} \sup_{0 \leq t \leq T} \sum_{j=0}^{\infty} j^2 (X_j^n(t))^2 \leq \kappa_{19} \exp \left[\kappa_{20} \int_0^T \left(1 + \sum_{j=1}^{\infty} j^2 \pi_j(s) \right) ds \right] < \infty.$$

This proves (82) and verifies Theorem 3(b) for $\{X^n(t)\}_{n \in \mathbb{N}}$ for each $t \in [0, T]$. Thus, $\{X^n(t)\}_{n \in \mathbb{N}}$ is a tight sequence of ℓ_2 -valued random variables for every $t \in [0, T]$.

We now show that condition (A) of Theorem 4 holds for $\{X^n\}_{n \in \mathbb{N}}$. Since $X^n(t) = X^n(0) + A^{n,c}(t) + M^{n,c}(t)$ and we have shown the condition is satisfied by $\{M^{n,c}\}_{n \in \mathbb{N}}$, it suffices to show that the condition holds for $\{A^{n,c}\}_{n \in \mathbb{N}}$. Let $T_0, \eta, \varepsilon, \theta > 0, T_0 \leq T - \theta$, and suppose that $\{\tau_n\}_{n \in \mathbb{N}}$ is a family of stopping times such that $\tau_n \leq T_0$. From the definition of $A^{n,c}$ (cf. (24)) and (48), we have

$$\|A^{n,c}(\tau_n + \theta) - A^{n,c}(\tau_n)\|_2 \leq \int_{\tau_n}^{\tau_n + \theta} \sqrt{n} \|F(\pi^n(t)) - F(\pi(t))\|_2 dt + \frac{\kappa_{21}}{\sqrt{n}}, \tag{86}$$

where κ_{21} is independent of the choices of τ_n and T_0 . Fix $n_0 \in \mathbb{N}$ such that $\eta - \kappa_{21}/\sqrt{n_0} > 0$ and let $\eta' = \eta - \kappa_{21}/\sqrt{n_0}$. Recall κ_7 and κ_8 introduced in (78) and (79), and B_M^n introduced below (74). Select $M \in (0, \infty)$ large enough that $M > \kappa_7 \vee \kappa_8$ and (74) holds. Then, for all $n \geq n_0$,

$$\begin{aligned} & \mathbb{P} \left\{ \left\| \int_{\tau_n}^{\tau_n + \theta} \sqrt{n} [F(\pi^n(t)) - F(\pi(t))] dt \right\|_2 > \eta' \right\} \\ & \leq \mathbb{P} \left\{ \left\| \int_{\tau_n}^{\tau_n + \theta} \sqrt{n} [F(\pi^n(t)) - F(\pi(t))] dt \right\|_2 > \eta', B_M^n \right\} + \mathbb{P} \left\{ \sup_{0 \leq t \leq T} \sum_{j=0}^{\infty} j \pi_j^n(t) > M \right\} \\ & \leq \mathbb{P} \left\{ \left\| \int_{\tau_n}^{\tau_n + \theta} \sqrt{n} [F(\pi^n(t)) - F(\pi(t))] dt \right\|_2 > \eta', B_M^n \right\} + \frac{\varepsilon}{2}. \end{aligned} \tag{87}$$

It follows from the Lipschitz property of F proved in Lemma 4 that

$$\begin{aligned} & \mathbb{P} \left\{ \int_{\tau_n}^{\tau_n + \theta} \sqrt{n} \|F(\pi^n(t)) - F(\pi(t))\|_2 dt > \eta', B_M^n \right\} \\ & \leq \mathbb{P} \left\{ C(M) \int_{\tau_n}^{\tau_n + \theta} \|X^n(t)\|_2 dt > \eta', B_M^n \right\}. \end{aligned} \tag{88}$$

Recall from (80) that, for some $\tilde{C}(M) \in (0, \infty)$ on the set B_M^n ,

$$C(M) \sup_{0 \leq t \leq T} \|X^n(t)\|_2 \leq \tilde{C}(M) \left(1 + \sup_{0 \leq t \leq T} \|M^{n,c}(t)\|_2 \right).$$

Thus, from (88), Markov’s inequality, and (81), we have

$$\begin{aligned} & \mathbb{P}\left\{\int_{\tau_n}^{\tau_n+\theta} \sqrt{n}\|F(\pi^n(t)) - F(\pi(t))\|_2 dt > \eta', B_M^n\right\} \\ & \leq \mathbb{P}\left\{\theta \tilde{C}(M)\left(1 + \sup_{0 \leq t \leq T} \|M^{n,c}(t)\|_2\right) > \eta'\right\} \\ & \leq \frac{\theta \tilde{C}(M)(1 + \mathbb{E} \sup_{0 \leq t \leq T} \|M^{n,c}(t)\|_2)}{\eta'} \\ & \leq \theta \tilde{C}(M)\kappa_{22}. \end{aligned} \tag{89}$$

Combining (87) and (89) gives, whenever $\theta \leq \delta$,

$$\sup_{0 \leq \theta \leq \delta} \mathbb{P}\left\{\left\|\int_{\tau_n}^{\tau_n+\theta} \sqrt{n}[F(\pi^n(t)) - F(\pi(t))] dt\right\|_2 > \eta'\right\} \leq C(M)\kappa_{22}\delta + \frac{\varepsilon}{2}.$$

Selecting δ small enough that the first term on the right-hand side is less than $\varepsilon/2$ we have

$$\sup_{0 \leq \theta \leq \delta} \mathbb{P}\left\{\left\|\int_{\tau_n}^{\tau_n+\theta} \sqrt{n}[F(\pi^n(t)) - F(\pi(t))] dt\right\|_2 > \eta'\right\} \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \tag{90}$$

Therefore, combining (86) and (90) gives

$$\sup_{n \geq n_0} \sup_{0 \leq \theta \leq \delta} \mathbb{P}\{\|A^{n,c}(\tau_n + \theta) - A^{n,c}(\tau_n)\|_2 > \eta\} \leq \varepsilon$$

which shows that condition (A) of Theorem 4 is satisfied for $\{A^{n,c}\}_{n \in \mathbb{N}}$. Therefore, as discussed earlier, $\{X^n\}_{n \in \mathbb{N}}$ is a tight sequence of $\mathbb{D}([0, T]: \ell_2)$ -valued random variables and, thus, $\{(X^n, M^{n,c})\}_{n \in \mathbb{N}}$ is a tight sequence of $\mathbb{D}([0, t]: (\ell_2)^2)$ -valued random variables.

Finally, the C-tightness of $\{(X^n, M^{n,c})\}_{n \in \mathbb{N}}$ is immediate from the estimate

$$j_T(X^n) = j_T(M^{n,c}) \leq \frac{2 + 2k}{\sqrt{n}}, \quad n \in \mathbb{N},$$

and the fact that there are almost surely no simultaneous jumps, which follows as in the proof of Proposition 3. □

5.3. Convergence

In this section we give the proofs of Proposition 2 and Theorem 2. Since we have shown tightness of $\{(X^n, M^{n,c})\}_{n \in \mathbb{N}}$ in Section 5.2, all that remains in order to complete the proof of Theorem 2 is to characterize the weak limit points of this sequence of processes. This will be argued by showing that the limit point of any weakly convergent subsequence of $\{X^n\}_{n \in \mathbb{N}}$ will be a solution to SDE (11) and that uniqueness holds for (11) in an appropriate class, which will also prove Proposition 2. We begin by establishing a uniform integrability property for the sequence $\{M^{n,c}\}_{n \in \mathbb{N}}$.

Lemma 5. *Suppose that $\{\pi^n\}_{n \in \mathbb{N}}$ satisfies the conditions of Proposition 4. Then the sequence $\{\sup_{0 \leq t \leq T} \sum_{j=0}^{\infty} |M_j^{n,c}(t)|^2\}_{n \in \mathbb{N}}$ is uniformly integrable.*

Proof. It follows from the Cauchy–Schwarz and Burkholder–Davis–Gundy inequalities that

$$\begin{aligned} \sup_{n \in \mathbb{N}} \mathbb{E} \sup_{0 \leq t \leq T} \left(\sum_{j=0}^{\infty} |M_j^{n,c}(t)|^2 \right)^2 &\leq \sup_{n \in \mathbb{N}} \left(\sum_{m=0}^{\infty} \frac{1}{m^2} \right) \sum_{j=0}^{\infty} \mathbb{E} \sup_{0 \leq t \leq T} j^2 |M_j^{n,c}(t)|^4 \\ &\leq \kappa_1 \sup_{n \in \mathbb{N}} \sum_{j=0}^{\infty} j^2 \mathbb{E}[M_j^{n,c}(T)]^2. \end{aligned} \tag{91}$$

Recalling the definition of M^n from (22), for each j , $\mathbb{E}[M_j^{n,c}(T)]^2$ can be written as

$$\begin{aligned} \mathbb{E}[M_j^{n,c}(T)]^2 &= \mathbb{E} \left\{ \sum_{\ell \in \Sigma} \frac{1}{n} \langle e_j, \Delta_\ell \Delta_\ell^T e_j \rangle_2 N_\ell \left(\frac{n\lambda}{\binom{n}{L}} \int_0^T \prod_{i=0}^{\infty} \binom{n\pi_i^n(s)}{\rho_i(\ell)} ds \right) \right. \\ &\quad \left. + \frac{1}{n} \left[D_j \left(k \int_0^T n\pi_j^n(s) ds \right) + D_{j+1} \left(k \int_0^T n\pi_{j+1}^n(s) ds \right) \right] \right\}^2. \end{aligned}$$

We now consider the first term in the expectation on the right-hand side of the above equation corresponding to the stream of incoming jobs assigned to queues of length j ,

$$\mathbb{E} \left(\sum_{\ell \in \Sigma} \frac{1}{n} \langle e_j, \Delta_\ell \Delta_\ell^T e_j \rangle_2 N_\ell \left(\frac{n\lambda}{\binom{n}{L}} \int_0^T \prod_{i=0}^{\infty} \binom{n\pi_i^n(s)}{\rho_i(\ell)} ds \right) \right)^2.$$

First, consider the diagonal terms in the above sum. It follows from Fubini’s theorem that

$$\begin{aligned} &\mathbb{E} \sum_{\ell \in \Sigma} \frac{1}{n^2} \langle e_j, \Delta_\ell \Delta_\ell^T e_j \rangle_2^2 N_\ell \left(\frac{n\lambda}{\binom{n}{L}} \int_0^T \prod_{i=0}^{\infty} \binom{n\pi_i^n(s)}{\rho_i(\ell)} ds \right)^2 \\ &\leq \mathbb{E} \sum_{\ell \in \Sigma} \frac{1}{n^2} \langle e_j, \Delta_\ell \Delta_\ell^T e_j \rangle_2^2 \left(\left(\frac{n\lambda}{\binom{n}{L}} \int_0^T \prod_{i=0}^{\infty} \binom{n\pi_i^n(s)}{\rho_i(\ell)} ds \right)^2 \right. \\ &\quad \left. + \frac{n\lambda}{\binom{n}{L}} \int_0^T \prod_{i=0}^{\infty} \binom{n\pi_i^n(s)}{\rho_i(\ell)} ds \right). \end{aligned} \tag{92}$$

The fact that $\pi_i^n(s) \in [0, 1]$ and (31) imply that, for all $i \in \mathbb{N}_0$,

$$\frac{\lambda}{\binom{n}{L}} \int_0^T \prod_{i=0}^{\infty} \binom{n\pi_i^n(s)}{\rho_i(\ell)} ds \leq \frac{\lambda n^L}{\binom{n}{L}} \int_0^T \prod_{i=0}^{\infty} \frac{\pi_i^n(s)^{\rho_i(\ell)}}{\rho_i(\ell)!} ds \leq \frac{\lambda n^L}{\binom{n}{L}} T \leq \kappa_3. \tag{93}$$

Combining (93), (92), (27), and (32) yields

$$\begin{aligned} & \mathbb{E} \sum_{\ell \in \Sigma} \frac{1}{n^2} \langle e_j, \Delta_\ell \Delta_\ell^T e_j \rangle_2 N_\ell \left(\frac{n\lambda}{\binom{n}{L}} \int_0^T \prod_{i=0}^{\infty} \binom{n\pi_i^n(s)}{\rho_i(\ell)} ds \right)^2 \\ & \leq \frac{\lambda(\kappa_3 + 1)}{\binom{n}{L}} \mathbb{E} \int_0^T Z(j, n\pi^n(s)) ds \\ & \leq \frac{\lambda(\kappa_3 + 1)c_Z n^L}{\binom{n}{L}} \mathbb{E} \int_0^T (\pi_{j-1}^n(s) + \pi_j^n(s)) ds \\ & \leq \kappa_4 \mathbb{E} \int_0^T (\pi_{j-1}^n(s) + \pi_j^n(s)) ds. \end{aligned} \tag{94}$$

We now analyze the cross terms from the sum. It follows from the independence of N_ℓ and $N_{\ell'}$ that

$$\begin{aligned} & \mathbb{E} \sum_{\substack{\ell, \ell' \in \Sigma \\ \ell \neq \ell'}} \frac{1}{n^2} \langle e_j, \Delta_\ell \Delta_\ell^T e_j \rangle_2 \langle e_j, \Delta_{\ell'} \Delta_{\ell'}^T e_j \rangle_2 N_\ell \left(\frac{n\lambda}{\binom{n}{L}} \int_0^T \prod_{i=0}^{\infty} \binom{n\pi_i^n(s)}{\rho_i(\ell)} ds \right) \\ & \quad \times N_{\ell'} \left(\frac{n\lambda}{\binom{n}{L}} \int_0^T \prod_{i=0}^{\infty} \binom{n\pi_i^n(s)}{\rho_i(\ell')} ds \right) \\ & = \mathbb{E} \sum_{\substack{\ell, \ell' \in \Sigma \\ \ell \neq \ell'}} \frac{1}{n^2} \langle e_j, \Delta_\ell \Delta_\ell^T e_j \rangle_2 \langle e_j, \Delta_{\ell'} \Delta_{\ell'}^T e_j \rangle_2 \left(\frac{n\lambda}{\binom{n}{L}} \int_0^T \prod_{i=0}^{\infty} \binom{n\pi_i^n(s)}{\rho_i(\ell)} ds \right) \\ & \quad \times \left(\frac{n\lambda}{\binom{n}{L}} \int_0^T \prod_{i=0}^{\infty} \binom{n\pi_i^n(s)}{\rho_i(\ell')} ds \right). \end{aligned} \tag{95}$$

Combining (95), (27), and (32) yields

$$\begin{aligned} & \mathbb{E} \sum_{\substack{\ell, \ell' \in \Sigma \\ \ell \neq \ell'}} \frac{1}{n^2} \langle e_j, \Delta_\ell \Delta_\ell^T e_j \rangle_2 \langle e_j, \Delta_{\ell'} \Delta_{\ell'}^T e_j \rangle_2 N_\ell \left(\frac{n\lambda}{\binom{n}{L}} \int_0^T \prod_{i=0}^{\infty} \binom{n\pi_i^n(s)}{\rho_i(\ell)} ds \right) \\ & \quad \times N_{\ell'} \left(\frac{n\lambda}{\binom{n}{L}} \int_0^T \prod_{i=0}^{\infty} \binom{n\pi_i^n(s)}{\rho_i(\ell')} ds \right) \\ & \leq \mathbb{E} \frac{\lambda^2}{\left(\binom{n}{L}\right)^2} \int_0^T Z(j, n\pi^n(s))^2 ds \\ & \leq \mathbb{E} \frac{\lambda^2 n^{2L}}{\left(\binom{n}{L}\right)^2} \int_0^T (\pi_{j-1}^n(s) + \pi_j^n(s))^2 ds \\ & \leq \kappa_5 \mathbb{E} \int_0^T (\pi_{j-1}^n(s) + \pi_j^n(s)) ds. \end{aligned} \tag{96}$$

Similarly,

$$\begin{aligned} & \mathbb{E} \left(\frac{1}{n} \left[D_j \left(k \int_0^T n \pi_j^n(s) \, ds \right) + D_{j+1} \left(k \int_0^T n \pi_{j+1}^n(s) \, ds \right) \right] \right)^2 \\ & \leq \kappa_6 \mathbb{E} \int_0^T [\pi_j^n(s) + \pi_{j+1}^n(s)] \, ds. \end{aligned}$$

Combining this estimate with (94), (96), and (50) gives

$$\sup_{n \in \mathbb{N}} \sum_{j=0}^{\infty} j^2 \mathbb{E}[M_j^{n,c}(T)]^2 \leq \kappa_7 \sup_{n \in \mathbb{N}} \mathbb{E} \int_0^T \sum_{j=1}^{\infty} j^2 (\pi_{j-1}^n(s) + \pi_j^n(s) + \pi_{j+1}^n(s)) \, ds < \infty,$$

which, in view of (91), gives the desired uniform integrability. □

The following lemma together with (82) shows that any weak limit point X of $\{X^n\}_{n \in \mathbb{N}}$ satisfies $X(t) \in \ell_2$ for all $t \in [0, T]$ almost surely.

Lemma 6. *Let z^n and z be $\mathbb{D}([0, T]; \ell_2)$ -valued random variables such that*

$$\sup_{0 \leq t \leq T} \|z^n(t) - z(t)\|_2 \rightarrow 0 \quad \text{in probability as } n \rightarrow \infty.$$

Suppose that $\sup_{n \in \mathbb{N}} \mathbb{E} \sup_{0 \leq t \leq T} \sum_{j=0}^{\infty} j^2 (z_j^n(t))^2 < \infty$. Then $\sup_{0 \leq t \leq T} \sum_{j=0}^{\infty} j^2 (z_j(t))^2 < \infty$ almost surely and $\sup_{0 \leq t \leq T} \left| \sum_{j=0}^{\infty} z_j^n(t) - \sum_{j=0}^{\infty} z_j(t) \right| \rightarrow 0$ in probability.

Proof. Let $\kappa = \sup_{n \in \mathbb{N}} \mathbb{E} \sup_{0 \leq t \leq T} \sum_{j=0}^{\infty} j^2 [z_j^n(t)]^2$. Note that

$$\sup_{n \in \mathbb{N}} \mathbb{E} \sup_{0 \leq t \leq T} \sum_{j=0}^{\infty} |z_j^n(t)| \leq \left(\sum_{j=1}^{\infty} \frac{1}{j^2} \right)^{1/2} \sqrt{\kappa} < \infty.$$

Also, by Fatou’s lemma, $\mathbb{E} \sup_{0 \leq t \leq T} \sum_{j=0}^{\infty} j^2 (z_j(t))^2 \leq \kappa$ and so we have $\sup_{0 \leq t \leq T} \sum_{j=0}^{\infty} |z_j(t)| < \infty$ almost surely as well. Now

$$\begin{aligned} & \mathbb{E} \left[\sup_{0 \leq t \leq T} \left| \sum_{j=0}^{\infty} z_j^n(t) - \sum_{j=0}^{\infty} z_j(t) \right| \wedge 1 \right] \\ & \leq \mathbb{E} \left[\sup_{0 \leq t \leq T} \left| \sum_{j=0}^m z_j^n(t) - \sum_{j=0}^m z_j(t) \right| \wedge 1 \right] + \mathbb{E} \left[\sup_{0 \leq t \leq T} \left| \sum_{j=m+1}^{\infty} z_j^n(t) \right| \wedge 1 \right] \\ & \quad + \mathbb{E} \left[\sup_{0 \leq t \leq T} \left| \sum_{j=m+1}^{\infty} z_j(t) \right| \wedge 1 \right] \\ & \equiv T_1^m(n) + T_2^m(n) + T_3^m(n). \end{aligned}$$

Then, for $\kappa_1 \in (0, \infty)$,

$$(T_2^m(n))^2 \leq \left(\sum_{j=m+1}^{\infty} \frac{1}{j^2} \right) \kappa_1 \quad \text{and} \quad (T_3^m(n))^2 \leq \left(\sum_{j=m+1}^{\infty} \frac{1}{j^2} \right) \kappa_1.$$

The result now follows on first sending $n \rightarrow \infty$ and then $m \rightarrow \infty$. □

The following result that shows that $\Phi(t)$ is a trace class operator will be useful in characterizing the martingale term in the limiting diffusion. Note that, from definition (14), $\Phi(t)$ is a nonnegative operator.

Lemma 7. *For each $t \in [0, T]$, $\Phi(t)$ is a nonnegative trace class operator. Denote by $a(t)$ the nonnegative square root of $\Phi(t)$. Then $\int_0^T \|a(s)\|_{\text{HS}}^2 ds < \infty$.*

Proof. We first show that $\Phi(t)$ is a trace class operator. Since $\Phi(t)$ is nonnegative (and hence self-adjoint), it suffices to show that

$$\sum_{j=0}^{\infty} \langle e_j, \Phi(s)e_j \rangle_2 < \infty.$$

Using an argument similar to that used in the derivation of (30), we can write $\langle e_j, \Phi(s)e_j \rangle_2$ as

$$\langle e_j, \Phi(s)e_j \rangle_2 = \lambda L! \bar{Z}(j, \pi(s)) + k(\pi_j(s) + \pi_{j+1}(s)), \tag{97}$$

where the definition of \bar{Z} is analogous to Z , given as

$$\begin{aligned} &\bar{Z}(j, \pi(s)) \\ &:= \sum_{i_1=0}^{k-2} \frac{(\sum_{m=0}^{j-1} \pi_m(s))^{i_1}}{i_1!} \sum_{i_2=0}^{L-i_1} \frac{\pi_{j-1}(s)^{i_2}}{i_2!} \sum_{i_3=0}^{L-i_1-i_2} [i_2 \wedge (k - i_1)_+ - i_3 \wedge (k - i_1 - i_2)_+]^2 \\ &\quad \times \frac{\pi_j(s)^{i_3}}{i_3!} \frac{(\sum_{m=j+1}^{\infty} \pi_m(s))^{L-i_1-i_2-i_3}}{(L - i_1 - i_2 - i_3)!}. \end{aligned} \tag{98}$$

For completeness, the proof of (97) is provided in Appendix A.5. Using similar arguments as in (32) and (61), it is easy to see that there exists $c_{\bar{Z}} \in (0, \infty)$ such that, for all $j \in \mathbb{N}_0$,

$$\bar{Z}(j, \pi(s)) \leq c_{\bar{Z}}(\pi_{j-1}(s) + \pi_j(s)). \tag{99}$$

Once again, details on this step are given in Appendix A.5. From (97) and (99), it follows that there exists a $\kappa_1 \in (0, M)$ such that

$$\sum_{j=0}^{\infty} \langle e_j, \Phi(t)e_j \rangle_2 \leq \kappa_2 \sum_{j=0}^{\infty} [\pi_{j-1}(t) + \pi_j(t) + \pi_{j+1}(t)] \leq 3\kappa_1.$$

Therefore, $\Phi(t)$ is a trace class operator. Finally, note that

$$\int_0^T \|a(s)\|_{\text{HS}}^2 ds = \int_0^T \sum_{j=0}^{\infty} \langle a(s)e_j, a(s)e_j \rangle_2 ds = \int_0^T \sum_{j=0}^{\infty} \langle e_j, \Phi(s)e_j \rangle_2 ds \leq 3\kappa_1 T,$$

which completes the proof. □

We now proceed with the proofs of Proposition 2 and Theorem 2.

Proof of Proposition 2. The existence of an $(X(t))_{0 \leq t \leq T}$ as in the statement of Proposition 2 will be proved as part of Theorem 2. We now consider the second statement in Proposition 2, and let $(X(t))_{0 \leq t \leq T}$ and $(\tilde{X}(t))_{0 \leq t \leq T}$ be two $\{\mathcal{F}_t\}$ -adapted processes solving (12) with sample

paths in $\mathbb{C}([0, T]: \ell_2)$ such that $X(t) \in \tilde{\ell}_2$ and $\tilde{X}(t) \in \tilde{\ell}_2$ for all t almost surely. In order to show that $X(t) = \tilde{X}(t)$ for all $t \in [0, T]$ almost surely, it suffices to show the following Lipschitz property on G . There exists a $C \in (0, \infty)$ such that, for all $x, \tilde{x} \in \tilde{\ell}_2$,

$$\sup_{0 \leq t \leq T} \|G(x, \pi(t)) - G(\tilde{x}, \pi(t))\|_2 \leq C\|x - \tilde{x}\|_2. \tag{100}$$

Note that from (4), (5), and (16), for $j \in \mathbb{N}_0$ and $(x, r) \in \tilde{\ell}_2 \times \mathcal{S}$,

$$G_j(x, r) = \lambda L! [\xi_{j-1}^1(x, r) - \xi_j^1(x, r) + \xi_{j-1}^2(x, r) - \xi_j^2(x, r) + \xi_{j-1}^3(x, r) - \xi_j^3(x, r)] + k\xi_j^4(x), \tag{101}$$

where

$$\begin{aligned} \xi_j^1(x, r) &:= \sum_{i_1=0}^{k-1} i_1 \frac{(\sum_{m=0}^{j-1} r_m)^{i_1-1}}{i_1!} \sum_{i_2=1}^{L-i_1} [i_2 \wedge (k - i_1)] \frac{(r_j)^{i_2}}{i_2!} \frac{(\sum_{m=j+1}^\infty r_m)^{L-i_1-i_2}}{(L - i_1 - i_2)!} \sum_{m=0}^{j-1} x_m, \\ \xi_j^2(x, r) &:= \sum_{i_1=0}^{k-1} \frac{(\sum_{m=0}^{j-1} r_m)^{i_1}}{i_1!} \sum_{i_2=1}^{L-i_1} i_2 [i_2 \wedge (k - i_1)] \frac{(r_j)^{i_2-1}}{i_2!} \frac{(\sum_{m=j+1}^\infty r_m)^{L-i_1-i_2}}{(L - i_1 - i_2)!} x_j, \\ \xi_j^3(x, r) &:= \sum_{i_1=0}^{k-1} \frac{(\sum_{m=0}^{j-1} r_m)^{i_1}}{i_1!} \sum_{i_2=1}^{L-i_1} (L - i_1 - i_2) [i_2 \wedge (k - i_1)] \frac{(r_j)^{i_2}}{i_2!} \frac{(\sum_{m=j+1}^\infty r_m)^{L-i_1-i_2-1}}{(L - i_1 - i_2)!} \\ &\quad \times \sum_{m=j+1}^\infty x_m, \end{aligned}$$

and

$$\xi_j^4(x) = [x_{j+1} - x_j].$$

Also, let $\xi^i := (\xi_j^i)_{j=0}^\infty$ for $i = 1, 2, 3, 4$. Using the triangle inequality, it suffices to show that (100) holds with G replaced with ξ^i , $i = 1, 2, 3, 4$. Since $\pi(t) \in \mathcal{S}$ for all $t \in [0, T]$,

$$\begin{aligned} \sup_{0 \leq t \leq T} \|\xi^1(x, \pi(t)) - \xi^1(\tilde{x}, \pi(t))\|_2^2 &\leq \kappa'_1 \sup_{0 \leq t \leq T} \sum_{j=0}^\infty \pi_j(t)^2 \left[\sum_{m=0}^{j-1} x_m - \sum_{m=0}^{j-1} \tilde{x}_m \right]^2 \\ &\leq \kappa'_1 \sup_{0 \leq t \leq T} \sum_{j=0}^\infty j \pi_j(t) \|x - \tilde{x}\|_2^2 \\ &\leq \kappa_1 \|x - \tilde{x}\|_2^2, \end{aligned} \tag{102}$$

where the last inequality is from (79). Also,

$$\sup_{0 \leq t \leq T} \|\xi^2(x, \pi(t)) - \xi^2(\tilde{x}, \pi(t))\|_2^2 \leq \kappa_2 \sum_{j=0}^\infty [x_j - \tilde{x}_j]^2 = \kappa_2 \|x - \tilde{x}\|_2^2.$$

Using the fact that $\sum_{m=0}^{\infty} x_m = \sum_{m=0}^{\infty} \tilde{x}_m = 0$ and the calculation in (102),

$$\begin{aligned} \sup_{0 \leq t \leq T} \|\xi^3(x, \pi(t)) - \xi^3(\tilde{x}, \pi(t))\|_2^2 &\leq \kappa'_3 \sup_{0 \leq t \leq T} \sum_{j=0}^{\infty} \pi_j(t)^2 \left[\sum_{m=j+1}^{\infty} x_m - \sum_{m=j+1}^{\infty} \tilde{x}_m \right]^2 \\ &= \kappa'_3 \sup_{0 \leq t \leq T} \sum_{j=0}^{\infty} \pi_j(t)^2 \left[\sum_{m=0}^j \tilde{x}_m - \sum_{m=0}^j x_m \right]^2 \\ &\leq \kappa_3 \|x - \tilde{x}\|_2^2. \end{aligned}$$

Finally,

$$\|\xi^4(x) - \xi^4(\tilde{x})\|_2^2 \leq \sum_{j=0}^{\infty} [x_j - \tilde{x}_j]^2 + \sum_{j=0}^{\infty} [x_{j+1} - \tilde{x}_{j+1}]^2 \leq 2\|x - \tilde{x}\|_2^2.$$

Combining the above Lipschitz estimates for ξ^i , $i = 1, 2, 3, 4$, we have (100) and the result follows. □

We now proceed to the proof of Theorem 2.

Proof of Theorem 2. From Proposition 4, $\{(X^n, M^{n,c})\}_{n \in \mathbb{N}}$ is \mathbb{C} -tight in $\mathbb{D}([0, T]: (\ell_2)^2)$. Suppose that (X, M^c) is a weak limit of a subsequence of $\{(X^n, M^{n,c})\}_{n \in \mathbb{N}}$ (also indexed by $\{n\}$) given on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $m \in \mathbb{N}$, and let $\mathcal{H}: (\ell_2 \times \ell_2)^m \rightarrow \mathbb{R}$ be a bounded and continuous function. For $s \leq t \leq T$ and $0 \leq t_1 \leq \dots \leq t_m \leq s$, we let $\xi_i^n = (X^n(t_i), M^{n,c}(t_i))$ and $\xi_i = (X(t_i), M^c(t_i))$. Then, for all $j \in \mathbb{N}_0$,

$$\mathbb{E}\mathcal{H}(\xi_1, \dots, \xi_m)[M_j^c(t) - M_j^c(s)] = \lim_{n \rightarrow \infty} \mathbb{E}\mathcal{H}(\xi_1^n, \dots, \xi_m^n)[M_j^{n,c}(t) - M_j^{n,c}(s)] = 0,$$

where the first equality follows from the uniform integrability property proved in Lemma 5 and the second equality follows from the fact that $M^{n,c}$ is a martingale for each $n \in \mathbb{N}$. It follows that M^c is an $\{\mathcal{F}_t\}$ -martingale, where $\mathcal{F}_t = \sigma\{X(s), M^c(s), s \leq t\}$.

As was shown in (60),

$$\begin{aligned} \langle M_i^{n,c}, M_j^{n,c} \rangle(t) &= n \langle M_i^n, M_j^n \rangle(t) \\ &= \frac{\lambda}{\binom{n}{L}} \int_0^t Z(i, j, n\pi^n(s)) \, ds - k \int_0^t \mathbf{1}_{\{i=j+1\}} \pi_i^n(s) \, ds \\ &\quad - k \int_0^t \mathbf{1}_{\{i+1=j\}} \pi_j^n(s) \, ds + k \int_0^t \mathbf{1}_{\{i=j\}} (\pi_j^n(s) + \pi_{j+1}^n(s)) \, ds \end{aligned} \tag{103}$$

(see (30) and (58) for the definition of Z). Using similar arguments as in (58), we have the estimate

$$\langle e_i, \Phi(s)e_j \rangle_2 = \lambda L! \bar{Z}(i, j, \pi(s)) - k \mathbf{1}_{\{i+1=j\}} \pi_j(s) - k \mathbf{1}_{\{i=j+1\}} \pi_i(s) + k \mathbf{1}_{\{i=j\}} (\pi_j(s) + \pi_{j+1}(s)),$$

where, for $i < j$,

$$\begin{aligned} \bar{Z}(i, j, \pi(s)) := & \sum_{i_1=0}^{k-2} \frac{(\sum_{m=0}^{i-2} \pi_m(s))^{i_1}}{i_1!} \sum_{i_2=0}^{k-i_1-1} \frac{\pi_{i-1}(s)^{i_2}}{i_2!} \sum_{i_3=0}^{k-i_1-i_2-1} [i_2 - i_3] \frac{\pi_i(s)^{i_3}}{i_3!} \\ & \times \sum_{i_4=0}^{k-i_1-i_2-i_3-1} \frac{(\sum_{m=i+1}^{j-2} \pi_m(s))^{i_4}}{i_4!} \sum_{i_5=0}^{L-\sum_{n=1}^4 i_n} \frac{\pi_{j-1}(s)^{i_5} \mathbf{1}_{\{j>i+1\}}}{i_5!} \\ & \times \sum_{i_6=0}^{L-\sum_{n=1}^5 i_n} \left[(\mathbf{1}_{\{j=i+1\}}(i_3 - i_5) + i_5) \wedge \left(k - \sum_{n=1}^4 i_n \right)_+ - i_6 \wedge \left(k - \sum_{n=1}^5 i_n \right)_+ \right] \\ & \times \frac{\pi_j(s)^{i_6}}{i_6!} \frac{(\sum_{m=j+1}^\infty \pi_m(s))^{L-\sum_{n=1}^6 i_n}}{(L - \sum_{n=1}^6 i_n)!} \end{aligned}$$

for $i > j$, $\bar{Z}(i, j, \pi(s)) := \bar{Z}(j, i, \pi(s))$, and, for $i = j$, $\bar{Z}(j, j, \pi(s)) := \bar{Z}(j, \pi(s))$, where $\bar{Z}(j, r)$ is defined in (98). Using arguments similar to those used in (46) and (47), we can write

$$\left| Z(i, j, n\pi^n(s)) - \frac{n!}{(n-L)!} \bar{Z}(i, j, \pi^n(s)) \right| \leq \kappa_1 n^{L-1}.$$

From this, (97), (103), and the fact that $\pi^n \rightarrow \pi$ in probability, it follows that

$$\sup_{0 \leq t \leq T} \left| \langle M_i^{n,c}(t), M_j^{n,c}(t) \rangle - \int_0^t \langle e_i, \Phi(s)e_j \rangle_2 ds \right| \rightarrow 0$$

in probability. A similar argument as in Lemma 5 shows that $\{ \langle M_i^{n,c}, M_j^{n,c} \rangle_t \}_{n \in \mathbb{N}}$ is uniformly integrable for each $t \in [0, T]$ and $i, j \in \mathbb{N}_0$. Applying the above convergence and uniform integrability properties,

$$\begin{aligned} & \mathbb{E} \mathcal{H}(\xi_1, \dots, \xi_m) [\langle M_i^c, M_j^c \rangle_t - \langle M_i^c, M_j^c \rangle_s - \int_s^t \langle e_i, \Phi(u)e_j \rangle_2 du] \\ &= \lim_{n \rightarrow \infty} \mathbb{E} \mathcal{H}(\xi_1^n, \dots, \xi_m^n) \left[\langle M_i^{n,c}, M_j^{n,c} \rangle_t - \langle M_i^{n,c}, M_j^{n,c} \rangle_s - \int_s^t \langle e_i, \Phi(u)e_j \rangle_2 du \right] \\ &= 0. \end{aligned}$$

Also, from Lemma 5 and Fatou’s lemma, $\mathbb{E} \sup_{0 \leq t \leq T} \sum_{j=0}^\infty |M_j^c(t)|^2 < \infty$. Thus, $M^c := (M_j^c)_{j \in \mathbb{N}_0}$ is a collection of square-integrable $\{\mathcal{F}_t\}$ -martingales with

$$\langle M_i^c, M_j^c \rangle(t) = \int_0^t \langle e_i, \Phi(s)e_j \rangle_2 ds, \quad t \in [0, T].$$

From Theorem 8.2 of [10], it now follows that there is an ℓ_2 -cylindrical Brownian motion $\{(W_t(h))_{0 \leq t \leq T} : h \in \ell_2\}$ on some extension $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbb{P}}, \{\bar{\mathcal{F}}_t\})$ of the filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\})$ such that

$$M^c(t) = \int_0^t a(s) dW(s). \tag{104}$$

Recall the representation of X^n in terms of $A^{n,c}$ and $M^{n,c}$ from (23). We now argue that, together with X^n and $M^{n,c}$, $A^{n,c}(\cdot)$ converges to $\int_0^t G(X(s), \pi(s)) ds$ in $\mathbb{D}([0, T]: \ell_2)$ in distribution as $n \rightarrow \infty$ (along the chosen subsequence). The definition of $A^{n,c}$ in (24) and the estimate in (48) imply that

$$\sup_{0 \leq t \leq T} \left\| A^{n,c}(t) - \int_0^t \sqrt{n} [F(\pi^n(s)) - F(\pi(s))] ds \right\|_2 \leq \frac{\kappa_2}{\sqrt{n}}. \tag{105}$$

For $r, \tilde{r} \in \mathcal{S}$ such that $(r - \tilde{r}) \in \tilde{\ell}_2$, the i th component of $F(r) - F(\tilde{r})$ can be written as

$$\begin{aligned} [F(r) - F(\tilde{r})]_i &= \int_0^1 \frac{\partial}{\partial u} F_i(ru + (1-u)\tilde{r}) du \\ &= \int_0^1 G_i((r - \tilde{r}), ru + (1-u)\tilde{r}) du \\ &= G_i(r - \tilde{r}, \tilde{r}) + \int_0^1 [G_i((r - \tilde{r}, ru + (1-u)\tilde{r}) - G_i(r - \tilde{r}, \tilde{r})] du. \end{aligned}$$

Therefore, observing that $cG_i(x, r) = G_i(cx, r)$ for $c \in \mathbb{R}$ and $(x, r) \in \tilde{\ell}_2 \times \mathcal{S}$, and noting from (82) that $X^n(s) \in \tilde{\ell}_2$ for every $s \in [0, T]$ almost surely, we can write

$$\sqrt{n}[F(\pi^n(s)) - F(\pi(s))]_i = G_i(X^n(s), \pi(s)) + R_i^n(s), \tag{106}$$

where

$$R_i^n(s) = \int_0^1 [G_i(X^n(s), \pi^n(s)u + (1-u)\pi(s)) - G_i(X^n(s), \pi(s))] du.$$

Thus,

$$\sqrt{n}[F(\pi^n(s)) - F(\pi(s))] = G(X^n(s), \pi(s)) + R^n(s),$$

where $R^n(s) := (R_i^n(s))_{i \in \mathbb{N}_0}$. We now show that $\int_0^T \|R^n(s)\|_2 ds \rightarrow 0$ in probability as $n \rightarrow \infty$. Since $\sum_{m=0}^j X_m^n(s) = -\sum_{m=j+1}^\infty X_m^n(s)$, it follows from (36) that, for $r, \tilde{r} \in \mathcal{S}$,

$$\begin{aligned} &\|\xi^i(X^n(s), r) - \xi^i(X^n(s), \tilde{r})\|_2^2 \\ &\leq \kappa'_3 \sum_{j=0}^\infty \left(\sum_{m=0}^j |X_m^n(s)| \right)^2 \left[[r - \tilde{r}]_j^2 + \tilde{r}_j \left(\sum_{i=0}^{j-1} [r - \tilde{r}]_i \right)^2 + \tilde{r}_j \left(\sum_{i=j+1}^\infty [r - \tilde{r}]_i \right)^2 \right] \\ &\leq \kappa_3 \left(\sum_{j=0}^\infty j^2 |X_j^n(s)|^2 \right) \sum_{j=0}^\infty [j\tilde{r}_j \|r - \tilde{r}\|_2^2 + [r - \tilde{r}]_j^2] \end{aligned}$$

for $i = 1, 2, 3$. The triangle inequality, (101), and the observation that $\sup_{0 \leq s \leq T} \sum_{j=0}^\infty j\pi_j(s) < \infty$ (see (79)) then imply that

$$\|G(X^n(s), \pi^n(s)u + (1-u)\pi(s)) - G(X^n(s), \pi(s))\|_2^2 \leq \kappa_3 \left(\sum_{j=0}^\infty j^2 |X_j^n(s)|^2 \right) \|\pi^n(s) - \pi(s)\|_2^2.$$

Since $\sup_{0 \leq s \leq T} \|\pi^n(s) - \pi(s)\|_2 \rightarrow 0$ in probability and, from (82), $\sup_{n \in \mathbb{N}} \mathbb{E} \sup_{0 \leq s \leq T} \sum_{j=0}^\infty j^2 |X_j^n(s)|^2 < \infty$, it follows that

$$\sup_{0 \leq u \leq 1} \sup_{0 \leq s \leq T} \|G(X^n(s), \pi^n(s)u + (1-u)\pi(s)) - G(X^n(s), \pi(s))\|_2 \rightarrow 0$$

in probability as $n \rightarrow \infty$, and, thus,

$$\int_0^T \|R_n(s)\|_2 ds \rightarrow 0 \quad \text{in probability.} \tag{107}$$

In view of (105), (106), and (107), it now suffices to show that, along the subsequence,

$$\left(X^n, M^{n,c}, \int_0^\cdot G(X^n(s), \pi(s)) ds \right) \Rightarrow \left(X, \bar{M}, \int_0^\cdot G(X(s), \pi(s)) ds \right)$$

in $\mathbb{D}([0, T]: (\ell_2)^3)$. By appealing to the Skorokhod representation theorem we can assume without loss of generality that $(X^n, M^{n,c})$ converges almost surely in $\mathbb{D}([0, T]: (\ell_2)^2)$ to (X, \bar{M}) . From (82) and Fatou’s lemma we also have

$$\sup_{0 \leq t \leq T} \sum_{j=0}^\infty j^2 (X_j(t))^2 < \infty \quad \text{almost surely.}$$

Also, since $\sum_{j=0}^\infty X_j^n(t) = 0$ for all $t \in [0, T]$ and $n \in \mathbb{N}$, by Lemma 6 and (82), we have $\sum_{j=0}^\infty X_j(t) = 0$ for all $t \in [0, T]$ almost surely as well. It then follows that $X^n(t), X(t) \in \tilde{\ell}_2$ for all $t \in [0, T]$ almost surely for all $n \in \mathbb{N}$. From the Lipschitz property in (100), it now follows that, as $n \rightarrow \infty$,

$$\int_0^T \|G(X^n(s), \pi(s)) - G(X(s), \pi(s))\|_2 ds \leq C \int_0^T \|X^n(s) - X(s)\|_2 ds \rightarrow 0,$$

which proves the desired convergence. Together with (23) and representation (104), it follows that the limit point (X, M^c) satisfies

$$X(t) = x_0 + \int_0^t G(X(s), \pi(s)) ds + \int_0^t a(s) dW(s)$$

almost surely for all $t \in [0, T]$. Since $X(t) \in \tilde{\ell}_2$ for all $t \in [0, T]$ almost surely, this in particular proves the existence part of Proposition 2. Finally, the uniqueness part of Proposition 2 (which was established earlier in this section) now says that X^n converges in distribution along the full sequence to the unique weak solution of (11) with values in $\tilde{\ell}_2$. The result follows. \square

6. Numerical results

In this section we present some simulation results comparing the prelimit n -server system with results of the corresponding law of large number and central limit approximations. We consider a system with $n = 10\,000$ servers. For all combinations of L and k in the set $\{(L, k) \in \mathbb{N} \times \mathbb{N}: 2 \leq L \leq 5, k < L\}$, we simulate 1000 realizations of both the n -server system and the diffusion approximation given in Theorem 2 using parameters $T = 10, \lambda = 0.9$, and $c = 1$.

TABLE 1: Empty queue coverage rate.

k	L (%)			
	2	3	4	5
1	95.1	96.3	97.7	95.9
2	–	96.5	95.3	95.6
3	–	–	96.8	97.5
4	–	–	–	97.1

TABLE 2: Large queue coverage rate.

k	L (%)			
	2	3	4	5
1	97.1	100	100	100
2	–	94.9	95.6	100
3	–	–	96.7	96.4
4	–	–	–	95.0

TABLE 3: Mean queue length coverage rate.

k	L (%)			
	2	3	4	5
1	95.2	94.8	94.8	95.4
2	–	94.7	92.9	94.9
3	–	–	96.8	95.1
4	–	–	–	94.8

Note that, since the limiting processes are infinite-dimensional, we must truncate to a finite-dimensional approximation in order to perform simulations. In our numerical approximations, we truncate to the first 20 coordinates. All computations were performed in MATLAB[®]. A numerical ODE solver (ode45) was used to compute the ODE corresponding to the law of large number limit. The limit diffusion was simulated using Euler’s method with step sizes of 0.1. The realizations of the diffusion were used to create 95% confidence intervals for the following metrics at time T ; the number of empty queues, the number of ‘large’ queues (queues with more than 5 jobs), and the mean queue length. The coverage rates (i.e. the proportion of the n -server system simulations which fall within the 95% confidence interval estimated by the diffusion approximation) can be found in Tables 1, 2, and 3. The diffusion approximation based confidence intervals for the first and third cases contain approximately 95% of the n -server simulated observations, as desired. However, the results given in Table 2 appear to be less satisfactory. This is not surprising since the events corresponding to large queues are rare, and, thus, their probabilities are harder to estimate.

TABLE 4: Average simulation times for (a) the finite system and (b) the limit diffusion.

k	L			
	2	3	4	5
1	22.6	23.8	25.4	19.2
2	-	39.1	38.0	33.1
3	-	-	44.5	45.5
4	-	-	-	57.4

(a)

k	L			
	2	3	4	5
1	0.29	0.50	0.79	0.79
2	-	2.4	3.7	4.6
3	-	-	6.0	10.0
4	-	-	-	16.3

(b)

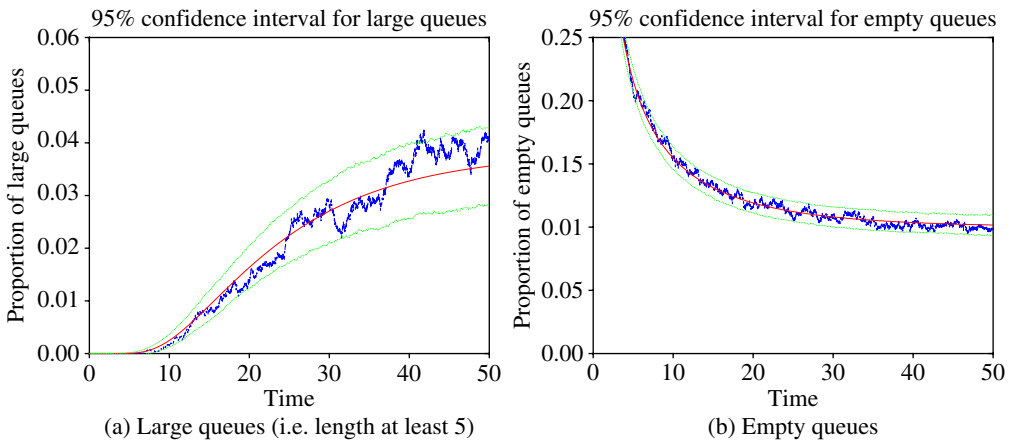


FIGURE 1: Empirical 95% confidence intervals obtained from the diffusion approximation, effectively capturing the fluctuations around the LLN ODE

In Figure 1 we present two graphs showing that the diffusion approximation captures the fluctuations of the underlying processes around the limiting ODE given via the LLN. We consider the supermarket model, i.e. $(L, k) = (2, 1)$ with $\lambda = 0.9$, $T = 50$, $c = 1$, and $n = 10\,000$ servers. Figures 1(a) and 1(b) present the results for large queues (i.e. queues of length at least 5) and empty queues, respectively. In each, the solid line represents the numerical solution to the limiting ODE, the dashed line represents the a single simulation of the underlying system, and the dotted lines represent empirical 95% confidence intervals obtained from the diffusion approximation. Namely, 1000 realizations of the diffusion were computed and the 2.5 and 97.5 percentiles were taken at each time point. The figures show that both the LLN and diffusion approximation are doing a good job of approximating the dynamics of the finite system over time.

The goal of this paper was to develop reliable approximations of the n -server system that are much quicker to simulate. In Table 4 we present the average time (in seconds) required to simulate one trial of the finite system and diffusion approximation. As is seen from these tables, the time required to simulate the diffusion approximations is substantially smaller than for the underlying n -server jump-Markov process. In addition, increasing n will further increase the amount of time required to simulate the n -server system. Indeed, $n = 10\,000$ is a small number compared to the size of typical data centers and server farms that have machines

which number in the hundreds of thousands. The point at which it becomes quicker to use the diffusion approximation will depend heavily on the system parameters. Indeed, simulation results indicate that this point occurs in the mid-hundreds for $(L, k) = (2, 1)$ while it is in the mid-thousands for $(L, k) = (5, 4)$. A further caveat is that this number will depend on the efficiency of implementations of both the numerical approximation and the simulation scheme.

Appendix A. Auxiliary results

A.1. Criterion for tightness of Hilbert-valued random variables

The following theorem gives sufficient conditions for tightness of a sequence of random variables taking values in a (possibly infinite-dimensional) Hilbert space. For a proof see Corollary 2.3.1 of [21].

Theorem 3. *Let \mathbb{H} be a separable Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and complete orthonormal system $\{e_i\}_{i=1}^\infty$. Suppose that $\{Y_n\}_{n \in \mathbb{N}}$ is a sequence of \mathbb{H} -valued random variables satisfying the following conditions:*

- (a) *for each $n_0 \in \mathbb{N}$, $\lim_{A \rightarrow \infty} \sup_{n \in \mathbb{N}} \mathbb{P}(\max_{1 \leq i \leq n_0} \langle Y_n, e_i \rangle^2 > A) = 0$;*
- (b) *for every $\delta > 0$, $\lim_{n_0 \rightarrow \infty} \sup_{n \in \mathbb{N}} \mathbb{P}(\sum_{j=n_0}^\infty \langle Y_n, e_j \rangle^2 > \delta) = 0$.*

Then $\{Y_n\}_{n \in \mathbb{N}}$ is a tight sequence of \mathbb{H} -valued random variables.

A.2. Criterion for tightness of RCLL processes

The following theorem gives a criterion for tightness of a sequence of RCLL processes with values in a Polish space; see [26].

Theorem 4. *Let \mathbb{S} be a Polish space and let $\{Y_n\}_{n \in \mathbb{N}}$ be a sequence of $\mathbb{D}([0, T]: \mathbb{S})$ -valued $\{\mathcal{F}_t^n\}$ -semimartingales satisfying the following conditions:*

- (T₁) *$\{Y_n(t)\}_{n \in \mathbb{N}}$ is tight for every t in a dense subset of $[0, T]$;*
- (A) *for each $\varepsilon > 0$, $\eta > 0$, and $T_0 \in [0, T - \varepsilon]$, there exists a $\delta > 0$ and n_0 with the property that, for every collection of stopping times $(\tau_n)_{n \in \mathbb{N}}$ (τ_n being an $\mathcal{F}_t^n := \sigma\{Y_n(s) : s \leq t\}$ -stopping time) with $\tau_n \leq T_0$,*

$$\sup_{n \geq n_0} \sup_{0 \leq \theta \leq \delta} \mathbb{P}\{d(Y_n(\tau_n + \theta), Y_n(\tau_n)) \geq \eta\} \leq \varepsilon,$$

where $d(\cdot, \cdot)$ is the distance on \mathbb{S} .

Then $\{Y_n\}_{n \in \mathbb{N}}$ is tight in $\mathbb{D}([0, T]: \mathbb{S})$.

A.3. Hilbert–Schmidt and trace class operators

Here we collect some elementary facts about trace class and Hilbert–Schmidt operators. We refer the reader to [34] for details. For a separable Hilbert space \mathbb{H} (with inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$), let $\mathcal{L}(\mathbb{H})$ be the collection of all bounded linear operators on \mathbb{H} . An operator $A \in \mathcal{L}(\mathbb{H})$ is called nonnegative if $\langle u, Au \rangle \geq 0$ for all $u \in \mathbb{H}$. Such an operator is called trace class if, for a complete orthonormal system (CONS) $\{e_i\}$ in \mathbb{H} , $\sum_i \langle Ae_i, e_i \rangle < \infty$ in which case the quantity is finite (and is the same) for every CONS $\{e_i\}$. An operator $A \in \mathcal{L}(\mathbb{H})$ is called Hilbert–Schmidt if there exists a CONS $\{e_i\}$ in \mathbb{H} such that $\sum_j \langle Ae_j, Ae_j \rangle = \sum_j \|Ae_j\|^2 < \infty$. In

that case, this quantity is the same for all CONS $\{e_i\}$ and its square root is called the Hilbert–Schmidt norm of A , denoted as $\|A\|_{\text{HS}}$. For a nonnegative operator $A \in \mathcal{L}(\mathbb{H})$, there is a unique nonnegative $B \in \mathcal{L}(\mathbb{H})$ referred to as the nonnegative square root of A such that $B^2 = A$. If A is a trace class operator then B is a Hilbert–Schmidt operator.

A.4. Cylindrical Brownian motion

A collection of continuous real stochastic processes $\{(W_t(h))_{0 \leq t \leq T} : h \in \ell_2\}$ given on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\})$ is called a ℓ_2 -cylindrical Brownian motion if, for every $h \in \ell_2$, $(W_t(h))_{0 \leq t \leq T}$ is a $\{\mathcal{F}_t\}$ -Brownian motion with variance $\|h\|_2^2$ and, for $h, k \in \ell_2$,

$$\langle W(h), W(k) \rangle_t = \langle h, k \rangle_{2t}, \quad 0 \leq t \leq T.$$

For a measurable map a from $[0, T]$ to the space of Hilbert–Schmidt operators from ℓ_2 to ℓ_2 such that $\int_0^T \|a(s)\|_{\text{HS}}^2 ds < \infty$, we denote by $\int_0^t a(s) dW(s)$ the ℓ_2 -valued martingale defined as the limit of

$$\sum_{i=1}^n \sum_{j=1}^n \phi_i \int_0^t \langle \phi_i, a(s)\phi_j \rangle_2 dW_s(\phi_j) \quad \text{as } n \rightarrow \infty,$$

where $\{\phi_i\}_{i \in \mathbb{N}}$ is a CONS in ℓ_2 . We refer the reader to Chapter 4 of [10] regarding the fact that the limit exists and is independent of the choice of the CONS.

A.5. Proofs of (97) and (99)

Recalling the definition of $\Phi(s)$ in (14) we can write

$$\begin{aligned} \langle e_j, \Phi(s)e_j \rangle_2 &= \lambda L! \sum_{\ell \in \Sigma} \langle e_j, \Delta_\ell \Delta_\ell^T e_j \rangle_2 \prod_{i=0}^\infty \frac{\pi_i(t)^{\rho_i(\ell)}}{\rho_i(\ell)!} \\ &\quad + k \sum_{i=1}^\infty \langle e_j, (e_{i-1} - e_i)(e_{i-1} - e_i)^T e_j \rangle_2 \pi_i(t). \end{aligned} \tag{A.1}$$

Recalling from (28) that, for $\ell \in \Sigma_j(i_1, i_2, i_3)$, (29) holds, we have from the decomposition,

$$\Sigma = \bigcup_{i_1=0}^{k-2} \bigcup_{i_2=0}^{L-i_1} \bigcup_{i_3=0}^{L-i_1-i_2} \Sigma_j(i_1, i_2, i_3),$$

and

$$\begin{aligned} &\lambda L! \sum_{\ell \in \Sigma} \langle e_j, \Delta_\ell \Delta_\ell^T e_j \rangle_2 \prod_{i=0}^\infty \frac{\pi_i(t)^{\rho_i(\ell)}}{\rho_i(\ell)!} \\ &= \lambda \sum_{i_1=0}^{k-2} \sum_{i_2=0}^{L-i_1} \sum_{i_3=0}^{L-i_1-i_2} \sum_{\ell \in \Sigma_j(i_1, i_2, i_3)} \langle e_j, \Delta_\ell \Delta_\ell^T e_j \rangle_2 \binom{L}{i_1, i_2, i_3, L-i_1-i_2-i_3} \\ &\quad \times \binom{i_1}{\rho_1(\ell), \dots, \rho_{j-2}(\ell)} \binom{L-i_1-i_2-i_3}{\rho_{j+1}(\ell), \rho_{j+2}(\ell), \dots} \prod_{i=0}^\infty \pi_i(t)^{\rho_i(\ell)}. \end{aligned}$$

$$\begin{aligned}
 &= \lambda \sum_{i_1=0}^{k-2} \sum_{i_2=0}^{L-i_1} \sum_{i_3=0}^{L-i_1-i_2} \binom{L}{i_1, i_2, i_3, L-i_1-i_2-i_3} \\
 &\quad \times \left(\sum_{m=0}^{j-2} \pi_m(s) \right)^{i_1} \pi_{j-1}(s)^{i_2} \pi_j(s)^{i_3} \left(\sum_{m=j+1}^{\infty} \pi_m(s) \right)^{L-i_1-i_2-i_3} \\
 &\quad \times [i_2 \wedge (k-i_1)_+ - i_3 \wedge (k-i_1-i_2)_+]^2 \\
 &= \bar{Z}(j, \pi(s)),
 \end{aligned}$$

where the second equality follows from the multinomial theorem. Furthermore, from (25), the second sum in (A.1) is

$$k \sum_{i=1}^{\infty} \langle e_j, (e_{i-1} - e_i)(e_{i-1} - e_i)^T e_j \rangle_2 \pi_i(t) = k(\pi_j(t) + \pi_{j+1}(t)).$$

Combining the last two equations with (A.1) gives (97).

For (99), note that

$$\begin{aligned}
 \bar{Z}(j, \pi(s)) &\leq \sum_{i_1=0}^{k-1} \frac{(\sum_{m=0}^{j-2} \pi_m(s))^{i_1}}{i_1!} \sum_{i_2=0}^{L-i_1} \frac{(\pi_{j-1}(s))^{i_2}}{i_2!} \\
 &\quad \times \sum_{i_3=0}^{L-i_1-i_2} k^2 \mathbf{1}_{\{i_2 \vee i_3 > 0\}} \frac{(\pi_j(s))^{i_3}}{i_3!} \frac{(\sum_{m=j+1}^{\infty} \pi_m(s))^{L-i_1-i_2-i_3}}{(L-i_1-i_2-i_3)!} \\
 &\leq \sum_{i_1=0}^{k-1} \sum_{i_2=0}^{L-i_1} \sum_{i_3=0}^{L-i_1-i_2} k^2 \mathbf{1}_{\{i_2 \vee i_3 > 0\}} (\pi_{j-1}(s))^{i_2} (\pi_j(s))^{i_3} \\
 &\leq c_{\bar{Z}} (\pi_{j-1}(s) + \pi_j(s))
 \end{aligned}$$

for some $c_{\bar{Z}} \in (0, \infty)$, where the second inequality follows from the fact that all factorials are greater than 1 and π is a probability measure. This proves (99).

Acknowledgements

The research was supported in part by the National Science Foundation (DMS-1016441, DMS-1305120, DMS-1814894), the Army Research Office (W911NF-14-1-0331), and DARPA (W911NF-15-2-0122).

References

- [1] AGHAJANI, R. AND RAMANAN, K. (2015). Ergodicity of an spde associated with a many-server queue. Preprint. Available at <https://arxiv.org/abs/1512.02929v1>.
- [2] ANDERSON, D. F. AND KURTZ, T. G. (2015). *Stochastic Analysis of Biochemical Systems*, Vol. 1. Springer, Cham.
- [3] ANTUNES, N., FRICKER, C., ROBERT, P. AND TIBI, D. (2008). Stochastic networks with multiple stable points. *Ann. Prob.* **36**, 255–278.
- [4] ATAR, R. AND SHIFRIN, M. (2014). An asymptotic optimality result for the multiclass queue with finite buffers in heavy traffic. *Stoch. Systems* **4**, 556–603.

- [5] BELL, S. L. AND WILLIAMS, R. J. (2001). Dynamic scheduling of a system with two parallel servers in heavy traffic with resource pooling: asymptotic optimality of a threshold policy. *Ann. Appl. Prob.* **11**, 608–649.
- [6] BRAMSON, M., LU, Y. AND PRABHAKAR, B. (2012). Asymptotic independence of queues under randomized load balancing. *Queueing Systems* **71**, 247–292.
- [7] BUDHIRAJA, A. AND FRIEDLANDER, E. (2017). Diffusion approximations for controlled weakly interacting large finite state systems with simultaneous jumps. *Ann. Appl. Prob.* **28**, 204–249.
- [8] BUDHIRAJA, A. AND GHOSH, A. P. (2012). Controlled stochastic networks in heavy traffic: convergence of value functions. *Ann. Appl. Prob.* **22**, 734–791.
- [9] BUDHIRAJA, A. GHOSH, A. P. AND LEE, C. (2011). Ergodic rate control problem for single class queueing networks. *SIAM J. Control Optimization* **49**, 1570–1606.
- [10] DA PRATO, G. AND ZABCZYK, J. (2014). *Stochastic Equations in Infinite Dimensions*. Cambridge University Press.
- [11] DAI, J. G. AND LIN, W. (2008). Asymptotic optimality of maximum pressure policies in stochastic processing networks. *Ann. Appl. Prob.* **18**, 2239–2299.
- [12] DECREUSEFOND, L. AND MOYAL, P. (2008). A functional central limit theorem for the $M/GI/\infty$ queue. *Ann. Appl. Prob.* **18**, 2156–2178.
- [13] ESCHENFELDT, P. AND GAMARNIK, D. (2015). Join the shortest queue with many servers. The heavy traffic asymptotics. Preprint. Available at <https://arxiv.org/abs/1502.00999>.
- [14] ETHIER, S. N. AND KURTZ, T. G. (2009). *Markov Processes: Characterization and Convergence*, Vol. 282. John Wiley, Hoboken.
- [15] FRIEDLANDER, E. (2018). Steady-state behavior of some load balancing mechanisms in cloud storage systems. Preprint. Available at <https://arxiv.org/abs/1801.02979>.
- [16] GRAHAM, C. (2000). Chaoticity on path space for a queueing network with selection of the shortest queue among several. *J. Appl. Prob.* **37**, 198–211.
- [17] HARRISON, J. M. (1988). Brownian models of queueing networks with heterogeneous customer populations. In *Stochastic Differential Systems, Stochastic Control Theory and Applications*, Springer, New York, pp. 147–186.
- [18] IKEDA, N. AND WATANABE, S. (1989). *Stochastic Differential Equations and Diffusion Processes*, 2nd edn. North-Holland Publishing, Amsterdam.
- [19] JACOD, J. AND SHIRYAEV, A. N. (1987). *Limit Theorems for Stochastic Processes*. Springer, Heidelberg.
- [20] JOFFE, A. AND MÉTIVIER, M. (1986). Weak convergence of sequences of semimartingales with applications to multitype branching processes. *Adv. Appl. Prob.* **18**, 20–65.
- [21] KALLIANPUR, G. AND XIONG, J. (1995). *Stochastic Differential Equations in Infinite-Dimensional Spaces* (IMS Lecture Notes Monogr. Ser. **26**). Institute of Mathematical Statistics, Hayward, CA.
- [22] KANG, H.-W. KURTZ, T. G. AND POPOVIC, L. (2014). Central limit theorems and diffusion approximations for multiscale Markov chain models. *Ann. Appl. Prob.* **24**, 721–759.
- [23] KASPI, H. AND RAMANAN, K. (2013). SPDE limits of many-server queues. *Ann. Appl. Prob.* **23**, 145–229.
- [24] KURTZ, T. G. (1971). Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. *J. Appl. Prob.* **8**, 344–356.
- [25] KURTZ, T. G. (1980). Representations of Markov processes as multiparameter time changes. *Ann. Prob.* **8**, 682–715.
- [26] KURTZ, T. G. (1981). *Approximation of Population Processes*. SIAM, Philadelphia.
- [27] KUSHNER, H. J. (2001). *Heavy Traffic Analysis of Controlled Queueing and Communication Networks* (Appl. Math. (New York) **47**). Springer, New York.
- [28] LI, B., RAMAMOORTHY, A. AND SRIKANT, R. (2016). Mean-field-analysis of coding versus replication in cloud storage systems. In *IEEE INFOCOM 2016*, IEEE, Piscataway, NJ.
- [29] LIN, S. AND COSTELLO, D. J. (2004). *Error Control Coding*. Pearson Education India.
- [30] MÉTIVIER, M. (1982). *Semimartingales: A Course on Stochastic Processes*, Vol. 2. Walter De Gruyter, Berlin.
- [31] MITZENMACHER, M. (2001). The power of two choices in randomized load balancing. *IEEE Trans. Parallel Distributed Systems* **12**, 1094–1104.
- [32] MUKHERJEE, D., BORST, S. C., VAN LEEUWAARDEN, J. S. H. AND WHITING, P. A. (2015). Universality of load balancing schemes on diffusion scale. Preprint. Available at <https://arxiv.org/abs/1510.02657v1>.
- [33] REED, J. AND TALREJA, R. (2015). Distribution-valued heavy-traffic limits for the $G/GI/\infty$ queue. *Ann. Appl. Prob.* **25**, 1420–1474.
- [34] REED, M. AND SIMON, B. (1980). *Functional Analysis*, Vol. 1. Academic Press.
- [35] STOLYAR, A. L. (2015). Pull-based load distribution in large-scale heterogeneous service systems. *Queueing Systems* **80**, 341–361.
- [36] SZNITMAN, A.-S. (1991). Topics in propagation of chaos. In *Ecole d'été de probabilités de Saint-Flour XIX* (Lecture Notes Math. **1464**). Springer, Berlin, pp. 165–251.
- [37] VVEDENSKAYA, N. D., DOBRUSHIN, R. L. AND KARPELEVICH, F. I. (1996). Queueing system with selection of the shortest of two queues: an asymptotic approach. *Probl. Peredachi Inf.* **32**, 20–34.
- [38] WHITT, W. (2002). *Stochastic-Process Limits: an Introduction to Stochastic-Process Limits and Their Application to Queues*. Springer, New York.