

NEAR EQUILIBRIUM FLUCTUATIONS FOR SUPERMARKET MODELS WITH GROWING CHOICES

BY SHANKAR BHAMIDI^{1,2,a}, AMARJIT BUDHIRAJA^{1,2,b} AND MIHEER DEWASKAR^{1,2,c}

¹Department of Statistics and Operations Research, University of North Carolina

²Department of Statistical Science, Duke University, ^abhamidi@email.unc.edu, ^bbudhiraja@email.unc.edu,
^cmiheer.dewaskar@duke.edu

We consider the supermarket model in the usual Markovian setting where jobs arrive at rate $n\lambda_n$ for some $\lambda_n > 0$, with n parallel servers each processing jobs in its queue at rate 1. An arriving job joins the shortest among $d_n \leq n$ randomly selected service queues. We show that when $d_n \rightarrow \infty$ and $\lambda_n \rightarrow \lambda \in (0, \infty)$, under natural conditions on the initial queues, the state occupancy process converges in probability, in a suitable path space, to the unique solution of an infinite system of constrained ordinary differential equations parametrized by λ . Our main interest is in the study of fluctuations of the state process about its near equilibrium state in the critical regime, namely when $\lambda_n \rightarrow 1$. Previous papers, for example, (Stoch. Syst. **8** (2018) 265–292) have considered the regime $\frac{d_n}{\sqrt{n \log n}} \rightarrow \infty$ while the objective of the current work is to develop diffusion approximations for the state occupancy process that allow for all possible rates of growth of d_n . In particular, we consider the three canonical regimes (a) $d_n/\sqrt{n} \rightarrow 0$; (b) $d_n/\sqrt{n} \rightarrow c \in (0, \infty)$ and, (c) $d_n/\sqrt{n} \rightarrow \infty$. In all three regimes, we show, by establishing suitable functional limit theorems, that (under conditions on λ_n) fluctuations of the state process about its near equilibrium are of order $n^{-1/2}$ and are governed asymptotically by a one-dimensional Brownian motion. The forms of the limit processes in the three regimes are quite different; in the first case, we get a linear diffusion; in the second case, we get a diffusion with an exponential drift; and in the third case we obtain a reflected diffusion in a half space. In the special case $d_n/(\sqrt{n \log n}) \rightarrow \infty$, our work gives alternative proofs for the universality results established in (Stoch. Syst. **8** (2018) 265–292).

CONTENTS

1. Introduction	2084
1.1. Organization of the paper	2087
1.2. Notation and setup	2087
2. Main results	2088
3. Poisson representation of state processes	2096
4. The law of large numbers	2098
4.1. Uniqueness of fluid limit equations	2098
4.2. Tightness and limit point characterization	2100
4.3. Completing the proof of LLN	2104
5. Properties of the near fixed point	2105
6. Preliminary estimates under diffusion scaling	2106
7. Proof of Theorem 2.2	2111
8. Proof of Theorem 2.3	2119
9. Proof of Theorem 2.4	2123
Appendix A: Proofs of results in Section 5	2132

Received June 2020; revised April 2021.

MSC2020 subject classifications. 90B15, 60F17, 90B22, 60C05.

Key words and phrases. Power of choice, join-the-shortest-queue, fluid limits, heavy traffic, Halfin–Whitt, load balancing, diffusion approximations, Skorohod problem, reflected diffusions, functional limit theorems.

A.1. Proof of Lemma 5.1	2132
A.2. Proof of Corollary 5.2	2133
A.3. Proof of Corollary 5.3	2133
A.4. Proof of Lemma 5.4	2133
A.5. Proof of Corollary 5.5	2134
A.6. Proof of Lemma 5.6	2135
A.7. Proof of Lemma 5.7	2135
A.8. Proof of Lemma 5.8	2135
Appendix B: Proof of Lemma 6.8	2135
Acknowledgments	2136
Funding	2137
References	2137

1. Introduction. In this work, we study the asymptotic behavior of a family of randomized load balancing schemes for many server systems. Consider a processing system with n parallel queues in which each queue's jobs are processed by the associated server at rate 1. Jobs arrive at rate $n\lambda_n$ and join the shortest queue among d_n randomly selected queues (without replacement), with $d_n \in [n] \doteq \{1, \dots, n\}$. The interarrival times and service times are mutually independent exponential random variables. This queuing system with the above described “join-the-shortest-queue amongst chosen queues” discipline is often denoted as $\text{JSQ}(d_n)$ and frequently referred to as the supermarket model (cf. [14, 23–25, 27, 31] and references therein). Note that when $d_n = n$ the above description corresponds to a policy where an incoming job joins the shortest of all queues in the system (see, e.g., [9]). The case $d_n = 1$ is the other extreme corresponding to incoming jobs joining a randomly chosen queue in which case the system is equivalent to one with n independent $M/M/1$ queues with arrival rate λ_n and service rate 1. The case $d_n = d$ where $d > 1$ is a fixed positive integer is sometimes also referred to as the power-of- d scheme. The analysis of $\text{JSQ}(d_n)$ schemes has been a focus of much recent research motivated by problems from large scale service centers, cloud computing platforms and data storage and retrieval systems (see, e.g., [1, 3, 8, 15, 26, 32, 33, 35]). The influential works of Mitzenmacher [29, 30] and Vvedenskaya et al. [36] showed by considering a fluid scaling that increasing d from 1 to 2 leads to significant improvement in performance in terms of steady-state queue length distributions in that the tails of the asymptotic steady-state distributions decay exponentially when $d = 1$ and superexponentially when $d = 2$. Limit theorems under a diffusion scaling for the $\text{JSQ}(d)$ system, with a fixed d , can be found in [7, 10]. Although $\text{JSQ}(d)$ for a fixed $d \geq 2$ leads to significant improvements over $\text{JSQ}(1)$, as observed in [12, 13], no fixed value of d provides the optimal waiting time properties of the join-the-shortest-queue system (i.e., $\text{JSQ}(n)$). See the survey [35] for an overview of the progress in this general area. This motivates the study of asymptotic behavior of a $\text{JSQ}(d)$ system in which the number of choices d increase with system size, namely n . Such an asymptotic study is the goal of this work.

The paper [31] studied the law of large numbers (LLN) behavior of a $\text{JSQ}(d_n)$ system, under a standard scaling, when $d_n \rightarrow \infty$. The precise result of [31] is as follows. For $i \in \mathbb{N}_0 \doteq \{0, 1, 2, \dots\}$ and $t \in [0, \infty)$, let $G_{n,i}(t)$ denote the fraction of queues with at least i customers at time t in the n th system. Note that $G_{n,0}(t) = 1$ for all $t \geq 0$. We will call $\mathbf{G}_n(t) \doteq \{G_{n,i}(t) : i > 0\}$ the state occupancy process. This process has sample paths in the space of summable nonnegative sequences. More precisely, for $p \geq 1$, let ℓ_p be the space of real sequences $\mathbf{x} \doteq (x_1, x_2, \dots)$ such that $\|\mathbf{x}\|_p \doteq (\sum_{i=1}^{\infty} |x_i|^p)^{1/p} < \infty$. Let

$$(1.1) \quad \ell_1^\downarrow \doteq \{\mathbf{x} \in \ell_1 : x_i \geq x_{i+1} \text{ and } x_i \in [0, 1] \text{ for all } i \in \mathbb{N}\}$$

be the space of nonincreasing sequences in ℓ_1 with values in $[0, 1]$, equipped with the topology generated by $\|\cdot\|_1$. Note that $\ell_1^\downarrow$ is a closed subset of ℓ_1 , and hence is a Polish space.

Then, whenever $\|\mathbf{G}_n(0)\|_1 < \infty$ a.s., it can be shown that $\{\mathbf{G}_n(t) : t \geq 0\}$ is a stochastic process with sample paths in $\mathbb{D}([0, \infty) : \ell_1^\downarrow)$ (the space of right-continuous functions with left limits from $[0, \infty)$ to $\ell_1^\downarrow$ equipped with the usual Skorohod topology); see Section 3. The paper [31] shows the following two facts under the assumption that $\mathbf{G}_n(0)$ converges in probability to some $\mathbf{r} \in \ell_1^\downarrow$:

- (a) When $d_n = n$ and $\lambda_n \rightarrow \lambda \in (0, \infty)$, $\mathbf{G}_n$ is a tight sequence in $\mathbb{D}([0, \infty) : \ell_1^\downarrow)$ and every weak limit point satisfies a certain set of “fluid limit equations” (see [31], Theorem 5, and equations (2.4)–(2.5) in the current work);
- (b) When d_n is an arbitrary sequence growing to ∞ and $\lambda_n \rightarrow \lambda \in (0, 1)$, then the statements in (a) hold once more for $\mathbf{G}_n$.

The current work begins by revisiting the above LLN results from [31]. In Theorem 2.1 of this work, we show that, when $\mathbf{G}_n(0)$ converges in probability to $\mathbf{r}$, for arbitrary sequences $d_n \rightarrow \infty$ and $\lambda_n \rightarrow \lambda \in (0, \infty)$, $\mathbf{G}_n$ converges in probability in $\mathbb{D}([0, \infty) : \ell_1^\downarrow)$ to a continuous trajectory $\mathbf{g}$ in $\ell_1^\downarrow$ that is characterized as the *unique* solution of an infinite system of constrained ordinary differential equations (ODE) (see (2.2) in Proposition 2.1). Using standard properties of the Skorohod reflection map, we observe in Remark 2.3 that a continuous trajectory in $\ell_1^\downarrow$ solves the fluid limit equations of [31] if and only if it solves (2.2). This together with Proposition 2.1 proves that the fluid limit equations in [31] in fact have a unique solution. In this manner, we complete and strengthen the result from [31]. Our proof of the LLN result is quite different from the arguments in [31]. The latter are based on sophisticated ideas of separation of time scales and weak convergence of measure valued processes from [16] to handle the convergence for $d_n = n$, and certain coupling techniques to treat the general case when $d_n < n$ and $d_n \rightarrow \infty$. In contrast, our approach is more direct and uses martingale estimates and well-known characterization properties of solutions of Skorohod problems (see, e.g., proof of Lemma 4.7).

Our main goal in this work is to study diffusion approximations for $\mathbf{G}_n$ in the heavy traffic regime, namely when $\lambda_n \rightarrow 1$. In the case when $d_n = n$ (JSQ(n) system), this problem has been studied in [9]. Their basic result is as follows. Suppose $d_n = n$ and $\sqrt{n}(1 - \lambda_n) \rightarrow \beta > 0$. Consider the unit vector $\mathbf{e}_1 = (1, 0, \dots)$ in ℓ_2 . Then under conditions on $\mathbf{G}_n(0)$, the process $\mathbf{Y}_n(\cdot) \doteq \sqrt{n}(\mathbf{G}_n(\cdot) - \mathbf{e}_1)$ converges in distribution in $\mathbb{D}([0, \infty) : \ell_2)$ to a continuous stochastic process $\mathbf{Y} = (Y_1, Y_2, \dots)$, described in terms of a one-dimensional Brownian motion, for which $Y_i = 0$ for $i > r$ for some $r \in \mathbb{N}$ (which depends on the conditions assumed on $\mathbf{G}_n(0)$). Specifically, when $r = 2$, the pair Y_1, Y_2 is given as a two-dimensional diffusion in the half-space $(-\infty, 0] \times \mathbb{R}$ with oblique reflection in the direction $(-1, 1)^t$ at the boundary $\{0\} \times \mathbb{R}$. (For the form of the limit in the general case, see Corollary 2.7). In [31] this result is extended to the case where $d_n < n$ and $\frac{d_n}{\sqrt{n \log n}} \rightarrow \infty$. Under the same assumptions on the initial condition as in [9], it is shown in [31] that $\mathbf{Y}_n$ converges to the same limit process as for the case $d_n = n$. The proof, as for the LLN result, proceeded by constructing a suitable coupling between a JSQ(d_n) and JSQ(n) system. The paper [31] also argued that when $\frac{d_n}{\sqrt{n \log n}} \rightarrow 0$, the process $\mathbf{Y}_n$ cannot be tight, and thus in this regime the above diffusion approximation cannot hold.

Our objective in this work is to develop diffusion approximations for $\mathbf{G}_n$ in the critical regime (i.e., when $\lambda_n \rightarrow 1$ in a suitable manner) that allow for possibly a slower growth of d_n than that permitted by the results in [31]. In fact, in contrast to [9, 31], we will prove diffusion limits when $d_n \rightarrow \infty$ in an arbitrary manner for choices of $\lambda_n \rightarrow 1$ constrained by the exact growth rate of d_n . See Table 1 for an overview of the regimes of (λ_n, d_n) that we cover, along with those covered by previous work. In the special case that $\frac{d_n}{\sqrt{n \log n}} \rightarrow \infty$, we will recover the results of [31] with a different proof. In order to motivate the type of limit theorems we

seek, we begin by observing that the centering e_1 used in the definition of Y_n is a stationary point of the fluid limit given in (2.2) with $\lambda = 1$, and thus the results of [9] and [31] give information on fluctuations of the state process G_n about this stationary point. However, e_1 is not the only stationary point of (2.2) (when $\lambda = 1$) and in fact this ODE has uncountably many fixed points given by $f_k^\gamma \doteq \sum_{j=1}^k e_j + \gamma e_{k+1} = (1, \dots, 1, \gamma, 0, 0, \dots) \in \ell_1^\downarrow$ for $k \in \mathbb{N}$ and $\gamma \in [0, 1)$, where e_j is the j th unit vector in ℓ_2 (with 1 at the j th coordinate and zeroes elsewhere). All of these stationary points arise in a natural fashion. Indeed, it turns out that the evolution of the state process G_n can be described via the equation (see Remark 3.1)

$$G_n(t) = G_n(0) + \int_0^t [a_n(G_n(s)) - b(G_n(s))] ds + M_n(t),$$

where M_n is a (infinite dimensional) martingale converging to zero in probability (see Lemma 4.1) and a_n, b are certain maps from $\ell_1^\downarrow$ to ℓ_1 (see Remark 3.1 for details). Thus for large n , trajectories of G_n will be close to solutions of the infinite dimensional ODE

$$\dot{g}_n = a_n(g_n) - b(g_n),$$

where $\dot{g}_n$ denotes the derivative of g_n . This equation has a unique stationary point μ_n , which is introduced in Definition 2. The fixed point μ_n corresponds to the point in the state space $\ell_1^\downarrow$ at which the inflow rate equals the outflow rate in the n th system, and thus it is of interest to explore system behavior in the neighborhood of this point. Since G_n is approximated by g_n (over any compact time interval), one can loosely interpret μ_n as a *near fixed point* of the state process G_n . Furthermore, it can be shown (see Remark 2.5(iv)) that, if $d_n \rightarrow \infty$ and $\lambda_n \rightarrow 1$ in a suitable manner, μ_n can converge to any specified fixed point f_k^γ of (2.2), and thus every fixed point of (2.2) arises from μ_n in a suitable asymptotic regime. In order to explore fluctuations of G_n close to different fixed points of (2.2), it is then natural to study the asymptotic behavior of

(1.2)
$$Z_n(t) \doteq \sqrt{n}(G_n(t) - \mu_n), \quad t \geq 0.$$

We note that in the regime considered in [31] where $\frac{d_n}{\sqrt{n} \log n} \rightarrow \infty$ and $\sqrt{n}(1 - \lambda_n) \rightarrow \alpha > 0$, $\sqrt{n}(e_1 - \mu_n) \rightarrow \alpha e_1$ and so in this case the asymptotic behavior of Z_n can be read off from that of Y_n (see Corollary 2.7 and Remark 2.8(v)). However, in general $\sqrt{n}(e_1 - \mu_n)$ (and more generally, $\sqrt{n}(f_k^\gamma - \mu_n)$) may not be bounded and so the asymptotic behavior of Z_n and Y_n may be very different.

In this work, we obtain limit theorems for Z_n as $d_n \rightarrow \infty$ in an arbitrary fashion and $\lambda_n \rightarrow 1$ in a suitable manner. Specifically, in Theorems 2.2, 2.3 and 2.4 we consider the three cases: (a) $d_n/\sqrt{n} \rightarrow 0$, (b) $d_n/\sqrt{n} \rightarrow c \in (0, \infty)$ and (c) $d_n/\sqrt{n} \rightarrow \infty$, respectively. In all three regimes, we consider initial conditions $G_n(0)$ such that for some $r \in \mathbb{N}$, $\sqrt{n}(G_{n,j}(0) - \mu_{n,j})$ converge to zero in probability for all $j > r$ and in each case (under conditions on λ_n), we obtain a limit process driven by a one-dimensional Brownian motion with continuous sample paths in ℓ_2 , which has all but finitely many coordinates 0. In particular, when $r = 2$ in the second and the third case and $r = k + 2$ for some $k \in \mathbb{N}$ in the first case (and d_n, λ_n depend on k in a suitable fashion), one can describe the limit through a two-dimensional diffusion driven by a one-dimensional Brownian motion. The form of this two-dimensional process in the three regimes is quite different; in the first case, we get a linear diffusion (i.e., the drift is of the form $b(y) = Ay$ for, $y \in \mathbb{R}^2$ and some 2×2 matrix A); in the second case, we get a diffusion with an exponential drift, and in the third case we obtain a reflected diffusion in the half space $(-\infty, \alpha] \times \mathbb{R}$ for some $\alpha \geq 0$.

Although the limit processes in Theorems 2.2 and 2.3 are quite different from those obtained in [10] and [31], the limit in Theorem 2.4 has a similar form (in that it is a reflected

diffusion in a half-space) as in the above papers. However here as well there are some differences. In particular, depending on how λ_n approaches 1, the reflection occurs at a different barrier $\alpha \in (0, \infty)$; in fact, $\alpha = \infty$ is possible as well in which case there is no reflection. Furthermore, recall that Z_n is defined by centering about μ_n . In general, $\sqrt{n}(\mu_n - e_1)$ will diverge, and thus the process Y_n considered in the above cited papers may not converge in this regime. However, as noted previously, when d_n grows sufficiently fast, namely $\frac{d_n}{\sqrt{n} \log n} \rightarrow \infty$ the process Y_n will indeed converge and in that case we recover the result in [31] (in fact a slight strengthening in that the drift parameter in Corollary 2.7 is allowed to be 0). In addition, Theorem 2.4 also covers the case $\frac{d_n}{\sqrt{n} \log n} \rightarrow c \in (0, \infty)$ and situations where $\lambda_n = 1 + O(n^{-1/2})$ (see Remark 2.8 (iv)). In such settings, once more both Z_n and Y_n converge and the limit of the latter has the same form as in [9, 31].

As is observed in Remarks 2.6 and 2.8, under conditions of Theorem 2.3 or Theorem 2.4, μ_n must converge to the fixed point $e_1 = f_1^0$. In contrast, Theorem 2.2 allows for a range of asymptotic behavior for μ_n . In particular, under the conditions of the theorem, with suitable λ_n, d_n, μ_n can converge to the fixed point f_k^0 for an arbitrary $k \in \mathbb{N}$ (see [5] for a similar observation). Here, k may then be considered as the average time spent by a job in the system, since asymptotically almost all (cf. [5]) queues will have length k under these conditions. In such a setting, the first $k - 1$ coordinates of the limit process are essentially 0 (see Theorem 2.2 for a precise statement) and the k th coordinate is the first one to exhibit stochastic variability. Thus a rather novel asymptotic behavior for the JSQ(d_n) system emerges when d_n approaches ∞ at significantly slower rates than those considered in [31] and λ_n approach 1 in a suitable manner (in relation to d_n).

1.1. Organization of the paper. Section 2 contains all our main results. The remaining sections starting with Section 3 contain proofs of the main results.

We now make some comments on the proofs of Theorems 2.2–2.4. The starting point is a convenient semimartingale representation for the centered state process Z_n in (6.1). In the study of the behavior of the drift term in this decomposition, an important ingredient is an analysis of the asymptotic properties of the near fixed point μ_n , and the asymptotic behavior of the function β_n (see Definition 1) in $O(n^{-\frac{1}{2}})$ sized neighborhoods around the coordinates of μ_n . This behavior, which is different in the three regimes considered above, determines the asymptotics of the drift $A_n(Z_n(s)) - b(Z_n(s))$, where $A_n(z) \doteq \sqrt{n}(a_n(\mu_n + n^{-1/2}z) - a_n(\mu_n))$. Properties of μ_n are also key in arguing that, in all three cases, under our conditions, $(Z_{n,r+1}, \dots)$ converges to 0 in probability in $\mathbb{D}([0, \infty) : \ell_2)$ (see Lemma 6.4). The rest of the work is in characterizing the asymptotics of the finite dimensional process $(Z_{n,1}, \dots, Z_{n,r})$. For this study, the three regimes require different approaches. In particular, Theorem 2.2 hinges on a detailed understanding of the asymptotic behavior of a tridiagonal matrix function $Q_n(s)$ (see, e.g., Lemmas 7.4 and 7.6); Theorem 2.3 requires an analysis of a stochastic differential equation with an exponential drift term (in particular, the drift does not satisfy the usual growth conditions); and Theorem 2.4 is based on a careful study of excursions of the prelimit processes above the limiting reflecting barrier and properties of Skorohod maps in order to characterize the reflection properties of the limit process.

1.2. Notation and setup. For $m \geq 1$, let $[m] \doteq \{1, 2, \dots, m\}$. We will denote finite-dimensional vectors in $\mathbb{R}^m$ as $\vec{x}, \vec{y}$, etc. and $\langle \vec{x}, \vec{y} \rangle$ will denote the standard inner product. Transpose of a vector $\vec{v}$ will be written as $\vec{v}^t$. The standard basis vectors in $\mathbb{R}^m$ will be denoted by $\vec{e}_i$ for $i = 1, 2, \dots, m$. Also, $\|\vec{x}\| \doteq \sqrt{\langle \vec{x}, \vec{x} \rangle}$ will denote the usual Euclidean norm.

We will often use bold symbols such as $\mathbf{x} \doteq (x_1, x_2, \dots)$ to denote a infinite dimensional vector or function. For $p \in \{1, 2, \dots, \infty\}$, let $\|\mathbf{x}\|_p \doteq (\sum_{i=1}^{\infty} |x_i|^p)^{1/p}$ denote the p -norm on

the space of infinite sequences and $\ell_p \doteq \{\mathbf{x} \in \mathbb{R}^\infty \mid \|\mathbf{x}\|_p < \infty\}$. Let $\ell_1^\downarrow$ be as in (1.1), which is a Polish space under $\|\cdot\|_1$. For $k \in \mathbb{N}$, let $\mathbf{f}_k \doteq (1, 1, \dots, 1, 0, 0, \dots) \in \ell_1^\downarrow$ denote the vector with first k indices equal to 1, and $\mathbf{e}_k \doteq (0, \dots, 0, 1, 0, \dots) \in \ell_1$ denote the vector with 1 in the k th coordinate. For any $k \in \mathbb{N}$ and $\gamma \in [0, 1)$, write $\mathbf{f}_k^\gamma \doteq \mathbf{f}_k + \gamma \mathbf{e}_{k+1} \in \ell_1^\downarrow$. For $\mathbf{z} = (z_1, z_2, \dots) \in \mathbb{R}^\infty$ and $r \in \mathbb{N}$, let $\mathbf{z}_{r+} \doteq (z_{r+1}, z_{r+2}, \dots) \in \mathbb{R}^\infty$ denote the vector shifted by r steps. Similar notation will be used for functions and processes with values in $\mathbb{R}^\infty$.

For a Polish space $\mathbb{S}$ and the interval $\mathcal{I} = [0, T]$ for $T > 0$ or $\mathcal{I} = [0, \infty)$, denote by $\mathbb{C}(\mathcal{I} : \mathbb{S})$ (resp. $\mathbb{D}(\mathcal{I} : \mathbb{S})$) the space of continuous functions (resp., right continuous functions with left limits) from $\mathcal{I}$ to $\mathbb{S}$, endowed with the topology defined by uniform convergence on compact sets (resp., Skorokhod topology). For $h \in \mathbb{D}([0, T] : \mathbb{R})$, $\mathbf{g} \in \mathbb{D}([0, T] : \ell_p)$ and $t \in [0, T]$, denote the size of the largest jump upto time t by $J_t(h) \doteq \sup_{s \in [0, t]} |h(s) - h(s-)|$ and $J_t(\mathbf{g}) \doteq \sup_{s \in [0, t]} \|\mathbf{g}(s) - \mathbf{g}(s-)\|_p$, and the supremum norms up to time t by $|h|_{*,t} \doteq \sup_{s \in [0, t]} |h(s)|$ and $\|\mathbf{g}\|_{p,t} \doteq \sup_{s \in [0, t]} \|\mathbf{g}(s)\|_p$. If h is absolutely continuous on $[0, T]$, then $\dot{h}(t)$ (or sometimes $dh(t)/dt$) will denote the derivative of h at $t \in [0, T]$ (defined almost everywhere).

We will use $\mathbb{I}_{\{cond\}}$ to denote the indicator function that takes the value 1 if *cond* is true, otherwise it takes the value 0. We will denote by id the identity map, $\text{id}(t) \doteq t$, on $[0, T]$ or $[0, \infty)$. We use $\mathbf{P}$ and $\mathbf{E}$ to denote the probability and expectation operators, respectively. For $x, y \in \mathbb{R}$, $x \wedge y$ denotes the minimum and $x \vee y$ the maximum of x and y , respectively. For any $x \in \mathbb{R}$, $x^+ \doteq x \vee 0$ and $x^- \doteq (-x) \vee 0$. We use $\xrightarrow{P}$ and $\Rightarrow$ to denote convergence in probability and convergence in distribution respectively on an appropriate Polish space which will depend on the context. For a sequence of real valued random variables $(X_n, n \geq 1)$, we write $X_n = o_p(b_n)$ when $|X_n|/b_n \xrightarrow{P} 0$ as $n \rightarrow \infty$. For nonnegative functions $h(\cdot), g(\cdot)$, we write $h(n) = O(g(n))$ when $h(n)/g(n)$ is uniformly bounded, and $h(n) = o(g(n))$ (or $h(n) \ll g(n)$) when $\lim_{n \rightarrow \infty} h(n)/g(n) = 0$. We write $h(n) \sim g(n)$ if $h(n)/g(n) \rightarrow 1$ as $n \rightarrow \infty$.

2. Main results. Recall the process $\mathbf{G}_n$ from Section 1. Our first result gives a law of large numbers (LLN) for the process $\mathbf{G}_n$ as $n \rightarrow \infty$. In order to state this result, we begin by recalling the one-dimensional Skorokhod map (cf. [18], Section 3.6.C, [19]) with reflecting barrier at $\alpha \in \mathbb{R}$. For $\alpha \in \mathbb{R}$ and $h \in \mathbb{D}([0, \infty) : \mathbb{R})$ with $h(0) \leq \alpha$, define $\Gamma_\alpha(h), \hat{\Gamma}_\alpha(h) \in \mathbb{D}([0, \infty) : \mathbb{R})$ as

$$(2.1) \quad \Gamma_\alpha(h)(t) = h(t) - \sup_{s \in [0, t]} (h(s) - \alpha)^+, \quad \hat{\Gamma}_\alpha(h)(t) = \sup_{s \in [0, t]} (h(s) - \alpha)^+.$$

The map Γ_α (and sometimes the pair $(\Gamma_\alpha, \hat{\Gamma}_\alpha)$) is referred to as the one-dimensional Skorokhod map (with reflection at α). We note that the above map is a modification of the usual definition to account for the fact that in our case reflection occurs from above (in order to prevent h from exceeding the level α). The following well-posedness result, which is proved in Section 4, will be used to characterize the LLN limit of $\mathbf{G}_n$.

PROPOSITION 2.1. Fix $\mathbf{r} \in \ell_1^\downarrow$. Then there is a unique $(\mathbf{g}, \mathbf{v}) \in \mathbb{C}([0, \infty) : \ell_1^\downarrow \times \ell_\infty)$ that solves the following system of equations

$$(2.2) \quad (g_i(t), v_i(t)) = (\Gamma_1, \hat{\Gamma}_1) \left(r_i - \int_0^\cdot (g_i(s) - g_{i+1}(s)) ds + v_{i-1}(\cdot) \right) (t) \quad \forall i \geq 1, t \geq 0,$$

where $v_0(t) = \lambda t$ for all $t \geq 0$.

REMARK 2.2. Using the well-known characterization of a one-dimensional Skorohod map, one can alternatively characterize $(\mathbf{g}, \mathbf{v})$ as the unique pair in $\mathbb{C}([0, \infty) : \ell_1^\downarrow \times \ell_\infty)$ such that v_i is nondecreasing,

$$(2.3) \quad \left. \begin{aligned} g_i(t) &= r_i - \int_0^t (g_i(s) - g_{i+1}(s)) ds + v_{i-1}(t) - v_i(t), \\ v_i(t) &\geq 0, \quad g_i(t) \leq 1, \quad \int_0^t (1 - g_i(s)) dv_i(s) = 0 \end{aligned} \right\} \quad \forall i \geq 1$$

and $v_0(t) = \lambda t$, for all $t > 0$ and $v_i(0) = 0$ for each $i \geq 0$.

We can now present the LLN result. The proof is given in Section 4.

THEOREM 2.1. Let $\mathbf{r} \in \ell_1^\downarrow$. Suppose that $\mathbf{G}_n(0) \xrightarrow{P} \mathbf{r}$ in $\ell_1^\downarrow$, $\lambda_n \rightarrow \lambda$ and $d_n \rightarrow \infty$, as $n \rightarrow \infty$. Then $\mathbf{G}_n \rightarrow \mathbf{g}$ in probability in $\mathbb{D}([0, \infty) : \ell_1^\downarrow)$ as $n \rightarrow \infty$, where $(\mathbf{g}, \mathbf{v}) \in \mathbb{C}([0, \infty) : \ell_1^\downarrow \times \ell_\infty)$ is the unique solution of (2.2).

REMARK 2.3. Note that Theorem 2.1 allows $d_n \rightarrow \infty$ in an arbitrary manner. The Skorohod reflection term v_i in (2.3), which increases only at time instants t when $g_i(t) = 1$, prevents g_i from exceeding the level 1. It arises as a result of the simple fact that an arriving job cannot join a queue of length $i - 1$ when all the queues in the selection are of length i or more. In [31], Theorem 1, it is shown that, under the assumptions of Theorem 2.1, $\mathbf{G}_n$ is a tight sequence of $\mathbb{D}([0, \infty) : \ell_1^\downarrow)$ valued random variables and that every subsequential weak limit $\hat{\mathbf{g}}$ satisfies a system of equations given as

$$(2.4) \quad \hat{g}_i(t) = r_i - \int_0^t (\hat{g}_i(s) - \hat{g}_{i+1}(s)) ds + \int_0^t p_{i-1}(\hat{\mathbf{g}}(s)) ds \quad \text{for } i \geq 1,$$

where

$$(2.5) \quad p_j(\hat{\mathbf{g}}(s)) = \begin{cases} \lambda - (\lambda - 1 + \hat{g}_{j+2}(s))^+ & \text{if } j = m(\hat{\mathbf{g}}(s)) - 1, \\ (\lambda - 1 + \hat{g}_{j+1}(s))^+ & \text{if } j = m(\hat{\mathbf{g}}(s)) > 0, \\ \lambda & \text{if } j = m(\hat{\mathbf{g}}(s)) = 0, \\ 0 & \text{otherwise,} \end{cases}$$

and for $\mathbf{x} \in \ell_1^\downarrow$, $m(\mathbf{x}) \doteq \inf\{i \mid x_{i+1} < 1\}$. (Note that $m(\mathbf{G}_n(t))$ is the length of the smallest queue at time t .) The uniqueness of the above system of equations was not shown in [31].

From (2.2) and the definition in (2.1), it follows that each v_i is absolutely continuous and, for a.e. t ,

$$\frac{dv_i(t)}{dt} = \left(\frac{dv_{i-1}(t)}{dt} - g_i(t) + g_{i+1}(t) \right)^+ \mathbb{I}_{\{g_i(t)=1\}}$$

for any $i \geq 1$. From this, we see that, for a.e. t ,

$$(2.6) \quad \frac{dv_i(t)}{dt} = \begin{cases} \lambda & \text{if } i = 0, \\ \frac{dv_{i-1}(t)}{dt} & \text{if } i < m(\mathbf{g}(t)) \text{ and } i \geq 1, \\ \left(\frac{dv_{i-1}(t)}{dt} - 1 + g_{i+1}(t) \right)^+ & \text{if } i = m(\mathbf{g}(t)) \text{ and } i \geq 1, \\ 0 & \text{if } i > m(\mathbf{g}(t)). \end{cases}$$

and consequently $p_j(\mathbf{g}(s)) = \frac{dv_j(s)}{ds} - \frac{dv_{j+1}(s)}{ds}$ for a.e. s . Substituting this back in (2.3) shows that $\mathbf{g}$ solves the system of equations in (2.4). Conversely, for any solution $\hat{\mathbf{g}}$ of (2.4), defining $\hat{\mathbf{v}}$ by the right side of (2.6) by replacing $\mathbf{g}$ with $\hat{\mathbf{g}}$, we see that $(\hat{\mathbf{g}}, \hat{\mathbf{v}})$ solves (2.3). From the uniqueness result in Lemma 2.1, it then follows that in fact there is only one solution to the system of equations in (2.4) and this solution equals $\mathbf{g}$ given in (2.2).

Consider now the time asymptotic behavior of $\mathbf{g}$ given in (2.2). When $\lambda < 1$, $(\lambda, 0, 0 \dots) \in \ell_1$ is the unique fixed point of (2.2), as can be seen by setting the derivative of the right-hand side of (2.4) to 0. In the critical case, that is, when $\lambda = 1$, the situation is very different and in fact there are uncountably many fixed points given by the collection $\{\mathbf{f} \in \ell_1^\downarrow \mid m(\mathbf{f}) > 0, f_{m(\mathbf{f})+2} = 0\} = \{\mathbf{f}_k^\gamma \mid k \in \mathbb{N}, \gamma \in [0, 1)\}$, which once more is seen by checking that the derivative on the right-hand side of (2.4) is 0 at exactly these points when $\lambda = 1$. In this work, we are interested in the fluctuations of $\mathbf{G}_n$ in the critical case when the system starts suitably close to one of the fixed points of (2.3). Thus for the remaining section we will assume that $\lambda_n < 1$ for every n and $\lambda_n \rightarrow 1$ as $n \rightarrow \infty$. In order to formulate precisely what we mean by “suitably close to the fixed point” we need some definitions and notation. The functions β_n in the next definition will play a central role.

DEFINITION 1. Given $d_n \in [n]$, define the function $\beta_n : [0, 1] \rightarrow [0, 1]$ by

(2.7)
$$\beta_n(x) \doteq \prod_{i=0}^{d_n-1} \left(\frac{x - \frac{i}{n}}{1 - \frac{i}{n}} \right)^+.$$

The function $\beta_n(\cdot)$ arises when sampling d_n random servers without replacement. Specifically, when $nx \in \mathbb{N}$, $\beta_n(x) = \mathbf{P}(\mathbb{A}_{n,d_n} \subseteq [nx]) = \binom{nx}{d_n} / \binom{n}{d_n}$, where $\mathbb{A}_{n,d_n}$ is a randomly chosen subset (without replacement) from $[n]$ of size d_n . Here, we adopt the convention that $\binom{m}{r} = 0$ when $m < r$. An alternative is to perform sampling with replacement, which corresponds to the simpler function $\gamma_n(x) \doteq x^{d_n}$ in place of β_n .

We now introduce the notion of a “near fixed point” of $\mathbf{G}_n$.

DEFINITION 2. For $n \in \mathbb{N}$, the *near fixed point* $\boldsymbol{\mu}_n$ of $\mathbf{G}_n$ is the vector in $\ell_1^\downarrow$ given as $\boldsymbol{\mu}_n = (\mu_{n,1}, \mu_{n,2} \dots)$ where $\mu_{n,i}$ are defined recursively as $\mu_{n,1} = \lambda_n$ and $\mu_{n,i+1} = \lambda_n \beta_n(\mu_{n,i})$ for $i \geq 1$.

Using $\beta_n(x) \leq x^{d_n} \leq x$ and $\lambda_n < 1$, it is easy to check that $\boldsymbol{\mu}_n \in \ell_1^\downarrow$. The reason $\boldsymbol{\mu}_n$ is referred to as a near fixed point of $\mathbf{G}_n$ is discussed in Remark 3.1. To study the fluctuations of the process around the near fixed point $\boldsymbol{\mu}_n$, we define the centered and scaled process, $\mathbf{Z}_n$ as in (1.2). We now present our three main results on fluctuations, which correspond to the three cases $d_n/\sqrt{n} \rightarrow 0$, $d_n/\sqrt{n} \rightarrow c \in (0, \infty)$ and $d_n/\sqrt{n} \rightarrow \infty$, respectively. In each of these cases, we will assume that the initial configuration starts sufficiently close to the near equilibrium point $\boldsymbol{\mu}_n$.

ASSUMPTION 2.4. Suppose that $\{\|\mathbf{Z}_n(0)\|_1\}_{n \in \mathbb{N}}$ is tight and $\mathbf{Z}_n(0) \xrightarrow{P} \mathbf{z}$ in ℓ_2 , where $\mathbf{z}_{r+} = \mathbf{0}$ for some $r \in \mathbb{N}$.

In the following, $\dot{\beta}_n(x)$ is as defined in (5.1) and in the convention noted below (5.1). In particular, for $x \in (0, 1) \setminus \{\frac{d_n-1}{n}\}$, $\dot{\beta}_n(x)$ is the derivative of β_n at x .

THEOREM 2.2. Suppose that, as $n \rightarrow \infty$, $1 \ll d_n \ll \sqrt{n}$, $\lambda_n \nearrow 1$, and there is a $k \in \mathbb{N}$ so that $\mu_{n,k} \rightarrow 1$ and $\dot{\beta}_n(\mu_{n,k}) \rightarrow \alpha \in [0, \infty)$ as $n \rightarrow \infty$. Further suppose that Assumption 2.4 holds for some $r > k$. Then for any $T \in (0, \infty)$,

(2.8)
$$\lim_{M \rightarrow \infty} \sup_n \mathbf{P}(\|\mathbf{Z}_n\|_{2,T} > M) = 0.$$

Furthermore, if $k > 1$, then $\sup_{t \in [\epsilon, T]} |Z_{n,i}(t)| \xrightarrow{P} 0$ as $n \rightarrow \infty$ for any $T < \infty$, $0 < \epsilon \leq T$ and $i \in [k - 1]$.

Consider the shifted process $\mathbf{Y}_n(t) \doteq (\sum_{i=1}^k Z_{n,i}(t), Z_{n,k+1}(t), Z_{n,k+2}(t), \dots)$ and $\mathbf{y} \doteq (\sum_{i=1}^k z_i, z_{k+1}, z_{k+2}, \dots)$. Then $\mathbf{Y}_n \Rightarrow \mathbf{Y}$ in $\mathbb{D}([0, \infty) : \ell_2)$, where $\mathbf{Y} \in \mathbb{C}([0, \infty) : \ell_2)$ is the unique pathwise solution to

$$\begin{aligned} Y_1(t) &= y_1 - (\alpha + \mathbb{I}_{\{k=1\}}) \int_0^t Y_1(s) ds + \int_0^t Y_2(s) ds + \sqrt{2}B(t), \\ Y_2(t) &= y_2 + \alpha \int_0^t Y_1(s) ds - \int_0^t Y_2(s) ds + \int_0^t Y_3(s) ds, \\ Y_i(t) &= y_i - \int_0^t Y_i(s) ds + \int_0^t Y_{i+1}(s) ds \quad \text{for } i \in \{3, \dots, r-k+1\}, \\ Y_i(t) &= 0 \quad \text{for } i > r-k+1, \end{aligned} \tag{2.9}$$

and $B(\cdot)$ is a one-dimensional standard Brownian motion.

REMARK 2.5.

- (i) Note that the convergence $\sup_{t \in [\epsilon, T]} |Z_{n,i}(t)| \xrightarrow{P} 0$ as $n \rightarrow \infty$ for any $0 < \epsilon \leq T$ is equivalent to the statement that $Z_{n,i} \rightarrow 0$ in probability in $\mathbb{D}((0, T] : \mathbb{R})$ where the latter space is equipped with the topology of uniform convergence on compacts. Note also that, since Theorem 2.2 allows $Z_{n,i}(0)$ to converge to a nonzero limit, the above convergence to 0 cannot be strengthened to a convergence in probability in $\mathbb{D}([0, T] : \mathbb{R})$.
- (ii) By Corollary 5.3 in Section 5, when $\mu_{n,k}$ is away from 0,

$$\dot{\beta}_n(\mu_{n,k}) = (1 + o(1)) \frac{d_n \mu_{n,k+1}}{\lambda_n \mu_{n,k}}$$

as $n \rightarrow \infty$. Hence the assumptions $d_n \rightarrow \infty$, $\lambda_n \rightarrow 1$, $\mu_{n,k} \rightarrow 1$ and $\dot{\beta}_n(\mu_{n,k}) \rightarrow \alpha < \infty$ in Theorem 2.2 say that $\mu_{n,k+1} \rightarrow 0$. Since $\mu_{n,k} \rightarrow 1$, this in fact shows that $\mu_n \rightarrow \mathbf{f}_k$ in $\ell_1^\downarrow$, where recall that $\mathbf{f}_k$ is one of the fixed points of the fluid-limit (2.2) when $\lambda = 1$. The fact that the convergence happens in $\ell_1^\downarrow$ can be seen on observing that if $\mu_{n,k+1} \leq \epsilon$ then, by (5.2), $\mu_{n,k+1+i} \leq \epsilon^{d_n^i}$.

We also note that in general $\sqrt{n}(\mu_n - \mathbf{f}_k)$ will diverge, and thus $\sqrt{n}(\mathbf{G}_n - \mathbf{f}_k)$ will typically not be tight, in this regime. Nevertheless, it may still be interesting to study the behavior of $n^\alpha(\mathbf{G}_n - \mathbf{f}_k)$ for some $\alpha \in (0, 1/2)$ and appropriate choices of $d_n \rightarrow \infty$ and $\lambda_n \rightarrow 1$. Note however that when $\alpha \in (0, 1/2)$, the martingale term in the semimartingale decomposition of $n^\alpha(\mathbf{G}_n - \mathbf{f}_k)$ will converge to zero (as can be seen from the convergence observed below (6.6)) and thus the limit behavior is expected to be different. We leave this for future work.

- (iii) In the special case when the system starts sufficiently close to the near fixed point μ_n so that $z_i = 0$ for $i > k+1$, the limit process $\mathbf{Y}$ simplifies to an essentially two-dimensional process given as, $Y_i(t) = 0$ for $i > 2$, and

$$\begin{aligned} Y_1(t) &= y_1 - (\alpha + \mathbb{I}_{\{k=1\}}) \int_0^t Y_1(s) ds + \int_0^t Y_2(s) ds + \sqrt{2}B(t) \\ Y_2(t) &= y_2 + \alpha \int_0^t Y_1(s) ds - \int_0^t Y_2(s) ds \end{aligned}$$

- (iv) The convergence behavior of $\mathbf{Z}_n$ is governed by the sequence of parameters (d_n, λ_n) . In Corollary 5.5 from Section 5, we show that if $1 \ll d_n^{k+1} \ll n$ and $1 - \lambda_n = \frac{\xi_n + \log d_n}{d_n^k}$ with $\xi_n \rightarrow -\log(\alpha) \in (-\infty, \infty]$ and $\frac{\xi_n^2}{d_n} \rightarrow 0$, then the conditions $\mu_{n,k} \rightarrow 1$ and $\dot{\beta}_n(\mu_{n,k}) \rightarrow \alpha \in [0, \infty)$ of Theorem 2.2 are satisfied. Using this fact, we make the following observations. For simplicity, consider $z = 0$.

(a) Suppose that $d_n = \log n$, $1 - \lambda_n = \frac{\log \log n}{(\log n)^k}$. In this case the assumptions of Theorem 2.2 are satisfied and one essentially sees nonzero fluctuations only in the k th and $k + 1$ -th coordinates. Note that as k becomes large, the traffic intensity increases and one sees more and more coordinates of the near fixed point approach 1.

(b) With the same d_n as in (a) but a somewhat lower traffic intensity given as $1 - \lambda_n = \frac{(\log n)^{1/2-\epsilon}}{(\log n)^k}$ for some $\epsilon \in (0, 1/2)$, one sees that condition of the theorem are satisfied with $\alpha = 0$ (i.e., $\dot{\beta}_n(\mu_{n,k}) \rightarrow 0$). Thus the limit process Y , in the case $k > 1$, simplifies to $Y_i = 0$ for $i > 1$ and $Y_1(t) = \sqrt{2}B(t)$. When $k = 1$, $Z_1 = Y_1$ is instead given as the following Ornstein–Uhlenbeck (OU) process

(2.10)
$$Z_1(t) = - \int_0^t Z_1(s) ds + \sqrt{2}B(t).$$

(c) With higher values of d_n , using Theorem 2.2, one can analyze fluctuations for systems with higher traffic intensity. For example, suppose that $d_n = \frac{\sqrt{n}}{\log n}$. Then the conditions of the theorem are satisfied with $k = 1$ and $1 - \lambda_n \sim (\log n)^2/\sqrt{n}$. In fact, in this case $\alpha = 0$ and the limit process is described by the one-dimensional OU process (2.10). With a slightly higher traffic intensity given as $1 - \lambda_n = ((\log n)^2 - 2 \log n \log \log n)/2\sqrt{n}$ one obtains a two-dimensional limit diffusion.

(d) The theorem allows for traffic intensity in the Halfin–Whitt scaling regime (i.e., $\sqrt{n}(1 - \lambda_n) \rightarrow \beta > 0$) as well. Specifically, for $k \geq 2$, if $d_n = (\sqrt{n} \log n)^{\frac{1}{k}}$ and $(1 - \lambda_n) = \frac{\beta + o(1)}{\sqrt{n}}$ for some $\beta > \beta_0 = 1/2k$, the conditions of the theorem are satisfied with $\alpha = 0$. With slightly higher traffic intensity (e.g., $\beta + o(1)$ replaced by $\beta_0 + (\frac{1}{k} \log \log n - \log \alpha)/\log n$) conditions of the theorem are met with a nonzero α .

(e) More generally, suppose we are interested in studying the fluctuation behavior when the traffic intensity is $\lambda_n = 1 - \gamma n^{-a}$ for some $a \in (0, 1)$ and $\gamma > 0$. The cases $a < 1/2$ and $a > 1/2$ correspond to the so-called sub and super Halfin–Whitt regimes, respectively. The asymptotic behavior of JSQ(d_n) schemes in steady state in these regimes has been studied in [5, 6, 21, 22]. In [5, 6], the authors prove the following: suppose $d_n = n^b$ for some $b \in (0, 1]$ that satisfies $a/b \notin \mathbb{N}$ and $2a < 1 + b(k - 1)$ where $k \doteq \lceil a/b \rceil$; then with high probability in equilibrium, the largest queue will have length k and a vanishingly small fraction of queues have length smaller than k . In [21, 22], the authors consider the case $a \in (0, 1/2)$ and show that for the JSQ(d_n) system with buffer size $b_n = O(\log n)$, in equilibrium, both the expected waiting time per job and the probability that a job is routed to a nonidle server are $O(b_n n^{-r(1/2-a)})$, whenever $d_n \geq \frac{r}{\gamma} n^a \log n$ for any positive integer $r \leq \frac{\log n}{72(b_n-1)^2}$. In the current work, we study the behavior of JSQ(d_n) over finite intervals of time. Our results, including Theorem 2.2, allow for both sub and super Halfin–Whitt regimes. To see this, choose any $a \in (0, 1)$, $v > 1$, an integer $k > a/(1 - a)$, and let $b \doteq a/k$. (In other words, a satisfies $2a < 1 + b(k - 1)$, which is the same condition as in [5]). Then Theorem 2.2 holds with k , $d_n = n^b (\frac{bv}{\gamma} \log n)^{1/k}$ and $\alpha = 0$.

(f) Recall that a fixed point of (2.2) when $\lambda = 1$ takes the form $f_k^\gamma \doteq f_k + \gamma e_{k+1} = (1, \dots, 1, \gamma, 0, \dots) \in \ell_1^\downarrow$, where $k \in \mathbb{N}$ and $\gamma \in [0, 1)$. Although Theorem 2.2 only considers settings where the near fixed point μ_n converges to $f_k^0 = f_k$ for some k , it is possible to give conditions under which μ_n converges to a different fixed point. Specifically, suppose that $1 \ll d_n^{k+1} \ll n$ and $1 - \lambda_n = \frac{a}{d_n^k}$ for some $a > 0$. Then it can be checked using Lemma 5.4 that $\mu_n \rightarrow f_k^\gamma$ with $\gamma = e^{-a}$. However, proving fluctuation results in this regime appears to be technically more involved, and we leave it for future work.

The next theorem describes the fluctuations of $\mathbf{Z}_n$ when d_n is of order $\sqrt{n}$.

THEOREM 2.3. *Suppose that $\frac{d_n}{\sqrt{n}} \rightarrow c \in (0, \infty)$ and $\lambda_n = 1 - (\frac{\log d_n}{d_n} + \frac{\alpha_n}{\sqrt{n}})$ with $\alpha_n \rightarrow \alpha \in (-\infty, \infty]$ and $\alpha_n = o(n^{1/4})$. Then $\mu_n \rightarrow f_1$ in $\ell_1^\downarrow$. Suppose further that Assumption 2.4 holds for some $r \geq 2$. Then, as $n \rightarrow \infty$, $\mathbf{Z}_n \Rightarrow \mathbf{Z}$ in $\mathbb{D}([0, \infty) : \ell_2)$, where $\mathbf{Z}$ is the unique pathwise solution to*

$$Z_1(t) = z_1 - \int_0^t (Z_1(s) - Z_2(s)) ds - (ce^{c\alpha})^{-1} \int_0^t (e^{cZ_1(s)} - 1) ds + \sqrt{2}B(t),$$

$$Z_2(t) = z_2 - \int_0^t (Z_2(s) - Z_3(s)) ds + (ce^{c\alpha})^{-1} \int_0^t (e^{cZ_1(s)} - 1) ds,$$

$$Z_i(t) = z_i - \int_0^t (Z_i(s) - Z_{i+1}(s)) ds \quad \text{for each } i \in \{3 \dots r\},$$

$$Z_i(t) = 0 \quad \text{for each } i > r,$$

and B is standard Brownian motion.

REMARK 2.6.

(i) Note that the coefficients in the above system of equations are only locally Lipschitz and have an exponential growth. However, since c is positive, the system of equations has a unique pathwise solution as is shown in Lemma 8.2.

(ii) Once more, when $z_i = 0$ for all $i > 2$, the system of equations simplifies to a two-dimensional system given as $Z_i = 0$ for all $i > 2$, and

$$Z_1(t) = z_1 - \int_0^t (Z_1(s) - Z_2(s)) ds - (ce^{c\alpha})^{-1} \int_0^t (e^{cZ_1(s)} - 1) ds + \sqrt{2}B(t),$$

$$Z_2(t) = z_2 - \int_0^t Z_2(s) ds + (ce^{c\alpha})^{-1} \int_0^t (e^{cZ_1(s)} - 1) ds.$$

(iii) In the regime considered in Theorem 2.3, the near fixed point μ_n can converge to only one particular fixed point of (2.2), namely f_1 . As before, the term $\sqrt{n}(\mu_n - f_1)$ may diverge and thus $\sqrt{n}(G_n(\cdot) - f_1)$ will in general not be tight.

(iv) Suppose that $d_n = c\sqrt{n}$ for some $c > 0$, $z = 0$ and $1 - \lambda_n = (\beta + o(1)) \log n / \sqrt{n}$ for some $\beta > \beta_0 = 1/2c$. Then the assumptions of the above theorem are satisfied with $\alpha = \infty$ and the limit system simplifies to a one-dimensional OU process given as $Z_i = 0$ for all $i > 1$, and Z_1 satisfies (2.10). If $(\beta + o(1)) \log n$ is replaced by $\beta_0 \log n + \gamma$ for some $\gamma \in \mathbb{R}$, we instead obtain a two-dimensional limit system given as $Z_i = 0$ for all $i > 2$, and

$$Z_1(t) = - \int_0^t (Z_1(s) - Z_2(s)) ds - e^{-c\gamma} \int_0^t (e^{cZ_1(s)} - 1) ds + \sqrt{2}B(t),$$

$$Z_2(t) = - \int_0^t Z_2(s) ds + e^{-c\gamma} \int_0^t (e^{cZ_1(s)} - 1) ds.$$

Finally, we consider the fluctuation behavior when $d_n \gg \sqrt{n}$. This time the limit system will involve reflected diffusion processes. Recall from (2.1) the definition of the Skorohod maps Γ_α and $\hat{\Gamma}_\alpha$ associated with a reflection barrier at $\alpha \in \mathbb{R}$. We will extend the definition of these maps to $\alpha = \infty$ by setting

$$(2.11) \quad \Gamma_\infty(h) = h, \quad \hat{\Gamma}_\infty(h) = 0 \quad \text{for } h \in \mathbb{D}([0, \infty) : \mathbb{R}).$$

THEOREM 2.4. Suppose that $\sqrt{n} \ll d_n$ and

$$(2.12) \quad \lambda_n = 1 - \left(\frac{\log d_n}{d_n} + \frac{\alpha_n}{\sqrt{n}} \right),$$

where $\alpha_n \rightarrow \alpha \in [0, \infty]$, with $\alpha_n^- = O(\sqrt{n}/d_n)$, and $\alpha_n = O(n^{1/6})$.

Then $\mu_n \rightarrow f_1$ in ℓ_1^+ . Suppose further that Assumption 2.4 holds for some $r \geq 2$ with $z_1 \leq \alpha$. Then, as $n \rightarrow \infty$, $Z_n \Rightarrow Z \in \mathbb{D}([0, \infty) : \ell_2)$, where Z is the first component of the pair (Z, η) , which is a $\ell_2 \times \mathbb{R}_+$ valued continuous process given as the unique solution to

$$(2.13) \quad \begin{aligned} (Z_1(t), \eta(t)) &= (\Gamma_\alpha, \hat{\Gamma}_\alpha) \left(z_1 - \int_0^t (Z_1(s) - Z_2(s)) ds + \sqrt{2}B(\cdot) \right)(t), \\ Z_2(t) &= z_2 - \int_0^t (Z_2(s) - Z_3(s)) ds + \eta(t), \\ Z_i(t) &= z_i - \int_0^t (Z_i(s) - Z_{i+1}(s)) ds \quad \text{for each } i \in \{3 \dots r\}, \\ Z_i(t) &= 0 \quad \text{for each } i > r, \end{aligned}$$

and B is a standard Brownian motion.

We note that given a standard Brownian motion B , there is a unique continuous process (Z, η) with values in $\ell_2 \times \mathbb{R}_+$, adapted to the filtration generated by B (See Remark 2.8(i)). As a corollary to this theorem, we obtain the specific regime considered in [31] (in fact we provide a slight strengthening in that, unlike [31], we allow $\alpha = 0$). See Remark 2.8(v) for further discussion.

COROLLARY 2.7. As $n \rightarrow \infty$, suppose that $d_n \gg \sqrt{n} \log n$ and $\sqrt{n}(1 - \lambda_n) \rightarrow \alpha \in [0, \infty)$, along with $\sqrt{n}(1 - \lambda_n) \geq (\sqrt{n} \log n)/d_n$ for large n if $\alpha = 0$. Let $Y_n(\cdot) \doteq \sqrt{n}(G_n(\cdot) - f_1)$ and assume that the sequence of random variables $\{\|Y_n(0)\|_1\}$ is tight, and as $n \rightarrow \infty$, $Y_n(0) \xrightarrow{P} y \in \ell_2$ with $y_{r+} = 0$ for some $r \geq 2$. Then $Y_n \Rightarrow Y$ in $\mathbb{D}([0, \infty) : \ell_2)$, where $(Y, \tilde{\eta})$ is the $\ell_2 \times [0, \infty)$ valued continuous process given by the unique solution to

$$\begin{aligned} (Y_1(t), \tilde{\eta}(t)) &= (\Gamma_0, \hat{\Gamma}_0) \left(y_1 - \alpha \text{id}(\cdot) - \int_0^t (Y_1(s) - Y_2(s)) ds + \sqrt{2}B(\cdot) \right)(t) \\ Y_2(t) &= y_2 - \int_0^t (Y_2(s) - Y_3(s)) ds + \tilde{\eta}(t), \\ Y_i(t) &= y_i - \int_0^t (Y_i(s) - Y_{i+1}(s)) ds \quad \text{for each } i \in \{3 \dots r\}, \\ Y_i(t) &= 0 \quad \text{for each } i > r, \end{aligned}$$

and B is a standard Brownian motion.

REMARK 2.8.

(i) The existence and uniqueness of solutions to the stochastic integral equations in (2.13) follows by standard fixed-point arguments on using the Lipschitz property of the map

Γ_α on $\mathbb{D}([0, \infty) : \mathbb{R})$. This system of equations can equivalently be written as

$$\begin{aligned}
 (2.14) \quad Z_1(t) &= z_1 - \int_0^t (Z_1(s) - Z_2(s)) ds + \sqrt{2}B(t) - \eta(t), \\
 Z_2(t) &= z_2 - \int_0^t (Z_2(s) - Z_3(s)) ds + \eta(t), \\
 Z_i(t) &= z_i - \int_0^t (Z_i(s) - Z_{i+1}(s)) ds \quad \text{for each } i \in \{3 \dots r\}, \\
 Z_i(t) &= 0 \quad \text{for each } i > r,
 \end{aligned}$$

where $\eta = 0$ when $\alpha = \infty$, and when $\alpha \in \mathbb{R}$, it satisfies

$$(2.15) \quad \left. \begin{aligned} \eta(0) &= 0 \text{ and } \eta \text{ is a nondecreasing function.} \\ Z_1(t) &\leq \alpha \\ \int_0^\infty (\alpha - Z_1(s)) d\eta(s) &= 0 \end{aligned} \right\}.$$

The system of equations (2.14) describes a constrained multidimensional diffusion driven by a one-dimensional Brownian motion. Existence and uniqueness for a similar system of equations and the convergence of $\mathbf{Y}_n$ to that system when $d_n = n$ is shown in [9]. However, note that unlike in [9] (where the reflection is at 0), the reflection in (2.14) occurs at a barrier $\alpha \in [0, \infty]$.

(ii) The convergence $\mu_n \rightarrow \mathbf{f}_1$ along with tightness of $\{\mathbf{Z}_n\}_{n \in \mathbb{N}}$ shows that, under the conditions of Theorems 2.3 or 2.4, most queues will be of length 1 on any fixed interval $[0, T]$.

(iii) The limit system in Theorem 2.4 simplifies when $z_i = 0$ for $i > 2$ and is given as $Z_i = 0$ for all $i > 2$, and

$$\begin{aligned}
 Z_1(t) &= z_1 - \int_0^t (Z_1(s) - Z_2(s)) ds + \sqrt{2}B(t) - \eta(t), \\
 Z_2(t) &= z_2 - \int_0^t Z_2(s) ds + \eta(t),
 \end{aligned}$$

where η is as in the statement of the theorem.

(iv) Suppose that $d_n = \sqrt{n} \log n / 2a$ for some $a > 0$ and $1 - \lambda_n = \frac{a}{\sqrt{n}} + \frac{2a(\log \log n + O(1))}{\sqrt{n} \log n}$. Then the assumptions in Theorem 2.4 are satisfied with $\alpha = 0$. In this case, the reflection barrier is at 0, namely $Z_1(t) \leq 0$ for all t . Also note that since $\sqrt{n}(1 - \lambda_n) \rightarrow a$, we have that $\mu_{n,1} = \lambda_n \rightarrow 1$. Since $d_n/\sqrt{n} \rightarrow \infty$, this shows that for $k \geq 2$,

$$\sqrt{n}\mu_{n,2} = \sqrt{n}\lambda_n\beta_n(\lambda_n) \leq \sqrt{n}\lambda_n\lambda_n^{d_n} = \sqrt{n}(1 - (1 - \lambda_n))^{d_n+1} \rightarrow 0.$$

Using $\mu_{n,i+1} \leq \mu_{n,i}^{d_n}$, see that $\sqrt{n}(\mu_n - \mathbf{f}_1) \rightarrow -a\mathbf{e}_1 \in \ell_1$, and hence the fluctuations of $\mathbf{G}_n$ about the fixed point $\mathbf{f}_1$ can be characterized as well. Specifically, letting $\mathbf{Y}_n(\cdot) = \sqrt{n}(\mathbf{G}_n(\cdot) - \mathbf{f}_1) = \mathbf{Z}_n(\cdot) + \sqrt{n}(\mu_n - \mathbf{f}_1)$, we see that, under the condition of the above theorem, $\mathbf{Y}_n \Rightarrow \mathbf{Y}$ in $\mathbb{D}([0, \infty) : \ell_2)$, where $\mathbf{Y} = \mathbf{Z} - a\mathbf{e}_1$, and hence, assuming $z_i = 0$ for $i > 2$, $(\mathbf{Y}, \tilde{\eta}) \in \mathbb{C}([0, \infty) : \ell_2 \times \mathbb{R}_+)$ is the unique solution to (2.15) with (Z_1, η, α) replaced with $(Y_1, \tilde{\eta}, -a)$, and the equations

$$\begin{aligned}
 Y_1(t) &= y_1 - at - \int_0^t (Y_1(s) - Y_2(s)) ds + \sqrt{2}B(t) - \tilde{\eta}(t), \\
 Y_2(t) &= y_2 - \int_0^t Y_2(s) ds + \tilde{\eta}(t),
 \end{aligned}$$

where $\mathbf{y} = \mathbf{z} - a\mathbf{e}_1$ and B is a standard Brownian motion. In particular, the limit $\mathbf{Y}$ takes the same form as in [9, 31] with a stronger constraint that $Y_1(t) \leq -a < 0$ for each $t > 0$.

TABLE 1
Analysis of JSQ(d_n) with $1 - \lambda_n = n^{-a}$, $d_n = \lceil n^{b_n} \rceil$ and $k \doteq \lim_n \lceil a/b_n \rceil$. The notation $b_n \in ((l_n, r])$ denotes the condition $n^{l_n} \ll d_n \leq n^r$

Reference	Regimes of a, b_n and k	Analysis type
Braverman [4]	$a = 0.5, b_n = 1, k = 1$	Convergence of stationary distribution
Eschenfeldt & Gamarnik [9]	$a = 0.5, b_n = 1, k = 1$	Functional central limit theorem
Mukherjee et al. [31]	$a = 0.5, b_n \in ((0.5 + \frac{\log \log n}{\log n}, 1], k = 1$	Functional central limit theorem
Theorem 2.4 ($\alpha = 0$)	$a \in (1/2, 1), b_n \in [a + \frac{\log \log n}{\log n}, 1], k = 1$	Functional central limit theorem
Theorem 2.4 ($\alpha \in (0, \infty)$)	$a = 0.5, b_n \in [0.5 + \frac{\log \log n}{\log n}, 1], k = 1$	Functional central limit theorem
Theorem 2.4 ($\alpha = \infty$)	$a \in [1/3, 1/2), b_n \in ((0.5, 1], k = 1$	Functional central limit theorem
Theorem 2.3 ($\alpha = \infty$)	$a \in (1/4, 1/2), b_n = 0.5, k = 1$	Functional central limit theorem
Theorem 2.2 ($\alpha = 0$)*	$a \in (0, 1), b_n = (a + \frac{\log \log n}{\log n})/k,$ $k > a/(1 - a), k \in \mathbb{N}$	Functional central limit theorem
Brightwell et al. [5]	$a \in (0, 1), b_n \rightarrow b \in (0, 1],$ $2a < 1 + b(k - 1), a/b \notin \mathbb{N}$	Equilibrium queue lengths
Liu & Ying [21, 22]	$a \in (0, 1/2), b_n \in [a + \frac{\log \log n}{\log n}, 1],$ $k = 1$	Equilibrium performance

(v) Suppose that $d_n \gg \sqrt{n} \log n$. Then it is easy to see that (2.12) holds with some $\alpha > 0$ if and only if $\sqrt{n}(1 - \lambda_n) \rightarrow \alpha > 0$. This regime was studied in [31]. Using the arguments as in (iv) above, it is easy to check that $\sqrt{n}(\mu_n - f_1) \rightarrow -\alpha e_1$ in ℓ_1 (and hence ℓ_2). Corollary 2.7 is immediate from this and Theorem 2.4. In particular, we recover [31], Theorem 3. However, the proof techniques in the current paper are different from the stochastic coupling techniques employed in [31].

(vi) Suppose $\sqrt{n} \ll d_n \ll \sqrt{n} \log n$ and that (2.12) holds with $\alpha < \infty$. Then, as observed in the proof of Theorem 4 in [31], Y_n will not be tight in this regime. But since $\sqrt{n}(1 - \mu_{n,1}) = (\sqrt{n} \log d_n)/d_n + \alpha_n \rightarrow \infty$, this does not preclude the convergence of $Z_n = Y_n - \sqrt{n}(\mu_n - f_1)$. Indeed, Theorem 2.4 shows that the process Z_n converges in distribution and the limit process has a reflecting barrier at α , that is, $Z_1 \leq \alpha$. In particular, unlike the case $d_n \gg \sqrt{n} \log n$, the barrier in this case does not come from the constraint $G_{n,1} \leq 1$.

(vii) Theorem 2.4 allows for a slower approach to criticality than $n^{-1/2}$, for example, λ_n such that $n^{1/3}(\lambda_n - 1) \rightarrow \gamma > 0$. In this case, $\alpha = \infty$ and there is no reflection. When $z_i = 0$ for all $i \geq 1$, this system reduces to the one-dimensional OU process given by (2.10) with $Z_i = 0$ for $i > 1$.

Table 1 summarizes some of the key regimes of (d_n, λ_n) that are covered by Theorems 2.2–2.4 and places them in the context of previous work on JSQ(d_n) systems in heavy traffic.

In order to make comparison with [5], note that the regime in * can equivalently be written as $a \in (0, 1), b_n = (a + \frac{\log \log n}{\log n})/k \rightarrow b, 2a < 1 + b(k - 1), a/b \in \mathbb{N}$.

3. Poisson representation of state processes. We now embark on the proofs of the main results. We start with a brief overview of the organization of the proofs. In this section, we describe a specific construction of the state process. Proof of the law of large numbers (Theorem 2.1) is given in Section 4. Section 5 describes fine-scaled (deterministic) properties of the function β_n and the near fixed points μ_n , which play a key technical role in the proofs of our diffusion approximations. Section 6 derives preliminary estimates required to prove

all the main results for the fluctuations of the state process. Sections 7, 8 and 9 complete the proofs of Theorems 2.2, 2.3 and 2.4, respectively.

We start with a specific construction of the state process through time changed Poisson processes (cf. [11, 20]). A similar representation has been used in previous work on JSQ(d) systems (cf. [9, 10, 31]). Let $\{N_{i,+}, N_{i,-} : i \geq 1\}$ be a collection of mutually independent rate one Poisson processes given on some probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Then G_n has the following (equivalent in distribution) representation. For $i \geq 1$ and $t \geq 0$,

$$(3.1) \quad \begin{aligned} G_{n,i}(t) = G_{n,i}(0) &- \frac{1}{n} N_{i,-} \left(n \int_0^t [G_{n,i}(s) - G_{n,i+1}(s)] ds \right) \\ &+ \frac{1}{n} N_{i,+} \left(\lambda_n n \int_0^t [\beta_n(G_{n,i-1}(s)) - \beta_n(G_{n,i}(s))] ds \right), \end{aligned}$$

where $G_{n,0}(t) = 1$ for all $t \geq 0$. Denoting

$$\begin{aligned} A_{n,i}(t) &\doteq N_{i,+} \left(\lambda_n n \int_0^t [\beta_n(G_{n,i-1}(s)) - \beta_n(G_{n,i}(s))] ds \right), \\ D_{n,i}(t) &\doteq N_{i,-} \left(n \int_0^t [G_{n,i}(s) - G_{n,i+1}(s)] ds \right), \end{aligned}$$

the above evolution equation can be rewritten as

$$(3.2) \quad G_{n,i}(t) = G_{n,i}(0) - \frac{1}{n} D_{n,i}(t) + \frac{1}{n} A_{n,i}(t), \quad i \in \mathbb{N}, t \geq 0.$$

The above equation can be intuitively understood as follows. The point process $D_{n,i}$ describes events that cause a decrease in $G_{n,i}$ owing to completion of service events for jobs in queues of length exactly i , which since each server works at rate 1, happens at rate $n[G_{n,i}(t) - G_{n,i+1}(t)]$ at the time instant t . On the other hand, $A_{n,i}$ describes events causing an increase in $G_{n,i}$, which only occur if the chosen queue of a new job has exactly $i - 1$ individuals; this occurs if among the d_n random choices made by this job, all of the chosen queues have load at least $i - 1$ but not all have load at least i . The probability of the latter event is exactly $\beta_n(G_{n,i-1}(t)) - \beta_n(G_{n,i}(t))$, and thus the rate at which $G_{n,i}$ increases at time instant t is given by $n\lambda_n(\beta_n(G_{n,i-1}(t)) - \beta_n(G_{n,i}(t)))$.

Let

$$\tilde{\mathcal{F}}_t^n = \sigma\{A_{n,i}(s), D_{n,i}(s), s \leq t, i \geq 1\},$$

and let $\mathcal{F}_t^n$ be the augmentation of $\tilde{\mathcal{F}}_t^n$ with $\mathbf{P}$ -null sets. It then follows that, for each $i \geq 1$,

$$(3.3) \quad \begin{aligned} M_{n,i,+}(t) &\doteq \frac{1}{n} N_{i,+} \left(\lambda_n n \int_0^t \beta_n(G_{n,i-1}(s)) - \beta_n(G_{n,i}(s)) ds \right) \\ &- \lambda_n \int_0^t \beta_n(G_{n,i-1}(s)) - \beta_n(G_{n,i}(s)) ds \end{aligned}$$

and

$$(3.4) \quad M_{n,i,-}(t) \doteq \frac{1}{n} N_{i,-} \left(n \int_0^t G_{n,i}(s) - G_{n,i+1}(s) ds \right) - \int_0^t (G_{n,i}(s) - G_{n,i+1}(s)) ds$$

are $\{\mathcal{F}_t^n\}$ -martingales with predictable (cross) quadratic variation processes given, for $t \geq 0$, as

$$\begin{aligned} \langle M_{n,i,+} \rangle_t &= \frac{\lambda_n}{n} \int_0^t (\beta_n(G_{n,i-1}(s)) - \beta_n(G_{n,i}(s))) ds, \quad i \geq 1, \\ \langle M_{n,i,-} \rangle_t &= \frac{1}{n} \int_0^t (G_{n,i}(s) - G_{n,i+1}(s)) ds, \quad i \geq 1, \end{aligned}$$

$$\begin{aligned} \langle M_{n,i,-}, M_{n,j,-} \rangle_t &= 0, & \langle M_{n,i,+}, M_{n,j,+} \rangle_t &= 0, & \text{for all } i, j \geq 1, i \neq j \text{ and} \\ \langle M_{n,i,+}, M_{n,k,-} \rangle_t &= 0 & \text{for all } i, k \geq 1. \end{aligned}$$

Using these martingales, the evolution of $\mathbf{G}_n$ can be rewritten as

$$(3.5) \quad \begin{aligned} G_{n,i}(t) &= G_{n,i}(0) - \int_0^t (G_{n,i}(s) - G_{n,i+1}(s)) ds \\ &\quad + \lambda_n \int_0^t (\beta_n(G_{n,i-1}(s)) - \beta_n(G_{n,i}(s))) ds + M_{n,i}(t), \quad i \geq 1, \end{aligned}$$

where $M_{n,i}(t) \doteq M_{n,i,+}(t) - M_{n,i,-}(t)$ and

$$(3.6) \quad \langle M_{n,i} \rangle_t = \frac{1}{n} \left(\int_0^t (G_{n,i}(s) - G_{n,i+1}(s)) ds + \lambda_n \int_0^t (\beta_n(G_{n,i-1}(s)) - \beta_n(G_{n,i}(s))) ds \right).$$

We will assume throughout that $\mathbf{G}_n(0) \in \ell_1^\downarrow$ a.s. Then it follows that, for every $t \geq 0$, $\|\mathbf{G}_n(t)\|_1 < \infty$ a.s. Indeed, since $n\|\mathbf{G}_n(t)\|_1$ equals the total number of jobs in the system at time t , and over any time interval $[0, t]$ finitely many jobs enter the system a.s., denoting by A_t the number of jobs that arrive over $[0, t]$, we see that $\|\mathbf{G}_n(t)\|_1 \leq \|\mathbf{G}_n(0)\|_1 + A_t/n < \infty$ a.s. Thus $\mathbf{G}_n$ is a stochastic process with sample paths in $\mathbb{D}([0, \infty) : \ell_1^\downarrow)$. Note that, for any $t > 0$, $\|\mathbf{G}_n(t) - \mathbf{G}_n(t-)\|_1 \leq \frac{1}{n}$.

REMARK 3.1. Let $\mathbf{a}_n, \mathbf{b} : \ell_1^\downarrow \rightarrow \ell_1$ be given by

$$\mathbf{a}_n(\mathbf{x})_i \doteq \lambda_n(\beta_n(x_{i-1}) - \beta_n(x_i)), \quad \mathbf{b}(\mathbf{x})_i \doteq x_i - x_{i+1}, \quad \mathbf{x} \in \ell_1^\downarrow, \quad i \geq 1,$$

where, by convention, for $\mathbf{x} \in \ell_1^\downarrow$, $x_0 = 1$. Then (3.5) can be rewritten as an evolution equation in ℓ_1 as

$$(3.7) \quad \mathbf{G}_n(t) = \mathbf{G}_n(0) + \int_0^t [\mathbf{a}_n(\mathbf{G}_n(s)) - \mathbf{b}(\mathbf{G}_n(s))] ds + \mathbf{M}_n(t),$$

where $\mathbf{M}_n(t) \doteq (M_{n,i}(t))_{i \geq 1}$ is a stochastic process with sample paths in $\mathbb{D}([0, \infty) : \ell_1)$ and the integral is a Bochner-integral [37]. Note that the near fixed point $\boldsymbol{\mu}_n$ from Definition 2 satisfies $\mathbf{a}_n(\boldsymbol{\mu}_n) = \mathbf{b}(\boldsymbol{\mu}_n)$. It is in fact the unique solution to

$$(3.8) \quad \mathbf{a}_n(\mathbf{x}) = \mathbf{b}(\mathbf{x}) \quad \text{for } \mathbf{x} \in \ell_1^\downarrow,$$

as is seen by adding up all the coordinates of (3.8) and using $\mathbf{x} \in \ell_1$. In Lemma 4.1, we will see that for any $T > 0$, as $n \rightarrow \infty$, $\sup_{t \leq T} \|\mathbf{M}_n(t)\|_2 \xrightarrow{P} 0$. Hence if $\mathbf{G}_n(0) = \boldsymbol{\mu}_n$, then by (3.7), we expect the process $\mathbf{G}_n(t)$ to stay close to $\boldsymbol{\mu}_n$ (over any compact time interval) as $n \rightarrow \infty$. In this sense, $\boldsymbol{\mu}_n$ can be viewed as a “near fixed point” of $\mathbf{G}_n(\cdot)$ and the terminology in Definition 2 is justified. Another reason for this terminology comes from the results in Theorems 2.2–2.4, which show that, under conditions, $\boldsymbol{\mu}_n$ converges to one of the fixed points of the fluid limit (2.2) when $\lambda = 1$.

4. The law of large numbers. In this section, we prove Proposition 2.1 and Theorem 2.1.

4.1. Uniqueness of fluid limit equations. In this subsection, we show that there is at most one solution of (2.2) in $\mathbb{C}([0, \infty) : \ell_1^\downarrow \times \ell_\infty)$. Results of Section 4.2 will provide existence of solutions to this equation. Suppose $(\mathbf{g}, \mathbf{v})$ and $(\mathbf{g}', \mathbf{v}')$ are two solutions to (2.2) in $\mathbb{C}([0, \infty) : \ell_1^\downarrow \times \ell_\infty)$. We will now argue that the two solutions are equal.

We claim that that v'_i and v_i are nonzero for only finitely many i 's. Indeed, since $\mathbf{g}, \mathbf{g}' \in \mathbb{C}([0, T] : \ell_1^\downarrow)$, there is a constant $C \in (0, \infty)$ so that $\sup_{s \leq T} \|\mathbf{g}(s)\|_1 \vee \sup_{s \leq T} \|\mathbf{g}'(s)\|_1 \leq C$. Since

$$(4.1) \quad x_i \leq \|\mathbf{x}\|_1 / i \quad \text{for any } \mathbf{x} \in \ell_1^\downarrow,$$

taking $M \doteq \lceil C + 1 \rceil \in \mathbb{N}$ shows that $\sup_{s \leq T} g_i(s) \vee g'_i(s) < 1$ for any $i \geq M$. But then by the equivalent representation of (2.2) given in (2.3) (in particular the second line), we must have $v_i = v'_i = 0$ for any $i \geq M$. This proves the claim.

Since $v_i = v'_i = 0$ for $i \geq M$, the first line of the equivalent formulation in (2.3) shows that both $\mathbf{x} = \mathbf{g}$ and $\mathbf{x} = \mathbf{g}'$ satisfy the integral equations

$$x_i(t) = r_i - \int_0^t (x_i(s) - x_{i+1}(s)) ds \quad \text{for } i \geq M + 1 \text{ and } t \in [0, T].$$

By standard arguments using Gronwall's lemma [11], Appendix 5, we then must have $g_i = g'_i$ for each $i \geq M + 1$. Indeed, letting $z_i(\cdot) \doteq g_i(\cdot) - g'_i(\cdot)$ for $i \geq M + 1$ and $v(t) \doteq \sum_{i=M+1}^\infty |z_i(t)|$ for $t \in [0, T]$, we have that

$$|z_i(t)| \leq \int_0^t (|z_i(s)| + |z_{i+1}(s)|) ds \quad \text{for all } i \geq M + 1, \text{ and } t \in [0, T]$$

and so

$$v(t) \leq 2 \int_0^t v(s) ds, \quad t \in [0, T],$$

which implies that $v(t) = 0$ for $t \in [0, T]$.

We now show that $g_i = g'_i$ for $i \leq M$. From the definition of the Skorohod map in (2.1), we see that for $h_1, h_2 \in \mathbb{D}([0, \infty) : \mathbb{R})$ with $h_i(0) \leq 1$, $i = 1, 2$, and $t \geq 0$

$$\|\Gamma_1(h_1) - \Gamma_1(h_2)\|_{*,t} \leq 2\|h_1 - h_2\|_{*,t}, \quad \|\hat{\Gamma}_1(h_1) - \hat{\Gamma}_1(h_2)\|_{*,t} \leq \|h_1 - h_2\|_{*,t}.$$

Thus, since $(\mathbf{g}, \mathbf{v})$ and $(\mathbf{g}', \mathbf{v}')$ solve (2.2),

$$(4.2) \quad \|g_i - g'_i\|_{*,t} \leq 2 \left(\int_0^t \|g_i - g'_i\|_{*,s} ds + \int_0^t \|g_{i+1} - g'_{i+1}\|_{*,s} ds + \|v_{i-1} - v'_{i-1}\|_{*,t} \right),$$

and

$$(4.3) \quad \|v_i - v'_i\|_{*,t} \leq \int_0^t \|g_i - g'_i\|_{*,s} ds + \int_0^t \|g_{i+1} - g'_{i+1}\|_{*,s} ds + \|v_{i-1} - v'_{i-1}\|_{*,t}$$

for any $i \geq 1$. Let $H_t \doteq \max_{i \in \{1, \dots, M\}} \|g_i - g'_i\|_{*,t}$. Note $g_{M+1} = g'_{M+1}$, and hence $H_t = \max_{i \in \{1, \dots, M+1\}} \|g_i - g'_i\|_{*,t}$. Then from (4.3), we have

$$(4.4) \quad \|v_i - v'_i\|_{*,t} \leq 2 \int_0^t H_s ds + \|v_{i-1} - v'_{i-1}\|_{*,t} \quad \text{for any } i \leq M.$$

Repeatedly using (4.4) along with $v_0 = v'_0$ shows that $\|v_i - v'_i\|_{*,t} \leq 2i \int_0^t H_s ds$ for any $i \leq M$. Using this bound in (4.2) shows for $1 \leq i \leq M$:

$$\|g_i - g'_i\|_{*,t} \leq 2 \left(2 \int_0^t H_s ds + 2(i-1) \int_0^t H_s ds \right) = 4i \int_0^t H_s ds.$$

Hence considering the maximum of $\|g_i - g'_i\|_{*,t}$ over $1 \leq i \leq M$ we get

$$0 \leq H_t \leq 4M \int_0^t H_s ds \quad \text{for each } t \in [0, T].$$

Gronwall's lemma now shows that $H_T = 0$, and hence $g_i = g'_i$ for $i = 1 \dots M$. Finally, since $v_0 = v'_0$, we see recursively from the second equation in (2.2) that $v_i = v'_i$ for all $i \geq 0$. $\square$

4.2. *Tightness and limit point characterization.* Some of the arguments in this section are similar to [31], however, in order to keep the presentation self-contained we provide details in a concise manner. The next result establishes the convergence of the martingale term $\mathbf{M}_n$ in the semimartingale decomposition in (3.7). Throughout this subsection and the next section, we assume that the conditions of Theorem 2.1 are satisfied, namely, $\mathbf{G}_n(0) \xrightarrow{P} \mathbf{r}$ in $\ell_1^\downarrow$, $\lambda_n \rightarrow \lambda$ and $d_n \rightarrow \infty$, as $n \rightarrow \infty$.

LEMMA 4.1. *For any $T > 0$, $\sup_{s \leq T} \|\mathbf{M}_n(s)\|_2 \xrightarrow{P} 0$.*

PROOF. It suffices to show that for any $T > 0$, $\lim_n E \sup_{s \leq T} \|\mathbf{M}_n(s)\|_2^2 = 0$. Applying Doob's maximal inequality, we have that

$$(4.5) \quad E \sup_{s \leq T} \|\mathbf{M}_n(s)\|_2^2 \leq 4E \|\mathbf{M}_n(T)\|_2^2 = 4E \sum_{i \geq 1} M_{n,i}(T)^2.$$

Since $EM_{n,i}^2(T) = E\langle M_{n,i} \rangle_T$, using the monotone convergence theorem in (4.5) shows

$$(4.6) \quad E \sup_{s \leq T} \|\mathbf{M}_n(s)\|_2^2 \leq 4E \sum_{i \geq 1} \langle M_{n,i} \rangle_T \leq 4 \frac{T(1 + \sup_n \lambda_n)}{n},$$

where the last inequality is from (3.6) on observing that

$$\sum_{i=1}^{\infty} \langle M_{n,i} \rangle_T \leq \frac{1}{n} \int_0^T G_{n,1}(s) + \frac{\lambda_n}{n} \int_0^T \beta_n(G_{n,0}(t)) \leq \frac{T(1 + \lambda_n)}{n}.$$

Sending $n \rightarrow \infty$ in (4.6) completes the proof of the lemma. $\square$

The next proposition characterizes compact sets in $\ell_1^\downarrow \subseteq \ell_1 \cap [0, 1]^\infty$. The proof is standard and can be found for example in [31].

PROPOSITION 4.2. *A subset $C \subseteq \ell_1^\downarrow$ is precompact if and only if*

$$\limsup_{M \rightarrow \infty} \sup_{\mathbf{x} \in C} \sum_{i > M} |x_i| = 0.$$

The estimates in the following lemmas will be useful when applying Aldous–Kurtz tightness criteria [17], Theorem 23.11, for proving tightness of $\{\mathbf{G}_n\}_{n \geq 1}$.

LEMMA 4.3. *For each $n \in \mathbb{N}$, there is a square integrable $\{\mathcal{F}_t^n\}$ -martingale $\{L_n(t)\}$ such that, for any $t \geq 0$,*

$$\|\mathbf{G}_n\|_{1,t} \doteq \sup_{s \in [0,t]} \|\mathbf{G}_n(s)\|_1 \leq \|\mathbf{G}_n(0)\|_1 + \lambda_n t + L_n(t).$$

Furthermore, $\langle L_n \rangle_t \leq \frac{\lambda_n t}{n}$, for all $t \geq 0$.

PROOF. For $i = 1, \dots, n$, let $X_i(t)$ denote the number of jobs in the i th server's queue at time t . Then

$$\|\mathbf{G}_n(t)\|_1 = \sum_{j=1}^{\infty} G_{n,j}(t) = \sum_{j=1}^{\infty} \sum_{i=1}^n \frac{\mathbb{I}_{\{X_i(t) \geq j\}}}{n} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{\infty} \mathbb{I}_{\{X_i(t) \geq j\}} = \frac{1}{n} \sum_{i=1}^n X_i(t).$$

Hence $\|\mathbf{G}_n(t)\|_1$ is the total number of jobs in the system at time t , divided by n .

Since the total number of jobs in the system at time t is bounded above by the sum of number of job arrivals by time t and the initial number of jobs, $\sup_{s \in [0, t]} \|\mathbf{G}_n(s)\| \leq \|\mathbf{G}_n(0)\|_1 + \frac{A_n(t)}{n}$, where $A_n(t)$ is the total number of arrivals to the system by time t . Since A_n is a Poisson process with arrival rate $\lambda_n n$, the result follows on setting $L_n(t) = \frac{A_n(t)}{n} - \lambda_n t$, $t \geq 0$. $\square$

LEMMA 4.4. *Fix $n \in \mathbb{N}$ and $\delta \in (0, \infty)$. Let τ be a bounded $\{\mathcal{F}_t^n\}$ -stopping time. Then*

$$\mathbf{E} \|\mathbf{G}_n(\tau + \delta) - \mathbf{G}_n(\tau)\|_1 \leq (\lambda_n + 1)\delta.$$

PROOF. From (3.2), for any $i \in \mathbb{N}$,

$$(4.7) \quad |G_{n,i}(\tau + \delta) - G_{n,i}(\tau)| \leq \frac{1}{n}(A_{n,i}(\tau + \delta) - A_{n,i}(\tau) + D_{n,i}(\tau + \delta) - D_{n,i}(\tau)).$$

From (3.3) and (3.4), we see that

$$\begin{aligned} \mathbf{E} \frac{1}{n}(A_{n,i}(\tau + \delta) - A_{n,i}(\tau)) &= \lambda_n \mathbf{E} \int_{\tau}^{\tau + \delta} (\beta_n(G_{n,i-1}(s)) - \beta_n(G_{n,i}(s))) ds, \\ \mathbf{E} \frac{1}{n}(D_{n,i}(\tau + \delta) - D_{n,i}(\tau)) &= \mathbf{E} \int_{\tau}^{\tau + \delta} (G_{n,i}(s) - G_{n,i+1}(s)) ds. \end{aligned}$$

Using the above identities in (4.7),

$$\begin{aligned} \mathbf{E} |G_{n,i}(\tau + \delta) - G_{n,i}(\tau)| &\leq \lambda_n \mathbf{E} \int_{\tau}^{\tau + \delta} (\beta_n(G_{n,i-1}(s)) - \beta_n(G_{n,i}(s))) ds \\ &\quad + \mathbf{E} \int_{\tau}^{\tau + \delta} (G_{n,i}(s) - G_{n,i+1}(s)) ds. \end{aligned} \quad (4.8)$$

Adding (4.8) over various values of $i \in \mathbb{N}$, we have

$$\begin{aligned} \mathbf{E} \|\mathbf{G}_n(\tau + \delta) - \mathbf{G}_n(\tau)\|_1 &\leq \lambda_n \sum_{i=1}^{\infty} \mathbf{E} \int_{\tau}^{\tau + \delta} (\beta_n(G_{n,i-1}(s)) - \beta_n(G_{n,i}(s))) ds \\ &\quad + \sum_{i=1}^{\infty} \mathbf{E} \int_{\tau}^{\tau + \delta} (G_{n,i}(s) - G_{n,i+1}(s)) ds \\ &\leq \mathbf{E} \int_{\tau}^{\tau + \delta} (\lambda_n \beta_n(G_{n,0}(s)) + G_{n,1}(s)) ds \leq (\lambda_n + 1)\delta. \quad \square \end{aligned}$$

The following lemma will be useful in verifying the tightness of $\{\mathbf{G}_n(t)\}$ in $\ell_1^\downarrow$ for each fixed $t \geq 0$.

LEMMA 4.5. *For every $n, m \in \mathbb{N}$ there is a square integrable $\{\mathcal{F}_t^n\}$ martingale $L_{n,m}(\cdot)$ so that, for all $t \geq 0$,*

$$\sup_{s \leq t} \sum_{i > m} G_{n,i}(s) \leq \sum_{i > m} G_{n,i}(0) + \frac{\lambda_n t}{m} \|\mathbf{G}_n\|_{1,t} + L_{n,m}(t)$$

and $\langle L_{n,m} \rangle_t \leq \frac{\lambda_n t}{nm} \|\mathbf{G}_n\|_{1,t}$.

PROOF. From (3.1), for any $i \in \mathbb{N}$ and $t \geq 0$:

$$(4.9) \quad G_{n,i}(t) \leq G_{n,i}(0) + \frac{1}{n} N_{+,i} \left(n \lambda_n \int_0^t \beta_n(G_{n,i-1}(s)) - \beta_n(G_{n,i}(s)) ds \right).$$

Consider the point-process given by

$$B_{n,m}(t) \doteq \sum_{i>m} N_{+,i} \left(n \lambda_n \int_0^t \beta_n(G_{n,i-1}(s)) - \beta_n(G_{n,i}(s)) ds \right).$$

Adding over $i > m$ in (4.9), we get

$$(4.10) \quad \sup_{s \leq t} \sum_{i>m} G_{n,i}(s) \leq \sum_{i>m} G_{n,i}(0) + \frac{1}{n} B_{n,m}(t).$$

It is easy to see that with

$$b_{n,m}(t) \doteq n \lambda_n \sum_{i>m} \int_0^t \beta_n(G_{n,i-1}(s)) - \beta_n(G_{n,i}(s)) ds, \quad t \geq 0,$$

$\tilde{L}_{n,m}(t) \doteq B_{n,m}(t) - b_{n,m}(t)$ is a $\mathcal{F}_t^n$ -martingale and

$$\begin{aligned} \langle \tilde{L}_{n,m} \rangle_t &= b_{n,m}(t) = n \lambda_n \int_0^t \beta_n(G_{n,m}(s)) ds \leq n \lambda_n \int_0^t G_{n,m}(s) ds \\ &\leq n \lambda_n t \left(\sup_{s \leq t} G_{n,m}(s) \right) \leq \frac{n \lambda_n t}{m} \|G_n\|_{1,t}, \end{aligned}$$

where, for the last inequality we have used (4.1). The lemma now follows on setting $L_{n,m}(t) = \tilde{L}_{n,m}(t)/n$ and using (4.10). $\square$

Recall that under our assumptions, $\lambda_n \rightarrow \lambda$ and $d_n \rightarrow \infty$ as $n \rightarrow \infty$. We are now ready to prove that the sequence of processes $\{G_n\}_{n \geq 1}$ is tight.

LEMMA 4.6. *Suppose that $\{G_n(0)\}_{n \geq 1}$ is a tight sequence of $\ell_1^\downarrow$ valued random variables. Then for any $T > 0$, $\{G_n\}_{n \geq 1}$ is a tight sequence of $\mathbb{D}([0, T] : \ell_1^\downarrow)$ valued random variables.*

PROOF. To show that $\{G_n\}_{n \geq 1}$ is tight, it suffices to show the following two conditions:

- (1) For any $t \in [0, T]$ and $\epsilon > 0$, there is a compact set $\Gamma \subseteq \ell_1^\downarrow$ so that $\inf_{n \in \mathbb{N}} P(G_n(t) \in \Gamma) \geq 1 - \epsilon$.
- (2) For every sequence of nonnegative numbers δ_n converging to 0 as $n \rightarrow \infty$, and every sequence of $\mathcal{F}_t^n$ -stopping times τ_n such that $\tau_n \leq T$, $\limsup_{n \rightarrow \infty} E \|G_n(\tau_n + \delta_n) - G_n(\tau_n)\|_1 = 0$.

To see this, note using the Aldous–Kurtz criteria [17], Theorem 23.11, that the above conditions (1) and (2), respectively, imply the conditions (i) and (ii) of [17], Theorem 23.8. Since it is immediate from Lemma 4.4 that condition (2) is satisfied, we now focus on proving condition (1).

Fix $\epsilon > 0$. Let $\bar{\lambda} = \sup_{n \geq 1} \lambda_n$. Since $G_n(0)$ is tight, there is a compact $K_1 \subset \ell_1^\downarrow$ such that

$$P(G_n(0) \in K_1) \geq 1 - \frac{\epsilon}{8} \quad \text{for all } n \in \mathbb{N}.$$

In particular, since $\|\cdot\|_1$ is a continuous function, there is a $\kappa_1 \in (0, \infty)$ such that $\sup_{\mathbf{x} \in K_1} \|\mathbf{x}\|_1 \leq \kappa_1$. For the martingale sequence $\{L_n\}_{n \in \mathbb{N}}$ defined in Lemma 4.3, we can find $\kappa_2 \in (0, \infty)$ so that

$$P(|L_n|_{*,T} > \kappa_2) \leq \frac{E\langle L_n \rangle_T}{\kappa_2^2} \leq \frac{\bar{\lambda}T}{\kappa_2^2} \leq \frac{\epsilon}{8}$$

for each $n \in \mathbb{N}$. Then, using the above estimates in Lemma 4.3, with $\kappa = \bar{\lambda}T + \kappa_1 + \kappa_2$, we see

$$P(\|G_n\|_{1,T} \geq \kappa) \leq \frac{\epsilon}{4}$$

for each $n \in \mathbb{N}$. Let $m_k \uparrow \infty$ be a sequence such that $4 \frac{\bar{\lambda}T(\kappa+1)}{m_k^{1/2}} \leq \frac{\epsilon}{2^{k+1}}$ for all $k \in \mathbb{N}$. Define

$$K_2 = \left\{ \mathbf{y} \in \ell_1^\downarrow : \text{for some } \mathbf{x} \in K_1, \sum_{i > m_k} y_i \leq \sum_{i > m_k} x_i + \frac{\bar{\lambda}T\kappa}{m_k} + \frac{1}{m_k^{1/4}}, \forall k \in \mathbb{N} \right\}.$$

Since K_1 is compact, it is immediate from Proposition 4.2 that K_2 is precompact in $\ell_1^\downarrow$. Also, using Lemma 4.5, for any $t \in [0, T]$,

$$\begin{aligned} P(G_n(t) \in K_2^c) &\leq P(\|G_n\|_{1,T} \geq \kappa) + P(G_n(0) \in K_1^c) \\ &\quad + P\left(|L_{n,m_k}|_{*,T} > \frac{1}{m_k^{1/4}} \text{ for some } k \in \mathbb{N}, \|G_n\|_{1,T} \leq \kappa\right) \\ &\leq \frac{\epsilon}{4} + \frac{\epsilon}{8} + 4(\kappa+1)\bar{\lambda}T \sum_{k=1}^{\infty} m_k^{1/2} \frac{1}{m_k} \leq \epsilon, \end{aligned}$$

where the second inequality follows from the application of Doob's maximal inequality with the stopping time $\tau \doteq \inf\{t \mid \|G_n(t)\|_1 > \kappa\}$ and from the expression of $\langle L_{n,m_k} \rangle$ in Lemma 4.5, and the third inequality follows from the choice of $\{m_k\}_{k \in \mathbb{N}}$. This proves (1) and completes the proof of the lemma. $\square$

The following lemma gives a characterization of the limit points of G_n .

LEMMA 4.7. *Fix $T \in (0, \infty)$. Suppose that, along some subsequence $\{n_k\}_{k \geq 1}$, $G_{n_k} \Rightarrow G$ in $\mathbb{D}([0, T] : \ell_1^\downarrow)$ as $k \rightarrow \infty$. Then $G \in \mathbb{C}([0, T] : \ell_1^\downarrow)$ a.s., and (2.2) is satisfied with (g_i, v_i) replaced with (G_i, V_i) , where V_i are defined recursively using the second equation in (2.2) with $V_0(t) = \lambda t$ for $t \geq 0$.*

PROOF. From Lemma 4.1, we see that $M_{n_k} \xrightarrow{P} 0$, in $\mathbb{D}([0, T] : \ell_2)$. By Skorohod embedding theorem, let us assume that G_{n_k}, M_{n_k}, G are all defined on the same probability space and

$$(G_{n_k}, M_{n_k}) \rightarrow (G, 0), \quad \text{a.s.}$$

in $\mathbb{D}([0, T] : \ell_1^\downarrow \times \ell_2)$. Since the jumps of G_n have size at most $1/n$, G is continuous and $\|G(s) - G_{n_k}(s)\|_{1,T} \rightarrow 0$ a.s. Similarly, $\|M_{n_k}(s)\|_{2,T} \rightarrow 0$ a.s. To simplify notation from now on, we will take $n_k = n$.

Let $V_{n,i}(t) \doteq \lambda_n \int_0^t \beta_n(G_{n,i}(s)) ds$ for $i \geq 1$ and $V_{n,0}(t) \doteq \lambda_n t$. From (3.5), for any $i \geq 1$,

$$(4.11) \quad G_{n,i}(t) = G_{n,i}(0) - \int_0^t (G_{n,i}(s) - G_{n,i+1}(s)) ds + V_{n,i-1}(t) - V_{n,i}(t) + M_{n,i}(t).$$

For $i \in \mathbb{N}$, $\sup_{s \leq T} |G_{n,i}(s) - G_i(s)| \leq \sup_{s \leq T} \|\mathbf{G}_n(s) - \mathbf{G}(s)\|_1 \rightarrow 0$ and $\sup_{s \leq T} |M_{n,i}(s)| \leq \sup_{s \leq T} \|\mathbf{M}_n(s)\|_2 \rightarrow 0$, a.s. as $n \rightarrow \infty$. We now show that, for each $i \in \mathbb{N}_0$, $V_{n,i}$ converges uniformly on $[0, T]$ (a.s.) to some limit process V_i . Clearly, this is true for $i = 0$ and in fact $V_0(t) = \lambda t$, $t \geq 0$. Proceeding recursively, suppose now that $V_{n,i-1} \rightarrow V_{i-1}$ on $[0, T]$ for some $i \geq 1$. Then, since all the terms in (4.11), except $V_{n,i}$, converge uniformly, $V_{n,i}$ must converge uniformly as well to some limit process V_i . Sending $n \rightarrow \infty$ in (4.11), we get for every $t \leq T$ and $i \geq 1$:

$$G_i(t) = G_i(0) - \int_0^t (G_i(s) - G_{i+1}(s)) ds + V_{i-1}(t) - V_i(t), \quad \text{a.s.}$$

This shows the first line in (2.3) is satisfied with (g_i, v_i) replaced with (G_i, V_i) .

We now show that the second line in (2.3) is satisfied as well. Since V_i is the limit of $\{V_{n,i}\}$, the following properties hold:

- (i) $V_0(t) = \lambda t$ for all $t \in [0, T]$.
- (ii) V_i is continuous, nondecreasing and $V_i(0) = 0$.
- (iii) For any $t \in [0, T]$, $\int_0^t (1 - G_i(s)) dV_i(s) = 0$. This is a consequence of the following identities:

$$\begin{aligned} \int_0^t (1 - G_i(s)) dV_i(s) &= \lim_n \int_0^t (1 - G_i(s)) dV_{n,i}(s) \\ &= \lim_n \int_0^t \lambda_n (1 - G_i(s)) \beta_n(G_{n,i}(s)) ds \\ &= \lambda \int_0^t \lim_{n \rightarrow \infty} (1 - G_i(s)) \beta_n(G_{n,i}(s)) ds \\ &= 0, \end{aligned}$$

where the first equality holds since G_i is a continuous and bounded function and $V_{n,i} \rightarrow V_i$ uniformly on $[0, T]$; the second equality uses the definition of $V_{n,i}$, the third is from the dominated convergence theorem and the fourth follows since $\beta_n(x) \leq x^{d_n}$, for $x \in [0, 1]$ and $d_n \rightarrow \infty$, $\beta_n(x) \rightarrow 0$ for every $x \in [0, 1)$.

Thus we have verified that the second line in (2.3) is satisfied with (G_i, V_i) as well. The result is now immediate from Remark 2.2. $\square$

4.3. Completing the proof of LLN. We can now complete the proofs of Proposition 2.1 and Theorem 2.1.

PROOF OF PROPOSITION 2.1. Fix $\mathbf{r} \in \ell_1^\downarrow$, $\lambda > 0$ and choose a sequence $\mathbf{r}_n \in \ell_1^\downarrow$ such that $\mathbf{r}_n \rightarrow \mathbf{r}$ in $\ell_1^\downarrow$ and for each i , $n\mathbf{r}_{n,i} \in \mathbb{N}_0$. Consider parameters $\lambda_n = \lambda$, $d_n = n$ and a JSQ(d_n) system initialized at $\mathbf{G}_n(0) = \mathbf{r}_n$. From Lemma 4.7, we have that there is at least one solution of (2.2), which is given as a limit point of an arbitrary weakly convergent subsequence of $\mathbf{G}_n$ (such a sequence exists in view of the tightness shown in Lemma 4.6). The fact that this equation can have at most one solution was shown in Section 4.1. The result follows. $\square$

PROOF OF THEOREM 2.1. Since $\mathbf{G}_n(0) \xrightarrow{P} \mathbf{r}$ in $\ell_1^\downarrow$, the hypothesis of Lemma 4.6 is satisfied, and thus the sequence $\{\mathbf{G}_n\}_{n \geq 1}$ is tight in $\mathbb{D}([0, T] : \ell_1^\downarrow)$ for any fixed $T > 0$. The result is now immediate from Lemma 4.7 and unique solvability of (2.2) shown in Proposition 2.1. $\square$

REMARK 4.8. We note that the proofs of Lemma 4.7 and Theorem 2.1 also show that, under the conditions of Theorem 2.1, for each $i \geq 1$,

$$\sup_{t \leq T} \left| \lambda_n \int_0^t \beta_n(G_{n,i}(s)) ds - v_i(t) \right| \xrightarrow{P} 0,$$

where (g_i, v_i) is the unique solution of (2.2).

5. Properties of the near fixed point. In this section, we give some important properties of the near fixed point μ_n that will be needed in the proofs of fluctuation theorems. Since μ_n is defined in terms of the function β_n , we begin by giving some results on the asymptotic behavior of β_n and its derivatives. Proofs follow via elementary algebra and Taylor's approximation and can be found in Appendix A. Roughly speaking, these results control the error between sampling with and without replacement of d_n servers from a collection of n servers. We first note that the function β_n is differentiable on $(0, 1) \setminus \{\frac{d_n-1}{n}\}$ and the derivative is given as

$$(5.1) \quad \dot{\beta}_n(x) = \sum_{j=0}^{d_n-1} (1 - j/n)^{-1} \prod_{\substack{i=0 \\ i \neq j}}^{d_n-1} \frac{x - i/n}{1 - i/n}$$

for $x \in \left(\frac{d_n-1}{n}, 1\right]$ and $\dot{\beta}_n(x) = 0$ for $x \in \left(0, \frac{d_n-1}{n}\right)$.

As a convention, we set $\dot{\beta}_n(x) = 0$ for $x = \frac{d_n-1}{n}$.

Note that $h(t) = \frac{a+t}{b+t}$ is an increasing function of t on $(-b, \infty)$ when $b > a$. Using this fact in (2.7) shows that, when $d_n \leq n$,

$$(5.2) \quad 0 \leq \beta_n(x) \leq x^{d_n} \doteq \gamma_n(x), \quad x \in [0, 1].$$

Using the same fact in (5.1) shows that, for $d_n < n$,

$$(5.3) \quad 0 \leq \dot{\beta}_n(x) \leq \frac{d_n x^{d_n-1}}{1 - \frac{d_n}{n}}, \quad x \in (0, 1).$$

The following lemma estimates the ratio between β_n and γ_n and its derivatives.

LEMMA 5.1. Assume $d_n \ll n$. Then for any $\epsilon \in (0, 1)$, as $n \rightarrow \infty$,

$$(5.4) \quad \sup_{x \in [\epsilon, 1]} \left| \frac{\dot{\beta}_n(x)/\beta_n(x)}{\dot{\gamma}_n(x)/\gamma_n(x)} - 1 \right| \rightarrow 0.$$

Furthermore, if $d_n \ll \sqrt{n}$, then

$$(5.5) \quad \sup_{x \in [\epsilon, 1]} \left| \frac{\beta_n(x)}{\gamma_n(x)} - 1 \right| \rightarrow 0 \quad \text{and} \quad \sup_{x \in [\epsilon, 1]} \left| \frac{\dot{\beta}_n(x)}{\dot{\gamma}_n(x)} - 1 \right| \rightarrow 0.$$

The next corollary follows from the proof of Lemma 5.1 (specifically the estimate (A.3) in the proof of the lemma).

COROLLARY 5.2. Assume $d_n \ll n$. Then for any $\epsilon \in (0, 1)$,

$$\sup_{x \in [\epsilon, 1]} |\log \beta_n(x) - \log \gamma_n(x)| = O\left(\frac{d_n^2}{n}\right).$$

Recall the near fixed points $\mu_n = (\mu_{n,i})_{i \geq 1}$ introduced in Definition 2.

COROLLARY 5.3. *Suppose that $d_n \ll n$. Let $i \in \mathbb{N}$ be such that $\liminf_n \mu_{n,i} > 0$. Then*

$$\lim_{n \rightarrow \infty} \frac{\lambda_n \mu_{n,i} \dot{\beta}_n(\mu_{n,i})}{d_n \mu_{n,i+1}} = 1.$$

LEMMA 5.4. *Assume $d_n \ll n$ and fix $\epsilon \in (0, 1)$. Then there is a $C \in (0, \infty)$ and $n_0 \in \mathbb{N}$ such that, if for some $k \in \mathbb{N}$ and $n_1 \in \mathbb{N}$, $\mu_{n,k} \geq \epsilon$ for all $n \geq n_1$, then for all $n \geq n_1 \vee n_0$,*

$$\left| \log \mu_{n,k+1} - (\log \lambda_n) \left(\sum_{i=0}^k d_n^i \right) \right| \leq \frac{C}{n} \sum_{i=1}^k d_n^{i+1}.$$

COROLLARY 5.5. *Suppose for some $k \in \mathbb{N}$, $1 \ll d_n^{k+1} \ll n$. Suppose also that $1 - \lambda_n = \frac{\xi_n + \log d_n}{d_n^k}$ where $\xi_n \rightarrow -\log(\alpha) \in (-\infty, \infty]$ and $\frac{\xi_n^2}{d_n} \rightarrow 0$ as $n \rightarrow \infty$. Then $\mu_{n,k} \rightarrow 1$ and $\dot{\beta}_n(\mu_{n,k}) \rightarrow \alpha$ as $n \rightarrow \infty$.*

LEMMA 5.6. *Suppose that $\lambda_n \nearrow 1$, and $1 \ll d_n \ll n$. Suppose also that, for some $k \geq 2$, $\mu_{n,k} \rightarrow 1$ and $\dot{\beta}_n(\mu_{n,k}) \rightarrow \alpha \in [0, \infty)$ as $n \rightarrow \infty$. Then as $n \rightarrow \infty$, $\dot{\beta}_n(\mu_{n,1}) \rightarrow \infty$ and for any $i \in [k - 1]$,*

$$\frac{\dot{\beta}_n(\mu_{n,i})}{\dot{\beta}_n(\mu_{n,1})} \rightarrow 1.$$

The following result is along the lines of Lemma 5.1. It allows for weaker assumptions on d_n but gives an approximation only in a neighborhood of 1.

LEMMA 5.7. *Suppose that $d_n \ll n^{2/3}$. Let $\{\epsilon_n\}$ be a sequence in $[0, 1]$ such that $\epsilon_n^2 \ll d_n^{-1}$. Then as $n \rightarrow \infty$:*

(5.6)
$$\sup_{x \in [1-\epsilon_n, 1]} \left| \frac{\beta_n(x)}{\gamma_n(x)} - 1 \right| \rightarrow 0,$$

and

(5.7)
$$\sup_{x \in [1-\epsilon_n, 1]} \left| \frac{\dot{\beta}_n(x)}{\dot{\gamma}_n(x)} - 1 \right| \rightarrow 0.$$

The next result shows that if $d_n \rightarrow \infty$, then the behavior of $\beta_n(x)$ is interesting only when x is sufficiently close to 1.

LEMMA 5.8. *Suppose that $d_n \gg 1$, and let $\epsilon_n \doteq \frac{2 \log d_n}{d_n}$. Then as $n \rightarrow \infty$, $\sup_{x \in [0, 1-\epsilon_n]} |\beta_n(x)| \rightarrow 0$. Furthermore, if $\limsup_n \frac{d_n}{n} < 1$ then we also have $\sup_{x \in [0, 1-\epsilon_n]} |\dot{\beta}_n(x)| \rightarrow 0$.*

6. Preliminary estimates under diffusion scaling. Recall the near fixed point μ_n from Definition 2 and the process Z_n introduced in (1.2). Also, recall the maps $\mathbf{a}_n$ and $\mathbf{b}$ from Remark 3.1. We will extend the definition of β_n and $\dot{\beta}_n$ to $\mathbb{R}$ by setting $\beta_n(x) = \dot{\beta}_n(x) = 0$ for $x < 0$. Further, in what follows, for $z < 0$ and real valued integrable function $h(\cdot)$, the integral $\int_{[0,z]} h(u) du = - \int_{[z,0]} h(u) du$. We start by giving a semimartingale decomposition for Z_n . The quantity $\mathbf{A}_n(z)$ defined in the following lemma can be viewed as a discrete derivative

of $\mathbf{a}_n$ at $\boldsymbol{\mu}_n$ in the direction $\mathbf{z}$. The function $\mathbf{A}_n$ is asymptotically linear under conditions of Theorem 2.2 (see Lemma 7.1), and is asymptotically nonlinear under conditions of Theorems 2.3 and 2.4 (see Lemma 6.7). The asymptotic analysis of this map and the resulting system $\mathbf{Z}_n$ is a key ingredient in our proofs.

LEMMA 6.1. *For $t \geq 0$, $\mathbf{Z}_n(t)$ satisfies*

$$(6.1) \quad \mathbf{Z}_n(t) = \mathbf{Z}_n(0) + \int_0^t \mathbf{A}_n(\mathbf{Z}_n(s)) ds - \int_0^t \mathbf{b}(\mathbf{Z}_n(s)) ds + \sqrt{n} \mathbf{M}_n(t),$$

where $\mathbf{A}_n : \ell_\infty \rightarrow \ell_\infty$, is defined as $\mathbf{A}_n(\mathbf{z}) \doteq \sqrt{n}\{\mathbf{a}_n(\boldsymbol{\mu}_n + n^{-1/2}\mathbf{z}) - \mathbf{a}_n(\boldsymbol{\mu}_n)\}$. Moreover,

$$(6.2) \quad \mathbf{A}_n(\mathbf{z})_i = q_{n,i-1}(z_{i-1}) - q_{n,i}(z_i), \quad i \in \mathbb{N}$$

where

$$(6.3) \quad \begin{aligned} q_{n,i}(z) &\doteq \lambda_n \int_{[0,z]} \dot{\beta}_n(\mu_{n,i} + y/\sqrt{n}) dy, \quad z \in \mathbb{R}, i \in \mathbb{N}, \\ q_{n,0}(z) &\doteq 0, \quad z \in \mathbb{R}. \end{aligned}$$

PROOF. From (3.7) and since $\mathbf{a}_n(\boldsymbol{\mu}_n) = \mathbf{b}(\boldsymbol{\mu}_n)$,

$$\begin{aligned} \sqrt{n}(\mathbf{G}_n(t) - \boldsymbol{\mu}_n) &= \sqrt{n}(\mathbf{G}_n(0) - \boldsymbol{\mu}_n) + \int_0^t \sqrt{n}\{\mathbf{a}_n(\mathbf{G}_n(s)) - \mathbf{a}_n(\boldsymbol{\mu}_n)\} ds \\ &\quad - \int_0^t \sqrt{n}\{\mathbf{b}(\mathbf{G}_n(s)) - \mathbf{b}(\boldsymbol{\mu}_n)\} ds + \sqrt{n} \mathbf{M}_n(t). \end{aligned}$$

Now (6.1) follows by using the definition of $\mathbf{Z}_n$ and $\mathbf{A}_n$, and the linearity of $\mathbf{b}$. Further, using the definition of $\mathbf{a}_n$, we see that (6.2) holds where

$$(6.4) \quad q_{n,i}(z) \doteq \begin{cases} \lambda_n \sqrt{n}\{\beta_n(\mu_{n,i} + z/\sqrt{n}) - \beta_n(\mu_{n,i})\} & \text{for } i \geq 1, \\ 0 & \text{if } i = 0. \end{cases}$$

Clearly, the $q_{n,i}$ defined in (6.4) is same as that given in (6.3). The result follows. $\square$

LEMMA 6.2. *Suppose that $d_n \rightarrow \infty$, $\lambda_n \rightarrow 1$, and for some $k \geq 1$, $\mathbf{G}_n(0) \xrightarrow{P} \mathbf{f}_k$ in ℓ_1 . Then there is a standard Brownian motion B so that $\sqrt{n} \mathbf{M}_n \Rightarrow \sqrt{2} B \mathbf{e}_k$ in $\mathbb{D}([0, \infty) : \ell_2)$.*

PROOF. Fix $T > 0$. Since $\mathbf{G}_n(0) \rightarrow \mathbf{f}_k$ and $\mathbf{f}_k$ is a fixed point of (2.2), by Theorem 2.1, $\mathbf{G}_n \xrightarrow{P} \mathbf{f}_k$ in $\mathbb{D}([0, T] : \ell_1)$, where $\mathbf{f}_k$ here is viewed as the function on $[0, T]$ that takes the constant value $\mathbf{f}_k \in \ell_1^\downarrow$. Moreover, by Remark 4.8, for every $i \geq 1$ $V_{n,i}(t) \doteq \lambda_n \int_0^t \beta_n(G_{n,i}(s)) ds$ converges uniformly on $[0, T]$ in probability to $v_i(t)$, where v_i solves

$$(6.5) \quad v_i = \hat{\Gamma}_1(f_{k,i} - (f_{k,i} - f_{k,i+1})\text{id} + v_{i-1}(\cdot)), \quad i \geq 1,$$

and $v_0(t) \doteq t$, where recall that ‘id’ denotes the identity map on $[0, T]$. Recalling the definition of $\mathbf{f}_k$ we see by a recursive argument that

$$(6.6) \quad v_i(t) \doteq \begin{cases} t & \text{if } i < k, \\ 0 & \text{if } i \geq k. \end{cases}$$

Combining this with (3.6), we have for each $i \geq 1$,

$$\begin{aligned} \langle \sqrt{n} \mathbf{M}_{n,i} \rangle &= \int_0^\cdot (G_{n,i}(s) - G_{n,i+1}(s)) ds + \lambda_n \int_0^\cdot (\beta_n(G_{n,i-1}(s)) - \beta_n(G_{n,i}(s))) ds \\ &\quad \rightarrow (f_{k,i} - f_{k,i+1})\text{id} + v_{i-1}(\cdot) - v_i(\cdot) = H(\cdot), \end{aligned}$$

in probability in $\mathbb{C}([0, T] : \mathbb{R})$ where

$$H(t) \doteq \begin{cases} 2t & \text{if } i = k, \\ 0 & \text{if } i \neq k, \end{cases} \quad t \in [0, T].$$

Adding (3.6) over i , we have for $t \in [0, T]$,

$$(6.7) \quad \sum_{i>k} \langle \sqrt{n} M_{n,i} \rangle_t \leq \int_0^t G_{n,k+1}(s) ds + \lambda_n \int_0^t \beta_n(G_{n,k})(s) ds.$$

The process on the right-hand side converges in probability in $\mathbb{C}([0, T] : \mathbb{R})$ to $f_{k,k+1} \text{id} + v_k(\cdot) = 0$ and thus $\sum_{i>k} \langle \sqrt{n} M_{n,i} \rangle_T$ converges to 0 in probability. By Doob's maximal inequality,

$$nE \sup_{t \leq T} \sum_{i>k} M_{n,i}^2(t) \leq 4E \sum_{i>k} \langle \sqrt{n} M_{n,i} \rangle_T \rightarrow 0, \quad \text{as } n \rightarrow \infty,$$

where the last convergence follows by the dominated convergence theorem on noting that the right-hand side of (6.7) is bounded above by $\sup_n (1 + \lambda_n) < \infty$. The result now follows on using the martingale central limit theorem (cf. [11], Theorem 7.1.4) for the k -dimensional martingale sequence $(\sqrt{n} M_{n,1}, \dots, \sqrt{n} M_{n,k})$. $\square$

Recall the functions $q_{n,i}$ from Lemma 6.1.

LEMMA 6.3. *Assume that for some $r \in \mathbb{N}$, $\limsup_{n \rightarrow \infty} \mu_{n,r} < 1$. Then for any $L > 0$,*

$$\limsup_{n \rightarrow \infty} \sup_{i \geq r} \sup_{0 < |z| \leq L} \left| \frac{q_{n,i}(z)}{z} \right| = 0.$$

PROOF. By (6.3),

$$\begin{aligned} \sup_{i \geq r} \sup_{0 < |z| \leq L} \left| \frac{q_{n,i}(z)}{z} \right| &\leq \lambda_n \sup_{i \geq r} \sup_{0 < |z| \leq L} \sup_{|y| \leq z} \left| \dot{\beta}_n \left(\mu_{n,i} + \frac{y}{\sqrt{n}} \right) \right| \\ &= \lambda_n \sup_{i \geq r} \sup_{|z| \leq L} \left| \dot{\beta}_n \left(\mu_{n,i} + \frac{z}{\sqrt{n}} \right) \right| \leq \lambda_n \sup_{0 \leq x \leq \mu_{n,r} + \frac{L}{\sqrt{n}}} \dot{\beta}_n(x) \end{aligned}$$

which converges to 0 by Lemma 5.8, since $\limsup_{n \rightarrow \infty} (\mu_{n,r} + \frac{L}{\sqrt{n}}) < 1$. $\square$

For $L \in (0, \infty)$, define the stopping time

$$(6.8) \quad \tau_{n,L} \doteq \inf \left\{ t \mid \|Z_n(t)\|_2 \geq L - \frac{1}{\sqrt{n}} \right\}.$$

Since the jumps of Z_n are of size $\frac{1}{\sqrt{n}}$, we see that, for any $T > 0$,

$$(6.9) \quad \|Z_n\|_{2,T \wedge \tau_{n,L}} \leq L.$$

Recall from Section 1.2 the vector $\mathbf{z}_{r+} \in \mathbb{R}^\infty$ associated with a vector $\mathbf{z} \in \mathbb{R}^\infty$.

LEMMA 6.4. *Suppose that as $n \rightarrow \infty$, $\mathbf{G}_n(0) \xrightarrow{P} \mathbf{f}_k$ in $\ell_1^\downarrow$ and $\mathbf{Z}_{n,r+}(0) \xrightarrow{P} \mathbf{0}$ in ℓ_2 for some $r > k$. Then for any $T, L > 0$, $\|\mathbf{Z}_{n,r+}\|_{2,T \wedge \tau_{n,L}} \xrightarrow{P} 0$.*

PROOF. For $i > k$ and $z \in \mathbb{R}$, let $\Delta_{n,i}(z) \doteq \frac{q_{n,i}(z)}{z} \mathbb{I}_{\{z \neq 0\}}$. Then, since $\lim_{n \rightarrow \infty} \mu_{n,k+1} = 0$, by Lemma 6.3,

$$(6.10) \quad \delta_{n,L} \doteq \sup_{i \geq k+1} \sup_{|z| \leq L} |\Delta_{n,i}(z)| \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

Next, from (6.1), for $i \geq r+1 > k+1$,

$$\begin{aligned} Z_{n,i}(t \wedge \tau_n) &= Z_{n,i}(0) + \int_0^{t \wedge \tau_n} \Delta_{n,i-1}(Z_{n,i-1}(s)) Z_{n,i-1}(s) ds \\ &\quad - \int_0^{t \wedge \tau_n} \Delta_{n,i}(Z_{n,i}(s)) Z_{n,i}(s) ds \\ &\quad - \int_0^{t \wedge \tau_n} (Z_{n,i}(s) - Z_{n,i+1}(s)) ds + \sqrt{n} M_{n,i}(t \wedge \tau_n), \end{aligned}$$

where we use τ_n instead of $\tau_{n,L}$ for notational simplicity. Then, observing from (6.10) that $\sup_{i \geq k+1} \sup_{t \in [0, \tau_n]} |\Delta_{n,i}(Z_{n,i}(t))| \leq \delta_{n,L}$, we have

$$(6.11) \quad \begin{aligned} |Z_{n,i}(t \wedge \tau_n)| &\leq |Z_{n,i}(0)| + \delta_{n,L} \int_0^{t \wedge \tau_n} (|Z_{n,i-1}(s)| + |Z_{n,i}(s)|) ds \\ &\quad + \int_0^{t \wedge \tau_n} (|Z_{n,i}(s)| + |Z_{n,i+1}(s)|) ds + |\sqrt{n} M_{n,i}(t \wedge \tau_n)|. \end{aligned}$$

Define maps $\mathbf{F}_1, \mathbf{F}_2 : \mathbb{R}^\infty \rightarrow \mathbb{R}^\infty$ by

$$\begin{aligned} (\mathbf{F}_1 \mathbf{x})_i &= \begin{cases} x_1 & i = 1, \\ x_{i-1} + x_i & i \geq 2, \end{cases} \\ (\mathbf{F}_2 \mathbf{x})_i &= x_i + x_{i+1}, \quad i \in \mathbb{N}. \end{aligned}$$

Then by collecting (6.11) over all $i \geq r+1$, we get

$$(6.12) \quad \begin{aligned} |Z_{n,r+}(t \wedge \tau_n)| &\leq |Z_{n,r+}(0)| + \delta_{n,L} \int_0^{t \wedge \tau_n} \mathbf{F}_1 |Z_{n,r+}(s)| ds \\ &\quad + \delta_{n,L} \int_0^{t \wedge \tau_n} |Z_{n,r}(s)| \mathbf{e}_1 ds \\ &\quad + \int_0^{t \wedge \tau_n} \mathbf{F}_2 |Z_{n,r+}(s)| ds + |\sqrt{n} \mathbf{M}_{n,r+}(t \wedge \tau_n)|, \end{aligned}$$

where the absolute values $|z| \in \mathbb{R}^\infty$ and integrals are interpreted as being coordinatewise for infinite dimensional vectors $\mathbf{z} \in \mathbb{R}^\infty$. Now noting that the maps $\mathbf{F}_i$, when considered from $\ell_2 \rightarrow \ell_2$, are bounded linear operators with norm bounded by 2, we have for $i = 1, 2$,

$$\left\| \int_0^{t \wedge \tau_n} \mathbf{F}_i |Z_{n,r+}(s)| ds \right\|_2 \leq \int_0^{t \wedge \tau_n} 2 \|Z_{n,r+}(s)\|_2 ds.$$

Using the triangle inequality in (6.12) shows for any $t \leq T$,

$$\begin{aligned} \|Z_{n,r+}(t \wedge \tau_n)\|_2 &\leq \|Z_{n,r+}(0)\|_2 + \|\sqrt{n} \mathbf{M}_{n,r+}\|_{2,T} + \delta_{n,L} L T \\ &\quad + 2(1 + \delta_{n,L}) \int_0^{t \wedge \tau_n} \|Z_{n,r+}(s)\|_2 ds, \end{aligned}$$

where we have used that $\int_0^{t \wedge \tau_n} |Z_{n,r}(s)| ds \leq L t$. Hence, using Gronwall's inequality,

$$\|Z_{n,r+}\|_{2,T \wedge \tau_n} \leq (\|Z_{n,r+}(0)\|_2 + \delta_{n,L} L T + \|\sqrt{n} \mathbf{M}_{n,r+}\|_{2,T}) e^{2(1+\delta_{n,L})T}.$$

Now, as $n \rightarrow \infty$, $\|\mathbf{Z}_{n,r+}(0)\|_2 \xrightarrow{P} 0$ by assumption, $\delta_{n,L} \rightarrow 0$ by (6.10), and $\|\sqrt{n} \times \mathbf{M}_{n,r+}\|_{2,T} \xrightarrow{P} 0$ by Lemma 6.2. The result follows. $\square$

The following elementary lemma will allow us to replace $\tau_{n,L} \wedge T$ with T in various convergence results. The proof is omitted.

LEMMA 6.5. *Fix $T \in [0, \infty)$. Suppose for each $n \in \mathbb{N}$ and $L > 0$ that $\tau_{n,L}$ is a $[0, T]$ valued random variable such that $\lim_{L \rightarrow \infty} \sup_n \mathbf{P}(\tau_{n,L} < T) \rightarrow 0$ for some $T > 0$. Suppose that there is a sequence of stochastic processes $\{F_n\}_{n \in \mathbb{N}}$ with sample paths in $\mathbb{D}([0, T] : \mathbb{R})$ such that for each $L > 0$ $|F_n|_{*,T \wedge \tau_{n,L}} \xrightarrow{P} 0$ as $n \rightarrow \infty$. Then in fact $|F_n|_{*,T} \xrightarrow{P} 0$ as $n \rightarrow \infty$.*

The next lemma gives conditions under which the near fixed point μ_n converges to $\mathbf{f}_1$.

LEMMA 6.6. *Let $0 \leq \epsilon_n \doteq 1 - \lambda_n$ be such that $\epsilon_n \rightarrow 0$ and $\epsilon_n d_n \rightarrow \infty$. Then $\mu_n \rightarrow \mathbf{f}_1$ in ℓ_1 as $n \rightarrow \infty$.*

PROOF. Using Definition 2 and (5.2) note that $0 \leq \mu_{n,i+1} = \lambda_n \beta_n(\mu_{n,i}) \leq \mu_{n,i}^{d_n}$ for each $i \geq 1$. Hence in order to show $\mu_n \rightarrow \mathbf{f}_1$ in ℓ_1 , it suffices to show that (1) $\mu_{n,1} \rightarrow 1$, and (2) $\mu_{n,2} \rightarrow 0$. This convergence is immediate on observing for (1) that $\mu_{n,1} = \lambda_n = 1 - \epsilon_n \rightarrow 1$, and for (2) that $\mu_{n,2} \leq \mu_{n,1}^{d_n} = (1 - \epsilon_n)^{d_n} \leq e^{-\epsilon_n d_n} \rightarrow 0$. $\square$

The following lemma gives a convenient approximation of the term $q_{n,1}$ introduced in (6.3) in terms of certain exponentials.

LEMMA 6.7. *Suppose $d_n \rightarrow \infty$ and $d_n \ll n^{2/3}$. Let $\lambda_n = 1 - (\frac{\log d_n}{d_n} + \frac{\alpha_n}{\sqrt{n}})$ for some real sequence $\{\alpha_n\}_{n \geq \mathbb{N}}$ satisfying $\frac{d_n \alpha_n^2}{n} \rightarrow 0$. Then, for any $L > 0$,*

$$(6.13) \quad \limsup_{n \rightarrow \infty} \sup_{0 < |z| \leq L} \left| \frac{\exp(\frac{d_n}{\sqrt{n}}(z - \alpha_n)) - \exp(-\frac{d_n}{\sqrt{n}}\alpha_n)}{q_{n,1}(z)d_n/\sqrt{n}} - 1 \right| = 0.$$

PROOF. We only consider the case $0 < z \leq L$. The case $-L \leq z < 0$ is treated similarly. Recall that $\mu_{n,1} = \lambda_n$. Noting that $d_n(1 - \lambda_n + \frac{L}{\sqrt{n}})^2 \leq 4d_n(\frac{\log^2 d_n}{d_n^2} + \alpha_n^2/n + L^2/n) \rightarrow 0$ we have on applying Lemma 5.7 with $\epsilon_n = (1 - \lambda_n + \frac{L}{\sqrt{n}})$ that, for any $|z| \leq L$,

$$\begin{aligned} q_{n,1}(z) &= (1 + o(1)) \int_0^z \dot{\gamma}_n \left(\lambda_n + \frac{y}{\sqrt{n}} \right) dy \\ &= (1 + o(1)) \int_0^z \exp \left((d_n - 1) \log \left\{ \lambda_n + \frac{y}{\sqrt{n}} \right\} + \log d_n \right) dy \\ &= (1 + o(1)) \int_0^z \exp \left(d_n \log \left\{ \lambda_n + \frac{y}{\sqrt{n}} \right\} + \log d_n \right) dy. \end{aligned}$$

Using expansion for $\log(1+h)$ around $h=0$ and once more the fact that $d_n(1 - \lambda_n + \frac{L}{\sqrt{n}})^2 \rightarrow 0$,

$$\begin{aligned} q_{n,1}(z) &= (1 + o(1)) \int_0^z \exp \left(d_n \left\{ \lambda_n - 1 + \frac{y}{\sqrt{n}} \right\} + \log d_n \right) dy \\ &= (1 + o(1)) \int_0^z \exp \left(\frac{d_n}{\sqrt{n}}(y - \alpha_n) \right) dy \end{aligned}$$

$$= (1 + o(1)) \frac{\exp(\frac{d_n}{\sqrt{n}}(z - \alpha_n)) - \exp(-\frac{d_n}{\sqrt{n}}\alpha_n)}{d_n/\sqrt{n}}$$

which proves (6.13). $\square$

Proof of the following lemma proceeds by standard arguments but we provide details in Appendix B.

LEMMA 6.8. Fix $T > 0$. Let g, h, M be three bounded measurable functions from $[0, T] \rightarrow \mathbb{R}$ and assume further that M is a right-continuous bounded variation function. Suppose that $m \doteq \inf_{s \in [0, T \wedge \tau]} h(s) > 0$ for some $\tau \geq 0$. Let $z : [0, T] \rightarrow \mathbb{R}$ be a bounded measurable function that satisfies for every $t \in [0, T]$,

$$(6.14) \quad z(t) = z(0) - \int_0^t h(s)z(s) ds + \int_0^t g(s) ds + M(t).$$

Then for any $t \in [0, T \wedge \tau]$,

$$|z(t)| \leq \frac{|g|_{*, T \wedge \tau}}{m} + 2|M|_{*, T \wedge \tau} + e^{-mt} |z(0)|.$$

LEMMA 6.9. Fix $T \in (0, \infty)$. For each n , let V_n be a martingale with respect to some filtration $\{\mathcal{G}_t^n\}$ such that $V_n(0) = 0$. Let $(r_n)_{n=1}^\infty$ be a positive sequence so that $\lim_{n \rightarrow \infty} r_n = +\infty$. Suppose that there is a $C \in (0, \infty)$ such that for all $n \in \mathbb{N}$ and $t \in [0, T]$, $\langle V_n \rangle_t \leq Ct$. Then for any $\epsilon > 0$,

$$P\left(\sup_{t \leq T} (V_n(t) - r_n t) > \epsilon\right) \rightarrow 0$$

as $n \rightarrow \infty$.

PROOF. Let $\delta_n \doteq \frac{1}{\sqrt{r_n}}$. Then

$$\begin{aligned} P\left(\sup_{0 \leq t \leq T} [V_n(t) - r_n t] > \epsilon\right) &\leq P\left(\sup_{0 \leq t \leq \delta_n} |V_n(t)| > \epsilon\right) + P\left(\sup_{\delta_n < t \leq T} |V_n(t)| > r_n \delta_n\right) \\ &\leq \frac{4E V_n(\delta_n)^2}{\epsilon^2} + \frac{4E V_n(T)^2}{(r_n \delta_n)^2} \\ &= \frac{4E \langle V_n \rangle_{\delta_n}}{\epsilon^2} + \frac{4E \langle V_n \rangle_T}{(r_n \delta_n)^2} \leq \frac{4C \delta_n}{\epsilon^2} + \frac{4CT}{(r_n \delta_n)^2} \rightarrow 0, \end{aligned}$$

where the inequality on the second line is from Doob's maximal inequality. $\square$

7. Proof of Theorem 2.2. Now we start with some preliminary lemmas. Recall from Remark 2.5(ii) that under the hypothesis of Theorem 2.2 we have $\mu_n \rightarrow f_k \in \ell_1^\downarrow$ as $n \rightarrow \infty$. Along with the tightness of $\{\|Z_n(0)\|_1\}_{n \in \mathbb{N}}$, this shows that $G_n(0) \xrightarrow{P} f_k \in \ell_1^\downarrow$ as $n \rightarrow \infty$.

LEMMA 7.1. Let $d_n \rightarrow \infty$, $\frac{d_n}{\sqrt{n}} \rightarrow 0$, and $\lambda_n \nearrow 1$. Assume that for some $k \in \mathbb{N}$, $\liminf_n \mu_{n,k} = \delta > 0$. Then for any $L > 0$ and $1 \leq i \leq k$, as $n \rightarrow \infty$,

$$(7.1) \quad \sup_{0 < |z| \leq L} \left| (\dot{\beta}_n(\mu_{n,i})z)^{-1} \lambda_n \int_0^z \dot{\beta}_n(\mu_{n,i} + y/\sqrt{n}) dy - 1 \right| \rightarrow 0.$$

PROOF. To prove (7.1), we will approximate $\dot{\beta}_n(x)$ by $\dot{\gamma}_n(x)$. Using Lemma 5.1,

$$\epsilon_n \doteq \sup_{x \in [\delta/2, 1]} \left| \frac{\dot{\beta}_n(x)}{\dot{\gamma}_n(x)} - 1 \right| \rightarrow 0.$$

Since $\liminf_n \mu_{n,k} > \delta/2$ and $j \mapsto \mu_{n,j}$ is decreasing, there is an N_0 so that for $n \geq N_0$, $\mu_{n,i} + \frac{y}{\sqrt{n}} \geq \frac{\delta}{2}$, for any $i \leq k$ and $y \in \mathbb{R}$ with $|y| \leq L$. Hence uniformly in $0 < |z| \leq L$ and $i \leq k$,

$$\begin{aligned} \frac{\lambda_n}{z} \int_0^z \frac{\dot{\beta}_n(\mu_{n,i} + \frac{y}{\sqrt{n}})}{\dot{\beta}_n(\mu_{n,i})} dy &= \frac{1 + o(1)}{z} \int_0^z \frac{\dot{\gamma}_n(\mu_{n,i} + \frac{y}{\sqrt{n}})}{\dot{\gamma}_n(\mu_{n,i})} dy \\ &= \frac{1 + o(1)}{z} \int_0^z \left(1 + \frac{y}{\sqrt{n}\mu_{n,i}} \right)^{d_n-1} dy \\ &= \frac{1 + o(1)}{z} \int_0^z \exp \left\{ (d_n - 1) \log \left(1 + \frac{y}{\sqrt{n}\mu_{n,i}} \right) \right\} dy \\ &= \frac{1 + o(1)}{z} \int_0^z \exp \left\{ O \left(\frac{d_n L}{\sqrt{n}\delta} \right) \right\} dy \rightarrow 1. \end{aligned}$$

This shows (7.1). $\square$

REMARK 7.2. Suppose that the hypothesis of Lemma 7.1 hold. Recall the definition of $\Delta_{n,i}$ for $i > k$ from the proof of Lemma 6.4. We extend this definition by setting

$$(7.2) \quad \Delta_{n,i}(z) \doteq q_{n,i}(z) / (\dot{\beta}_n(\mu_{n,i})z) \mathbb{I}_{\{z \neq 0\}} - 1 \quad \text{if } 1 \leq i \leq k,$$

where $q_{n,i}$ is defined by (6.3). With this extension,

$$(7.3) \quad q_{n,i}(z) = \begin{cases} \dot{\beta}_n(\mu_{n,i})(1 + \Delta_{n,i}(z))z & \text{if } 1 \leq i \leq k, \\ \Delta_{n,i}(z)z & \text{if } i > k. \end{cases}$$

Using this notation, Lemma 7.1 and Lemma 6.3 show that, for any $L > 0$,

$$(7.4) \quad \gamma_{n,L} \doteq \sup_{i \in \mathbb{N}} \sup_{0 < |z| \leq L} |\Delta_{n,i}(z)| \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

The following corollary is an immediate consequence of Remark 7.2 and Lemma 6.1.

COROLLARY 7.3. Under the hypothesis of Lemma 7.1, Z_n satisfies the following integral equations.

For $i = 1$,

$$\begin{aligned} Z_{n,1}(t) &= Z_{n,1}(0) - \int_0^t \dot{\beta}_n(\mu_{n,1})(1 + \Delta_{n,1}(Z_{n,1}(s)))Z_{n,1}(s) ds \\ &\quad - \int_0^t (Z_{n,1}(s) - Z_{n,2}(s)) ds + \sqrt{n}M_{n,1}(t). \end{aligned}$$

For $i \in \{2, \dots, k\}$

$$\begin{aligned} Z_{n,i}(t) &= Z_{n,i}(0) + \int_0^t \dot{\beta}_n(\mu_{n,i-1})(1 + \Delta_{n,i-1}(Z_{n,i-1}(s)))Z_{n,i-1}(s) ds \\ &\quad - \int_0^t \dot{\beta}_n(\mu_{n,i})(1 + \Delta_{n,i}(Z_{n,i}(s)))Z_{n,i}(s) ds \\ &\quad - \int_0^t (Z_{n,i}(s) - Z_{n,i+1}(s)) ds + \sqrt{n}M_{n,i}(t). \end{aligned}$$

For $i = k + 1$,

$$\begin{aligned} Z_{n,k+1}(t) &= Z_{n,k+1}(0) + \int_0^t \dot{\beta}_n(\mu_{n,k})(1 + \Delta_{n,k}(Z_{n,k}(s)))Z_{n,k}(s) ds \\ &\quad - \int_0^t \Delta_{n,k+1}(Z_{n,k+1}(s))Z_{n,k+1}(s) \\ &\quad - \int_0^t (Z_{n,k+1}(s) - Z_{n,k+2}(s)) ds + \sqrt{n}M_{n,k+1}(t). \end{aligned}$$

For $i > k + 1$,

$$\begin{aligned} Z_{n,i}(t) &= Z_{n,i}(0) + \int_0^t \Delta_{n,i-1}(Z_{n,i-1}(s))Z_{n,i-1}(s) ds - \int_0^t \Delta_{n,i}(Z_{n,i}(s))Z_{n,i}(s) ds \\ &\quad - \int_0^t (Z_{n,i}(s) - Z_{n,i+1}(s)) ds + \sqrt{n}M_{n,i}(t), \end{aligned}$$

where $\Delta_{n,i}$ is as in Remark 7.2.

Finally, if $Y_{n,1} \doteq \sum_{i=1}^k Z_{n,i}$, then

$$\begin{aligned} (7.5) \quad Y_{n,1}(t) &= Y_{n,1}(0) - \int_0^t \dot{\beta}_n(\mu_{n,k})(1 + \Delta_{n,k}(Z_{n,k}(s)))Z_{n,k}(s) ds \\ &\quad - \int_0^t (Z_{n,1}(s) - Z_{n,k+1}(s)) ds + \sum_{i=1}^k \sqrt{n}M_{n,i}(t). \end{aligned}$$

LEMMA 7.4. Suppose $\lambda_n \nearrow 1$ and $1 \ll d_n \ll n$. Assume that for some $k \geq 2$, $\mu_{n,k} \rightarrow 1$ and $\dot{\beta}_n(\mu_{n,k}) \rightarrow \alpha \in [0, \infty)$ as $n \rightarrow \infty$. Define the $k - 1 \times k - 1$ tridiagonal matrix $Q_n(s)$ as

$$\begin{aligned} (7.6) \quad Q_n(s)[j, j] &= \dot{\beta}_n(\mu_{n,j})(1 + \Delta_{n,j}(Z_{n,j}(s))) + 1, \quad 1 \leq j \leq k - 1, \\ Q_n(s)[j, j + 1] &= -1, \quad 1 \leq j \leq k - 2, \\ Q_n(s)[j, j - 1] &= -\dot{\beta}_n(\mu_{n,j-1})(1 + \Delta_{n,j-1}(Z_{n,j-1}(s))), \quad 2 \leq j \leq k - 1, \end{aligned}$$

and for all other j, k , $Q_n(s)[j, k] = 0$. Then for any $T, L \in (0, \infty)$,

$$\lim_{n \rightarrow \infty} \inf_{s \in [0, T \wedge \tau_{n,L}]} \inf_{\vec{x} \in \mathbb{R}^{k-1} \setminus \{0\}} \frac{\vec{x}^t Q_n(s) \vec{x}}{\|\vec{x}\|^2} = +\infty \quad a.s.$$

PROOF. Let $h_{n,i}(s) \doteq \dot{\beta}_n(\mu_{n,i})(1 + \Delta_{n,i}(Z_{n,i}(s))) + 1$ and $H_n(s) \doteq Q_n(s) + Q_n(s)^t$. Then $H_n(s)$ is a symmetric tridiagonal matrix with entries

$$\begin{aligned} (7.7) \quad H_n(s)[j, j] &= 2h_{n,j}(s), \quad 1 \leq j \leq k - 1, \\ H_n(s)[j, j + 1] &= -h_{n,j}(s), \quad 1 \leq j \leq k - 2, \\ H_n(s)[j, j - 1] &= -h_{n,j-1}(s), \quad 2 \leq j \leq k - 1. \end{aligned}$$

Let $h_n \doteq \dot{\beta}_n(\mu_{n,1})$. By Lemma 5.6, $h_n \rightarrow \infty$ and by the uniform convergence in (7.4) and Lemma 5.6 once more

$$\max_{i \leq k-1} \sup_{s \in [0, T \wedge \tau_{n,L}]} \left| \frac{h_{n,i}(s)}{h_n} - 1 \right| \rightarrow 0 \quad \text{as } n \rightarrow \infty \text{ a.s.}$$

This in particular shows that

$$(7.8) \quad \sup_{s \in [0, T \wedge \tau_{n,L}]} \left\| \frac{1}{h_n} H_n(s) - H \right\|_F \rightarrow 0 \quad \text{a.s.,}$$

where $\|\cdot\|_F$ is the Frobenius norm and H is the $(k-1) \times (k-1)$ tridiagonal matrix given as

$$\begin{aligned} H[j, j] &= 2, \quad 1 \leq j \leq k-1, \\ H[j, j+1] &= -1, \quad 1 \leq j \leq k-2, \\ H[j, j-1] &= -1, \quad 2 \leq j \leq k-1. \end{aligned}$$

Note for any $\vec{x} = (x_1, x_2, \dots, x_{k-1}) \in \mathbb{R}^{k-1}$ by completing squares

$$\vec{x}^t H \vec{x} = x_1^2 + (x_2 - x_1)^2 + (x_3 - x_2)^2 + \dots + (x_{k-2} - x_{k-1})^2 + x_{k-1}^2,$$

which is strictly positive if $\vec{x} \neq 0$. Let $c \doteq \inf_{\|\vec{x}\|=1} \vec{x}^t H \vec{x}$. Since the unit sphere is compact, the infimum is attained, and hence $c > 0$. This shows that H is a positive definite matrix.

Finally, note that for any $s \geq 0$,

$$\begin{aligned} \vec{x}^t \frac{1}{h_n} H_n(s) \vec{x} &= \vec{x}^t H \vec{x} + \vec{x}^t \left(\frac{1}{h_n} H_n(s) - H \right) \vec{x} \\ &\geq \vec{x}^t H \vec{x} - \|h_n^{-1} H_n(s) - H\|_F \|\vec{x}\|^2 \geq (c - \|h_n^{-1} H_n(s) - H\|_F) \|\vec{x}\|^2. \end{aligned}$$

On taking infimum and using $\vec{x}^t H_n(s) \vec{x} = 2\vec{x}^t Q_n(s) \vec{x}$, this shows

$$\inf_{s \in [0, T \wedge \tau_L]} \inf_{x \in \mathbb{R}^k \setminus \{0\}} \frac{2\vec{x}^t Q_n(s) \vec{x}}{\|\vec{x}\|^2} \geq \left(c - \sup_{s \in [0, T \wedge \tau_L]} \|h_n^{-1} H_n(s) - H\|_F \right) h_n.$$

As $n \rightarrow \infty$, the convergence in (7.8) and the divergence $h_n \rightarrow +\infty$ now completes the proof. $\square$

REMARK 7.5. For every $s > 0$, the $(k-1) \times (k-1)$ matrix $Q_n(s)$ appearing in the previous lemma is the drift operator that appears in the right-hand side of the first $k-1$ coordinates in Corollary 7.3. More precisely, for each $t \geq 0$,

$$(7.9) \quad \vec{X}_n(t) = \vec{X}_n(0) - \int_0^t Q_n(s) \vec{X}_n(s) ds + \vec{e}_{k-1} \int_0^t Z_{n,k}(s) ds + \vec{W}_n(t),$$

where $\vec{X}_n \doteq (Z_{n,1}, Z_{n,2}, \dots, Z_{n,k-1})$, $\vec{W}_n \doteq (\sqrt{n}M_{n,1}, \dots, \sqrt{n}M_{n,k-1})$ and $\vec{e}_{k-1}$ is the vector $(0, 0, \dots, 0, 1)^t \in \mathbb{R}^{k-1}$.

LEMMA 7.6. Suppose that the hypothesis of Theorem 2.2 holds with $k \geq 2$ and let $\vec{X}_n \doteq (Z_{n,1}, Z_{n,2}, \dots, Z_{n,k-1})$. Then for $L, T, \epsilon \in (0, \infty)$,

$$(7.10) \quad P\left(\sup_{s \in [0, T \wedge \tau_{n,L}]} \|\vec{X}_n(s)\| > \|\vec{X}_n(0)\| + \epsilon\right) \rightarrow 0,$$

and

$$(7.11) \quad \sup_{s \in [\epsilon, T \wedge \tau_{n,L}]} \|\vec{X}_n(s)\| \xrightarrow{P} 0,$$

as $n \rightarrow \infty$.

PROOF. Applying Itô's formula (see [34], Section II.7) to the function $h(\vec{x}) = \|\vec{x}\|^2$ and the semimartingale representation of $\vec{X}_n$ from (7.9) in Remark 7.5, we get

$$\begin{aligned} \|\vec{X}_n(t)\|^2 &= \|\vec{X}_n(0)\|^2 + 2 \int_{0+}^t \langle \vec{X}_n(s-), d\vec{X}_n(s) \rangle + [\vec{W}_n]_t \\ (7.12) \quad &= \|\vec{X}_n(0)\|^2 - 2 \int_0^t \langle \vec{X}_n(s), Q_n(s) \vec{X}_n(s) \rangle ds + 2 \int_0^t Z_{n,k}(s) \langle \vec{X}_n(s), \vec{e}_{k-1} \rangle ds \\ &\quad + 2 \int_0^t \langle \vec{X}_n(s-), d\vec{W}_n(s) \rangle + [\vec{W}_n]_t, \end{aligned}$$

where $[\vec{W}_n]_t \doteq \sum_{i=1}^{k-1} [\sqrt{n} M_{n,i}]_t$. Let

$$\begin{aligned} h_n(s) &\doteq \frac{\vec{X}_n(s)^t Q_n(s) \vec{X}_n(s)}{\|\vec{X}_n(s)\|^2} \mathbb{I}_{\{\vec{X}_n(s) \neq 0\}} + n \mathbb{I}_{\{\vec{X}_n(s) = 0\}}, \\ g_n(s) &\doteq 2Z_{n,k}(s)Z_{n,k-1}(s), \\ R_n(s) &\doteq 2 \int_0^t \langle \vec{X}_n(s-), d\vec{W}_n(s) \rangle + [\vec{W}_n]_t \end{aligned}$$

then (7.12) becomes

$$(7.13) \quad \|\vec{X}_n(t)\|^2 = \|\vec{X}_n(0)\|^2 - 2 \int_0^t h_n(s) \|\vec{X}_n(s)\|^2 ds + \int_0^t g_n(s) ds + R_n(t).$$

Further, by Lemma 7.4,

$$(7.14) \quad m_n \doteq \inf_{s \in [0, T \wedge \tau_{n,L}]} h_n(s) \rightarrow +\infty \quad \text{a.s. as } n \rightarrow \infty,$$

and by Doob's inequality and Itô's isometry (see, e.g., [34], Corollary 3, Section II.7), for $i \leq k-1$,

$$\begin{aligned} \mathbf{E} \sup_{s \in [0, T \wedge \tau_{n,L}]} \left| \int_0^t Z_{n,i}(s-) d(\sqrt{n} M_{n,i})(s) \right|^2 &\leq 4\mathbf{E} \int_0^{T \wedge \tau_{n,L}} Z_{n,i}^2(s-) d[\sqrt{n} M_{n,i}]_s \\ &\leq 4L^2 \mathbf{E} [\sqrt{n} M_{n,i}]_T = 4L^2 \mathbf{E} \langle \sqrt{n} M_{n,i} \rangle_T, \end{aligned}$$

where the second to last inequality is obtained by using $\|Z_n\|_{2, T \wedge \tau_{n,L}} \leq L$. From the proof of Lemma 6.2, we see that for any $i \leq k-1$, $\mathbf{E} \langle \sqrt{n} M_{n,i} \rangle_T = \mathbf{E} [\sqrt{n} M_{n,i}]_T \rightarrow 0$ as $n \rightarrow \infty$. Along with the above display, this shows that the two terms appearing in the definition of R_n are converging to zero, and hence

$$(7.15) \quad |R_n|_{*, T \wedge \tau_{n,L}} \xrightarrow{P} 0 \quad \text{as } n \rightarrow \infty.$$

Applying Lemma 6.8 to (7.13) with $z(t) = \|\vec{X}_n(t)\|^2$, $h = 2h_n$, $g = g_n$, $M = R_n$, and $\tau = \tau_{n,L}$ shows for any $t \in [0, T \wedge \tau_{n,L}]$

$$\|\vec{X}_n(t)\|^2 \leq \frac{|g_n|_{*, T \wedge \tau_{n,L}}}{2m_n} + 2|R_n|_{*, T \wedge \tau_{n,L}} + e^{-2m_n t} \|\vec{X}_n(0)\|^2.$$

Taking $t = \epsilon_n \doteq 1/\sqrt{m_n}$ and using (7.14), (7.15), $|g_n|_{*, T \wedge \tau_{n,L}} \leq 2L^2$ and $\vec{X}_n(0) \xrightarrow{P} (z_1, \dots, z_{k-1})^t$, we see that

$$(7.16) \quad \sup_{t \in [\epsilon_n, T \wedge \tau_{n,L}]} \|\vec{X}_n(t)\| \xrightarrow{P} 0.$$

Since $\epsilon_n \rightarrow 0$, this shows (7.11) for any fixed $\epsilon > 0$. Finally, from (7.13), we see that

$$\sup_{t \in [0, \epsilon_n \wedge \tau_{n,L} \wedge T]} \|\vec{X}_n(t)\|^2 \leq \|\vec{X}_n(0)\|^2 + |g_n|_{*, T \wedge \tau_{n,L}} \epsilon_n + |R_n|_{*, T \wedge \tau_{n,L}}.$$

Since we have already shown (7.16), the convergence in (7.10) is now immediate on using that $\epsilon_n \rightarrow 0$, $|g_n|_{*, T \wedge \tau_{n,L}} \leq 2L^2$ and that (7.15) holds. $\square$

COROLLARY 7.7. *Under the assumptions of Lemma 7.6, for each $i < k$, $\int_0^{T \wedge \tau_{n,L}} |Z_{n,i}(s)| ds \xrightarrow{P} 0$, as $n \rightarrow \infty$.*

PROOF. For any $\epsilon > 0$,

$$\begin{aligned} \int_0^{T \wedge \tau_{n,L}} |Z_{n,i}(s)| ds &\leq \int_{[0, \epsilon \wedge \tau_{n,L}]} |Z_{n,i}(s)| ds + \int_{[\epsilon, T \wedge \tau_{n,L}]} |Z_{n,i}(s)| ds \\ &\leq L\epsilon + \sup_{s \in [\epsilon, T \wedge \tau_{n,L}]} |Z_{n,i}(s)| T. \end{aligned}$$

Now fix $\delta > 0$ and let $\epsilon = \frac{\delta}{2L}$. Then for any $i < k$,

$$(7.17) \quad P\left(\int_0^{T \wedge \tau_{n,L}} |Z_{n,i}(s)| ds > \delta\right) \leq P\left(\sup_{s \in [\epsilon, T \wedge \tau_{n,L}]} |Z_{n,i}(s)| > \frac{\delta}{2T}\right),$$

which from (7.11) converges to 0 as $n \rightarrow \infty$. Since $\delta > 0$ was arbitrary, this completes the proof. $\square$

PROOF OF THEOREM 2.2. Recall the conditions in the theorem. By Remark 2.5(ii) and the tightness of $\{\|Z_n(0)\|_1\}_{n \in \mathbb{N}}$, the hypothesis of Lemma 6.2 holds. Hence by Skorokhod's embedding theorem, we can assume that $\{(Z_n(0), M_n)\}_{n \in \mathbb{N}}$ and a standard Brownian motion B are defined on a common probability space such that for any $T > 0$,

$$(7.18) \quad \sup_{t \leq T} \|\sqrt{n}M_n(t) - \sqrt{2}B(t)e_k\|_2 \rightarrow 0$$

and

$$(7.19) \quad \|Z_n(0) - z\|_2 \rightarrow 0 \quad \text{a.s.,}$$

as $n \rightarrow \infty$. Let Y and Y_n be as in the statement of the theorem. Taking $m \doteq r - k + 1$, let $\tilde{Y}_n \doteq (\sum_{i=1}^k Z_{n,i}, Z_{n,k+1}, \dots, Z_{n,r})$ be the stochastic process with sample paths in $\mathbb{D}([0, T] : \mathbb{R}^m)$ corresponding to the first m coordinates of Y_n . Note $Y_{n,m+} = Z_{n,r+}$, $Z_{n,k} = Y_{n,1} - \sum_{i=1}^{k-1} Z_{n,i}$, and for $k=1$, $Y_{n,1} = Z_{n,1}$. Hence by Corollary 7.3, $\tilde{Y}_n$ satisfy

$$\begin{aligned} (7.20) \quad Y_{n,1}(t) &= Y_{n,1}(0) - \int_0^t a_{n,k}(s)Y_{n,1}(s) ds - \mathbb{I}_{\{k=1\}} \int_0^t Y_{n,1}(s) ds \\ &\quad + \int_0^t Y_{n,2}(s) ds + \sqrt{n}M_{n,k}(t) \\ &\quad + \sum_{i=1}^{k-1} \int_0^t a_{n,k}(s)Z_{n,i}(s) ds - \mathbb{I}_{\{k>1\}} \int_0^t Z_{n,1}(s) ds + \sum_{i=1}^{k-1} \sqrt{n}M_{n,i}(t), \end{aligned}$$

$$\begin{aligned} (7.21) \quad Y_{n,2}(t) &= Y_{n,2}(0) + \int_0^t a_{n,k}(s)Y_{n,1}(s) ds - \int_0^t Y_{n,2}(s) ds + \int_0^t Y_{n,3}(s) ds \\ &\quad - \sum_{i=1}^{k-1} \int_0^t a_{n,k}(s)Z_{n,i}(s) ds - \int_0^t \delta_{n,k+1}(s)Y_{n,2}(s) ds \\ &\quad + \sqrt{n}M_{n,k+1}(t), \end{aligned}$$

and for $i \in \{3, 4, \dots, m\}$,

$$\begin{aligned} (7.22) \quad Y_{n,i}(t) &= Y_{n,i}(0) - \int_0^t Y_{n,i}(s) ds + \int_0^t Y_{n,i+1}(s) ds \\ &\quad + \int_0^t \delta_{n,k+i-2}(s)Y_{n,i-1}(s) ds - \int_0^t \delta_{n,k+i-1}(s)Y_{n,i}(s) ds \\ &\quad + \sqrt{n}M_{n,k+i-1}(t), \end{aligned}$$

where $a_{n,k}(s) \doteq \dot{\beta}_n(\mu_{n,k})(1 + \Delta_{n,k}(Z_{n,k}(s)))$ and $\delta_{n,i}(s) \doteq \Delta_{n,i}(Z_{n,i}(s))$ for $i \in \mathbb{N}$.

Since $\|Z_n\|_{2,T \wedge \tau_{n,L}} \leq L$, we have by (7.4) that, for any $i \in \mathbb{N}$,

$$(7.23) \quad |\delta_{n,i}|_{*,T \wedge \tau_{n,L}} \leq \gamma_{n,L} \rightarrow 0 \quad \text{a.s. as } n \rightarrow \infty.$$

Moreover, since $\dot{\beta}_n(\mu_{n,k}) \rightarrow \alpha \in [0, \infty)$, this also shows that

$$(7.24) \quad \sup_{s \in [0, T \wedge \tau_{n,L}]} |a_{n,k}(s) - \alpha| \rightarrow 0 \quad \text{a.s. as } n \rightarrow \infty.$$

We now show that

$$(7.25) \quad \|Y_n - Y\|_{2,T \wedge \tau_{n,L}} \xrightarrow{P} 0 \quad \text{as } n \rightarrow \infty.$$

To see this, note that, by Remark 2.5(ii), the hypothesis of Lemma 6.4 is satisfied, and hence

$\|Z_{n,r+}\|_{2,T \wedge \tau_{n,L}} \xrightarrow{P} 0$. Since $Y_{n,m+} = Z_{n,r+}$ and $Y_{m+} = 0$, this shows that

$$(7.26) \quad \|Y_{n,m+} - Y_{m+}\|_{2,T \wedge \tau_{n,L}} \xrightarrow{P} 0.$$

Thus in order to prove (7.25) it suffices to show that $\sum_{i=1}^m |Y_{n,i} - Y_i|_{*,T \wedge \tau_{n,L}} \xrightarrow{P} 0$ as $n \rightarrow \infty$. To show this, we consider $U_{n,i} \doteq Y_{n,i} - Y_i$. Subtracting (2.9) from (7.20), (7.21) and (7.22), we see

$$(7.27) \quad \begin{aligned} U_{n,1}(t) &= U_{n,1}(0) - (\alpha + \mathbb{I}_{\{k=1\}}) \int_0^t U_{n,1}(s) ds + \int_0^t U_{n,2}(s) ds \\ &\quad + \sqrt{n}M_{n,k}(t) - \sqrt{2}B(t) + W_{n,1}(t), \\ U_{n,2}(t) &= U_{n,2}(0) + \alpha \int_0^t U_{n,1}(s) - \int_0^t U_{n,2}(s) ds + \int_0^t U_{n,3}(s) ds + W_{n,2}(t), \\ U_{n,i}(t) &= U_{n,i}(0) - \int_0^t U_{n,i}(s) ds + \int_0^t U_{n,i+1}(s) ds \\ &\quad + W_{n,i}(t) \quad \text{for } i \in \{3, 4, \dots, m\}, \end{aligned}$$

where

$$\begin{aligned} W_{n,1}(t) &\doteq \int_0^t (\alpha - a_{n,k}(s))Y_{n,1}(s) ds + \sum_{i=1}^{k-1} \int_0^t a_{n,k}(s)Z_{n,i}(s) ds \\ &\quad - \mathbb{I}_{\{k>1\}} \int_0^t Z_{n,1}(s) ds + \sum_{i=1}^{k-1} \sqrt{n}M_{n,i}(t), \\ W_{n,2}(t) &\doteq \int_0^t (a_{n,k}(s) - \alpha)Y_{n,1}(s) ds - \sum_{i=1}^{k-1} \int_0^t a_{n,k}(s)Z_{n,i}(s) ds \\ &\quad - \int_0^t \delta_{n,k+1}(s)Y_{n,2}(s) ds + \sqrt{n}M_{n,k+1}(t), \\ W_{n,i}(t) &\doteq \int_0^t \delta_{n,k+i-2}(s)Y_{n,i-1}(s) ds \\ &\quad - \int_0^t \delta_{n,k+i-1}(s)Y_{n,i}(s) ds + \sqrt{n}M_{n,k+i-1}(t) \quad \text{for } i \in \{3, \dots, m\}. \end{aligned}$$

Note that, for each n , $\|Y_n\|_{2,T \wedge \tau_{n,L}} \leq k\|Z_n\|_{2,T \wedge \tau_{n,L}}$, which by (6.9) is bounded above by kL . Hence by (7.24), (7.23), (7.18) and Corollary 7.7,

$$(7.28) \quad |W_{n,i}|_{*,T \wedge \tau_{n,L}} \xrightarrow{P} 0 \quad \text{as } n \rightarrow \infty$$

for each $i \in [m]$. Let $\|U_n\|_{1,t} \doteq \sup_{s \in [0,t]} \sum_{i=1}^m |U_{n,i}(s)|$. Then, from (7.27), for any $t \in [0, T \wedge \tau_{n,L}]$,

$$\|U_n\|_{1,t} \leq \sum_{i=1}^m (|U_{n,i}(0)| + |W_{n,i}|_{*,T \wedge \tau_{n,L}}) + |\sqrt{n}M_{n,k} - \sqrt{2}B|_{*,T} + R \int_0^t \|U_n\|_{1,s} ds$$

with $R \doteq \max(2\alpha + \mathbb{I}_{\{k=1\}}, 2)$. Hence by Gronwall's inequality,

$$\|U_n\|_{1,T \wedge \tau_{n,L}} \leq \left(|\sqrt{n}M_{n,k} - \sqrt{2}B|_{*,T} + \sum_{i=1}^m (|U_{n,i}(0)| + |W_{n,i}|_{*,T \wedge \tau_{n,L}}) \right) e^{RT}.$$

By our hypothesis, as $n \rightarrow \infty$, $|U_{n,i}(0)| = |Z_{n,k+i-1}(0) - z_{n,k+i-1}| \xrightarrow{P} 0$ for each $i \in [m]$. Hence by (7.28) and (7.18), $\|U_n\|_{1,T \wedge \tau_{n,L}} = \sum_{i=1}^m |Y_{n,i} - Y_i|_{*,T \wedge \tau_{n,L}} \xrightarrow{P} 0$ as $n \rightarrow \infty$. Combined with (7.26), this completes the proof of (7.25).

Next, we prove (2.8). Fix $\delta > 0$. Since $\mathbf{Y}$ has sample paths in $\mathbb{C}([0, T] : \ell_2)$, we can find $L_1 \in (0, \infty)$ so that

$$(7.29) \quad P(\|\mathbf{Y}\|_{2,T} > L_1) \leq \frac{\delta}{2}.$$

Also, since $\mathbf{Z}_n(0) \xrightarrow{P} \mathbf{z}$, we can find a $L_2 \in (0, \infty)$ so that

$$(7.30) \quad \sup_n P(\|\mathbf{Z}_n(0)\|_2 > L_2) \leq \frac{\delta}{2}.$$

Let $L \doteq (L_1 + 1) + k(L_2 + 1) + 1$. Also, let $\vec{X}_n$ be as in Lemma 7.6 when $k > 1$. For $k = 1$, we set $\vec{X}_n \doteq 0$. Then

$$\begin{aligned} \|\mathbf{Z}_n\|_{2,T \wedge \tau_{n,L}} &\leq \|\vec{X}_n\|_{2,T \wedge \tau_{n,L}} + \left\| \mathbf{Y}_n - \mathbf{e}_1 \sum_{i=1}^{k-1} Z_{n,i} \right\|_{2,T \wedge \tau_{n,L}} \\ &\leq k\mathbb{I}_{\{k>1\}} \|\vec{X}_n\|_{2,T \wedge \tau_{n,L}} + \|\mathbf{Y}_n\|_{2,T \wedge \tau_{n,L}}. \end{aligned}$$

Hence for each $n \in \mathbb{N}$,

$$\begin{aligned} P(\tau_{n,L} \leq T) &\leq P(\|\mathbf{Z}_n\|_{2,T \wedge \tau_{n,L}} > L - 1) \\ &\leq P(\|\mathbf{Y}_n\|_{2,T \wedge \tau_{n,L}} > L_1 + 1) + P(\|\vec{X}_n\|_{2,T \wedge \tau_{n,L}} > L_2 + 1), \\ &\leq \delta + P(\|\mathbf{Y}_n - \mathbf{Y}\|_{2,T \wedge \tau_{n,L}} > 1) + P(\|\vec{X}_n\|_{2,T \wedge \tau_{n,L}} > \|\vec{X}_n(0)\| + 1), \end{aligned}$$

where the last inequality uses (7.29) and (7.30). From Lemma 7.6 and (7.25), we see

$$\limsup_{n \rightarrow \infty} P(\|\mathbf{Z}_n\|_{2,T} \geq L) \leq \limsup_{n \rightarrow \infty} P(\tau_{n,L} \leq T) \leq \delta.$$

Since $\delta > 0$ is arbitrary, the convergence in (2.8) is now immediate.

This convergence in particular says that $\lim_{L \rightarrow \infty} \sup_n P(\tau_{n,L} \leq T) = 0$. Using Lemma 6.5 with $F_n(t) = \|\mathbf{Y}_n - \mathbf{Y}\|_{2,t}$, we now see from (7.25) that $\|\mathbf{Y}_n - \mathbf{Y}\|_{2,T} \xrightarrow{P} 0$ as $n \rightarrow \infty$. Similarly, if $k > 1$, then taking $F_n(t) = \sup_{s \in [\epsilon, t]} |Z_{n,i}(s)|$ in Lemma 6.5 we conclude from Lemma 7.6 that for each $i \in [k-1]$ and $\epsilon > 0$ $\sup_{s \in [\epsilon, T]} |Z_{n,i}(s)| \xrightarrow{P} 0$ as $n \rightarrow \infty$. This completes the proof of Theorem 2.2. $\square$

8. Proof of Theorem 2.3. In this section, we give the proof of Theorem 2.3. We begin by giving a convenient representation for $\mathbf{Z}_n$ under the assumptions of Theorem 2.3 and establishing some a priori convergence properties.

LEMMA 8.1. Suppose $c_n = \frac{d_n}{\sqrt{n}} \rightarrow c \in (0, \infty)$ and $\lambda_n = 1 - (\frac{\log d_n}{d_n} + \frac{\alpha_n}{\sqrt{n}})$ where $\alpha_n \in \mathbb{R}$, $\liminf_{n \rightarrow \infty} \alpha_n > -\infty$ and $\frac{\alpha_n}{n^{1/4}} \rightarrow 0$. Suppose also that $\{\|\mathbf{Z}_n(0)\|_1\}_{n \in \mathbb{N}}$ is a tight sequence of random variables and $\mathbf{Z}_{n,r+1}(0) \xrightarrow{P} \mathbf{0}$ in ℓ_2 for some $r \geq 2$. Then there are stochastic processes $\delta_n, \{W_{n,i}\}_{i=2}^r$ with sample paths in $\mathbb{D}([0, \infty) : \mathbb{R})$ such that for any $t \geq 0$,

$$\begin{aligned} Z_{n,1}(t) &= Z_{n,1}(0) - \int_0^t Z_{n,1}(s) ds + \int_0^t Z_{n,2}(s) ds + \sqrt{n}M_{n,1}(t) \\ &\quad - (c_n e^{c_n \alpha_n})^{-1} \int_0^t (1 + \delta_n(s))(e^{c_n Z_{n,1}(s)} - 1) ds, \\ (8.1) \quad Z_{n,2}(t) &= Z_{n,2}(0) - \int_0^t Z_{n,2}(s) ds + \int_0^t Z_{n,3}(s) ds + W_{n,2}(t) \\ &\quad + (c_n e^{c_n \alpha_n})^{-1} \int_0^t (1 + \delta_n(s))(e^{c_n Z_{n,1}(s)} - 1) ds, \\ Z_{n,i}(t) &= Z_{n,i}(0) - \int_0^t Z_{n,i}(s) ds + \int_0^t Z_{n,i+1}(s) ds + W_{n,i}(t) \quad \text{for } i \in \{3, \dots, r\} \end{aligned}$$

and for any fixed $L, T \in (0, \infty)$:

- (1) $\sqrt{n}M_{n,1} \Rightarrow \sqrt{2}B$ in $\mathbb{D}([0, \infty) : \mathbb{R})$ where B is a standard Brownian motion,
- (2) $|\delta_n|_{*,T_n} \rightarrow 0$ a.s.
- (3) $|W_{n,i}|_{*,T_n} \xrightarrow{P} 0$ for $i \in \{2, \dots, r\}$,
- (4) $\|\mathbf{Z}_{n,r+1}\|_{2,T_n} \xrightarrow{P} 0$,

where $T_n \doteq T \wedge \tau_{n,L}$ and $\tau_{n,L}$ is defined as in (6.8).

PROOF. Recall the definition of $q_{n,i}$ from Lemma 6.1. Define

$$\delta_n(s) \doteq q_{n,1}(Z_{n,1}(s))c_n(e^{c_n[Z_{n,1}(s)-\alpha_n]} - e^{-c_n\alpha_n})^{-1} - 1$$

so that

$$q_{n,1}(Z_{n,1}(s)) = (1 + \delta_n(s))c_n^{-1}(e^{c_n[Z_{n,1}(s)-\alpha_n]} - e^{-c_n\alpha_n}).$$

Since $\sup_{s \leq T \wedge \tau_{n,L}} |Z_{n,1}(s)| \leq L$, Lemma 6.7 shows that $|\delta_n|_{*,T_n} \rightarrow 0$ a.s. Define

$$W_{n,2}(t) \doteq - \int_0^t q_{n,2}(Z_{n,2}(s)) ds + \sqrt{n}M_{n,2}(t),$$

$$W_{n,i}(t) \doteq \int_0^t q_{n,i-1}(Z_{n,i-1}(s)) ds - \int_0^t q_{n,i}(Z_{n,i}(s)) ds + \sqrt{n}M_{n,i}(t) \quad \text{for } i \in \{3, \dots, r\}.$$

From Lemma 6.1, it follows that (8.1) is satisfied. Lemma 6.6 shows that $\mu_n \rightarrow \mathbf{f}_1 \in \ell_1^\downarrow$. Along with the assumed tightness of $\{\|\mathbf{Z}_n(0)\|_1\}_{n \in \mathbb{N}}$, this shows $\mathbf{G}_n(0) = \mu_n + \frac{\mathbf{Z}_n(0)}{\sqrt{n}} \xrightarrow{P} \mathbf{f}_1$ in $\ell_1^\downarrow$. Hence by Lemma 6.2 and Lemma 6.4,

$$(8.2) \quad \sqrt{n}\mathbf{M}_n \Rightarrow \sqrt{2}B\mathbf{e}_1 \text{ in } \mathbb{D}([0, \infty) : \ell_2)$$

and $\|\mathbf{Z}_{n,r+1}\|_{2,T \wedge \tau_{n,L}} \xrightarrow{P} 0$ as $n \rightarrow \infty$. Since $|Z_{n,i}|_{*,T \wedge \tau_{n,L}} \leq L$ and $\mu_{n,2} \rightarrow 0$, Lemma 6.3, together with (8.2), shows that $|W_{n,i}|_{*,T_n} \xrightarrow{P} 0$ for each $i \in \{2, \dots, r\}$, as $n \rightarrow \infty$. $\square$

The next lemma gives pathwise existence and uniqueness of solutions to a system of stochastic differential equations in which the drift fails to satisfy a linear growth condition.

LEMMA 8.2. Suppose $c \in (0, \infty)$, $\alpha \in (0, \infty]$ and B is a standard Brownian motion. Then for any $r \geq 2$ the system of equations

$$\begin{aligned} Z_1(t) &= z_1 - \int_0^t Z_1(s) ds + \int_0^t Z_2(s) ds + \sqrt{2}B(t) - (ce^{c\alpha})^{-1} \int_0^t (e^{cZ_1(s)} - 1) ds, \\ Z_2(t) &= z_2 - \int_0^t Z_2(s) ds + \int_0^t Z_3(s) ds + (ce^{c\alpha})^{-1} \int_0^t (e^{cZ_1(s)} - 1) ds, \\ Z_i(t) &= z_i - \int_0^t Z_i(s) ds + \int_0^t Z_{i+1}(s) ds \quad \text{for } i \in \{3, \dots, r\}, \\ Z_i(t) &= 0 \quad \text{for } i > r \end{aligned} \quad (8.3)$$

has a unique pathwise solution $\mathbf{Z}$ with sample paths in $\mathbb{C}([0, \infty) : \ell_2)$ for any $(z_1, \dots, z_r) \in \mathbb{R}^r$.

PROOF. The case when $\alpha = \infty$ is standard and is thus omitted. Consider now the case $\alpha < \infty$. It is straightforward to see that there is a unique $\mathbf{Z}_{2+} \doteq (Z_3, Z_4, \dots)$ in $\mathbb{C}([0, \infty) : \ell_2)$ that solves the last two equations in (8.3). Hence it suffices to show that, the system of equations

$$\begin{aligned} Z_1(t) &= z_1 - (ce^{c\alpha})^{-1} \int_0^t (e^{cZ_1(s)} - 1) ds + \int_0^t (Z_2(s) - Z_1(s)) ds + \sqrt{2}B(t), \\ Z_2(t) &= z_2 + (ce^{c\alpha})^{-1} \int_0^t (e^{cZ_1(s)} - 1) ds - \int_0^t Z_2(s) ds + \int_0^t h(s) ds \end{aligned} \quad (8.4)$$

has a unique pathwise solution (Z_1, Z_2) with sample paths in $\mathbb{C}([0, \infty) : \mathbb{R}^2)$ where $h \doteq Z_3 \in \mathbb{C}([0, \infty) : \mathbb{R})$ is a given (nonrandom) continuous trajectory and $(z_1, z_2) \in \mathbb{R}^2$.

Define $y_1 = z_1$, $y_2 = z_1 + z_2$ and consider the equation:

$$\begin{aligned} Y_1(t) &= y_1 - (ce^{c\alpha})^{-1} \int_0^t (e^{cY_1(s)} - 1) ds + \int_0^t (Y_2(s) - 2Y_1(s)) ds + \sqrt{2}B(t), \\ Y_2(t) &= y_2 - \int_0^t Y_1(s) ds + \int_0^t h(s) ds + \sqrt{2}B(t). \end{aligned} \quad (8.5)$$

Note that (Z_1, Z_2) solve (8.4) if and only if (Y_1, Y_2) , with $Y_1 = Z_1$ and $Y_2 = Z_1 + Z_2$ solve (8.5). Thus it suffices to prove existence and uniqueness of solutions for (8.5).

For $L \in (0, \infty)$, let $\eta_L : \mathbb{R} \rightarrow [0, 1]$ be such that η_L is smooth, $\eta_L(x) = 1$ for $|x| \leq L$ and $\eta_L(x) = 0$ for $|x| \geq L + 1$. Consider the equation

$$\begin{aligned} Y_1^L(t) &= y_1 - (ce^{c\alpha})^{-1} \int_0^t e^{cY_1^L(s)} \eta_L(Y_1^L(s)) ds \\ &\quad + (ce^{c\alpha})^{-1} t + \int_0^t (Y_2^L(s) - 2Y_1^L(s)) ds + \sqrt{2}B(t), \\ Y_2^L(t) &= y_2 - \int_0^t Y_1^L(s) ds + \int_0^t h(s) ds + \sqrt{2}B(t). \end{aligned} \quad (8.6)$$

Since for each L (8.6) is an equation with (globally) Lipschitz coefficients, by standard results, it has a unique pathwise continuous solution.

Fix $T \in (0, \infty)$ and let $\tau_L \doteq \inf\{t \geq 0 : |Y_1^L(t)| \geq L\} \wedge T$ for any $L > 0$. Then by pathwise uniqueness of (8.6), for $0 \leq t \leq \tau_L \wedge \tau_{L+1}$,

$$Y^L(t) = Y^{L+1}(t).$$

This in particular shows that, $\tau_L \leq \tau_{L+1}$ a.s.

We now estimate the second moment of $|Y_1^L(t)|$. By Itô's formula,

$$\begin{aligned} (Y_1^L(t))^2 &= (y_1)^2 - 2(ce^{c\alpha})^{-1} \int_0^t Y_1^L(s) e^{cY_1^L(s)} \eta_L(Y_1^L(s)) ds + 2(ce^{c\alpha})^{-1} \int_0^t Y_1^L(s) ds \\ &\quad + 2 \int_0^t Y_1^L(s) (Y_2^L(s) - 2Y_1^L(s)) ds + 2\sqrt{2} \int_0^t Y_1^L(s) dB(s) + 2t, \\ (Y_2^L(t))^2 &= (y_2)^2 - 2 \int_0^t Y_1^L(s) Y_2^L(s) ds + 2 \int_0^t Y_2^L(s) h(s) ds + 2\sqrt{2} \int_0^t Y_2^L(s) dB(s) + 2t. \end{aligned}$$

Thus

$$\begin{aligned} (Y_1^L(t))^2 + (Y_2^L(t))^2 &= (y_1)^2 + (y_2)^2 - 2(ce^{c\alpha})^{-1} \int_0^t Y_1^L(s) e^{cY_1^L(s)} \eta_L(Y_1^L(s)) ds \\ &\quad + 2(ce^{c\alpha})^{-1} \int_0^t Y_1^L(s) ds + 2 \int_0^t Y_2^L(s) h(s) ds \\ &\quad - 4 \int_0^t (Y_1^L(s))^2 ds + 2\sqrt{2} \int_0^t (Y_1^L(s) + Y_2^L(s)) dB(s) + 4t. \end{aligned}$$

Since $c > 0$, we have on using the inequality $|x| \leq 1 + |x|^2$ that $-xe^{cx} \eta_L(x) \leq (1 + |x|^2)$ for all $x \in \mathbb{R}$. Thus with $\|Y^L\|_{*,t} \doteq \sup_{s \in [0,t]} \|Y^L(s)\|$,

$$\begin{aligned} \|Y^L\|_{*,t}^2 &\leq \|y\|^2 + 4(ce^{c\alpha})^{-1} \int_0^t (1 + \|Y^L\|_{*,s}^2) ds + 2 \int_0^t (1 + \|Y^L\|_{*,s}^2) |h(s)| ds \\ &\quad + 2\sqrt{2} \left(1 + \sup_{0 \leq s \leq t} \left| \int_0^s (Y_1^L(u) + Y_2^L(u)) dB(u) \right|^2 \right) + 4t. \end{aligned}$$

Taking expectations and using Doob's inequality and Itô's isometry to compute the expectation over the supremum:

$$\begin{aligned} \mathbf{E} \|Y^L\|_{*,t}^2 &\leq \|y\|^2 + (4(ce^{c\alpha})^{-1} + 2|h|_{*,T}) \int_0^t (1 + \mathbf{E} \|Y^L\|_{*,s}^2) ds \\ &\quad + 2\sqrt{2} \left(1 + 4\mathbf{E} \int_0^t |Y_1^L(u) + Y_2^L(u)|^2 du \right) + 4t \\ &\leq (\|y\|^2 + K(T+1)) + K \int_0^t \mathbf{E} \|Y^L\|_{*,s}^2 ds \end{aligned}$$

with $K \doteq 4(ce^{c\alpha})^{-1} + 2|h|_{*,T} + 16\sqrt{2}$ for any $t \in [0, T]$. By Gronwall lemma, for every $L \in \mathbb{N}$,

$$\mathbf{E} \|Y^L\|_{*,T}^2 \leq (\|y\|^2 + K(T+1))e^{KT} \doteq c_1.$$

Thus, as $L \rightarrow \infty$

$$P(\tau_L < T) \leq P(\|Y^L\|_{*,T} \geq L) \leq c_1/L^2 \rightarrow 0,$$

and consequently $\tau_L \uparrow T$ a.s. as $L \rightarrow \infty$. Now define $Y(t) \doteq Y^L(t)$ for $0 \leq t \leq \tau_L$. Then Y is a solution of (8.5) on $[0, T)$. The same argument as before shows that this is the unique pathwise solution on $[0, T)$. Since T is arbitrary, we get a unique pathwise solution of (8.5) on $[0, \infty)$. This completes the proof of the lemma. $\square$

LEMMA 8.3. *Suppose the assumptions of Theorem 2.3 hold. Suppose further that $\mathbf{Z}_n(0)$, $\mathbf{M}_n$ and a standard Brownian motion B are given on a common probability space such that*

$\mathbf{Z}_n(0) \rightarrow \mathbf{z}$ in $\ell_1^\perp$ and $\mathbf{M}_n \rightarrow \sqrt{2}B\mathbf{e}_1$ in $\mathbb{D}([0, \infty) : \ell_2)$ a.s. Let $\mathbf{Z}$ be as defined in Lemma 8.2. Then for any $T, L \in (0, \infty)$,

$$(8.7) \quad \|\mathbf{Z}_n - \mathbf{Z}\|_{2, T \wedge \tau_{n,L} \wedge \tau_L} \xrightarrow{P} 0 \quad \text{as } n \rightarrow \infty,$$

where $\tau_L \doteq \inf\{t \mid \|\mathbf{Z}(t)\|_{2,t} > L\}$.

PROOF. Fix $L, T \in (0, \infty)$ and let $T_n \doteq T \wedge \tau_{n,L} \wedge \tau_L$. Using the estimate $|e^{ax} - e^{ay}| \leq ae^{a(x \vee y)}|x - y|$ for $x, y \in \mathbb{R}$, $a \geq 0$ and since $|Z_{n,1}(s)|, |Z_1(s)| \leq L$ for any $s \in [0, T_n]$, note

$$\begin{aligned} & |a_n(s)e^{c_n Z_{n,1}(s)} - ae^{c Z_1(s)}| \\ & \leq |a_n(s)e^{c_n Z_{n,1}(s)} - a_n(s)e^{c_n Z_1(s)}| + |a_n(s)e^{c_n Z_1(s)} - a_n(s)e^{c Z_1(s)}| \\ & \quad + |e^{c Z_1(s)}||a_n(s) - a| \\ & \leq |a_n(s)|c_n e^{c_n L}|U_{n,1}(s)| + |a_n(s)|Le^{L(c_n \vee c)}|c_n - c| + e^{cL}|a_n(s) - a|, \end{aligned}$$

where $a_n(s) \doteq (c_n e^{c_n \alpha_n})^{-1}(1 + \delta_n(s))$, $c_n \doteq d_n/\sqrt{n} \rightarrow c$, δ_n is as in Lemma 8.1, $a \doteq (ce^{c\alpha})^{-1}$, and $U_{n,i} \doteq Z_{n,i} - Z_i$ for $i \in \mathbb{N}$. Since $c_n \rightarrow c$ and $|\delta_n|_{*, T_n} \rightarrow 0$ a.s. by Lemma 8.1, $|a_n - a|_{*, T_n} \rightarrow 0$ a.s. Hence for any $s \in [0, T_n]$,

$$(8.8) \quad |a_n e^{c_n Z_{n,1}} - ae^{c Z_1}|_{*,s} \leq K|U_{n,1}|_{*,s} + r_n,$$

where $K \doteq \sup_n (c_n e^{c_n L}|a_n|_{*, T_n}) < \infty$ a.s. and

$$r_n \doteq |a_n|_{*, T_n} Le^{L(c_n \vee c)}|c_n - c| + e^{cL}|a_n - a|_{*, T_n} \rightarrow 0 \quad \text{a.s.}$$

Subtracting (8.3) from (8.1), for any $t > 0$,

$$\begin{aligned} U_{n,1}(t) &= U_{n,1}(0) - \int_0^t (U_{n,1}(s) - U_{n,2}(s)) ds + \sqrt{n}M_{n,1}(t) - \sqrt{2}B(t) \\ &\quad - \int_0^t (a_{n,1}(s)e^{c_n Z_{n,1}(s)} - ae^{c Z_1(s)}) ds + \int_0^t (a_n(s) - a) ds, \\ (8.9) \quad U_{n,2}(t) &= U_{n,2}(0) - \int_0^t (U_{n,2}(s) - U_{n,3}(s)) ds + W_{n,2}(t) \\ &\quad + \int_0^t (a_{n,1}(s)e^{c_n Z_{n,1}(s)} - ae^{c Z_1(s)}) ds - \int_0^t (a_n(s) - a) ds, \\ U_{n,i}(t) &= U_{n,i}(0) - \int_0^t (U_{n,i}(s) - U_{n,i+1}(s)) ds + W_{n,i}(t) \quad \text{for } i \in \{3, \dots, r\}. \end{aligned}$$

Let $H_t \doteq \sup_{s \in [0,t]} \sum_{i=1}^r |U_{n,i}(s)|$. Then from (8.8) and (8.9), for any $t \in [0, T_n]$,

$$\begin{aligned} H_t &\leq H_0 + |\sqrt{n}M_{n,1} - \sqrt{2}B|_{*,T} + 2T(|a_n - a|_{*, T_n} + r_n) \\ &\quad + \sum_{i=2}^r |W_{n,i}|_{*, T_n} + |U_{n,r+1}|_{*, T_n} + 2(1+K) \int_0^t H_s ds. \end{aligned}$$

Hence by Gronwall's lemma,

$$\begin{aligned} H_{T_n} &\leq \left(H_0 + |\sqrt{n}M_{n,1} - \sqrt{2}B|_{*,T} + 2T(|a_n - a|_{*, T_n} + r_n) + \sum_{i=2}^r |W_{n,i}|_{*, T_n} + |U_{n,r+1}|_{*, T_n} \right) \\ &\quad \times e^{2(1+K)T}. \end{aligned}$$

Note $U_{n,r+} = Z_{n,r+}$ and $U_{n,i}(0) = Z_{n,i}(0) - z_i$ for $i \leq r$; hence using Lemma 8.1 and the assumed convergences, it follows that $\|U_{n,r+}\|_{2,T_n} \xrightarrow{P} 0$ and, based on the above display, that $H_{T_n} \xrightarrow{P} 0$. Together these show $\|U\|_{2,T_n} = \|Z_n - Z\|_{2,T_n} \xrightarrow{P} 0$ as $n \rightarrow \infty$. $\square$

COROLLARY 8.4. *Under assumptions of Lemma 8.3, $\{\|Z_n\|_{2,T}\}_{n \in \mathbb{N}}$ is a tight sequence of random variables and*

$$(8.10) \quad \lim_{L \rightarrow \infty} \sup_n P(\tau_{n,L} \leq T) = 0.$$

PROOF. Fix $\delta > 0$. Since Z has sample paths in $\mathbb{C}([0, T] : \ell_2)$, we can find $L \in (0, \infty)$ so that $P(\|Z\|_{2,T} > L) \leq \delta$. With $\tau_{L+2} \doteq \inf\{t \mid \|Z(t)\|_2 > L+2\}$, note the inclusion $\{\|Z\|_{2,T} \leq L\} \subseteq \{\tau_{L+2} > T\}$ which will be used in the next display. Now, by the right continuity of Z_n , note for each $n \in \mathbb{N}$,

$$\begin{aligned} P(\tau_{n,L+2} \leq T) &\leq P(\|Z_n\|_{2,T \wedge \tau_{n,L+2}} > L+1) \\ &\leq P(\|Z_n - Z\|_{2,T \wedge \tau_{n,L+2}} > 1 \text{ or } \|Z\|_{2,T} > L) \\ &\leq P(\|Z_n - Z\|_{2,T \wedge \tau_{n,L+2} \wedge \tau_{L+2}} > 1) + P(\|Z\|_{2,T} > L) \\ &\leq P(\|Z_n - Z\|_{2,T \wedge \tau_{n,L+2} \wedge \tau_{L+2}} > 1) + \delta. \end{aligned}$$

Sending $n \rightarrow \infty$ and using Lemma 8.3 shows $\limsup_n P(\tau_{n,L+2} \leq T) \leq \delta$. Therefore,

$$\limsup_n P(\|Z_n\|_{2,T} > L+2) \leq \limsup_n P(\tau_{n,L+2} \leq T) \leq \delta.$$

Since $\delta > 0$ is arbitrary, this shows that $\{\|Z_n\|_{2,T}\}_{n \in \mathbb{N}}$ is tight. The convergence in (8.10) now follows since $\{\tau_{n,L+1} \leq T\} \subseteq \{\|Z_n\|_{2,T} > L\}$. $\square$

PROOF OF THEOREM 2.3. Using Lemma 8.1 and Skorohod embedding theorem, we can assume without loss of generality that $Z_n(0)$, M_n and B are given on a common probability space, $Z_n(0) \rightarrow z$ in $\ell_1^\downarrow$, and $M_n \rightarrow \sqrt{2}Be_1$ in $\mathbb{D}([0, \infty) : \ell_2)$ a.s. From Lemma 8.3, we now have that for every $T, L \in (0, \infty)$ (8.7) holds. In fact, this shows $\|Z_n - Z\|_{2,T} \xrightarrow{P} 0$ as $n \rightarrow \infty$ using Lemma 6.5, (8.10) and the fact $\lim_{L \rightarrow \infty} P(\tau_L \leq T) = 0$ observed in the proof of Corollary 8.4. $\square$

9. Proof of Theorem 2.4. In this section, we give the proof of Theorem 2.4. As for the proof of Theorem 2.3, we begin with a convenient representation for Z_n and by establishing some useful convergence properties.

LEMMA 9.1. *Let λ_n, α_n, d_n be as in the statement of Theorem 2.4. Suppose that $\{\|Z_n(0)\|_1\}_{n \in \mathbb{N}}$ is a tight sequence of random variables and $Z_{n,r+}(0) \xrightarrow{P} \mathbf{0}$ in ℓ_2 for some $r \geq 2$. Then there are real stochastic processes $\{W_{n,i}\}_{i=1}^r$ and η_n with sample paths in $\mathbb{D}([0, \infty) : \mathbb{R})$ so that, $W_{n,1}, \eta_n$ have absolutely continuous paths a.s., $W_{n,1}(0) = \eta_n(0) = 0$, and for any $t \geq 0$,*

$$(9.1) \quad Z_{n,1}(t) = Z_{n,1}(0) - \int_0^t Z_{n,1}(s) ds + \int_0^t Z_{n,2}(s) ds + \sqrt{n}M_{n,1}(t) + W_{n,1}(t) - \eta_n(t),$$

$$(9.2) \quad Z_{n,2}(t) = Z_{n,2}(0) - \int_0^t Z_{n,2}(s) ds + \int_0^t Z_{n,3}(s) ds + W_{n,2}(t) + \eta_n(t),$$

$$(9.3) \quad Z_{n,i}(t) = Z_{n,i}(0) - \int_0^t Z_{n,i}(s) ds + \int_0^t Z_{n,i+1}(s) ds + W_{n,i}(t) \quad \text{for } i \in \{3, \dots, r\}.$$

Furthermore, η_n is nondecreasing process with $\eta_n(0) = 0$ that satisfies

$$(9.4) \quad \eta_n(t) = \int_0^t \mathbb{I}_{\{Z_{n,1}(s) \geq \theta_n\}} d\eta_n(s) \quad a.s.$$

for some constants $\theta_n = \alpha_n + O(\sqrt{n}/d_n) \geq 0$ as $n \rightarrow \infty$. Also for any $L, T \in (0, \infty)$, as $n \rightarrow \infty$:

- (1) $\sqrt{n}M_{n,1} \Rightarrow \sqrt{2}B$ in $\mathbb{D}([0, T]: \mathbb{R})$
- (2) $\text{tv}(W_{n,1}; [0, T_n]) \doteq \int_0^{T_n} |\dot{W}_{n,1}(s)| ds \xrightarrow{P} 0$
- (3) $|W_{n,i}|_{*, T_n} \xrightarrow{P} 0$ for $i \in \{2, \dots, r\}$
- (4) $\|Z_{n,r+}\|_{2, T_n} \xrightarrow{P} 0$.

Here, B is a standard Brownian motion and $T_n \doteq T \wedge \tau_{n,L}$.

PROOF. By our assumptions on α_n , we can find a $\kappa \in (0, \infty)$ such that $\theta_n \doteq \alpha_n + \frac{\kappa\sqrt{n}}{d_n} \geq 0$ for every n . Note that $\theta_n \rightarrow \alpha$ as $n \rightarrow \infty$. Recall the functions $q_{n,i}$ defined in (6.4). Define

$$W_{n,1}(t) \doteq - \int_0^t q_{n,1}(Z_{n,1}(s)) \mathbb{I}_{\{Z_{n,1}(s) < \theta_n\}} ds,$$

$$\eta_n(t) \doteq \int_0^t q_{n,1}(Z_{n,1}(s)) \mathbb{I}_{\{Z_{n,1}(s) \geq \theta_n\}} ds$$

so that $\eta_n(t) = \int_0^t \mathbb{I}_{\{Z_{n,1}(s) \geq \theta_n\}} d\eta_n(s)$, and

$$(9.5) \quad \int_0^t q_{n,1}(Z_{n,1}(s)) ds = \eta_n(t) - W_{n,1}(t).$$

From Lemma 6.1, it then follows that (9.1) is satisfied. Recall from (6.4) that $q_{n,1}(z) = \lambda_n \sqrt{n} \{\beta_n(\lambda_n + z/\sqrt{n}) - \beta_n(\lambda_n)\}$. Then, by monotonicity of β_n , $q_{n,1}(z) \geq 0$ whenever $z \geq 0$. The condition $\theta_n \geq 0$ shows that η_n is nondecreasing and

$$\begin{aligned} \sup_{z \leq \theta_n} |q_{n,1}(z)| &\leq \sqrt{n} \beta_n(\lambda_n + \theta_n/\sqrt{n}) \leq \sqrt{n}(\lambda_n + \theta_n/\sqrt{n})^{d_n} \\ &= \sqrt{n}(1 - ((\log d_n)/d_n + (\alpha_n - \theta_n)/\sqrt{n}))^{d_n} = \sqrt{n}(1 - (\log d_n - \kappa)/d_n)^{d_n} \\ &\leq \exp\left(-\log \frac{d_n}{\sqrt{n}} + \kappa\right) \rightarrow 0 \quad \text{as } n \rightarrow \infty. \end{aligned}$$

This shows that $\text{tv}(W_{n,1}; [0, T]) \rightarrow 0$ a.s.

Next, since $d_n(1 - \lambda_n) \rightarrow \infty$, Lemma 6.6 shows that

$$(9.6) \quad \mu_n \rightarrow f_1 \in \ell_1^\downarrow \quad \text{as } n \rightarrow \infty.$$

Therefore, $G_n(0) = \mu_n + \frac{Z_n(0)}{\sqrt{n}} \rightarrow f_1$ in $\ell_1^\downarrow$. Then by Lemma 6.2,

$$(9.7) \quad \sqrt{n}M_n \Rightarrow \sqrt{2}B e_1 \text{ in } \mathbb{D}([0, \infty]: \ell_2),$$

and by Lemma 6.4, $\|Z_{n,r+}\|_{2, T \wedge \tau_{n,L}} \xrightarrow{P} 0$ as $n \rightarrow \infty$. Define

$$W_{n,2}(t) \doteq - \int_0^t q_{n,2}(Z_{n,2}(s)) ds + \sqrt{n}M_{n,2}(t) - W_{n,1}(t).$$

Using (9.5) and Lemma 6.1 once more, we see that (9.2) is satisfied. Finally, for $i \in \{3, \dots, r\}$, define

$$W_{n,i}(t) \doteq \int_0^t q_{n,i-1}(Z_{n,i-1}(s)) ds - \int_0^t q_{n,i}(Z_{n,i}(s)) + \sqrt{n}M_{n,i}(t).$$

Then, from Lemma 6.1 again, it follows that (9.3) is satisfied with the above choice of $W_{n,i}$. Lemma 6.3 along with (9.6), (9.7) and $|Z_{n,i}|_{*,T \wedge \tau_{n,L}} \leq L$ show that, as $n \rightarrow \infty$, and $|W_{n,i}|_{*,T_n} \xrightarrow{P} 0$ for each $i \in \{2, \dots, r\}$. $\square$

COROLLARY 9.2. *Suppose that the assumptions in Lemma 9.1 are satisfied. Assume further that $d_n \ll n^{2/3}$. Then the conclusions of Lemma 9.1 hold with $\theta_n = \alpha_n$ and*

$$(9.8) \quad \eta_n(t) \doteq \int_0^t \gamma_n^{-1} (1 + \delta_n(s))^+ e^{\gamma_n(Z_{n,1}(s) - \alpha_n)} \mathbb{I}_{\{Z_{n,1}(s) \geq \alpha_n\}} ds,$$

where $\gamma_n \doteq \frac{d_n}{\sqrt{n}}$ and δ_n is a process with sample paths in $\mathbb{D}([0, \infty), \mathbb{R})$ such that $|\delta_n|_{*,T \wedge \tau_{n,L}} \rightarrow 0$ a.s. for each $L > 0$.

PROOF. Since $d_n \ll n^{2/3}$ and $\alpha_n = O(n^{1/6})$, the hypothesis of Lemma 6.7 is satisfied. Define

$$\delta_n(s) \doteq q_{n,1}(Z_{n,1}(s)) \gamma_n (e^{\gamma_n[Z_{n,1}(s) - \alpha_n]} - e^{-\gamma_n \alpha_n})^{-1} - 1.$$

Since $\sup_{s \leq T \wedge \tau_{n,L}} |Z_{n,1}(s)| \leq L$, Lemma 6.7 shows that $|\delta_n|_{*,T_n} \rightarrow 0$ a.s. as $n \rightarrow \infty$. Next, define

$$\begin{aligned} W_{n,1}(t) &\doteq \gamma_n^{-1} \int_0^t (1 + \delta_n(s)) (e^{-\gamma_n \alpha_n} - e^{\gamma_n(Z_{n,1}(s) - \alpha_n)} \mathbb{I}_{\{Z_{n,1}(s) < \alpha_n\}}) ds \\ &\quad + \int_0^t \gamma_n^{-1} (1 + \delta_n(s))^- e^{\gamma_n(Z_{n,1}(s) - \alpha_n)} \mathbb{I}_{\{Z_{n,1}(s) \geq \alpha_n\}} ds. \end{aligned}$$

Then $W_{n,1}(0) = 0$, $W_{n,1}$ is absolutely continuous and, with $\kappa = \sup_n \frac{d_n}{\sqrt{n}} \alpha_n^- < \infty$,

$$\begin{aligned} \text{tv}(W_{n,1}; [0, T_n]) \mathbb{I}_{\{|\delta_n|_{*,T_n} < 1\}} &= \gamma_n^{-1} \int_0^{T_n} |1 + \delta_n(s)| |e^{-\gamma_n \alpha_n} - e^{\gamma_n(Z_{n,1}(s) - \alpha_n)} \mathbb{I}_{\{Z_{n,1}(s) < \alpha_n\}}| ds \\ &\leq \frac{2(1 + e^\kappa)T}{\gamma_n} \rightarrow 0 \quad \text{as } n \rightarrow \infty. \end{aligned}$$

Hence, since $|\delta_n|_{*,T_n} \rightarrow 0$, we have that $\text{tv}(W_{n,1}; [0, T_n]) \xrightarrow{P} 0$ as $n \rightarrow \infty$. By rearranging terms, we see that, with the above definitions of $W_{n,1}$ and η_n , (9.5) is satisfied. The result follows. $\square$

Since $\gamma_n \rightarrow \infty$ and $\theta_n \rightarrow \alpha$ as $n \rightarrow \infty$, the previous lemma suggests a connection to the Skorokhod map Γ_α defined in (2.1). In order to make this connection precise, we begin with the following lemma.

LEMMA 9.3. *Under the assumptions of Theorem 2.4, for any $L \in (0, \infty)$,*

$$(9.9) \quad \sup_{t \in [0, T \wedge \tau_{n,L}]} (Z_{n,1}(t) - \alpha_n)^+ \xrightarrow{P} 0 \quad \text{as } n \rightarrow \infty.$$

PROOF. Consider first the case when $d_n \gg \sqrt{n} \log n$. For this case, $\epsilon_n \doteq \frac{\sqrt{n} \log d_n}{d_n} \rightarrow 0$, and since

$$Z_{n,1}(t) = \sqrt{n}(G_{n,1}(t) - \lambda_n) \leq \sqrt{n}(1 - \lambda_n) = \frac{\sqrt{n} \log d_n}{d_n} + \alpha_n,$$

we have that (9.9) holds. Now consider the complementary case, namely $d_n \gg \sqrt{n}$ but $d_n \gg \sqrt{n} \log n$ does not hold. In this case, we may find an increasing subsequence $\{n_k\}_{k \in \mathbb{N}} \subseteq \mathbb{N}$ so

that $\sqrt{n} \ll d_n \ll n^{2/3}$ holds when $n \in \{n_k\}_{k \in \mathbb{N}}$ and $d_n \gg \sqrt{n} \log n$ holds when $n \notin \{n_k\}_{k \in \mathbb{N}}$ (e.g., take $\{n_k\}_{k \in \mathbb{N}} \doteq \{n \in \mathbb{N} \mid d_n \leq n^{0.6}\}$). The argument above shows the convergence of (9.9) along the latter subsequence. Therefore, it suffices to show the convergence of (9.9) along the subsequence $\{n_k\}_{k \in \mathbb{N}}$ where $\sqrt{n} \ll d_n \ll n^{2/3}$.

We will use Corollary 9.2. Since $Z_{n,1}(0) \xrightarrow{P} z_1 \in \mathbb{R}$ with $z_1 \leq \alpha$, we have $(Z_{n,1}(0) - \alpha_n)^+ \xrightarrow{P} 0$ as $n \rightarrow \infty$. It now suffices to show that for any $\epsilon > 0$,

$$P\left(\sup_{t \in [0, T \wedge \tau_{n,L}]} Z_{n,1}(t) > \alpha_n + 6\epsilon\right) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Let $\vartheta_n \doteq \inf\{t \geq 0 \mid Z_{n,1}(t) > \alpha_n + 6\epsilon\}$ and, as before, $T_n \doteq T \wedge \tau_{n,L}$. It is then enough to show that $P(\vartheta_n \leq T_n) \rightarrow 0$ as $n \rightarrow \infty$. For this, inductively define stopping times, $\sigma_{n,0} = 0$,

$$\begin{aligned} \sigma_{n,2k-1} &= \inf\{t > \sigma_{n,2k-2} \mid Z_{n,1}(t) > \alpha_n + 3\epsilon\}, \\ \sigma_{n,2k} &= \inf\{t > \sigma_{n,2k-1} \mid Z_{n,1}(t) < \alpha_n + 2\epsilon\}, \end{aligned} \quad k \in \mathbb{N}.$$

Note that for each $n \in \mathbb{N}$, $\sigma_{n,r} \rightarrow \infty$ as $r \rightarrow \infty$, almost surely. Also, henceforth, without loss of generality, we consider only n that are large enough so that $1/\sqrt{n} < \epsilon$. Hence on the set $\{\vartheta_n < \infty\}$, $\vartheta_n \in [\sigma_{n,2k-1}, \sigma_{n,2k})$ for some $k \in \mathbb{N}$. Then for every $K \in \mathbb{N}$,

$$P(\vartheta_n \leq T_n) \leq \sum_{k=1}^K P(\vartheta_n \in [\sigma_{n,2k-1}, \sigma_{n,2k} \wedge T_n]) + P(\sigma_{n,2K+1} \leq T_n).$$

Hence to complete the proof it is enough to show that:

- (1) For each $k \in \mathbb{N}$, $\lim_{n \rightarrow \infty} P(\vartheta_n \in [\sigma_{n,2k-1}, \sigma_{n,2k} \wedge T_n]) = 0$,
- (2) $\lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} P(\sigma_{n,2K+1} \leq T_n) = 0$.

Consider (1) first. Note that on the set $C_{n,1} \doteq \{Z_{n,1}(0) \leq \alpha_n + 3\epsilon\}$, for any $k \in \mathbb{N}$,

$$\begin{aligned} \alpha_n + 3\epsilon &\leq Z_{n,1}(\sigma_{n,2k-1}) = Z_{n,1}(\sigma_{n,2k-1}-) + Z_{n,1}(\sigma_{n,2k-1}) - Z_{n,1}(\sigma_{n,2k-1}-) \\ (9.10) \quad &\leq Z_{n,1}(\sigma_{n,2k-1}-) + \epsilon \leq \alpha_n + 4\epsilon. \end{aligned}$$

Similarly,

$$(9.11) \quad Z_{n,1}(t) \geq \alpha_n + \epsilon \quad \text{for each } t \in [\sigma_{n,2k-1}, \sigma_{n,2k}].$$

Let $H_n(t) \doteq \sqrt{n}M_{n,1}(t + \sigma_{n,2k-1}) - \sqrt{n}M_{n,1}(\sigma_{n,2k-1})$ for $t \geq 0$ and consider the sets

$$C_{n,2} \doteq \{\vartheta_n \vee \sigma_{n,2k-1} \leq T_n\}, \quad C_{n,3} \doteq \left\{|W_{n,1}|_{*,T_n} \leq \epsilon/2, |\delta_n|_{*,T_n} \leq \frac{1}{2}\right\}.$$

Then on the set $C_n = \bigcap_{i=1}^3 C_{n,i}$, using Corollary 9.2, for any $t \in [0, (T_n \wedge \sigma_{n,2k}) - \sigma_{n,2k-1}]$,

$$\begin{aligned} &Z_{n,1}(t + \sigma_{n,2k-1}) - Z_{n,1}(\sigma_{n,2k-1}) \\ &= - \int_{\sigma_{n,2k-1}}^{\sigma_{n,2k-1}+t} (Z_{n,1}(s) - Z_{n,2}(s)) ds + H_n(t) + W_{n,1}(t + \sigma_{n,2k-1}) - W_{n,1}(\sigma_{n,2k-1}) \\ &\quad - \int_{\sigma_{n,2k-1}}^{\sigma_{n,2k-1}+t} \gamma_n^{-1} (1 + \delta_n(s))^+ e^{\gamma_n(Z_{n,1}(s) - \alpha_n)} \mathbb{I}_{\{Z_{n,1}(s) \geq \alpha_n\}} ds. \end{aligned}$$

Since for t in the above interval $\sigma_{n,2k-1} + t \leq T_n \leq \tau_{n,L}$, $|Z_{n,1}(s)| + |Z_{n,2}(s)| \leq 2L$ for any $s \leq \sigma_{n,2k-1} + t$. Also, since $\sigma_{n,2k-1} + t \leq \sigma_{n,2k}$, by (9.11), $Z_{n,1}(s) - \alpha_n \geq \epsilon$ for any $s \in [\sigma_{n,2k-1}, \sigma_{n,2k-1} + t]$. Thus on C_n we have

$$(9.12) \quad Z_{n,1}(t + \sigma_{n,2k-1}) - Z_{n,1}(\sigma_{n,2k-1}) \leq 2Lt + H_n(t) + \epsilon - \frac{t}{2\gamma_n} \exp(\gamma_n \epsilon) \doteq Y_n(t).$$

Using (9.10), on C_n , $Z_{n,1}(\vartheta_n) - Z_{n,1}(\sigma_{n,2k-1}) \geq \alpha_n + 6\epsilon - \alpha_n - 4\epsilon = 2\epsilon$. Hence

$$\begin{aligned}
 & \mathbf{P}(\vartheta_n \in [\sigma_{n,2k-1}, \sigma_{n,2k} \wedge T_n]) \\
 (9.13) \quad & \leq \mathbf{P}(\vartheta_n \in [\sigma_{n,2k-1}, \sigma_{n,2k} \wedge T_n], C_n) + \mathbf{P}(C_{n,1}^c) + \mathbf{P}(C_{n,3}^c) \\
 & \leq \mathbf{P}\left(\sup_{t \in [0, T]} Y_n(t) \geq 2\epsilon\right) + \mathbf{P}(C_{n,1}^c) + \mathbf{P}(C_{n,3}^c),
 \end{aligned}$$

where the second inequality is on observing that on the set $\{\vartheta_n \in [\sigma_{n,2k-1}, \sigma_{n,2k} \wedge T_n]\}$, (9.12) holds with t replaced by $\vartheta_n - \sigma_{n,2k-1}$. Next, note that H_n is a $\{\mathcal{G}_t^n\}$ martingale, where $\mathcal{G}_t^n = \mathcal{F}_{t+\sigma_{n,2k-1}}^n$ and

$$\begin{aligned}
 \langle H_n \rangle_t &= \langle \sqrt{n} M_{n,1} \rangle_{t+\sigma_{n,2k-1}} - \langle \sqrt{n} M_{n,1} \rangle_{\sigma_{n,2k-1}} \\
 &= \int_{\sigma_{n,2k-1}}^{\sigma_{n,2k-1}+t} [G_{n,1}(s) - G_{n,2}(s) + \lambda_n - \lambda_n \beta_n(G_{n,1}(s))] ds \\
 &\leq 2t,
 \end{aligned}$$

where the second equality is from (3.6).

Since $\gamma_n \rightarrow \infty$, we can apply Lemma 6.9 to conclude

$$\mathbf{P}\left(\sup_{t \in [0, T]} Y_n(t) \geq 2\epsilon\right) = \mathbf{P}\left(\sup_{t \in [0, T]} \left[H_n(t) - \left(\frac{\exp(\gamma_n \epsilon)}{2\gamma_n} - 2L\right)t\right] \geq \epsilon\right) \rightarrow 0$$

as $n \rightarrow \infty$. We also have $\lim_n \mathbf{P}(C_{n,i}^c) = 0$ for $i = 1, 3$ since, as noted earlier $(Z_{n,1}(0) - \alpha_n)^+ \xrightarrow{P} 0$, and by Corollary 9.2, respectively. From these observations, it follows that the right-hand side of (9.13) converges to 0 as $n \rightarrow \infty$, which completes the proof of (1).

Now we prove (2). Let $\rho_{n,i} \doteq \sigma_{n,i} \wedge \tau_{n,L}$ and define

$$Y_{n,K}(t) \doteq \sum_{i=0}^K (Z_{n,1}(t \wedge \rho_{n,2i+1}) - Z_{n,1}(t \wedge \rho_{n,2i})).$$

Note that $\{\sigma_{n,2K+1} \leq T_n\} \subseteq \{Y_{n,K}(T) \geq K\epsilon\}$, and hence to prove (2) it is sufficient to show that

$$(9.14) \quad \limsup_{n \rightarrow \infty} \mathbf{P}(Y_{n,K}(T) \geq K\epsilon) \rightarrow 0 \quad \text{as } K \rightarrow \infty.$$

From Corollary 9.2, we have that on the set $C_{n,4} \doteq \{\text{tv}(W_{n,1}; [0, T_n]) \leq 1\}$,

$$\begin{aligned}
 & Y_{n,K}(T) \\
 &= \sum_{i=0}^K \int_{T \wedge \rho_{n,2i}}^{T \wedge \rho_{n,2i+1}} (Z_{n,2}(s) - Z_{n,1}(s)) ds \\
 &+ \sum_{i=0}^K \sqrt{n} M_{n,1}(T \wedge \rho_{n,2i+1}) - \sqrt{n} M_{n,1}(T \wedge \rho_{n,2i}) \\
 &+ \sum_{i=0}^K W_{n,1}(T \wedge \rho_{n,2i+1}) - W_{n,1}(T \wedge \rho_{n,2i}) \\
 &- \sum_{i=0}^K \int_{T \wedge \rho_{n,2i}}^{T \wedge \rho_{n,2i+1}} \gamma_n^{-1} (1 + \delta_n(s))^+ e^{\gamma_n(Z_{n,1}(s) - \alpha_n)} \mathbb{I}_{\{Z_{n,1}(s) \geq \alpha_n\}} ds
 \end{aligned}$$

$$\begin{aligned} &\leq 2LT + \sum_{i=0}^K (\sqrt{n}M_{n,1}(T \wedge \rho_{n,2i+1}) - \sqrt{n}M_{n,1}(T \wedge \rho_{n,2i})) + \text{tv}(W_{n,1}; [0, T]) \\ &\leq 2LT + 1 + H_{n,K}(T), \end{aligned}$$

where we have used the facts that $\sup_{s \leq \tau_{n,L}} |Z_{n,1}(s)| \leq L$, and that the rightmost term in the third line is nonpositive. Also, here

$$H_{n,K}(t) \doteq \sum_{i=0}^K (\sqrt{n}M_{n,1}(t \wedge \rho_{n,2i+1}) - \sqrt{n}M_{n,1}(t \wedge \rho_{n,2i})).$$

Using (3.6), we see that $H_{n,K}$ is a $\mathcal{F}_t^n$ -martingale with quadratic variation given by

$$\begin{aligned} \langle H_{n,K} \rangle_t &= \sum_{i=0}^K (\langle \sqrt{n}M_{n,1} \rangle_{t \wedge \rho_{n,2i+1}} - \langle \sqrt{n}M_{n,1} \rangle_{t \wedge \rho_{n,2i}}) \\ &= \sum_{i=0}^K \int_{t \wedge \rho_{n,2i}}^{t \wedge \rho_{n,2i+1}} (G_{n,1}(s) - G_{n,2}(s) + \lambda_n - \lambda_n \beta_n(G_{n,1}(s))) ds \leq 2t. \end{aligned}$$

Hence

$$\begin{aligned} \mathbf{P}(Y_{n,K}(T) \geq K\epsilon) &\leq \mathbf{P}(Y_{n,K}(T) \geq K\epsilon, C_{n,4}) + \mathbf{P}(C_{n,4}^c) \\ &\leq \mathbf{P}(H_{n,K}(T) > K\epsilon - (2LT + 1)) + \mathbf{P}(C_{n,4}^c) \\ &\leq \frac{EH_{n,K}^2(T)}{(K\epsilon - (2LT + 1))^2} + \mathbf{P}(C_{n,4}^c) \\ &\leq \frac{2T}{(K\epsilon - (2LT + 1))^2} + \mathbf{P}(C_{n,4}^c). \end{aligned}$$

From Corollary 9.2, $\mathbf{P}(C_{n,4}^c) \rightarrow 0$ as $n \rightarrow \infty$. This together with the above display shows $\lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbf{P}(Y_{n,K}(T) \geq K\epsilon) = 0$. Thus we have shown (9.14) and the proof of (2) is complete. The result follows. $\square$

LEMMA 9.4. *Suppose the hypothesis of Theorem 2.4 holds, then for each $n \in \mathbb{N}$, there is a real constant $\theta_n = \alpha_n + O(\sqrt{n}/d_n) \geq 0$ and processes $\tilde{W}_{n,1}$, $W_{n,2}$ with sample paths in $\mathbb{D}([0, \infty) : \mathbb{R})$ such that with $\tilde{Z}_{n,1} \doteq Z_{n,1} \wedge \theta_n$,*

$$\begin{aligned} \tilde{Z}_{n,1}(t) &= \Gamma_{\theta_n} \left(\tilde{Z}_{n,1}(0) - \int_0^\cdot (\tilde{Z}_{n,1}(s) - Z_{n,2}(s)) + \sqrt{n}M_{n,1}(\cdot) + \tilde{W}_{n,1}(\cdot) \right)(t), \quad \text{and} \\ (9.15) \quad Z_{n,2}(t) &= Z_{n,2}(0) - \int_0^t (Z_{n,2}(s) - Z_{n,3}(s)) ds + W_{n,2}(t) + \eta_n(t) \quad \text{for all } t > 0, \end{aligned}$$

where

$$(9.16) \quad \eta_n = \hat{\Gamma}_{\theta_n} \left(\tilde{Z}_{n,1}(0) - \int_0^\cdot (\tilde{Z}_{n,1}(s) - Z_{n,2}(s)) + \sqrt{n}M_{n,1}(\cdot) + \tilde{W}_{n,1}(\cdot) \right).$$

Furthermore, for any $L, T \in (0, \infty)$, the random variables $|(Z_{n,1} - \theta_n)^+|_{*, T \wedge \tau_{n,L}}$, $|\tilde{W}_{n,1}|_{*, T \wedge \tau_{n,L}}$ and $|W_{n,2}|_{*, T \wedge \tau_{n,L}}$ converge to zero in probability as $n \rightarrow \infty$.

PROOF. Let θ_n be as in Lemma 9.1. Since $d_n \gg \sqrt{n}$, $\theta_n = \alpha_n + o(1)$ and Lemma 9.3 shows

$$(9.17) \quad |(Z_{n,1} - \theta_n)^+|_{*, T \wedge \tau_{n,L}} \rightarrow 0.$$

Note that $\tilde{Z}_{n,1} = Z_{n,1} - (Z_{n,1} - \theta_n)^+$. Hence we can rewrite (9.1) and (9.2) as

$$\begin{aligned} \tilde{Z}_{n,1}(t) &= \tilde{Z}_{n,1}(0) - \int_0^t \tilde{Z}_{n,1}(s) ds \\ (9.18) \quad &+ \int_0^t Z_{n,2}(s) ds + \sqrt{n}M_{n,1}(t) + \tilde{W}_{n,1}(t) - \eta_n(t), \\ Z_{n,2}(t) &= Z_{n,2}(0) - \int_0^t (Z_{n,2}(s) - Z_{n,3}(s)) ds + W_{n,2}(t) + \eta_n(t), \end{aligned}$$

where

$$\tilde{W}_{n,1}(t) \doteq W_{n,1}(t) - \int_0^t (Z_{n,1}(s) - \theta_n)^+ ds - (Z_{n,1}(t) - \theta_n)^+ + (Z_{n,1}(0) - \theta_n)^+.$$

The properties of η_n from Lemma 9.1 (and Corollary 9.2) say that η_n is a nondecreasing process, with $\eta_n(0) = 0$ and $\eta_n(t) = \int_0^t \mathbb{I}_{\{\tilde{Z}_{n,1}(s) = \theta_n\}} d\eta_n(s)$. Since $\tilde{Z}_{n,1} \leq \theta_n$, (9.18) and the characterizing properties of the Skorokhod map show (9.15) and (9.16). Finally, by Lemma 9.1, Corollary 9.2 and Lemma 9.3,

$$|W_{n,1}|_{*,T \wedge \tau_{n,L}} \xrightarrow{P} 0, \quad \text{and} \quad |W_{n,2}|_{*,T \wedge \tau_{n,L}} \xrightarrow{P} 0$$

as $n \rightarrow \infty$. Hence, using (9.17), $|\tilde{W}_{n,1}|_{*,T \wedge \tau_{n,L}} \xrightarrow{P} 0$ as $n \rightarrow \infty$, and the result follows. $\square$

The following lemma will be needed in order to prove the tightness of $\mathbf{Z}_n$.

LEMMA 9.5. *Under the hypothesis of Theorem 2.4, the collection of random variables $\{\|\mathbf{Z}_n\|_{2,T}\}_{n \in \mathbb{N}}$ is tight for any $T \in (0, \infty)$.*

PROOF. Fix $T \in (0, \infty)$. In Lemma 9.4, using the definition of the Skorokhod map Γ_{θ_n} for $\theta_n \geq 0$ (see (2.1)), we have, for any $t > 0$ that

$$\eta_n(t) \leq |\tilde{Z}_{n,1}(0)| + \int_0^t |\tilde{Z}_{n,1}(s)| ds + \int_0^t |Z_{n,2}(s)| ds + |\sqrt{n}M_{n,1}|_{*,t} + |\tilde{W}_{n,1}|_{*,t}.$$

This shows that for any $t \geq 0$,

$$|\tilde{Z}_{n,1}|_{*,t} \leq 2 \left(|\tilde{Z}_{n,1}(0)| + \int_0^t |\tilde{Z}_{n,1}|_{*,s} ds + \int_0^t |Z_{n,2}|_{*,s} ds + |\sqrt{n}M_{n,1}|_{*,t} + |\tilde{W}_{n,1}|_{*,t} \right),$$

$$\begin{aligned} |Z_{n,2}|_{*,t} &\leq |\tilde{Z}_{n,1}(0)| + |Z_{n,2}(0)| + \int_0^t |\tilde{Z}_{n,1}|_{*,s} ds + \int_0^t (2|Z_{n,2}|_{*,s} + |Z_{n,3}|_{*,s}) ds \\ &\quad + |\sqrt{n}M_{n,1}|_{*,t} + |\tilde{W}_{n,1}|_{*,t} + |W_{n,2}|_{*,t}, \end{aligned}$$

and

$$|Z_{n,i}|_{*,t} \leq |Z_{n,i}(0)| + \int_0^t |Z_{n,i}|_{*,s} ds + \int_0^t |Z_{n,i+1}|_{*,s} ds + |W_{n,i}|_{*,t} \quad \text{for } i \in \{3, \dots, r\},$$

where the last line is from Lemma 9.1. Let $H_t \doteq |\tilde{Z}_{n,1}|_{*,t} + |Z_{n,2}|_{*,t} + \dots + |Z_{n,r}|_{*,t}$. By adding over equations in the above display, we have for $t \in [0, \tau]$ and $\tau \in [0, T]$ that

$$0 \leq H_t \leq 4 \left(H_0 + |\sqrt{n}M_{n,1}|_{*,\tau} + |\tilde{W}_{n,1}|_{*,\tau} + \sum_{i=2}^r |W_{n,i}|_{*,\tau} + \int_0^t H_s ds \right).$$

By Gronwall's inequality, for all $\tau \in [0, T]$,

$$(9.19) \quad H_\tau \leq 4 \left(H_0 + |\sqrt{n}M_{n,1}|_{*,\tau} + |\tilde{W}_{n,1}|_{*,\tau} + \sum_{i=2}^r |W_{n,i}|_{*,\tau} \right) e^{4\tau}.$$

Let $\vec{Z}_n \doteq (\tilde{Z}_{n,1}, Z_{n,2}, \dots, Z_{n,r})$. Since $\vec{Z}_n(0) \xrightarrow{P} (z_1, \dots, z_r)$, and $\sqrt{n}\mathbf{M}_n \Rightarrow B\mathbf{e}_1$, for every $\epsilon > 0$, there is a $L_1 \in (0, \infty)$ such that for every $n \in \mathbb{N}$,

$$P(C_{n,1}) \leq \frac{\epsilon}{2}, \text{ where } C_{n,1} \doteq \{H_0 + |\sqrt{n}M_{n,1}|_{*,T} \geq L_1\}.$$

Applying Lemmas 9.1 and Lemma 9.4 with $L = 4(L_1 + 1)e^{4T} + 2$, we can find an $n_0 \in \mathbb{N}$ so that $P(C_{n,2}) \leq \frac{\epsilon}{2}$ for $n \geq n_0$, where

$$C_{n,2} \doteq \left\{ |\tilde{W}_{n,1}|_{*,T_n} + \sum_{i=2}^r |W_{n,i}|_{*,T_n} + |(Z_{n,1} - \theta_n)^+|_{*,T_n} + \|\mathbf{Z}_{n,r+}\|_{2,T_n} \geq 1 \right\}$$

and $T_n \doteq T \wedge \tau_{n,L}$. On the event $(C_{n,1} \cup C_{n,2})^c$,

$$\|\vec{Z}_n\|_{1,T_n} = H_{T_n} < 4(L_1 + 1)e^{4T}$$

by (9.19), and hence by triangle inequality (and noting $\|\vec{x}\|_2 \leq \|\vec{x}\|_1$),

$$(9.20) \quad \begin{aligned} \|\mathbf{Z}_n\|_{2,T_n} &\leq \|\vec{Z}_n\|_{1,T_n} + |(Z_{n,1} - \theta_n)^+|_{*,T_n} + \|\mathbf{Z}_{n,r+}\|_{2,T_n} \\ &< 4(L_1 + 1)e^{4T} + 1 = L - 1. \end{aligned}$$

Also, by the definition of $\tau_{n,L}$, $\|\mathbf{Z}_n(\tau_{n,L})\|_2 \geq L - \frac{1}{\sqrt{n}}$ on the set $\tau_{n,L} < T$. Hence we must have that $\tau_{n,L} > T$ whenever (9.20) holds, and hence

$$\|\mathbf{Z}_n\|_{2,T} < L - 1 \quad \text{on the event } (C_{n,1} \cup C_{n,2})^c.$$

This shows that

$$P(\|\mathbf{Z}_n\|_{2,T} \geq L) \leq P(C_{n,1} \cup C_{n,2}) \leq \epsilon \quad \forall n \geq n_0.$$

Since $\epsilon > 0$ is arbitrary, the result follows. $\square$

The following result is immediate from Lemmas 6.5, 9.1, 9.4 and 9.5.

COROLLARY 9.6. *Under the hypothesis of Theorem 2.4, for any $T > 0$, $\lim_{L \rightarrow \infty} \sup_n P(\tau_{n,L} \leq T) = 0$. In particular, the processes $\tilde{W}_{n,1}$, $\{W_{n,i}\}_{i=2}^r$, $\|\mathbf{Z}_{n,r+}\|_2$, $(Z_{n,1} - \theta_n)^+$ converge in probability to zero in $\mathbb{D}([0, \infty) : \mathbb{R})$ as $n \rightarrow \infty$.*

COROLLARY 9.7. *Under the hypothesis of Theorem 2.4, the sequence of processes $\{\mathbf{Z}_n\}_{n \in \mathbb{N}}$ is tight in $\mathbb{D}([0, \infty) : \ell_2)$.*

PROOF. Let θ_n be as in Lemma 9.4. For the sequence $\{\vec{Z}_n\}_{n \in \mathbb{N}}$ introduced in the proof of Lemma 9.5, note that

$$(9.21) \quad \mathbf{Z}_n = \mathbf{P}\vec{Z}_n + (Z_{n,1} - \theta_n)^+ \mathbf{e}_1 + \mathbf{S}_r \mathbf{Z}_{n,r+},$$

where $\mathbf{P} : \mathbb{R}^r \rightarrow \ell_2$ is given by $\mathbf{P}(x_1, \dots, x_r) = (x_1, \dots, x_r, \mathbf{0})$ while $\mathbf{S}_r : \ell_2 \rightarrow \ell_2$ is given by $\mathbf{S}_r \mathbf{y} = (\vec{0}, \mathbf{y})$ where $\vec{0}$ is the zero vector in $\mathbb{R}^r$. Since these maps are continuous, the tightness of $\{\vec{Z}_n\}_{n \in \mathbb{N}}$ in $\mathbb{D}([0, T] : \mathbb{R}^k)$, the tightness of $\{\mathbf{Z}_{n,r+}\}_{n \in \mathbb{N}}$ in $\mathbb{D}([0, T] : \ell_2)$ and the tightness of $\{(Z_{n,1} - \theta_n)^+\}_{n \in \mathbb{N}}$ in $\mathbb{D}([0, T] : \mathbb{R})$ will show the tightness of the sequence $\{\mathbf{Z}_n\}_{n \in \mathbb{N}}$ in $\mathbb{D}([0, T] : \ell_2)$. Note by Corollary 9.6, for each fixed $T < \infty$, $\|\mathbf{Z}_{n,r+}\|_{2,T} \xrightarrow{P} 0$ and $|(Z_{n,1} - \theta_n)^+|_{*,T} \xrightarrow{P} 0$. Hence it is sufficient to show that $\{\vec{Z}_n\}_{n \in \mathbb{N}}$ is tight in $\mathbb{D}([0, T] : \mathbb{R}^r)$. From Lemma 9.5, the convergence of $W_{n,i}$ in Corollary 9.6, and equations for $Z_{n,j}$,

$j = 3, \dots, r$ in Lemma 9.1, it is immediate that $(Z_{n,3}, \dots, Z_{n,r})$ is tight in $\mathbb{D}([0, \infty) : \mathbb{R}^{r-2})$. Finally, consider the pair $(\tilde{Z}_{n,1}, Z_{n,2})$. Note that

$$|\tilde{Z}_{n,1} - Z_{n,2}|_{*,T} \leq |(Z_{n,1} - \theta_n)^+|_{*,T} + |Z_{n,1}|_{*,T} + |Z_{n,2}|_{*,T} \leq |(Z_{n,1} - \theta_n)^+|_{*,T} + 2\|Z_n\|_{2,T}$$

and the right-hand side in the above display is tight in $\mathbb{R}_+$. This shows the tightness of

$$\int_0^\cdot (\tilde{Z}_{n,1}(s) - Z_{n,2}(s)) ds$$

in $\mathbb{C}([0, \infty) : \mathbb{R})$. Combining this observation with Lemma 9.5, the convergence of $\sqrt{n}M_{n,1}$ in Lemma 9.1, and the convergence of $\tilde{W}_{n,1}$ in Corollary 9.6, it follows that

$$(9.22) \quad R_n(\cdot) \doteq \tilde{Z}_{n,1}(0) - \int_0^\cdot (\tilde{Z}_{n,1}(s) - Z_{n,2}(s)) ds + \sqrt{n}M_{n,1}(\cdot) + \tilde{W}_{n,1}(\cdot)$$

is tight in $\mathbb{D}([0, \infty) : \mathbb{R})$. Using the identity,

$$\Gamma_{\theta_n}(R_n)(t) = \Gamma_{\theta_n}(\Gamma_{\theta_n}(R_n)(s) + R_n(\cdot + s) - R_n(s))(t - s)$$

for $0 \leq s \leq t \leq T$, we see from the definition of the Skorohod map that

$$|\Gamma_{\theta_n}(R_n)(t) - \Gamma_{\theta_n}(R_n)(s)| \leq 2 \sup_{s \leq u \leq t} |R_n(u) - R_n(s)|.$$

Together with the tightness of R_n , this immediately implies the tightness of $\tilde{Z}_{n,1} = \Gamma_{\theta_n}(R_n)$ and of $\hat{\Gamma}_{\theta_n}(R_n)$. Finally, the tightness of $Z_{n,2}$ is now immediate from Lemma 9.5, the convergence of $W_{n,2}$ in Corollary 9.6 and the tightness of $\hat{\Gamma}_{\theta_n}(R_n)$ noted above. The result follows. $\square$

PROOF OF THEOREM 2.4. From Lemma 6.6 and from the tightness of $\{\|Z_n(0)\|_1\}_{n \in \mathbb{N}}$, it follows under the conditions of the theorem that $\mu_n \xrightarrow{P} f_1$ and $G_n(0) \xrightarrow{P} f_1$ in $\ell_1^\downarrow$. This proves the first statement in the theorem. Now consider the second statement. Fix $T < \infty$. From Corollary 9.7, $\{Z_n\}_{n \in \mathbb{N}}$ is tight in $\mathbb{D}([0, \infty) : \ell_2)$. Also from Lemma 9.1, $\sqrt{n}M_{n,1}$ converges in distribution to $\sqrt{2}B$ where B is a standard Brownian motion and from Corollary 9.6,

$$(\tilde{W}_{n,1}, \{W_{n,i}\}_{i=2}^r, (Z_{n,1} - \theta_n)^+) \xrightarrow{P} \mathbf{0} \quad \text{in } \mathbb{D}([0, T] : \mathbb{R}^{r+1}).$$

Suppose that along a subsequence

$$(Z_n, \sqrt{n}M_{n,1}, \tilde{W}_{n,1}, \{W_{n,i}\}_{i=2}^r, (Z_{n,1} - \theta_n)^+) \Rightarrow (Z, \sqrt{2}B, \mathbf{0})$$

in $\mathbb{D}([0, \infty) : \ell_2 \times \mathbb{R}^{r+2})$ and for notational simplicity label the subsequence once more as $\{n\}$. Also by appealing to Skorohod embedding theorem, we assume that all the processes in the above display are given on a common probability space and the above convergence holds a.s. Since $J_T(Z_n) \doteq \sup_{0 \leq t \leq T} \|Z_n(t) - Z_n(t-)\|_2$ is at most $\frac{1}{\sqrt{n}}$ and $Z_n(0) \xrightarrow{P} z$, we have $J_T(Z) = 0$ and $Z(0) = z$ a.s. In particular, Z has sample paths in $\mathbb{C}([0, \infty) : \ell_2)$ and $(Z_n, \sqrt{n}M_{n,1}) \rightarrow (Z, \sqrt{2}B)$ uniformly over compact time intervals in $\ell_2 \times \mathbb{R}$. Since by Corollary 9.6, for every $T < \infty$, $\|Z_{n,r+}\|_{2,T} \xrightarrow{P} 0$, it suffices to show that $(Z_1, \dots, Z_r)$ along with B satisfy (2.13).

From the equations of $(Z_{n,3}, \dots, Z_{n,r})$ in Lemma 9.1, uniform convergence of Z_n to Z , and the uniform convergence of $\{W_{n,i}\}_{i=3}^r$ to 0, it is immediate that $(Z_3, \dots, Z_r)$ satisfy (2.13). Finally, consider the equations for (Z_1, Z_2) . From (9.22) and uniform convergence properties observed above, it is immediate that R_n converges uniformly, a.s., to R given as

$$R(\cdot) = Z_1(0) - \int_0^\cdot (Z_1(s) - Z_2(s)) + \sqrt{2}B(\cdot).$$

Since $\theta_n = \alpha_n + O(\sqrt{n}/d_n) \rightarrow \alpha$, this shows that, for every $T < \infty$,

$$\begin{aligned}\Gamma_{\theta_n}(R_n)(t) &= R_n(t) - \sup_{s \in [0, t]} (R_n(t) - \theta_n)^+ \\ &\rightarrow R(t) - \sup_{s \in [0, t]} (R(t) - \alpha)^+ = \Gamma_\alpha(R)(t)\end{aligned}$$

uniformly for $t \in [0, T]$, a.s., where $(R(t) - \alpha)^+$ is taken to be 0 when $\alpha = \infty$. Similarly,

$$\hat{\Gamma}_{\theta_n}(R_n)(t) \rightarrow \hat{\Gamma}_\alpha(R)(t)$$

uniformly for $t \in [0, T]$, a.s. Here, when $\alpha = \infty$, Γ_α and $\hat{\Gamma}_\alpha$ are as introduced in (2.11). The fact that (Z_1, Z_2) solve the first two equations in (2.13) is now immediate from Lemma 9.4, the convergence $\tilde{Z}_{n,1} - Z_{n,1} \xrightarrow{P} 0$, and the uniform convergence of $W_{n,2}$ to 0 noted previously. The result follows. $\square$

APPENDIX A: PROOFS OF RESULTS IN SECTION 5

A.1. Proof of Lemma 5.1. PROOF. Fix $\epsilon \in (0, 1)$. First suppose $\frac{d_n}{n} \rightarrow 0$. Consider $x \in (\epsilon, 1]$. Let $\Delta_n(x) \doteq \log \beta_n(x) - \log \gamma_n(x)$. Let $n_0 \in \mathbb{N}$ be such that for all $n \geq n_0$, $d_n/n < \epsilon/2$. Then, for $n \geq n_0$,

$$\begin{aligned}\Delta_n(x) &\doteq \sum_{i=0}^{d_n-1} \log\left(\frac{x - i/n}{1 - i/n}\right) - \log x^{d_n} = \sum_{i=0}^{d_n-1} \left\{ \log\left(\frac{x - i/n}{1 - i/n}\right) - \log x \right\} \\ (A.1) \quad &= \sum_{i=0}^{d_n-1} \log\left(\frac{1 - i/(nx)}{1 - i/n}\right) = \sum_{i=0}^{d_n-1} \log\left(1 - (i/n) \frac{1/x - 1}{1 - i/n}\right).\end{aligned}$$

Differentiating Δ_n gives,

$$\dot{\Delta}_n(x) = \sum_{i=0}^{d_n-1} \left(\frac{1}{x - i/n} - \frac{1}{x} \right) = \sum_{i=0}^{d_n-1} \frac{i/n}{x(x - i/n)}.$$

Since $n \geq n_0$ and $x \in [\epsilon, 1]$ we have $x(x - \frac{i}{n}) \geq \epsilon^2/2$ for $i \leq d_n - 1$. Hence,

$$|\dot{\Delta}_n(x)| \leq \frac{2}{\epsilon^2} \sum_{i=0}^{d_n-1} (i/n) \leq \frac{1}{\epsilon^2} \frac{d_n^2}{n}.$$

From the definition of Δ_n , we also have,

$$(A.2) \quad \dot{\Delta}_n(x) = \frac{\dot{\beta}_n(x)}{\beta_n(x)} - \frac{\dot{\gamma}_n(x)}{\gamma_n(x)} = \frac{\dot{\gamma}_n(x)}{\gamma_n(x)} \left(\frac{\dot{\beta}_n(x)}{\dot{\gamma}_n(x)} \frac{\gamma_n(x)}{\beta_n(x)} - 1 \right).$$

Since $\frac{\dot{\gamma}_n(x)}{\gamma_n(x)} = \frac{d_n}{x} \geq d_n$ for $x \in [\epsilon, 1]$, from (A.2) we have,

$$\sup_{x \in [\epsilon, 1]} \left| \frac{\dot{\beta}_n(x)}{\dot{\gamma}_n(x)} \frac{\gamma_n(x)}{\beta_n(x)} - 1 \right| \leq \frac{1}{d_n} \sup_{x \in [\epsilon, 1]} |\dot{\Delta}_n(x)| \leq \frac{1}{\epsilon} \frac{d_n}{n} \rightarrow 0.$$

This proves (5.4).

Now assume $\frac{d_n}{\sqrt{n}} \rightarrow 0$. Once more consider $x \in (\epsilon, 1]$ and $n \geq n_0$. Let $C \doteq \sup_{n \geq n_0} \frac{1/\epsilon - 1}{1 - d_n/n} < \infty$ and let $n_1 > n_0$ be such that $d_n C/n < 1/2$ for all $n \geq n_1$. Then for $n \geq n_1$ and $x \in [\epsilon, 1]$:

$$(A.3) \quad |\Delta_n(x)| \leq \sum_{i=0}^{d_n-1} 2 \left| (i/n) \frac{1/x - 1}{1 - i/n} \right| \leq 2C \sum_{i=0}^{d_n-1} i/n \leq C \frac{d_n^2}{n},$$

where the first inequality is from (A.1) and the inequality $|\log(1+h)| \leq 2|h|$ for $|h| \leq 1/2$. This shows $\sup_{x \in [\epsilon, 1]} |\Delta_n(x)| \rightarrow 0$, hence showing the first convergence in (5.5). Finally the second convergence (5.5) is immediate on combining the first convergence with (5.4). $\square$

A.2. Proof of Corollary 5.2. This is an immediate consequence of the estimate in (A.3).

A.3. Proof of Corollary 5.3. PROOF. Let $\epsilon > 0$ and $n_0 \in \mathbb{N}$ be such that $\mu_{n,i} > \epsilon$ for all $n \geq n_0$. By Lemma 5.1, as $n \rightarrow \infty$

$$(A.4) \quad \frac{\dot{\beta}_n(\mu_{n,i})}{\beta_n(\mu_{n,i})} = (1 + o(1)) \frac{\dot{\gamma}_n(\mu_{n,i})}{\gamma_n(\mu_{n,i})}.$$

Recall that $\mu_{n,i+1} \doteq \lambda_n \beta_n(\mu_{n,i})$ and $\gamma_n(x) \doteq x^{d_n}$. Hence (A.4) gives

$$(A.5) \quad \frac{\dot{\beta}_n(\mu_{n,i})}{\mu_{n,i+1}/\lambda_n} = (1 + o(1)) \frac{d_n}{\mu_{n,i}}$$

completing the proof. $\square$

A.4. Proof of Lemma 5.4. PROOF. From Corollary 5.2, there is a $n_0 \in \mathbb{N}$ and $C \in (0, \infty)$ such that for all $n \geq n_0$

$$\sup_{x \in [\epsilon, 1]} |\log \beta_n(x) - \log \gamma_n(x)| \leq \frac{C d_n^2}{n}.$$

Thus, if for $n \geq n_0$ and $i \in \mathbb{N}$, $\mu_{n,i} \geq \epsilon$, then

$$(A.6) \quad \begin{aligned} \log \mu_{n,i+1} &= \log \lambda_n + \log \beta_n(\mu_{n,i}) = \log \lambda_n + \log \gamma_n(\mu_{n,i}) + \gamma_{n,i} \\ &= \log \lambda_n + d_n \log \mu_{n,i} + \gamma_{n,i}, \end{aligned}$$

where $|\gamma_{n,i}| \leq \frac{C d_n^2}{n}$. Now let $k \in \mathbb{N}$ and $n_1 \in \mathbb{N}$ be such that for all $n \geq n_1$, $\mu_{n,k} \geq \epsilon$. We will show that for $n \geq n_0 \vee n_1$ and $j \in \{1, \dots, k\}$ that

$$(A.7) \quad \log \mu_{n,j+1} = (\log \lambda_n) \left(\sum_{i=0}^j d_n^i \right) + \beta_{n,j},$$

where $|\beta_{n,j}| \leq \frac{C}{n} \sum_{i=1}^j d_n^{i+1}$. Note the lemma is immediate from (A.7) on taking $j = k$.

To prove (A.7) we argue inductively. First note that since $\mu_n \in \ell_1^\downarrow$, $\mu_{n,i} \geq \mu_{n,k} \geq \epsilon$ for each $i \leq k$ and $n \geq n_1$. Hence (A.6) holds for each $i \leq k$ and $n \geq n_0 \vee n_1$. Taking $i = 1$ in (A.6) and noting that $\mu_{n,1} = \lambda_n$ proves (A.7) for the case $j = 1$.

Suppose now (A.7) holds for some $j \leq k - 1$. Then, using $i = j + 1$, in (A.6)

$$\log \mu_{n,j+2} = \log \lambda_n + d_n \log \mu_{n,j+1} + \gamma_{n,j+1},$$

where $|\gamma_{n,j+1}| \leq \frac{C d_n^2}{n}$. By the induction hypothesis, (A.7) holds for j . Hence

$$\begin{aligned} \log \mu_{n,j+2} &= \log \lambda_n + d_n \left\{ (\log \lambda_n) \left(\sum_{i=0}^j d_n^i \right) + \beta_{n,j} \right\} \\ &= (\log \lambda_n) \left(\sum_{i=0}^{j+1} d_n^i \right) + d_n \beta_{n,j} + \gamma_{n,j} \end{aligned}$$

and hence $\beta_{n,j+1} = d_n \beta_{n,j} + \gamma_{n,j}$. This shows

$$|\beta_{n,j+1}| = |d_n \beta_{n,j} + \gamma_{n,j}| \leq d_n \frac{C}{n} \sum_{i=1}^j d_n^{i+1} + \frac{C d_n^2}{n} = \frac{C}{n} \sum_{i=1}^{j+1} d_n^{i+1}$$

which shows that (A.7) holds for $j+1$. This completes the proof. $\square$

A.5. Proof of Corollary 5.5. PROOF. Since $d_n \rightarrow \infty$, the assumption $\frac{\xi_n^2}{d_n} \rightarrow 0$ shows that $\frac{|\xi_n|}{d_n} \leq \frac{1+\xi_n^2}{d_n} \rightarrow 0$. This shows that $\epsilon_n \doteq 1 - \lambda_n = \frac{\xi_n + \log d_n}{d_n^k}$ also converges to 0.

We first show that $\mu_{n,i} \rightarrow 1$ for each $i \in \{1, \dots, k\}$. We will argue inductively. Since $\mu_{n,1} \doteq \lambda_n = 1 - \epsilon_n$, we have $\mu_{n,1} \rightarrow 1$. Suppose now that $\mu_{n,i} \rightarrow 1$ for some $i \leq k-1$. Hence eventually $\mu_{n,i} \geq \frac{1}{2}$. Applying Lemma 5.4 with $k=i$ and $\epsilon = \frac{1}{2}$ and simplifying the resulting expression, we get

$$(A.8) \quad \log \mu_{n,i+1} = (\log \lambda_n) \frac{d_n^{i+1} - 1}{d_n - 1} + O\left(\frac{d_n^2(d_n^i - 1)}{n(d_n - 1)}\right)$$

$$(A.9) \quad \begin{aligned} &= O(\epsilon_n) \frac{d_n^{i+1} - 1}{d_n - 1} + O\left(\frac{d_n^2(d_n^i - 1)}{n(d_n - 1)}\right) \\ &= O\left(\frac{\xi_n + \log d_n}{d_n^{k-i}}\right) + O\left(\frac{d_n^{i+1}}{n}\right), \end{aligned}$$

where the second equality uses $\log \lambda_n = \log(1 - \epsilon_n) = O(\epsilon_n)$ and the third follows on recalling that $d_n \rightarrow \infty$. Since $i \leq k-1$, $\frac{|\xi_n|}{d_n^{k-i}} \leq \frac{1+\xi_n^2}{d_n} \rightarrow 0$. Using this along with $d_n^{k+1} \ll n$ in (A.9) shows that $\mu_{n,i+1} \rightarrow 1$. Hence, by induction, $\mu_{n,i} \rightarrow 1$ for $i \leq k$.

Next we argue that $\beta_n(\mu_{n,k}) \rightarrow \alpha$. Since $\lambda_n \rightarrow 1$ and $\mu_{n,k} \rightarrow 1$, from Corollary 5.3 we have that

$$\lim_{n \rightarrow \infty} \frac{\dot{\beta}_n(\mu_{n,k})}{d_n \mu_{n,k+1}} = 1.$$

Hence it suffices to show that $d_n \mu_{n,k+1} \rightarrow \alpha$. For this note that

$$\begin{aligned} \log(d_n \mu_{n,k+1}) &= \log \mu_{n,k+1} + \log d_n \\ &= \log(1 - \epsilon_n) \left(\frac{d_n^{k+1} - 1}{d_n - 1} \right) + O\left(\frac{d_n^2 d_n^k - 1}{n d_n - 1}\right) + \log d_n \\ &= (-\epsilon_n + O(\epsilon_n^2)) d_n^k (1 + O(1/d_n)) + \log d_n + O\left(\frac{d_n^{k+1}}{n}\right), \end{aligned}$$

where the second equality is from (A.8) and last equality is by using Taylor's expansion for $\log(1 - \epsilon_n)$. Using $d_n^{k+1} \ll n$ and $|\epsilon_n^2 d_n^k| \leq \frac{2(\xi_n^2 + (\log d_n)^2)}{d_n^k} \rightarrow 0$, we now have

$$\begin{aligned} \log(d_n \mu_{n,k+1}) &= (-\epsilon_n d_n^k + o(1))(1 + O(1/d_n)) + \log d_n + o(1) \\ &= (-\xi_n - \log d_n)(1 + O(1/d_n)) + \log d_n + o(1) \\ &= -\xi_n - \log d_n + \log d_n + O\left(\frac{\xi_n + \log d_n}{d_n}\right) + o(1) \\ &= -\xi_n + o(1) \rightarrow \log(\alpha), \end{aligned}$$

where the last equality once more uses the observation that $\frac{|\xi_n|}{d_n} \rightarrow 0$. Thus we have $d_n \mu_{n,k+1} \rightarrow \alpha$ as $n \rightarrow \infty$ which completes the proof. $\square$

A.6. Proof of Lemma 5.6. PROOF. Since $\mu_{n,k} \rightarrow 1$ and $j \mapsto \mu_{n,j}$ is nonincreasing, we have $\mu_{n,i} \rightarrow 1$ for each $i \leq k$. Additionally, since $\lambda_n \rightarrow 1$, Corollary 5.3 shows that for any $i \in [k]$ $\lim_{n \rightarrow \infty} \frac{\dot{\beta}_n(\mu_{n,i})}{d_n \mu_{n,i+1}} = 1$. As a consequence, $\dot{\beta}_n(\mu_{n,k-1}) \rightarrow \infty$ as $n \rightarrow \infty$, and for any $j \in [k-2]$

$$\lim_{n \rightarrow \infty} \frac{\dot{\beta}_n(\mu_{n,j})}{\dot{\beta}_n(\mu_{n,j+1})} = \lim_{n \rightarrow \infty} \frac{d_n \mu_{n,j+1}}{d_n \mu_{n,j+2}} = \lim_{n \rightarrow \infty} \frac{\mu_{n,j+1}}{\mu_{n,j+2}} = 1.$$

This completes the proof of the lemma. $\square$

A.7. Proof of Lemma 5.7. PROOF. By the first part of Lemma 5.1, (5.7) is immediate from (5.6). Now consider (5.6). Taking logarithms in (2.7), for $x > d_n/n$,

$$\begin{aligned} \log \beta_n(x) &= \sum_{i=0}^{d_n-1} \left(\log \left(x - \frac{i}{n} \right) - \log \left(1 - \frac{i}{n} \right) \right) \\ &= \sum_{i=0}^{d_n-1} \left(\log \left(1 - \frac{i}{n} - (1-x) \right) - \log \left(1 - \frac{i}{n} \right) \right). \end{aligned}$$

Let $\delta_n = \epsilon_n + \frac{d_n}{n}$. For large n , $\delta_n \leq \frac{1}{2}$, and hence, using the expansion $\log(1-h) = -h + O(h^2)$ for $|h| \leq \frac{1}{2}$, for any $x \in [1 - \epsilon_n, 1]$:

$$\begin{aligned} \log \beta_n(x) &= \sum_{i=0}^{d_n-1} \left\{ -\frac{i}{n} - (1-x) + \frac{i}{n} + O(\delta_n^2) \right\} = -d_n(1-x) + O(d_n \delta_n^2) \\ &= d_n \log(1 - (1-x)) + O(d_n \delta_n^2) = \log \gamma_n(x) + O(d_n \delta_n^2). \end{aligned}$$

Note that $\delta_n^2 = (\epsilon_n + d_n/n)^2 \leq 2(\epsilon_n^2 + \frac{d_n^2}{n^2})$. Hence by our assumptions $d_n \delta_n^2 \rightarrow 0$. This proves (5.6) and completes the proof of the lemma. $\square$

A.8. Proof of Lemma 5.8. PROOF. By (5.2)

$$\sup_{x \in [0, 1-\epsilon_n]} |\beta_n(x)| \leq (1 - \epsilon_n)^{d_n} = e^{-d_n \epsilon_n + o(1)} \rightarrow 0.$$

Similarly, by (5.3), under the assumption $\limsup_n \frac{d_n}{n} < 1$, for large n ,

$$\sup_{x \in [0, 1-\epsilon_n]} |\dot{\beta}_n(x)| \leq (1 - d_n/n)^{-1} d_n (1 - \epsilon_n)^{d_n-1} = e^{-d_n \epsilon_n + \log d_n + O(1)} \rightarrow 0. \quad \square$$

APPENDIX B: PROOF OF LEMMA 6.8

For a right continuous bounded variation function $F : [0, T] \rightarrow \mathbb{R}$, let dF denote the signed measure on $(0, T]$ given by $dF(a, b] = F(b) - F(a)$ for $0 \leq a < b \leq T$, and $d\lambda$ denote the Lebesgue measure on $(0, T]$. Bounded measurable functions $h : [0, T] \rightarrow \mathbb{R}$ act on signed measure $d\mu$ on $(0, T]$ on the left as follows: $h d\mu$ denotes the signed measure $A \mapsto \int_A h(x) d\mu(x)$, $A \in \mathcal{B}(0, T]$.

Let $H(t) \doteq \int_0^t h(s) d\lambda(s)$ for $t \in [0, T]$. Note that z defined in (6.14) is a right continuous function with bounded variations. The corresponding measure dz on $(0, T]$ satisfies the identity

$$dz = -hz d\lambda + g d\lambda + dM,$$

namely

$$dz + hz d\lambda = g d\lambda + dM.$$

Acting on the left in the above identity by the bounded continuous function $e^H(t) \doteq e^{H(t)}$ we get

$$e^H dz + e^H hz d\lambda = e^H g d\lambda + e^H dM.$$

Since $dH = h d\lambda$, by the change of variable formula (cf. [28], Theorem VI.8.3) $de^H = he^H d\lambda$. Hence

$$e^H dz + z de^H = e^H g d\lambda + e^H dM.$$

Two applications of the integration by parts formula (cf. [2], Theorem 18.4) show that

$$d(e^H z) = e^H g d\lambda + d(e^H M) - M de^H.$$

Computing the total measure on $(0, t]$ for $t \leq T$:

$$e^{H(t)} z(t) - z(0) = \int_0^t e^{H(s)} g(s) d\lambda(s) + e^{H(t)} M(t) - M(0) - \int_0^t M(s) de^H(s).$$

Rearranging terms and multiplying by $e^{-H(t)}$ on both sides and noting, from (6.14), that $M(0) = 0$:

$$(B.1) \quad z(t) = \int_0^t e^{H(s)-H(t)} g(s) d\lambda(s) + M(t) - e^{-H(t)} \int_0^t M(s) de^H(s) + e^{-H(t)} z(0).$$

We now estimate the various terms on the right-hand side of (B.1). The first term on the right-hand side of (B.1) satisfies for $t \in [0, T \wedge \tau]$

$$\begin{aligned} \left| \int_0^t e^{H(s)-H(t)} g(s) d\lambda(s) \right| &\leq |g|_{*, T \wedge \tau} \int_0^t e^{H(s)-H(t)} d\lambda(s) \\ &= |g|_{*, T \wedge \tau} \int_0^t e^{-\int_s^t h(u) du} d\lambda(s) \\ (B.2) \quad &\leq |g|_{*, T \wedge \tau} \int_0^t e^{-m(t-s)} d\lambda(s) \\ &= |g|_{*, T \wedge \tau} \frac{1 - e^{-tm}}{m} \leq \frac{|g|_{*, T \wedge \tau}}{m}. \end{aligned}$$

Next we estimate the third term in the right-hand side of (B.1). Since h is nonnegative on $[0, T \wedge \tau]$, de^H in a positive measure on $(0, T \wedge \tau]$. Hence for $t \in [0, T \wedge \tau]$

$$(B.3) \quad \left| e^{-H(t)} \int_0^t M(s) de^H(s) \right| \leq |M|_{*, T \wedge \tau} e^{-H(t)} \int_0^t de^H(s) \leq |M|_{*, T \wedge \tau}.$$

Finally, the last term in the right-hand side of (B.1) for any $t \in [0, \tau \wedge T]$ can be bounded as

$$(B.4) \quad |e^{-H(t)} z(0)| \leq |z(0)| e^{-mt}.$$

Using (B.2), (B.3) and (B.4) in (B.1) completes the proof of the lemma. $\square$

Acknowledgments. We thank the two anonymous referees for a very careful review of the manuscript that led to numerous improvements in the work.

Funding. Research of SB is supported in part by NSF Grants DMS-1613072, DMS-1606839, DMS-2113662 and ARO Grant W911NF-17-1-0010.

Research of AB is supported in part by the National Science Foundation (DMS-1814894 and DMS-1853968).

Research of MD is supported by the NSF Grant DMS-1613072 and NIH R01 grant HG009125-01.

AB is grateful for the support from Nelder Fellowship from Imperial College, London, where part of this research was completed.

REFERENCES

- [1] ALTMAN, E., AYESTA, U. and PRABHU, B. J. (2011). Load balancing in processor sharing systems. *Telecommun. Syst.* **47** 35–48.
- [2] BILLINGSLEY, P. (1995). *Probability and Measure. Wiley Series in Probability and Mathematical Statistics.* Wiley, New York.
- [3] BRAMSON, M., LU, Y. and PRABHAKAR, B. (2012). Asymptotic independence of queues under randomized load balancing. *Queueing Syst.* **71** 247–292. [MR2943660](#) <https://doi.org/10.1007/s11134-012-9311-0>
- [4] BRAVERMAN, A. (2020). Steady-state analysis of the join-the-shortest-queue model in the Halfin-Whitt regime. *Math. Oper. Res.* **45** 1069–1103. [MR4135843](#) <https://doi.org/10.1287/moor.2019.1023>
- [5] BRIGHTWELL, G., FAIRTHORNE, M. and LUCZAK, M. J. (2018). The supermarket model with bounded queue lengths in equilibrium. *J. Stat. Phys.* **173** 1149–1194. [MR3876921](#) <https://doi.org/10.1007/s10955-018-2044-7>
- [6] BRIGHTWELL, G. and LUCZAK, M. (2012). The supermarket model with arrival rate tending to one. Preprint. Available at [arXiv:1201.5523](#).
- [7] BUDHIRAJA, A. and FRIEDLANDER, E. (2019). Diffusion approximations for load balancing mechanisms in cloud storage systems. *Adv. in Appl. Probab.* **51** 41–86. [MR3984010](#) <https://doi.org/10.1017/apr.2019.3>
- [8] CARDELLINI, V., CASALICCHIO, E., COLAJANNI, M. and YU, P. S. (2002). The state of the art in locally distributed web-server systems. *ACM Comput. Surv.* **34** 263–311.
- [9] ESCHENFELDT, P. and GAMARNIK, D. (2018). Join the shortest queue with many servers. The heavy-traffic asymptotics. *Math. Oper. Res.* **43** 867–886. [MR3846076](#) <https://doi.org/10.1287/moor.2017.0887>
- [10] ESCHENFELDT, P. and GAMARNIK, D. (2016). Supermarket queueing system in the heavy traffic regime. Short queue dynamics. Preprint. Available at [arXiv:1610.03522](#).
- [11] ETHIER, S. N. and KURTZ, T. G. (2009). *Markov Processes: Characterization and Convergence* **282**. Wiley, New York.
- [12] GAMARNIK, D., TSITSIKLIS, J. N. and ZUBELDIA, M. (2016). Delay, memory, and messaging tradeoffs in distributed service systems. *ACM SIGMETRICS Perform. Eval. Rev.* **44** 1–12.
- [13] GAMARNIK, D., TSITSIKLIS, J. N. and ZUBELDIA, M. (2018). Delay, memory, and messaging tradeoffs in distributed service systems. *Stoch. Syst.* **8** 45–74. [MR3775991](#) <https://doi.org/10.1287/stsy.2017.0008>
- [14] GRAHAM, C. (2000). Chaoticity on path space for a queueing network with selection of the shortest queue among several. *J. Appl. Probab.* **37** 198–211. [MR1761670](#) <https://doi.org/10.1017/s0021900200015345>
- [15] GUPTA, V., BALTER, M. H., SIGMAN, K. and WHITT, W. (2007). Analysis of join-the-shortest-queue routing for web server farms. *Perform. Eval.* **64** 1062–1081.
- [16] HUNT, P. J. and KURTZ, T. G. (1994). Large loss networks. *Stochastic Process. Appl.* **53** 363–378. [MR1302919](#) [https://doi.org/10.1016/0304-4149\(94\)90071-X](https://doi.org/10.1016/0304-4149(94)90071-X)
- [17] KALLENBERG, O. (2021). *Foundations of Modern Probability. Probability Theory and Stochastic Modelling* **99**. Springer, Cham. [MR4226142](#) <https://doi.org/10.1007/978-3-030-61871-1>
- [18] KARATZAS, I. and SHREVE, S. E. (1998). *Brownian Motion and Stochastic Calculus*. Springer, New York.
- [19] KRUK, L., LEHOCZKY, J., RAMANAN, K. and SHREVE, S. (2007). An explicit formula for the Skorokhod map on $[0, a]$. *Ann. Probab.* **35** 1740–1768. [MR2349573](#) <https://doi.org/10.1214/009117906000000890>
- [20] KURTZ, T. G. (1977/78). Strong approximation theorems for density dependent Markov chains. *Stochastic Process. Appl.* **6** 223–240. [MR0464414](#) [https://doi.org/10.1016/0304-4149\(78\)90020-0](https://doi.org/10.1016/0304-4149(78)90020-0)
- [21] LIU, X. and YING, L. (2019). A simple steady-state analysis of load balancing algorithms in the sub-Halfin-Whitt regime. *ACM SIGMETRICS Perform. Eval. Rev.* **46** 15–17.

- [22] LIU, X. and YING, L. (2020). Steady-state analysis of load-balancing algorithms in the sub-Halfin-Whitt regime. *J. Appl. Probab.* **57** 578–596. [MR4125466](#) <https://doi.org/10.1017/jpr.2020.13>
- [23] LUCZAK, M. J. and MCDIARMID, C. (2006). On the maximum queue length in the supermarket model. *Ann. Probab.* **34** 493–527. [MR2223949](#) <https://doi.org/10.1214/009117905000000710>
- [24] LUCZAK, M. J. and MCDIARMID, C. (2007). Asymptotic distributions and chaos for the supermarket model. *Electron. J. Probab.* **12** 75–99. [MR2280259](#) <https://doi.org/10.1214/EJP.v12-391>
- [25] LUCZAK, M. J. and NORRIS, J. (2005). Strong approximation for the supermarket model. *Ann. Appl. Probab.* **15** 2038–2061. [MR2152252](#) <https://doi.org/10.1214/105051605000000368>
- [26] MAGULURI, S. T., SRIKANT, R. and YING, L. (2012). Stochastic models of load balancing and scheduling in cloud computing clusters. In 2012 *Proceedings IEEE Infocom* 702–710. IEEE, Orlando, FL.
- [27] MARTIN, J. B. and SUHOV, YU. M. (1999). Fast Jackson networks. *Ann. Appl. Probab.* **9** 854–870. [MR1722285](#) <https://doi.org/10.1214/aoap/1029962816>
- [28] MCSHANE, E. J. and BOTTS, T. A. (2013). *Real Analysis*. Van Nostrand, Princeton, NJ.
- [29] MITZENMACHER, M. (2001). The power of two choices in randomized load balancing. *IEEE Trans. Parallel Distrib. Syst.* **12** 1094–1104.
- [30] MITZENMACHER, M., RICH, A. W. and SITARAMAN, R. (2001). The power of two random choices: A survey of techniques and results. In *Handbook of Randomized Computing, Vol. I, II. Comb. Optim.* **9** 255–312. Kluwer Academic, Dordrecht. [MR1966907](#) https://doi.org/10.1007/978-1-4615-0013-1_9
- [31] MUKHERJEE, D., BORST, S. C., VAN LEEUWAARDEN, J. S. H. and WHITING, P. A. (2018). Universality of power-of- d load balancing in many-server systems. *Stoch. Syst.* **8** 265–292. [MR3899726](#) <https://doi.org/10.1287/stsy.2018.0016>
- [32] MUKHOPADHYAY, A. and MAZUMDAR, R. R. (2016). Analysis of randomized join-the-shortest-queue (JSQ) schemes in large heterogeneous processor-sharing systems. *IEEE Trans. Control Netw. Syst.* **3** 116–126. [MR3514587](#) <https://doi.org/10.1109/TCNS.2015.2428331>
- [33] ONGARO, D., RUMBLE, S. M., STUTSMAN, R., OUSTERHOUT, J. and ROSENBLUM, M. (2011). Fast crash recovery in RAMCloud. In *Proceedings of the Twenty-Third ACM Symposium on Operating Systems Principles* 29–41.
- [34] PROTTER, P. (2013). *Stochastic Integration and Differential Equations* **21**. Springer, Berlin.
- [35] VAN DER BOOR, M., BORST, S. C., VAN LEEUWAARDEN, J. S. and MUKHERJEE, D. (2018). Scalable load balancing in networked systems: A survey of recent advances. Preprint. Available at [arXiv:1806.05444](#).
- [36] VVEDENSKAYA, N. D., DOBRUSHIN, R. L. and KARPELEVICH, F. I. (1996). A queueing system with a choice of the shorter of two queues—an asymptotic approach. *Problemy Peredachi Informatsii* **32** 20–34. [MR1384927](#)
- [37] YOSIDA, K. (2012). *Functional Analysis*. Springer, Berlin.