

Power-of- d Choices Load Balancing in the Sub-Halfin-Whitt Regime

Sushil Mahavir Varma

Industrial and Systems Engineering, Georgia Institute of Technology, sushil@gatech.edu,
<https://sites.google.com/view/sushil-varma/home>

Francisco Castro

Anderson School of Management, University of California, Los Angeles, francisco.castro@anderson.ucla.edu,
<https://fcocastro.github.io>

Siva Theja Maguluri

Industrial and Systems Engineering, Georgia Institute of Technology, siva.theja@gatech.edu,
<https://sites.google.com/site/sivatheja/>

We consider the load balancing system under Poisson arrivals, exponential services, and homogeneous servers. Upon arrival, a job is to be routed to one of the servers, where it is queued until service. We consider the Power-of- d choices routing algorithm, which chooses the queue with minimum length among d randomly sampled queues. We study this system in the many-server heavy-traffic regime where the number of servers goes to infinity simultaneously when the load approaches the capacity. In particular, we consider a sequence of systems with n servers, where the arrival rate of the n^{th} system is $\lambda = n - n^{1-\gamma}$ for some $\gamma \in (0, 0.5)$, known as the sub-Halfin-Whitt regime. It was shown by [Liu Ying (2020)] that under Power-of- d choices routing with $d \geq n^\gamma \log n$, the queue length behaves similarly to that of JSQ and that there are asymptotically zero queueing delays.

The focus of this paper is to characterize the behavior when d is below this threshold. We obtain high probability bounds on the queue lengths for various values of d and large enough n . In particular, we show that when d grows polynomially in n but slower than in [Liu Ying (2020)], i.e., if d is $\Theta\left((n^\gamma \log n)^{1/m}\right)$ for some integer $m > 1$, then the asymptotic queue length is m with high probability. This finite queue length behavior is similar to JSQ in the so-called nondegenerate slowdown regime (where $\gamma = 1$). Moreover, if d grows polylog in n , i.e., slower than any polynomial, but is at least $\Omega(\log(n)^3)$, the queue length blows up to infinity asymptotically. Such behavior is similar to that under JSQ in the so-called super slowdown regime ($\gamma > 1$). We obtain these results by using an iterative state space collapse approach. We first establish a weak state-space collapse (SSC) on the queue lengths. Then, we bootstrap on weak SSC to iteratively narrow down the region of the collapse. After enough steps, this inductive refinement provides the bounds we seek. We establish these sequences of collapse using Lyapunov drift arguments.

Key words: Load Balancing, Sub-Halfin-Whitt, Many Server Heavy Traffic, Iterative State Space Collapse

1. Introduction

We study a load-balancing queuing system in which a single stream of jobs arrives governed by a Poisson process and is routed to one of the n homogeneous servers, operating with a service rate equal to one. Each server is endowed with a queue of maximum buffer size b .

The job dispatcher uses a load balancing or routing algorithm to route arriving jobs to the queues. The literature considers many possible routing algorithms ranging from random routing to Joining the Shortest Queue (JSQ). In random routing, a new job joins a queue selected uniformly at random. On the other hand, a new job joins the shortest queue under JSQ. While random routing has no informational requirements—the dispatcher does not need to know any information about the system primitives and state—it does not provide optimal delay performance. In contrast, JSQ

has more informational requirements—the dispatcher needs to know the system state to determine the shortest queue—but it has a proven near-optimal delay performance, e.g., see [13]. In this paper, we consider an in-between policy known as Power-of- d choices, in which once a job arrives, d queues are sampled uniformly at random from the n queues. Then, the job joins the shortest among the d sampled ones. Note that $d = 1$ is the same as random routing, and $d = n$ is the same as JSQ.

For a tractable analysis of the performance of the routing algorithm, the literature considers different asymptotic regimes, where the number of servers goes to infinity, the load on the system approaches its capacity, or both happen simultaneously. As we explain below in Section 1.1, the performance of JSQ has been studied extensively under these regimes and combinations thereof. On the other hand, the performance analysis of Power-of- d choices is comparatively limited. We contribute towards this deficiency by analyzing Power-of- d under the sub-Halfin-Whitt asymptotic regime (see Section 1.1). In this regime, the arrival rate of jobs increases with the number of servers at a rate of $\lambda = n - n^{1-\gamma}$ with $\gamma \in (0, 0.5)$. Under this scaling, our goal is to characterize the system’s asymptotic delay and steady-state behavior for growing choices, i.e. $d \rightarrow \infty$ as $n \rightarrow \infty$.

It is known by [22] that if d is sufficiently large ($d \geq n^\gamma \log n$), Power-of- d behaves like JSQ, and the jobs experience zero asymptotic delays in steady-state. In particular, [22] shows that the asymptotic queue lengths at each server are either zero or one. However, for smaller values of d , one expects the delay to be higher. In particular, we will later show that the queue lengths can be finite but greater than one or even asymptotically infinite depending on how d scales with n . Thus the asymptotic queue lengths are qualitatively different from JSQ, i.e., they are not just zero-one but exhibit a rich steady-state distribution. Characterizing such behavior under all the scenarios warrants a new approach. In this paper, we aim to provide a unified framework for almost all scalings of d . Before presenting our main contributions, we briefly outline prior work on various asymptotic regimes

1.1. Many-Server-Heavy-Traffic Regimes

In general, it is challenging to determine the exact delay under a routing policy. So, it has been studied in various asymptotic regimes to gain insights into the optimality of routing policies. We now provide a comprehensive overview of these asymptotic regimes and the insights obtained in the literature for different routing policies. In turn, we explain how our results fit into the literature.

Mean field. In this regime, the number of servers increases to infinity while maintaining a constant load on each server. It has been shown in the literature [26, 25, 28, 30] that under the Power-of- d choices algorithm, even for $d = 2$, the steady-state queue lengths exhibits a double exponential tail as opposed to an exponential tail for random routing. On the other hand, under JSQ, it was shown [27] that almost all the queues have length zero or one. Thus, all the jobs experience asymptotically zero delay. In addition, it was shown that the same behavior holds true for Power-of- d with growing choices, i.e. $d \rightarrow \infty$, as $n \rightarrow \infty$.

Classical heavy-traffic . Another popular regime considered in the literature is the classical heavy-traffic regime. In this regime, the load converges to the capacity while the number of servers is constant. In particular, let $\epsilon \triangleq 1 - \lambda/n$ be such that $1 - \epsilon$ quantifies the load on the system. Then, $\epsilon \downarrow 0$ is the heavy-traffic regime. This regime allows one to analyze the bottlenecks in the system. Under any routing algorithm, the queue lengths in this regime increase to infinity asymptotically. Under JSQ [13, 10, 18], an appropriately scaled queue length converges to an exponential distribution with a mean depending on the variance of the arrivals and services. In addition, the limiting behavior of Power-of- d for all $d \geq 2$ is identical to that of JSQ [24], while that of random routing is worse by a factor of n .

Many-server-heavy-traffic. One can also consider a hybrid of the mean field and the classical heavy-traffic regime, i.e., many-server-heavy-traffic regime, wherein the load increases to capacity

simultaneously while the number of servers increases to infinity. Depending on the relative rate at which the load and the number of servers converge to their asymptotes, one can obtain different viewpoints on the performance of the routing algorithms. In particular, the mean-field and classical heavy traffic are two extreme ways to scale the system and provide different perspectives. For instance, there is a distinction between the performance of JSQ and Power-of- d for small d in the mean-field as opposed to an identical limiting behavior in classical heavy traffic. Considering many-server-heavy-traffic regimes provides us with a more comprehensive understanding of the performance of various load balancing algorithms, allows us to differentiate between their performance, and enables us to pick the right d in Power-of- d type algorithms. Studying such regimes was first initiated by Halfin-Whitt [15] in an M/M/ n queue.

More precisely, the parameterization of the arrival rate as $\lambda = n - \beta n^{1-\gamma}$ for some $\gamma \in (0, \infty)$ is defined as the many-server-heavy-traffic regimes. The parameter γ determines the relative rate at which λ and n converge to their asymptotes. As γ increases from 0 to ∞ , the load on each server is more prominent, resulting in higher delays. Note that the mean-field regime is a special case with $\gamma = 0$ and $\beta < 1$, and the classical heavy traffic is interpreted as $\gamma \rightarrow \infty$.

Now, we discuss the performance of JSQ for $\gamma \in (0, \infty)$ as summarized in Fig. 1. For the sub-Halfin-Whitt regime, i.e., $\gamma \in (0, 0.5)$, similar to the mean-field regime, the delay experienced by the jobs is asymptotically zero. A phase transition occurs in the Halfin-Whitt regime, i.e., $\gamma = 0.5$. In this regime [12, 6, 1, 2], a vanishing fraction of jobs experiences a constant delay bounded away from zero. Similar results were proved by [23, 31] for the super-Halfin-Whitt regime, i.e., $\gamma \in (0.5, 1)$. Another phase transition occurs at the nondegenerate slowdown (NDS) regime, i.e., $\gamma = 1$. In this case [14], incoming jobs experience a non-zero, finite delay. When γ increases beyond one, it is called the super slowdown regime, and the limiting queue length at each server increases to infinity. The authors of [18] analyze the limiting stationary distribution of appropriately scaled queue lengths for $\gamma \geq 2$ and show that its behavior is similar to the classical heavy-traffic regime. The case $\gamma \in (1, 2)$ is still an open problem. Figure 1 (left) summarizes this discussion by illustrating the delay performance of the JSQ policy under different asymptotic regimes.

As the load in the system increases, one expects the delay under any routing algorithm to be higher. Consistent with the intuition, increasing delay with γ is observed under JSQ, as previously discussed. On the other hand, one can fix a γ and consider the delay performance as d is varied in the Power-of- d choices routing. Similar to how JSQ exhibits higher delay for more loaded regimes, one would expect the delay to increase as d reduces. In this paper, we quantify such behavior exhibited by Power-of- d for all $\gamma \in (0, 0.5)$. We now present our main contributions.

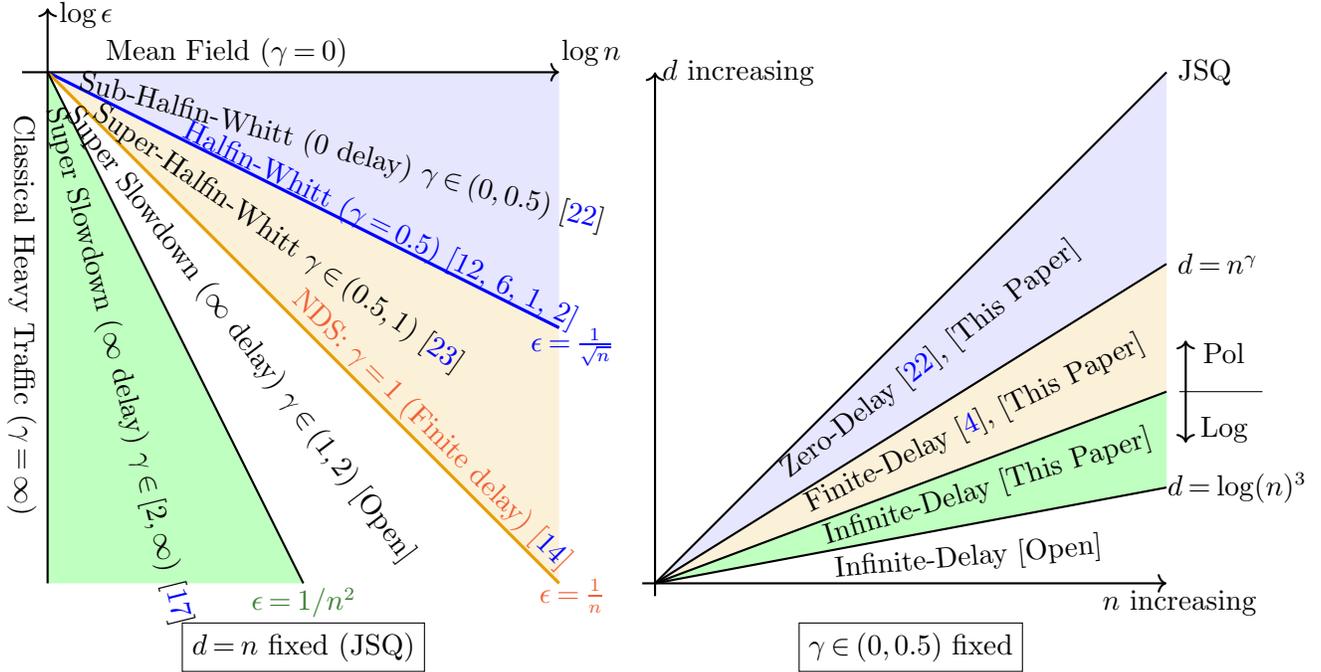
1.2. Main Contributions

Our focus is on understanding the performance of the Power-of- d for different choices of d . Note that Figure 1 (left) provides the performance of JSQ, and augmenting it with Power-of- d would correspond to adding a third dimension for d as a function of n . The special case of JSQ as depicted in Figure 1 (left) corresponds to one slice of the three dimensional figure with $d = n$.

In this paper, we restrict ourselves to the sub-Halfin-Whitt regime, i.e. $\gamma \in (0, 0.5)$, and consider a broad range of values of d . It was shown in [22] that Power-of- d with $d \geq n^\gamma \log n$ has an identical limiting behavior as JSQ. We go beyond this range and provide a quantitative distinction between JSQ and Power-of- d by characterizing the performance of the Power-of- d for $d < n^\gamma$. A summary of the results in this paper is given in Figure 1 (right).

Finite Delay: First, we consider the case when $d = (n^\gamma \log n)^{1/m}$ for some positive integer m . We show that the queue lengths exhibits the following behavior with high probability: most of the queues are of length m and a vanishing fraction are either longer or shorter. In particular, we show that the fraction of queues with length less than i is equal to $n^{-\gamma} d^{i-1} (1 + o(1))$ for $i \leq m$ and the fraction of queues with length more than m is at most $o(n^{-\gamma} d^{m-1})$ which is $o(1)$. It is worth

Figure 1 Performance of JSQ ($d = n$) under many-server-heavy-traffic regimes ($\gamma \in [0, \infty]$), where $\epsilon = n^{-\gamma}$ (left) and performance of Power-of- d for different choices of d under the sub-Halfin-Whitt regime, i.e. $\gamma \in (0, 0.5)$ (right).



noting that these results are applicable for the pre-limit system as well, i.e. for all finite, large enough n (and we provide explicit expressions for all the $o(\cdot)$ terms). These results imply that when $m \geq 2$, the queue lengths are non-zero but finite, behaving qualitatively similar to that of JSQ in NDS regime. However, a fundamental difference in behavior is that while our results show that the queue lengths are essentially concentrated around m for Power-of- d in sub-Halfin-Whitt regime, the limiting queue lengths of JSQ in NDS are spread over multiple values and the distribution has a nontrivial support. Also note that, when we pick $m = 1$, our result implies that the jobs experience zero asymptotic delay and the queue lengths are either zero or one. The result in this special case was first established in [22].

Infinite Asymptotic Delay: Now, we consider the case when d is Poly-Log(n) but is at least $\Omega(\log(n)^3)$. Note that, Poly-Log(n) is smaller than $d = (n^\gamma \log n)^{1/m}$ for any $m \in \mathbb{Z}_+$. We show that all the queue lengths are $\Theta(\log n / \log d)$ with high probability. This implies that the asymptotic queue lengths are infinite. Similar to the finite delay case, we characterize the fraction of queue lengths smaller or larger than m for the pre-limit system. Note that, such a behavior is qualitatively similar to that of JSQ in the super slowdown regime. However, there is again a fundamental difference in behavior because while we show that the queue lengths concentrate around $\Theta(\log n / \log d)$ for Power-of- d , JSQ in the super slowdown regime has a large support. Extending the result to the case when $d < \log(n)^3$ is an open future research direction.

Methodological Contribution: In contrast to the prior work on load balancing that is based on fluid and diffusion limits (e.g. see: [12, 1]), Stein's method (e.g. see: [22, 30]), transform method (e.g. see: [18]), and a combination of iterative SSC and Stein's method [21, 20], our approach uses iterative SSC alone without the use of Stein's method. We first obtain a crude bound on the possible values of the queue lengths, i.e., a weak state space collapse. We then iteratively bootstrap from this weak SSC to obtain more and more refined SSC. This iterative refinement is inductively repeated (m times) until a tight characterization of the steady-state queue lengths as described above is obtained. Lyapunov drift-based arguments achieve each step of the refinement.

Table 1 Power-of- d choices for $\gamma \in (0, 0.5)$

Value of d	Regime	Queue Length	References
$d \geq n^\gamma \log n$	Zero-Delay	≈ 1	This paper and [22]
Polynomial ($d = (n^\gamma \log n)^{1/m}$)	Finite-Delay	$\approx m = \Theta\left(\frac{\log n}{\log d}\right)$	This paper and [7]
Poly-log and $d \geq \log(n)^3$	Infinite-Delay	$\Theta\left(\frac{\log n}{\log d}\right)$	This paper
$d \leq \log(n)^3$	Infinite-Delay	$\Theta\left(\frac{\log n}{\log d}\right)$	Open

Iterative SSC was used as an intermediate step in characterizing the limiting distribution of queue lengths in [21, 20] to study the case of $m = 1$ with Coxian service times. Using their SSC methodology directly in our setting does not suffice as it results in only a crude bound (see Section 5.1 for detailed discussion), so further refinement is required to obtain tight queue length bounds. The novelty of our approach lies in independently constructing a sequence of Lyapunov functions which allows us to obtain tight queue length bounds simply by applying iterative SSC enough times. The main takeaway of our methodology is that iterative SSC is a powerful tool to analyze queueing systems in mean field types of regimes, i.e., the queueing system concentrates around the fixed point of the corresponding deterministic, dynamical system.

1.3. Related Work

The prior work that is closest to ours are [4, 22, 7]. In [4], the analysis for Power-of- d was carried out for the ‘finite delay’ regime, i.e. $d = (n^\gamma \log n)^{1/m}$ for $m \in \mathbb{Z}_+$, and a process level law of large numbers is established to show convergence of the queue length process to its mean. By observing the mean, it was noted that most of the queue lengths are m in the limit. However, the lower order terms, i.e. the fraction of queues with length larger or smaller than m is not characterized in this result. Although [4] also provides a diffusion process that characterizes further fluctuations around the mean, the steady-state distribution of the diffusion process is not characterized. In addition, to conclude that these results holds for the steady-state of the pre-limit process, interchange of limits is required which is not established. If these two steps were completed, then the approach in [4] would obtain the lower order terms. In contrast, by directly working with the steady-state quantities (as opposed to process level convergence), we characterize the dominant lower order term and show that it is exactly $n^{-\gamma} d^{i-1}$ thereby obtaining a sharper characterization of the steady-state queue lengths distribution. Moreover, we also obtain bounds on the fluctuations around these lower-order terms.

In addition, [22] is also closely related to our result which shows that the asymptotic delay experienced by the jobs is zero under Power-of- d for $d \geq n^\gamma \log n$. A key difference is that [22] characterizes all the moments of the total number of jobs in the system, whereas we present high probability tail bounds on the queue length distribution. Another difference is in the methodology. The proof technique used in [22] is SSC combined with Stein’s method, whereas we use iterative SSC without the use of Stein’s method. In particular, we carefully construct a sequence of Lyapunov functions that results in tight queue length bounds as opposed to obtaining crude bounds as an intermediary step in [22].

The main result of [7] and its previous (arXiv) version [8] are closely related to our result. In addition to a few technical differences, the papers differ on the proof methodology. In particular, [7] considers the case of $m \notin \mathbb{Z}_+$, whereas we consider $m \in \mathbb{Z}_+$ along with a logarithmic function for our choice of d . Also, while [7] considers a broader range of values of γ , we consider a broader range of values of d , that is, we allow for d to be smaller than any polynomial (poly-log). Note that [8] considers the same range of γ and d as ours but their result requires $n \geq 10^{15}$, so it is essentially an asymptotic characterization. On the other hand, our result provides tight bounds on the queue lengths that are applicable in the pre-limit as well. We also verify that our suggested

bounds are tight by conducting simulations on the pre-limit system. Lastly, our proof methodology is more simplified and can be thought of as a steady-state version of [7]. In particular, [7] analyzes the drift of a sequence of Lyapunov functions in a finite time. As the analysis is carried out in finite time, [7] have to ensure that the Lyapunov function stays small while the other Lyapunov functions in the sequence decrease. Such a complication is circumvented in our methodology as we take a steady-state approach. Also, our sequence of Lyapunov functions is completely different from that of [7] providing an alternative proof and revealing geometric insights. In particular, while [7] provides an elegant algebraic construction of the sequence of Lyapunov functions, our sequence is based on geometric intuition via the trajectory of the ODE approximation.

Now, we present a non-exhaustive overview of the literature on load balancing under the sub-Halfin-Whitt regime. JSQ and Power-of- d for $d \geq n^\gamma \log n$ was analyzed in [22]. This result was extended for more general settings in the literature: coxian-2 service distribution in [21], coxian- k service distribution in [20], and parallel jobs arriving in the system in [29]. In addition, Power-of- d choices has been analyzed in [30, 11] for $d = 2$, and [4] provides a transient analysis for growing sequences of d . The reader can refer to the survey paper [9] for a holistic review of the literature.

Lastly, we would like to point out that our results have a similar qualitative flavor as in [19]. In particular, for a given $m \in \mathbb{Z}_+$, [19] characterizes the critical load below which the blocking probability is very small for a load balancing system with finite buffer equal to m ; under a state-dependent random routing policy.

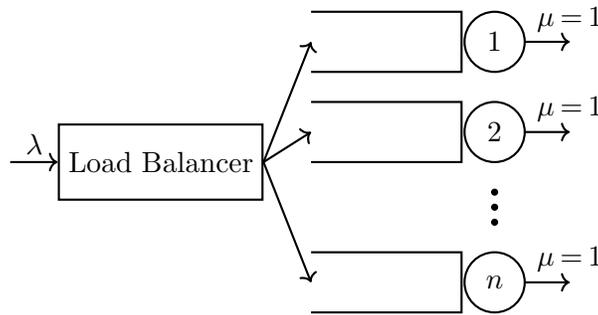
1.4. Notation

The set of all positive integers (excluding zero) is denoted by \mathbb{Z}_+ . For some $k \in \mathbb{Z}_+$, the set of numbers $\{1, 2, \dots, k\}$ is denoted by $[k]$. We use the shorthand *w.h.p.* to denote “with high probability”.

2. Model

Consider a load balancing system with n homogeneous servers. A single stream of jobs arrive, governed by a Poisson process with rate $\lambda < n$. Upon arrival, the job is routed to one of the servers, where it waits in a queue before getting served. Each queue has a maximum buffer size $b \in \mathbb{Z}_+$. Preemption is not allowed and the job cannot move within queues. The service times for all servers are i.i.d. exponential random variables with rate $\mu = 1$. An illustration of the model is given in Fig. 2. A natural state descriptor for the system is the number of jobs in each queue. However, it

Figure 2 A homogeneous load balancing model with n -servers.



is mathematically more convenient to consider $\mathbf{s} \in (\mathbb{Z}_+ \cup \{0\})^b$ as the state descriptor. Here, s_i is the number of queues with length at least i , and b is the maximum buffer size. The state space is given by

$$\mathcal{S} := \{ \mathbf{s} \in (\mathbb{Z}_+ \cup \{0\})^b : s_{i_1} \leq s_{i_2} \leq n \quad \forall i_1 \geq i_2 \in [b] \}.$$

In addition, we also denote the number of queues with at least 0 jobs by $s_0 = n$. Once a job arrives, d queues are sampled uniformly at random, with replacement from n queues. Then, the job is routed to the smallest among the d sampled queues. This algorithm is known as Power-of- d choices in the literature. Under this routing scheme, the process $\{\mathbf{s}(t) : t \geq 0\}$ is a finite state-space, irreducible, continuous time Markov chain. Thus, the CTMC $\{\mathbf{s}(t) : t \geq 0\}$ is positive recurrent and exhibits a unique stationary distribution. Denote by $\bar{\mathbf{s}}$ a random variable with the same distribution as the stationary distribution of the CTMC.

As the exact analysis is challenging, we consider a many-server-heavy-traffic asymptotic regime, wherein the number of servers are scaled to infinity ($n \rightarrow \infty$) and the arrival rate increases to the capacity ($\lambda/n \rightarrow 1$). In particular, consider a sequence of load balancing systems parameterized by n . The arrival rate for the n^{th} system is given by $\lambda = n - n^{1-\gamma}$. In this paper, we are interested in the case of $\gamma \in (0, 0.5)$, known as the sub-Halfin-Whitt regime. In addition, our focus is on growing choices in Power-of- d , i.e. $d \rightarrow \infty$ as $n \rightarrow \infty$. The goal is to characterize the limiting steady-state distribution $\bar{\mathbf{s}}^{(n)}$ as $n \rightarrow \infty$. In the rest of the paper, we suppress the dependence of $\bar{\mathbf{s}}$ on n for notational convenience.

3. Results and Insights

In this section, we develop intuition by considering an ODE approximation of the load-balancing system. This will guide the limiting behavior of the stochastic model. In particular, we expect the limiting stationary distribution of the stochastic model to concentrate around the fixed point of the ODE approximation. Note that, we do not directly work with the ODE to prove the result for the stochastic model. We only leverage intuition from the ODE approximation. To prove such a result, we find the region where the steady-state stochastic system resides with high-probability by iteratively narrowing down the possible regions of the state-space [21, 20].

3.1. Intuition: ODE approximation

To simplify the arguments in this section, consider $b = \infty$, i.e., the queues have infinite buffer capacity. Now, the evolution of n^{th} system can be approximated by abstracting out the stochasticity to obtain the following ODE:

$$\frac{ds_i}{dt} = \lambda \left(\left(\frac{s_{i-1}}{n} \right)^d - \left(\frac{s_i}{n} \right)^d \right) - (s_i - s_{i+1}) \quad \forall i \in \mathbb{Z}_+. \quad (1)$$

The rate of change of s_i is the difference of the rate at which it increases and decreases. The first term on the RHS is the product of arrival rate of the customers (λ) and the probability that the incoming customer will join a queue with length equal to $i - 1$. This is equal to the rate at which s_i is increasing. The rest of the terms is the product of the service rate ($\mu = 1$) and the number of queues ($s_i - s_{i+1}$) with queue length equal to i . This is equal to the rate at which s_i is decreasing.

To obtain the fixed point of the dynamical system, substitute $ds_i/dt = 0$ for all $i \in \mathbb{Z}_+$ resulting in a set of non-linear equations. After solving these equations, one obtains the following solution:

$$\frac{s_i}{n} = \left(\frac{\lambda}{n} \right)^{\frac{d^i - 1}{d - 1}} \quad \forall i \in \mathbb{Z}_+. \quad (2)$$

The above suggests a candidate stationary distribution of the stochastic system. In [26, 28], this intuition is made formal in the mean field regime, i.e., for $\lambda = (1 - \beta)n$ with $\beta \in (0, 1)$, and $d = 2$. Now, we consider the regime where $\lambda = n - n^{1-\gamma}$ for $\gamma \in (0, 0.5)$ and the routing is governed by Power-of- d with $d \rightarrow \infty$ as $n \rightarrow \infty$. Then, (2) can be approximated as follows:

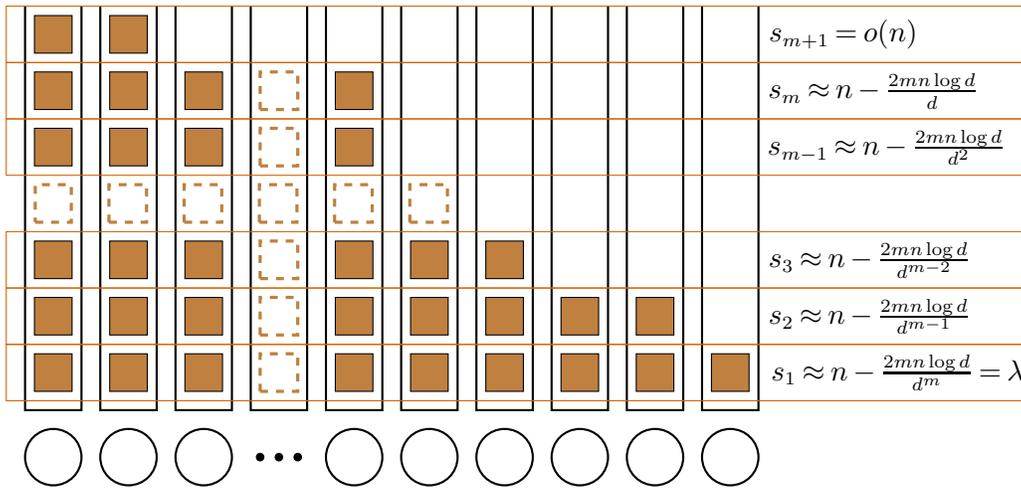
$$s_i = n (1 - n^{-\gamma})^{\frac{d^i - 1}{d - 1}} \stackrel{(a)}{\approx} n (1 - n^{-\gamma})^{d^{i-1}} \stackrel{(b)}{\approx} \begin{cases} n - n^{1-\gamma} d^{i-1} & \forall i \in [b] : n^{1-\gamma} d^{i-1} = o(n) \\ o(n) & \text{otherwise.} \end{cases} \quad (3)$$

where (a) follows by approximating $d^i - 1 \approx d^i$ and $d - 1 \approx d$ as d scales to infinity. Next, (b) follows by Taylor's series expansion up to the first order term. Now, define $m \in \mathbb{Z}_+$ to be the smallest integer such that the number of queues with length at least $m + 1$ is $o(n)$, i.e. $s_{m+1} = o(n)$ and $s_m = \Theta(n)$. Then, we must have $n^{1-\gamma}d^m \approx n$ implying that $d \approx n^{\gamma/m}$. More precisely, our choice of d is such that it satisfies $d = (2mn^\gamma)^{1/m} \log(d)^{1/m}$ which is approximately equivalent to $d \approx n^{\gamma/m}$ when the lower order terms are ignored. For this value of d , from (3), we get

$$s_i \approx \begin{cases} n - \frac{2mn \log d}{d^{m-i+1}} & \forall i \in [m] \\ o(n) & \text{otherwise.} \end{cases} \quad (4)$$

Fig. 3 illustrates the fixed point in terms of queue occupancy. We expect the stationary distribution of the stochastic model to concentrate around the above fixed point. The main contribution of our paper is to prove that the heuristic argument is indeed correct. We state the formal result in the next sub-section.

Figure 3 Illustration of the fixed point of the ODE approximation in terms of queue occupancy.



3.2. Main Result

Now, we present the main results of the paper below.

THEOREM 1. *Let $\{m_n \in \mathbb{Z}_+ : n \in \mathbb{Z}_+\}$ be a sequence such that either $m_n \equiv m \in \mathbb{Z}_+$ or $m_n \rightarrow \infty$. Consider a load balancing model operating under Power-of- d routing algorithm with $d = (2m_n n^\gamma)^{1/m_n} \log(d)^{1/m_n}$. If further $d = \Omega(\log(n)^3)$ and $b = O(\log(n)^3)$, then with probability at least $1 - (\frac{1}{n})^{(m_n \log n)^9}$, for large enough n , we have*

$$\bar{s}_i = \begin{cases} n - \frac{2m_n n \log d}{d^{m_n-i+1}} + o\left(\frac{2m_n n \log d}{d^{m_n-i+1}}\right) & \forall i \in [m_n] \\ o(n) & \text{for } i = m_n + 1 \\ o(1) & \text{otherwise.} \end{cases}$$

The result in Theorem 1 is obtained by proving a high probability lower bound and a high probability upper bound separately on \mathbf{s} in Theorem 2 and Theorem 3, respectively. These are presented at the end of this section, wherein, we also explicitly characterize the $o(\cdot)$ terms.

Note that the above theorem considers d as a solution of an implicit equation. However, one can obtain upper and lower bounds on d , matching up to a logarithmic term. In particular, we have

$(2mn^\gamma)^{1/m} \leq d \leq (2mn^\gamma)^{1/m} \log(n)^{1/m}$. Also note that, since d is the number of sampled queues, it should be an integer. However, we ignore this issue and in our results and proofs we do not explicitly impose the constraint that d should be an integer. This can be easily fixed by introducing floors and ceilings at appropriate places, which we do not do for ease of exposition. Moreover, as $d \rightarrow \infty$ as $n \rightarrow \infty$, the integrality approximation error will be negligible for large n .

The above result shows that the stationary distribution of the stochastic model concentrates around the fixed point of the ODE approximation given by (4). To further understand the result, consider the limit as $n \rightarrow \infty$ to get

$$\lim_{n \rightarrow \infty} \frac{\bar{s}_i}{n} = \begin{cases} 1 & \forall i \in [m] \\ 0 & \text{otherwise.} \end{cases}$$

Thus, most of the queues have lengths equal to m which implies that an incoming customer joins a queue with length $m - 1$ with high probability. In particular, we have

$$m = \gamma \frac{\log n}{\log d} + \frac{\log(2m \log d)}{\log d} = \gamma \frac{\log n}{\log d} (1 + o(1)). \quad (5)$$

So, if d is a polynomial in n , then m is finite. On the other hand, if d is smaller than any polynomial, then m will increase to infinity as $n \rightarrow \infty$. We discuss such a phase transition in the queue length behavior below.

3.2.1. Phase Transitions The limiting steady-state performance in this regime exhibits phase transitions as the values of different parameters like γ and d are varied. We outline two such phase transitions below.

Phase Transition as d varies: It was shown in [22] that if $d \geq n^\gamma \log n$, the delay experienced by the customers is asymptotically zero. This is consistent with our result as $d = n^\gamma \log n$ corresponds to the case of $m = 1$ implying asymptotically zero-delay. Now, if the value of d decreases beyond n^γ , one would expect the steady-state queue length distribution to be higher. According to Theorem 1, if d is of the form $d = n^{\gamma/m}$ for some $m \geq 2$, the steady-state queue lengths are equal to m . In particular, if d is a polynomial less than n^γ , then the waiting times are non-zero but finite. As discussed before, such a qualitative behavior is similar to that of JSQ in the NDS regime. Now, consider the case when d is smaller than any polynomial, for instance, Poly-Log(n). This corresponds to $d \approx n^{\gamma/m}$ with $m \approx \gamma \log n / \log d$ which increases to infinity as $n \rightarrow \infty$. In this case, the queue lengths are asymptotically infinite, which corresponds to infinite delay. As discussed before, such behavior is qualitatively similar to that of JSQ in the super slowdown regime.

To summarize, we characterize the limiting steady-state behavior of the stochastic model in Theorem 1. Based on this, we show that different orders of delays can emerge depending on the choice of d . In particular, zero-delay for $d \geq n^\gamma \log n$, finite-delay for polynomial $d(n)$, and infinite delay for poly-log $d(n)$. Such a phase transition is reminiscent of the phase transition of JSQ as the load increases from the mean field to the classical heavy traffic regime. To summarize, we characterize the limiting steady-state behavior of the stochastic model in Theorem 1. Based on this, we show that different orders of delays can emerge depending on the choice of d . In particular, zero-delay for $d \geq n^\gamma \log n$, finite-delay for polynomial $d(n)$, and infinite delay for poly-log $d(n)$. Such a phase transition is reminiscent of the phase transition of JSQ as the load increases from the mean field to the classical heavy traffic regime.

Phase Transition as γ varies: It is well known [28, 26] that the queue length under the power of d choices for $d \geq 2$ is approximately $\Theta(\log \log n / \log d)$ for the mean field regime, i.e. $\gamma = 0$. On the other hand, we show that the queue lengths are $m = \Theta(\gamma \log n / \log d)$ when $\gamma > 0$. In particular, we observe a phase transition in the queue length behavior as the value of γ moves from 0 to > 0 . In other words, as the value of γ increases, the arrival rate increases resulting in a higher load on the system, which results in larger queue lengths.

3.2.2. Matching Upper and Lower Bounds We present two theorems characterizing matching high probability lower and upper bounds on \mathbf{s} . Taken together, these two theorems give us Theorem 1.

THEOREM 2. *Consider the same setup as Theorem 1. Then, there exists $n_{LB} \in \mathbb{Z}_+$ such that for all $n \geq n_{LB}$, we have*

$$\bar{s}_i \geq n - 2m \frac{n \log d}{d^{m-i+1}} - 4md^{i-1} \sqrt{mn} \log n - 16m^3 \frac{n \log(d)^2}{d^{m-i+2}} \text{ w.p. at least } 1 - \left(\frac{1}{n}\right)^{(m \log n)/5} \quad \forall i \in [m].$$

THEOREM 3. *Consider the same setup as Theorem 1. Then, there exists $n_{UB} \in \mathbb{Z}_+$, such that for all $n \geq n_{UB}$, we have*

$$\begin{aligned} \bar{s}_i &\leq n - 2m \frac{n \log d}{d^{m-i+1}} + 19md^{i-1} \sqrt{mn} \log n + 39m^3 \frac{n \log(d)^2}{d^{m-i+2}} + \frac{n^{1-\gamma}}{d^{m-i}} \mathbb{1}\{m > 1\} \quad \forall i \in [m] \\ \bar{s}_{m+1} &\leq 18md^{m-1} \sqrt{mn} \log n + 36m^3 \frac{n \log(d)^2}{d^2} + n^{1-\gamma} \mathbb{1}\{m > 1\} \\ \sum_{l=m+2}^b \bar{s}_l &\leq 1. \end{aligned}$$

each one with probability at least $\left(\frac{1}{n}\right)^{(m \log n)/9}$.

We prove Theorem 2 in Section 4.2 and Theorem 3 in Section 5.2. Our result only holds true for $b \leq \log(n)^3$ for technical reasons discussed in Section 5.2. However, note that, since the queue lengths are of size m w.h.p. which is $o(\log n)$, the finite buffer requirement $b \leq \log(n)^3$ is inconsequential. So, we expect that the result holds true for all $b \in \mathbb{Z}_+$. In particular, by (5), we have

$$m = \Theta\left(\frac{\log n}{\log d}\right) = o(\log n),$$

where the last equality follows as $d = \Omega(\log(n)^3)$.

The terms $4md^{i-1} \sqrt{mn} \log n$, $16m^3 \frac{n \log(d)^2}{d^{m-i+2}}$, and $\frac{n^{1-\gamma}}{d^{m-i}} \mathbb{1}\{m > 1\}$ in Theorem 2 and similar terms in Theorem 3 are lower order terms compared to the leading term $2m \frac{n \log d}{d^{m-i+1}}$. We verify it by considering their ratio as follows:

$$\begin{aligned} \frac{4md^{i-1} \sqrt{mn} \log n + 16m^3 \frac{n \log(d)^2}{d^{m-i+2}} + \frac{n^{1-\gamma}}{d^{m-i}} \mathbb{1}\{m > 1\}}{2m \frac{n \log d}{d^{m-i+1}}} &= \frac{2\sqrt{m} d^m \log n}{\sqrt{n} \log d} + \frac{8m^2 \log d}{d} + \frac{dn^{-\gamma} \{m > 1\}}{2m \log d} \\ &\stackrel{(a)}{\leq} 4n^{\gamma-0.5} \log(n)^{2.5} + \frac{8 \log(n)^2 \log d}{d} + (2m \log d)^{1/m-1} n^{-\gamma(1-1/m)} \{m > 1\} \stackrel{(b)}{\rightarrow} 0, \end{aligned} \quad (6)$$

where (a) follows as $d^m = 2mn^\gamma \log d$ and $m = o(\log n)$. Next, (b) follows by substituting $d = \Omega(\log(n)^3)$ and noting that $\gamma < 0.5$. This shows that the rate of increase of d should be large enough to assure that the bound obtained in Theorem 2 concentrates around the fixed point (4). One can observe that even $d \approx \log(n)^2$ suffices to ensure (b) holds. However, we require a slightly stronger condition due to certain intermediate bounds in the proof. To extend the result to $d < \log(n)^2$, one needs to improve the lower order terms obtained in Theorem 2. We leave this as a possible future work.

3.3. Iterative State Space Collapse

The central idea to prove Theorem 1 is the iterative state space collapse framework based on drift analysis which was first developed in [21, 20]. As the stochastic system is expected to concentrate around the fixed ‘point’ of the ODE approximation, we keep slicing off the state space until the stationary distribution is implied to live in a ball around the fixed point. In particular, the stationary distribution is iteratively shown to collapse to smaller regions of the state space. Each step of the iteration is achieved by analyzing drift of a carefully engineered Lyapunov function. We present an intuitive overview of the framework and defer the details to Appendix B.2

Weak SSC: We start by proving a weak SSC. Let $V_1(\cdot)$ be a Lyapunov functions with $\Delta V_1(\mathbf{s}) \leq -\gamma$ for $\mathbf{s} \in \mathcal{S}$ such that $V_1(\mathbf{s}) \geq B$ for some $\gamma, B \in \mathbb{R}_+$. Then, by standard drift arguments [3, Theorem 1], we obtain that $\mathbb{P}(V_1(\bar{\mathbf{s}}) \geq B + j)$ is very small for a large enough j . This implies that $\bar{\mathbf{s}}$ collapses to the set $\mathcal{E}_1 = \{V_1(\mathbf{s}) \leq B + j\} \subseteq \mathcal{S}$.

Refining the SSC: Now, consider the subset of the state space $\mathcal{E}_1 \subseteq \mathcal{S}$ where $\bar{\mathbf{s}}$ resides with high probability as shown in the previous step. We analyze the drift of another Lyapunov function $V_2(\cdot)$ restricted to the set $\mathbf{s} \in \mathcal{E}_1$ such that $V_2(\mathbf{s}) \geq B$. As $\mathcal{E}_1 \subseteq \mathcal{S}$, one would obtain a stronger upper bound on the drift of $V_2(\cdot)$ as opposed to the bound over all whole state-space. Iterative SSC [21, Lemma 10] framework says that, as long as $\mathbb{P}(\bar{\mathbf{s}} \notin \mathcal{E}_1) \approx 0$, negative drift of $V_2(\cdot)$ over \mathcal{E}_1 implies a high probability bound on the steady-state distribution. This implies that $\bar{\mathbf{s}}$ now collapses to the set $\mathcal{E}_2 = \mathcal{E}_1 \cap \{V_2(\mathbf{s}) \leq B + j\}$. A key takeaway is that, the SSC framework poses a trade-off between relaxing the negative drift condition to $\mathbf{s} \in \mathcal{E}_1$ and the steady-state probability $\mathbb{P}(\bar{\mathbf{s}} \notin \mathcal{E}_1)$. One can carefully negotiate this trade-off to obtain meaningful results.

Further Refinements: The refinement can be repeated multiple times. Consider a family of Lyapunov functions $\{V_k(\mathbf{s}) : k \in \mathbb{Z}_+\}$ and define

$$\mathcal{E}_k = \mathcal{E}_{k-1} \cap \{V_k(\mathbf{s}) \leq B + j\}.$$

Now, we inductively analyze the drift of $V_{k+1}(\mathbf{s})$ for $\mathbf{s} \in \mathcal{E}_k$ such that $V_{k+1}(\mathbf{s}) \geq B$. By iterative SSC framework, negative drift of $V_{k+1}(\mathbf{s})$ implies the high probability bound $\mathbb{P}(V_{k+1}(\mathbf{s}) \leq B + j) \approx 1$. Thus, after $k + 1$ iterations, the region of SSC shrinks to $\mathcal{E}_{k+1} = \mathcal{E}_k \cap \{V_{k+1}(\mathbf{s}) \leq B + j\}$. In further sections, we use this inductive framework to prove Theorem 2 and Theorem 3.

Lemma 14 in Appendix B.2 formalizes the above intuition and is the workhorse of our proof, used to show each step of the SSC refinement. Our proof is tailored to analyze a queueing system in the steady state allowing us to greatly simplify the methodology compared to [7]. In particular, each step of our refinement directly implies high probability bounds on the steady-state distribution of the queue length (\mathbf{s}) process. On the other hand, [7] (see Lemma 3 in their paper) essentially works with a transient version of Lemma 14, so they need to worry about any given Lyapunov function to stay small while the other Lyapunov functions decrease, further complicating the proof. Another difference is that our sequence of Lyapunov functions is completely different from that of [7] providing an alternative proof and revealing geometric insights. In particular, while [7] provides an elegant algebraic construction of the sequence of Lyapunov functions, our sequence is based on geometric intuition via the trajectory of the ODE approximation.

4. Lower Bound

In this section, we prove Theorem 2, by iteratively showing high probability lower bounds on $\{s_i\}_{i \in [m]}$ based on iterative SSC described in the previous section. At a high level, starting with an empty system, we prove a high probability lower bound on s_1 as most of the incoming jobs would join an empty queue. Further, using the lower bound on s_1 , we obtain a lower bound on s_2 . This procedure is continued for $\{s_i\}_{i \in [m]}$. As most of the proof is algebraic, we first focus on the special case of $m = 2$ to provide intuition behind the methodology.

4.1. Special case ($m = 2$)

4.1.1. Approximate ODE Trajectory We start by analyzing the trajectory of the ODE approximation for $m = 2$. The main idea is that the stochastic model would follow a noisy sample path around the ODE trajectory. We present a cartoon of the approximate ODE trajectory based on the differential equations given by (1) in Fig. 4. The model is initialized by an all-empty system.

Trajectory (o) – (a): Each incoming job joins an empty queue until $s_1 = n - n/d$, as there is a surplus of empty queues. In particular, if $s_1 \leq n - n \log d/d$, then the probability of sampling at least one empty queue $(1 - (s_1/n)^d)$ is almost one. On the other hand, if $s_1 \geq n - n/(d \log d)$, then the probability is almost zero. Mathematically, we have

$$\begin{aligned} 1 - \left(1 - \frac{\log d}{d}\right)^d &\approx 1 - \frac{1}{d} \rightarrow 1 \\ 1 - \left(1 - \frac{1}{d \log d}\right)^d &\approx \frac{1}{\log d} \rightarrow 0. \end{aligned}$$

Consistent with the intuition, one can confirm that $ds_1/dt > 0$ when $s_1 \leq n - n \log d/d$. We have

$$\begin{aligned} \frac{ds_1}{dt} &= \lambda \left(1 - \left(\frac{s_1}{n}\right)^d\right) - s_1 + s_2 \geq \lambda \left(1 - \left(1 - \frac{\log d}{d}\right)^d\right) - n + \frac{n \log d}{d} \\ &\approx n \left(1 - \frac{1}{d}\right) - n + \frac{n \log d}{d} = \frac{n(\log d - 1)}{d} > 0. \end{aligned}$$

Note that, one can show that s_2 increases as well but at a smaller rate compared to s_1 by observing that $0 < ds_2/dt \ll ds_1/dt$. We omit the details here for brevity. We approximately represent the trajectory as horizontal from (o) to (a) in Fig. 4.

Trajectory (a) – (b): Once $s_1 \approx n - n/d$, most of the incoming jobs start joining a queue with length one, as the probability of sampling an empty queue is asymptotically zero. Mathematically, for $s_1 \geq n - n/(d \log d)$ and $s_2 = o(n)$, we have

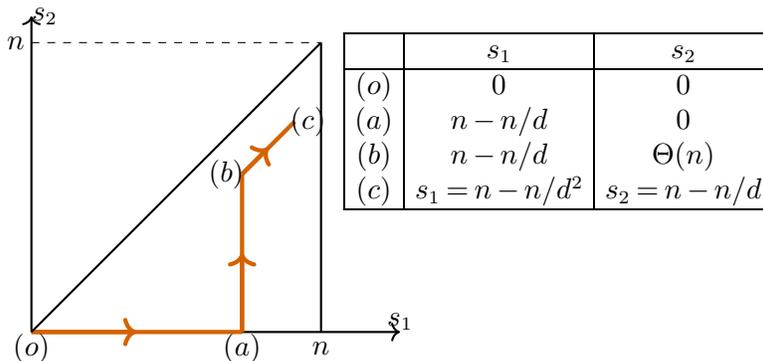
$$\frac{ds_2}{dt} = \lambda \left(\left(\frac{s_1}{n}\right)^d - \left(\frac{s_2}{n}\right)^d\right) - s_2 + s_3 \geq \lambda \left(\left(1 - \frac{1}{d \log d}\right)^d - o(1)^d\right) - o(n) \approx n - \frac{n}{\log d} - o(n) > 0.$$

Thus, s_2 increases until $s_2 = \Theta(n)$. Similar to (o) – (a), s_1 increases as well but at a smaller rate compared to s_2 as $ds_1/dt \ll ds_2/dt$. We approximately represent the trajectory as vertical from (a) to (b) in Fig. 4.

Trajectory (b) – (c): As s_2 increases, ds_1/dt increases and ds_2/dt decreases. After a critical point, ds_1/dt and ds_2/dt are comparable resulting in both s_1 and s_2 increasing at a similar rate. This is approximately represented as a tilted trajectory from (b) to (c). Once $s_1 = \lambda \approx n - n/d^2$, the arrival rate is equal to the effective service rate (number of busy servers), and the ODE trajectory converges to that point. One can verify that (c) is indeed the fixed point (ignoring the logarithmic terms) of the ODE approximation.

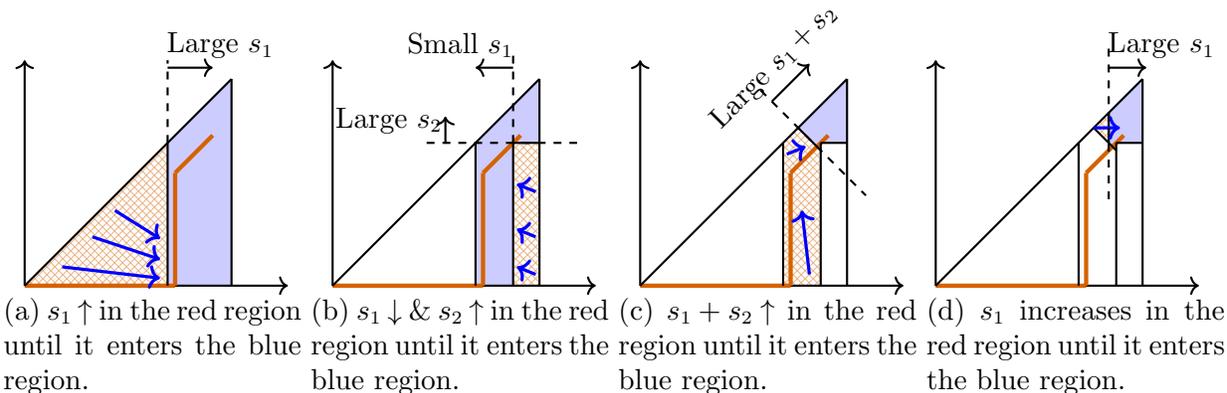
Note that Fig. 4 is the transient behavior of the ODE approximation starting from an all-empty system. On the other hand, our goal is to characterize the steady-state behavior of the stochastic system. Steady-state corresponds to a fixed distribution invariant with time and it doesn't follow a transient trajectory as shown in Fig. 4. However, the sequence of SSC we establish in the next sub-section is inspired by the transience of the ODE approximation. Such an interpretation of the ODE approximation is also consistent with the fundamentals of the Lyapunov drift arguments. In particular, drift of a Lyapunov function is a transient quantity as it depends on the state of the system. In turn, it implies a high probability bound on the steady-state distribution. Lastly, note that we do not have a cartoon for the ODE approximation for the general case as it would be m dimensional. However, we carry forward the algebraic intuitions developed using $m = 2$ for the general case.

Figure 4 Approximate ODE trajectory with the initial condition equal to $s_1 = 0$.



4.1.2. Stochastic Analysis The form of iterative SSC is inspired by the ODE trajectory depicted in Fig. 5. In particular, there is a drift that pulls the system closer to the ODE trajectory. We use this idea to show that the states that are far from the trajectory are experienced with low steady-state probability. Now, we elucidate the steps to prove the theorem that are outlined in Fig.

Figure 5 Graphical representation of iterative SSC for lower bound for $m = 2$: The red hatched region is shown to have low steady-state probability leading to the collapse into the solid blue region. In addition, the arrows represent the drift - $(ds_1/dt, ds_2/dt)$.



6.

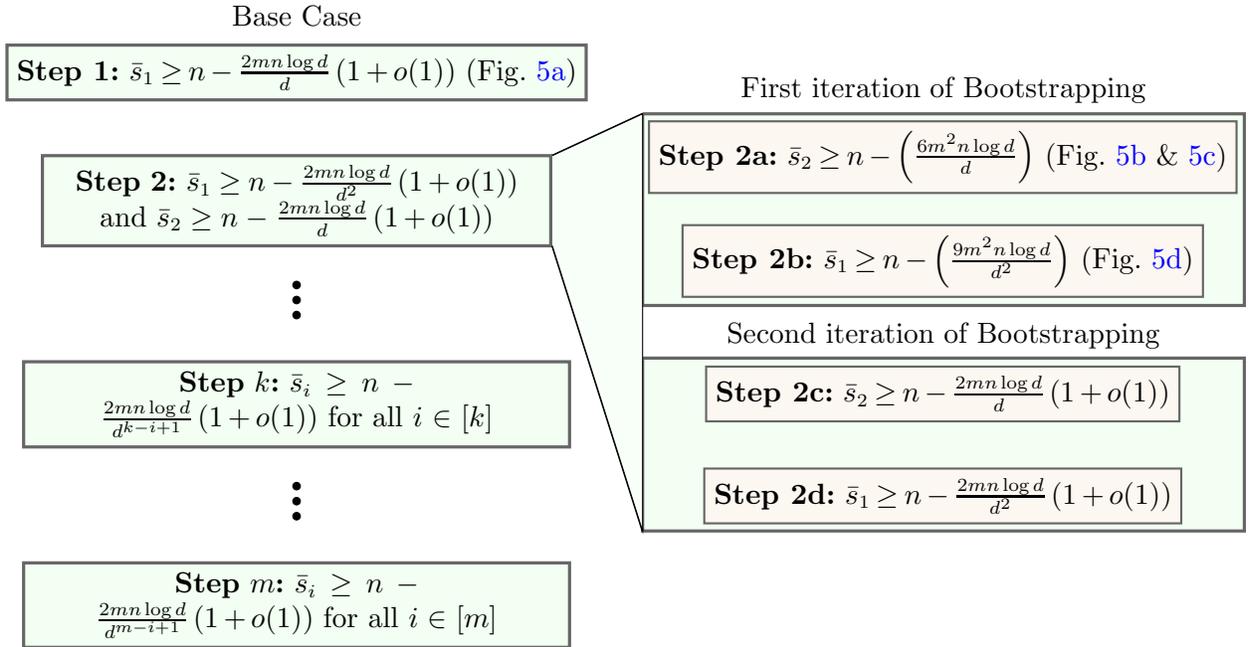
- **Step 1:** We first show that $\bar{s}_1 \geq n - 2mn \log d/d$ w.h.p. corresponding to trajectory (o) – (a) in the ODE approximation. Intuitively, all incoming jobs join empty queues due to their availability in surplus.
- **Step 2a:** Given the lower bound on \bar{s}_1 , we obtain $\bar{s}_2 \geq n - 6m^2n \log d/d$ w.h.p. corresponding to trajectory (a) – (b) in the ODE approximation. This is carried out in two-steps as depicted in Fig. 5b and 5c.
- **Step 2b:** We improve the lower bound on \bar{s}_1 using the lower bound on \bar{s}_2 to get $\bar{s}_1 \geq n - 9m^2n \log d/d^2$ that loosely corresponds to trajectory (b) – (c) in the ODE approximation. This step is illustrated in Fig. 5d.
- **Step 2c:** This step improves the lower bound on \bar{s}_2 previously obtained in Step 2a. In particular, we leverage the newly obtained lower bound on \bar{s}_1 and repeat the same steps as in Step 2a, to obtain a better lower bound on \bar{s}_2 . We call this a bootstrapping step as a weaker lower bound on \bar{s}_2 results in a stronger lower bound on itself.

- **Step 2d:** Similar to the previous step, we improve the lower bound on \bar{s}_1 previously obtained in Step 2b. In particular, using a better lower bound on \bar{s}_2 obtained in Step 2c, we obtain a better lower bound on \bar{s}_1 by following the outline of Step 2b.

Why Iterative SSC?: Ideally, we would like to construct a single Lyapunov function whose drift analysis would reveal that the stochastic system concentrates around the fixed point of the ODE. However, it is not clear if such a Lyapunov function exists, or at least, we were not successful in constructing such a function. In our approach, we construct a sequence of Lyapunov functions that mimics the trajectory of the ODE. The ODE trajectory in Fig. 4 reveals that the drift of s_1 and s_2 are of different orders in different parts of the state space. For example, the upward drift of s_1 is very large compared the drift of s_2 when the system is close to an empty system. Such a behavior is reminiscent of the state space collapse result in classical heavy-traffic regime [10] and two timescale algorithms in reinforcement learning (e.g. see [5, Chapter 6] and [16]). Such a two timescale behavior warrants the need of constructing multiple Lyapunov functions to analyze the collapse of s_1 and s_2 .

In the next sub-section, we build upon the intuition to extend the proof to the general case as shown in Fig. 6. We do not separately present the proof for $m = 2$ as it is subsumed in the general case.

Figure 6 Proof Outline: High Probability Lower Bounds.



4.2. General Case

The proof is mainly divided in five steps and we present five lemmas corresponding to these steps. The proof outline is presented in Fig. 6 and we naturally prove it using induction. In step k , we provide a high probability lower bound on \bar{s}_k and improve the previous bounds obtained for $\{\bar{s}_i\}_{i \in [k-1]}$. The proof of all the lemmas mentioned in this section is deferred to Appendix C.

Base Case: We start by obtaining a high probability lower bound on \bar{s}_1 (Step 1) in the following lemma.

LEMMA 1 (Base Case). *Consider the same setup as Theorem 1. Then, there exists $n_0 \in \mathbb{Z}_+$ such that for all $n \geq n_0$,*

$$\mathbb{P} \left(\bar{s}_1 \leq n - \frac{2mn \log d}{d} - \frac{2n \log d}{d^2} - 2\sqrt{mn} \log n \right) \leq \left(\frac{1}{n} \right)^{m \log n/4}.$$

The proof of the lemma follows by showing that the drift of the Lyapunov function,

$$V_1(\mathbf{s}) = n - \frac{2mn \log d}{d} - \frac{2n \log d}{d^2} - \sqrt{mn} \log n - s_1,$$

is negative when $V_1(\mathbf{s}) \geq 0$. Thus, s_1 increases w.h.p. when $s_1 \leq n - \frac{2mn \log d}{d}(1 + o(1))$, implying that $\bar{s}_1 > n - \frac{2mn \log d}{d}(1 + o(1))$ w.h.p. This completes the base case.

Induction Step: For some $k \in \{2, \dots, m\}$, we define our induction hypothesis as follows. There exists $n_{IH} \in \mathbb{Z}_+$, independent of k such that for all $n \geq n_{IH}$, we have

$$\mathbb{P} \left(\bar{s}_i \leq n - 2m \frac{n \log d}{d^{k-i}} - 4md^{i-1} \sqrt{mn} \log n - 16m^3 \frac{n \log(d)^2}{d^{k-i+1}} \right) \leq \left(\frac{1}{n} \right)^{\frac{m \log n}{4} - 4(k-1)m} \quad \forall i \in [k-1]. \quad (\text{IH})$$

Note that, (IH) is equivalent to Step $k-1$ as in the proof outline given by Fig. 6. After completing the induction step, we improve the bound on $(n - \bar{s}_i)$ by a factor of d for all $i \in [k-1]$ and moreover, we introduce a lower bound on \bar{s}_k of the order $n - 2nm \log d/d$. Similar to the way Step 2 is proved in four parts as in Fig. 6, we present four lemmas that together completes the induction step. The first of the four lemmas corresponds to part *a* of the induction step (Steps 2 to m in Fig. 6), wherein we obtain a lower bound on \bar{s}_k .

LEMMA 2 (Induction Part a). *Consider the same setup as Theorem 1. Also, assume that (IH) holds for some $k \leq m$. Then, there exists $n_1 \in \mathbb{Z}_+$ such that for all $n \geq n_1$, we have*

$$\mathbb{P} \left(\bar{s}_k \leq n - 6m^2 \frac{n \log d}{d} \right) \leq \left(\frac{1}{n} \right)^{\frac{m \log n}{4} - 4(k-\frac{3}{4})m}.$$

Note that, the above lower bound is weak, as the term accompanying $n \log n/d$ is $6m^2$ which is larger than the required $2m$ in the induction step. We improve this bound in part *c* (Lemma 4). Now, we present the proof sketch.

Proof Sketch of Lemma 2 The proof of this lemma is induction based. To state the induction hypothesis, consider a Lyapunov function of the following form (the exact expression is presented in the proof):

$$L_{ik}(\mathbf{s}) = \min \left\{ s_i - n + \Theta \left(\frac{m^2 n \log d}{d^{k-i+1}} \right), (k-i+1)n - \Theta \left(\frac{m^2 n \log d}{d} \right) - \sum_{l=i+1}^k s_l \right\} \quad \forall i \in [k-1].$$

Now, for some $i \in [k-1]$, the induction hypothesis is given as follows, which runs backwards on the index i : $L_{ik}(\bar{\mathbf{s}}) = O(\sqrt{mn} \log n)$. To prove the induction step, we show that the drift of the Lyapunov function $L_{i-1,k}(\mathbf{s})$ is negative for all states such that $L_{i,k}(\mathbf{s}) = O(\sqrt{mn} \log n)$ and $L_{i-1,k}(\mathbf{s}) \geq 0$. Thus, by applying the iterative SSC framework with $V = L_{ik}$ and, $\mathcal{E} = \{L_{i+1,k}(\mathbf{s}) = O(\sqrt{mn} \log n)\}$, we conclude that $L_{i-1,k}(\bar{\mathbf{s}}) = O(\sqrt{mn} \log n)$ w.h.p. Thus, the induction implies that $L_{1k}(\bar{\mathbf{s}}) = O(\sqrt{mn} \log n)$ w.h.p. Note that $L_{i-1,k}(\mathbf{s})$ is small implies that either s_{i-1} is small or $\sum_{l=i}^k s_l$ is large. This form of state space collapse is reminiscent of the first sub-task of Step 2a as depicted in Fig. 5b in which we showed that either \bar{s}_1 is small or $\sum_{l=2}^b \bar{s}_l$ is large.

To complete the proof, we further consider the following Lyapunov function:

$$L_{0k}(\mathbf{s}) = kn - \Theta\left(\frac{m^2 n \log d}{d}\right) - \sum_{l=1}^k s_l.$$

Using the fact that $L_{1k}(\bar{\mathbf{s}}) = O(\sqrt{mn} \log n)$ w.h.p., we show that $L_{0k}(\bar{\mathbf{s}}) = O(\sqrt{mn} \log n)$ w.h.p. In particular, we show that the drift of $L_{0k}(\mathbf{s})$ is negative for all states such that $L_{1k}(\mathbf{s}) = O(\sqrt{mn} \log n)$ and $L_{0k}(\mathbf{s}) \geq 0$. Thus, by Lemma 14 with $V = L_{0k}$, $\mathcal{E} = \{L_{1k} = O(\sqrt{mn} \log n)\}$, $B = 0$, and $j = \sqrt{mn} \log n$, we conclude that $\sum_{l=1}^k \bar{s}_l \geq kn - \Theta(m^2 n \log d/d)$ w.h.p. Observing that $\bar{s}_l \leq n$ w.p. 1, we get $\bar{s}_k \geq n - \Theta(m^2 n \log d/d)$ w.h.p. This completes the proof. Note that, analyzing the drift of $L_{0k}(\mathbf{s})$ is equivalent to the second sub-task of Step 2a as depicted in Fig. 5c. \square

In the next lemma, we prove weak lower bounds on \bar{s}_i for all $i \in [k-1]$ completing part *b* of steps $2-m$, analogous to Step 2b as outlined in Fig. 6.

LEMMA 3 (Induction Part b). *Consider the same setup as Theorem 1. Also, assume that $m \geq 2$ and (IH) holds for some $k \leq m$. Then, there exists $n_2 \in \mathbb{Z}_+$ such that for all $n \geq n_2$, we have*

$$\mathbb{P}\left(\bar{s}_i \leq n - 9m^2 \frac{n \log d}{d^{k-i+1}}\right) \leq \left(\frac{1}{n}\right)^{\frac{m \log n}{4} - 4(k-\frac{1}{2})m} \quad \forall i \in [k].$$

Note that, the bound obtained in the above lemma is weaker compared to the required bound for the induction step as $9m^2 > 2m$. We improve on this bound in part *d* to obtain the induction step. We now present the sketch of the proof.

Proof Sketch of Lemma 3 We again use induction on i to prove bounds on \bar{s}_i for all $i \in [k]$ by going backwards over the index i . The base case ($i = k$) is already proved in Lemma 2. To prove the induction step, we analyze the drift of the following Lyapunov function:

$$W_i(\mathbf{s}) = n - \Theta\left(\frac{m^2 n \log d}{d^{k-i+1}}\right) - s_i.$$

We consider the induction hypothesis $W_{i+1}(\bar{\mathbf{s}}) = O(\sqrt{mn} \log n)$ w.h.p. which is equivalent to a high probability lower bound on \bar{s}_{i+1} . We show that the drift of $W_i(\mathbf{s})$ is negative when $W_{i+1}(\mathbf{s}) = O(\sqrt{mn} \log n)$ and $W_i(\mathbf{s}) \geq 0$. Thus, by Lemma 14, we conclude that $W_i(\bar{\mathbf{s}}) = O(\sqrt{mn} \log n)$ w.h.p. This provides a lower bound on \bar{s}_i which completes the proof. Note that, the bound on $W_{i+1}(\mathbf{s})$ is crucial in obtaining the negative drift for $W_i(\mathbf{s})$. Thus, the iterative version of SSC is an integral part of the proof. \square

Now, we prove the required lower bound on \bar{s}_k completing part *c* of steps $2-m$, analogous to Step 2c as in Fig. 6.

LEMMA 4 (Induction Part c). *Consider the same setup as Theorem 1. Also, assume that $m \geq 2$ and (IH) holds for some $k \leq m$. Then, there exists $n_3 \in \mathbb{Z}_+$ such that for all $n \geq n_3$, we have*

$$\mathbb{P}\left(\bar{s}_k \leq n - 2m \frac{n \log d}{d} - 3md^{k-1} \sqrt{mn} \log n - 12m^2 \frac{n \log d}{d^2}\right) \leq \left(\frac{1}{n}\right)^{\frac{m \log n}{4} - 4(k-\frac{1}{4})m}.$$

Note that the above lemmas replaces the coefficient of $n \log d/d$ from $6m^2$ in Lemma 2 with the required $2m$ for the induction step. The proof of the above lemma follows similarly to the proof of Lemma 2 but leverages the lower bound on \bar{s}_{k-1} proved in Lemma 3. As the lower bound provided by Lemma 3 is sharper than the induction hypothesis (IH), we obtain a better lower bound on \bar{s}_k . Turns out, applying this bootstrapping step once suffices to prove the induction step for \bar{s}_k . Now, using this bound, we obtain the required lower bound on \bar{s}_i for all $i \in [k-1]$ which completes the part *d* of steps $2-m$, analogous to Step 2d as outlined in Fig. 6.

LEMMA 5 (**Induction Part d**). Consider the same setup as Theorem 1. Also, assume that $m \geq 2$ and (IH) holds for some $k \leq m$. Then, there exists $n_4 \in \mathbb{Z}_+$ such that for all $n \geq n_4$, we have

$$\mathbb{P}\left(\bar{s}_i \leq n - 2m \frac{n \log d}{d^{k-i+1}} - 4md^{i-1} \sqrt{mn} \log n - 16m^3 \frac{n \log(d)^2}{d^{k-i+2}}\right) \leq \left(\frac{1}{n}\right)^{\frac{m \log n}{4} - 4km} \quad \forall i \in [k]. \quad (7)$$

Note that the above lemma proves the coefficient of $n \log d / d^{k-i+1}$ from $9m^2$ in Lemma 3 to the required $2m$ for the induction step. The proof follows similarly to the proof of Lemma 3 but leverages a sharper lower bounds on \bar{s}_i for all $i \in [k-1]$ proved in Lemma 3. Note that the above lemma essentially completes the induction because (7) is same as (IH) with $k-1$ replaced by k . Now, we use the above lemmas below to prove Theorem 2.

Proof of Theorem 2 Consider the induction hypothesis as defined in (IH). The base case follows by Lemma 1 for $n_{IH} \geq \max_{l \in [4]} \{n_l\}$, and the induction step follows by Lemma 5. Thus, for all $n \geq n_{IH}$, we have

$$\begin{aligned} \mathbb{P}\left(\bar{s}_i \leq n - 2m \frac{n \log d}{d^{m-i+1}} - 4md^{i-1} \sqrt{mn} \log n - 16m^3 \frac{n \log(d)^2}{d^{m-i+2}}\right) &\leq \left(\frac{1}{n}\right)^{\frac{m \log n}{4} - 4m^2} \\ &\leq \left(\frac{1}{n}\right)^{\frac{m \log n}{5}} \quad \forall i \in [m], \end{aligned}$$

where the last inequality follows for all $n \geq n_5$ for some $n_5 \in \mathbb{Z}_+$ by noting that $m = o(\log n)$. Now, by defining $n_{LB} \triangleq \max\{n_5, n_{IH}\}$, the proof is complete. \square

5. Upper Bound

Similar to the previous section, the analysis for the upper bound follows the iterative SSC framework. We first prove high probability upper bounds on \bar{s}_{m+1} and $\sum_{l=m+2}^b \bar{s}_l$ and then inductively prove high probability upper bounds on $\{\bar{s}_i\}_{i \in [m]}$. We start by focusing on the special case of $m=1$ to provide intuition behind the methodology.

5.1. Special Case ($m=1$)

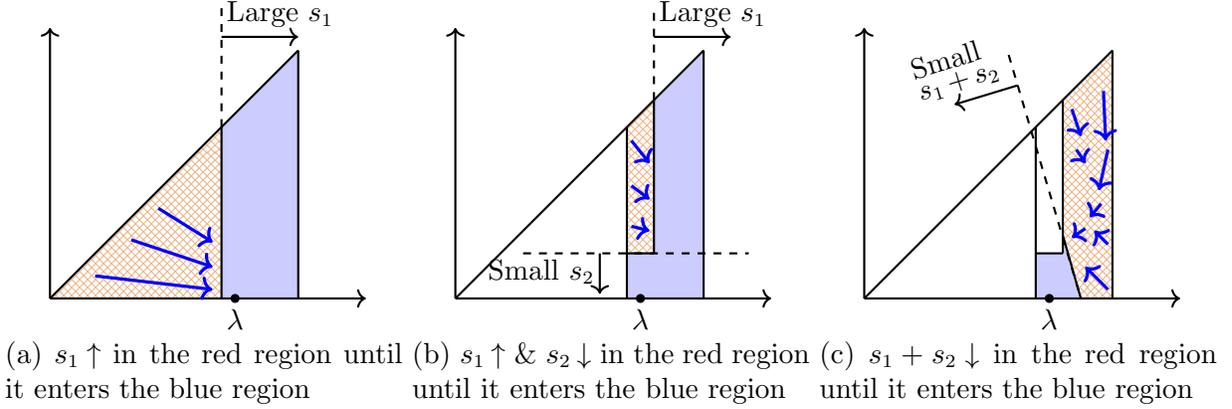
The geometric intuition for the steps to prove the special case is given in Fig. 7. In particular, Fig. 7a corresponds to Theorem 2 (lower bound on s_1) for the special case of $m=1$. The other two sub-figures correspond to proving matching upper bounds. Now, we elucidate the steps to obtain the upper bound for $m=1$, as outlined in Fig. 8.

- **Step 1a:** We first show that $\sum_{l=2}^b \bar{s}_l = O(b\sqrt{n} \log n)$ w.h.p. Noting that $\bar{s}_i \geq 0$ for all $i \in \{3, \dots, b\}$, we get $\bar{s}_2 = O(b\sqrt{n} \log n) = o(n)$ w.h.p. This is a weaker bound than what is required for Theorem 3. We improve it further in Step 1c by first proving a high probability upper bound on $\sum_{l=m+2}^b \bar{s}_l$ in the next step.
- **Step 1b:** The probability of an incoming job joining a queue with length at least two is equal to the probability of sampling d queues with length at least two. This is equal to $\left(\frac{\bar{s}_2}{n}\right)^d = o(1)^d$ w.h.p. which converges to zero very fast. Consistent with the intuition, we show that $\sum_{l=3}^b \bar{s}_l = o(1)$ w.h.p.
- **Steps 1c and 1d:** Using the upper bound proved in Step 1b, we show that $\bar{s}_2 = O(\sqrt{n} \log n)$ and $\bar{s}_1 \leq \lambda + O(\sqrt{n} \log n)$ w.h.p. This is obtained by first showing (8a) holds, which allows us to prove (8b) as shown below.

$$\min \left\{ \sum_{l=2}^b \bar{s}_l - O(\sqrt{n} \log n), \lambda + O(\sqrt{n} \log n) - \bar{s}_1 \right\} \leq 0 \quad w.h.p. \quad (8a)$$

$$\sum_{l=1}^b \bar{s}_l \leq \lambda + O(\sqrt{n} \log n) \quad w.h.p. \quad (8b)$$

Figure 7 Graphical representation of iterative SSC for lower and upper bound for $m = 1$: The red hatched region is shown to have low steady-state probability leading to the collapse into the solid blue region. In addition, the arrows represents the drift - $(ds_1/dt, ds_2/dt)$.



Eq. (8b) immediately implies that $\bar{s}_1 \leq \lambda + O(\sqrt{n} \log n)$ as $\sum_{l=2}^b \bar{s}_l \geq 0$. In addition, by using the lower bound on \bar{s}_1 given by Theorem 2 in (8b), we obtain that $\sum_{l=2}^b \bar{s}_l \leq O(\sqrt{n} \log n)$ completing Step 1c and 1d. The SSC corresponding to (8a) is depicted in Fig. 7b which shows that either s_1 is large or s_2 is small. In particular, if s_1 is small and s_2 is large, then most of incoming jobs will join an empty queue resulting in s_1 increasing and s_2 decreasing. Further, the SSC corresponding to (8b) is depicted in Fig. 7c which upper bounds $\sum_{l=1}^b \bar{s}_l$ w.h.p.

Tight characterization of $\bar{\mathbf{s}}$ for the special case of $m = 1$ was first obtained in [22]. The authors used Stein's method in conjunction with SSC to prove the result. In particular, the SSC proved in [22] is equivalent to (8a) as depicted in Fig. 7b. This alone is not sufficient to characterize the complete stationary distribution. So, Stein's method was used along with (8a) to obtain the stationary distribution. On the other hand, we circumvent the use of Stein's method by using the iterative SSC framework, to further improve the SSC and obtain tight bounds on the stationary distribution. A takeaway from our paper is that one can simply use the iterative SSC approach to obtain tight bounds on the queue lengths if the stochastic system lives close to the fixed point of the corresponding dynamical system given by a set of ODEs.

5.2. General Case

To present the required intermediate results in a concise manner, define the following:

$$B_i = 18md^{i-1}\sqrt{mn} \log n + 36m^3 \frac{n \log(d)^2}{d^{m-i+2}} + \frac{n^{1-\gamma}}{d^{m-i}} \mathbb{1}\{m > 1\}. \quad (9)$$

Note that, B_i consists of lower order terms. In particular, $B_i = o\left(\frac{mn \log d}{d^{m-i+1}}\right)$ by (6). Now, corresponding to Step 1a, we provide a high probability upper bound on \bar{s}_{m+1} in the following lemma.

LEMMA 6 (Step 1a). *There exists $\tilde{n}_1 \in \mathbb{Z}_+$ such that for all $n \geq \tilde{n}_1$, we have*

$$\mathbb{P}(\bar{s}_{m+1} \geq bB_m) \leq \left(\frac{1}{n}\right)^{(m \log n)/6}.$$

Proof Sketch of Lemma 6 To prove the lemma, we first consider a family of functions defined as follows:

$$U_i(\mathbf{s}) = \min \left\{ \sum_{l=i+1}^b s_l - (m-i)n - o\left(\frac{2mn \log d}{d}\right), n - \frac{2mn \log d}{d^{m-i+1}}(1 - o(1)) - s_i \right\} \quad \forall i \in [m-1].$$

The proof is induction based. The induction hypothesis assumes that $U_{i+1}(\bar{\mathbf{s}}) = O(\sqrt{mn} \log n)$ w.h.p. and the induction runs backward in the index i . To prove the induction step, we show that the drift of $U_i(\mathbf{s})$ is negative for $\mathbf{s} \in \{U_{i+1}(\mathbf{s}) = O(\sqrt{mn} \log n)\}$ and $U_i(\mathbf{s}) \geq 0$. By the iterative SSC framework, we get $U_i(\bar{\mathbf{s}}) = O(\sqrt{mn} \log n)$ w.h.p. This completes the induction step which shows that $U_1(\bar{\mathbf{s}}) = O(\sqrt{mn} \log n)$ w.h.p. Note that, $U_i(\mathbf{s})$ is small implies that either s_i is large or $\sum_{l=i+1}^b s_l$ is small. Such a form of SSC is a generalization of the special case of $m = 1$ as depicted in Fig. 7b, where we show that either s_1 is large or $\sum_{l=2}^b s_l$ is small as in (8a).

To complete the proof, we further consider the following Lyapunov function:

$$U_0(\mathbf{s}) = \sum_{l=1}^b s_l - mn - o\left(\frac{2mn \log d}{d} b\right).$$

Now, using the fact that $U_1(\bar{\mathbf{s}}) = O(\sqrt{mn} \log n)$ w.h.p., by the iterative SSC framework, we show that $U_0(\bar{\mathbf{s}}) = O(\sqrt{mn} \log n)$ w.h.p. Lastly, by lower bounding $\{\bar{s}_i : i \in [m]\}$ using Theorem 2, we obtain

$$\sum_{l=m+1}^b \bar{s}_l = o\left(\frac{2mn \log d}{d} b\right).$$

This completes the proof of the lemma. □

Note that, the bound in Lemma 6 only makes sense if $bB_m = o(n)$. One can verify that $bB_m = o(n)$ by substituting $b \leq \log(n)^3$ and $d \geq \log(n)^3$. This implies that an appropriate upper bound on b is essential for the proof to work. It turns out that $b \leq \log(n)^3$ works for all $d \geq \log(n)^3$. The result can be easily extended for $b \leq \min\{n^{0.5-\gamma}, d\}$. Relaxing the requirements on b and d are part of future work.

Now, the next lemma corresponds to Step 1b in Fig. 8 and proves an $o(1)$ high probability upper bound on $\sum_{l=m+2}^b \bar{s}_l$ by using the high probability upper bound $\bar{s}_{m+1} = o(n)$ that was proved in Lemma 6.

LEMMA 7 (Step 1b). *There exists $\tilde{n}_2 \in \mathbb{Z}_+$ such that for all $n \geq \tilde{n}_2$, we have*

$$\mathbb{P}\left(\sum_{l=m+2}^b \bar{s}_l \geq 1\right) \leq \left(\frac{1}{n}\right)^{(m \log n)/7}.$$

As $\bar{s}_{m+1} = o(n)$ w.h.p., we have $\bar{s}_{m+1} \leq n/2$ w.h.p. for large enough n . Thus, the probability that an incoming customer will be matched with a queue with at least $m+1$ customers is $0.5^d \leq (1/n)^{\log(n)^2}$ w.h.p. as $d \geq \Omega(\log(n)^3)$. We analyze the drift of $\sum_{l=m+2}^b \bar{s}_l$ to obtain a high probability upper bound on itself.

Next, we employ this bound to improve the upper bound on \bar{s}_{m+1} , corresponding to Step 1c in the following lemma:

LEMMA 8 (Step 1c). *There exists $\tilde{n}_3 \in \mathbb{Z}_+$ such that for all $n \geq \tilde{n}_3$, we have*

$$\mathbb{P}(\bar{s}_{m+1} \geq B_m) \leq \left(\frac{1}{n}\right)^{(m \log n)/8}.$$

The proof of the lemma is a more fine tuned version of the proof of Lemma 6. In particular, the proof of Lemma 6 uses the coarse bound $\sum_{l=m+1}^b s_l \leq bs_{m+1}$. Lemma 7 improves this bound by showing $\sum_{l=m+2}^b \bar{s}_l = o(1)$ w.h.p., which implies $\sum_{l=m+1}^b s_l \leq s_{m+1} + o(1)$. With this refinement, we repeat the steps of the proof of Lemma 6, to get Lemma 8.

The rest of the proof of Theorem 3 is based on induction. In particular, we inductively prove upper bounds on $\{\bar{s}_i : i \in [m]\}$, where the induction on i runs backward. The induction hypothesis is given as follows: There exists $\tilde{n}_{IH} \in \mathbb{Z}_+$ such that for all $n \geq \tilde{n}_{IH}$, we have

$$\mathbb{P}\left(\bar{s}_i \geq n - 2m \frac{n \log d}{d^{m+i-1}} + B_i + 2(m-i)m \frac{n \log d}{d^{m-i+2}} + \sqrt{mn} \log n\right) \leq \left(\frac{1}{n}\right)^{(m \log n)/8 - (m-i)}. \quad (\text{IH2})$$

The base case is to prove an upper bound on \bar{s}_m that corresponds to Step 1d in Fig. 8. This is done in the lemma below.

LEMMA 9 (Step 1d). *There exists $\tilde{n}_4 \in \mathbb{Z}_+$ such that for all $n \geq \tilde{n}_4$, we have*

$$\mathbb{P}\left(\bar{s}_m \geq n - 2m \frac{n \log d}{d} + B_m\right) \leq \left(\frac{1}{n}\right)^{(m \log n)/8}.$$

Further, we prove the induction step, corresponding to Steps 2 to m in Fig. 8.

LEMMA 10 (Steps 2 to m). *Assume that (IH2) holds true for $i+1$. Then, (IH2) holds true for i .*

Proof Sketch of Lemma 10 Consider the following family of functions:

$$W_i(\mathbf{s}) = s_i - n + \frac{2mn \log d}{d^{m-i+1}} + o\left(\frac{2mn \log d}{d^{m-i+1}}\right) \quad \forall i \in [m].$$

We consider the induction hypothesis $W_{i+1}(\bar{\mathbf{s}}) = O(\sqrt{mn} \log n)$ w.h.p. and show that the drift of $W_i(\mathbf{s})$ is negative when $W_{i+1}(\mathbf{s}) = O(\sqrt{mn} \log n)$, $s_{i-1} \geq n - \Theta(2mn \log d / d^{m-i+2})$, and $W_i(\mathbf{s}) \geq 0$. Now, by applying Lemma 14 with $V = W_{i+1}$ and $\mathcal{E} = \{W_{i+1}(\mathbf{s}) = O(\sqrt{mn} \log n)\} \cap \{s_{i-1} \geq n - \Theta(2mn \log d / d^{m-i+2})\}$ and noting that $\mathbb{P}(\bar{\mathbf{s}} \in \mathcal{E}) \approx 1$ by the induction hypothesis and Theorem 2, we obtain that $W_i(\bar{\mathbf{s}}) = O(\sqrt{mn} \log n)$ w.h.p. This provides the required bound on \bar{s}_i . \square

We conclude this section by presenting the proof of Theorem 3 using the results outlined above.

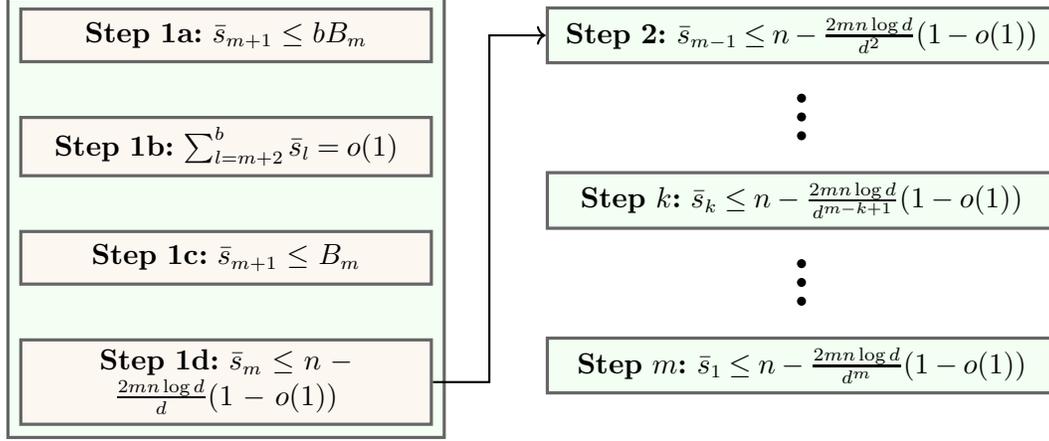
Proof of Theorem 3 Consider the induction hypothesis defined in (IH2). By setting $\tilde{n}_{IH} \geq \tilde{n}_4$, the base case is complete by Lemma 9. In addition, the induction step is complete by Lemma 10. Thus, for all $n \geq \tilde{n}_{IH}$ and $i \in [m]$,

$$\begin{aligned} & \mathbb{P}\left(\bar{s}_i \geq n - 2m \frac{n \log d}{d^{m-i+1}} + 19md^{i-1} \sqrt{mn} \log n + 39m^3 \frac{n \log(d)^2}{d^{m-i+2}} + \frac{n^{1-\gamma}}{d^{m-i}} \mathbb{1}\{m > 1\}\right) \\ & \leq \mathbb{P}\left(\bar{s}_i \geq n - 2m \frac{n \log d}{d^{m+i-1}} + B_i + 2(m-i)m \frac{n \log d}{d^{m-i+2}} + \sqrt{mn} \log n\right) \\ & \leq \left(\frac{1}{n}\right)^{(m \log n)/8 - (m-i)} \leq \left(\frac{1}{n}\right)^{(m \log n)/9}, \end{aligned}$$

where the last inequality follows for all $n \geq \tilde{n}_6$ for some $\tilde{n}_6 \in \mathbb{Z}_+$ as $m = o(\log n)$. Thus, by setting $n_{UB} \triangleq \max_{k \in \{2,3,6\}} \{\tilde{n}_k, \tilde{n}_{IH}\}$, and using Lemma 7 and 8, completes the proof of Theorem 3. \square

6. Simulations

In this section, we simulate the load balancing model for various values of n and d . To avoid confusion, all the simulated variables are denoted with a dagger in the super-script.

Figure 8 Outline of the proof to establish high probability upper bounds, where B_m is defined in (9)

6.1. Setup

Load balancing under power-of- d choices routing algorithm is governed by a continuous time Markov chain (CTMC) with transition rates $q_{\mathbf{s}^1, \mathbf{s}^2}$ given by

$$q_{\mathbf{s}^1, \mathbf{s}^2} = \begin{cases} \lambda \left(\left(\frac{s_i^1}{n} \right)^d - \left(\frac{s_{i+1}^1}{n} \right)^d \right) & \text{for } \mathbf{s}^2 = \mathbf{s}^1 + \mathbf{e}_{i+1} \quad \forall i \in [b-1] \\ s_i^1 - s_{i+1}^1 & \text{for } \mathbf{s}^2 = \mathbf{s}^1 - \mathbf{e}_i \quad \forall i \in [b] \\ 0 & \text{otherwise.} \end{cases}$$

We fix the initial state $\mathbf{s}(0) = \mathbf{0}_b$, buffer size $b = 10$, arrival rate $\lambda = n - n^{0.6}$ and carry out the simulation for various values of n and d . For each tuple, (n, d) , the CTMC is simulated until it approximately reaches the steady state. This is verified by plotting the evolution of $\mathbf{s}^\dagger(t)$ as a function of time, as in Fig. 9. The mean estimate, i.e. $\mathbb{E}[\bar{\mathbf{s}}^\dagger]$ is calculated by considering only the last 40% of the data to ensure that a steady state has already been reached. In addition, to ensure repeatability of results, this procedure is repeated 5 times for each (n, d) tuple, and the mean of the 5 estimates is reported in Table 2.

The goal of this section is to compare the theoretical bounds with the simulated steady state expectations ($\mathbb{E}[\bar{s}_i]$). We massage the bounds given by Theorem 2 and Theorem 3 to hypothesize the following expression for \bar{s}_i for a finite n : define $m^* = \max\{i : n - n^{1-\gamma}d^{i-1} > 0\}$ and consider

$$\frac{\bar{s}_i}{n} \approx \begin{cases} \left(\frac{\lambda}{n} \right)^{\frac{d^i-1}{d-1}} + \frac{d^{i-1}}{\sqrt{n}} & \forall i \leq m^* \\ 0 & \forall i \geq m^* + 2. \end{cases} \quad \text{w.h.p.} \quad (10)$$

The advantage of (10) is that it is defined for all $d \in \mathbb{Z}_+$ as opposed to the result in Theorem 1 that requires $d \approx n^{\gamma/m}$ for $m \in \mathbb{Z}_+$. Note that, m^* can be viewed as a generalization of m as $m^* = m$ for $d \approx n^{\gamma/m}$ and m^* is defined for all $d \in \mathbb{Z}_+$. Equation 10 is inspired from Theorem 2 and 3 and it is consistent with it except for $i = m^* + 1$, in which case the upper bound in Theorem 3 turns out to be loose for the simulated pre-limit system. Thus, we omit it from (10) and focus on $i \neq m^* + 1$. For the case of $i \geq m^* + 2$, by Theorem 3, we have $\bar{s}_i \leq 1$ w.h.p. Thus, $\bar{s}_i/n \approx 0$ w.h.p. For the case of $i \leq m^*$, consider the bounds given by Theorem 2 and Theorem 3. By dropping the constants and logarithmic terms in the bounds, we get

$$\frac{\bar{s}_i}{n} \asymp \underbrace{1 - \frac{2m \log d}{d^{m+1-i}}}_{\left(\frac{\lambda}{n} \right)^{\frac{d^i-1}{d-1}}} \pm \underbrace{\frac{1}{d^{m+2-i}}}_{\text{Dropped in (10)}} \pm \underbrace{\frac{d^{i-1}}{\sqrt{n}}}_{\text{Preserved in (10)}}.$$

By Taylor's series expansion, observe that the sum of the first two terms in the above equation is approximately equal to the ODE fixed point as in (3), i.e. $1 - (2m \log d)/d^{m+1-i} \approx (\lambda/n)^{(d^i-1)/(d-1)}$. Next, we believe that the lower order term $\frac{\log n}{d^{m+2-i}}$ is an artefact of our proof but it can be improved by further bootstrapping the result of Theorem 2, so we drop it in (10). As it is already $o(\frac{\log n}{d^{m+1-i}})$, we do not attempt to improve it in the theoretical analysis. Lastly, the lower order term d^{i-1}/\sqrt{n} is also an artefact of our proof but cannot be improved using our proof technique, and so, we keep it in (10). Note that d^{i-1}/\sqrt{n} has a non-negligible contribution in (10) for large i . For example, when $i = m$ and $\gamma = 0.4$, we have $d^{m-1}/\sqrt{n} \geq 0.05$ for all $n \leq 20^{10}$ as $d^{m-1}/\sqrt{n} \sim n^{-0.5+\gamma-\gamma/m} \approx n^{-0.1}$.

Observe that the error from the fixed point, i.e., $\frac{\bar{s}_i}{n} - \left(\frac{\lambda}{n}\right)^{\frac{d^i-1}{d-1}}$, is a different order of n for different indices $i \in [m^*]$ in (10). We observe such a qualitative behavior in the simulations as well and outline the results in Appendix A.

6.2. Results

We simulate the system for $n = (10^2, 10^3, 10^4, 10^5)$ and set d to be the greatest integer such that $d \leq (2mn^\gamma \log d)^{1/m}$ for $m = (1, 2, 3, 4, 5)$. We summarize the results in Table 2. In particular, we

Table 2 Simulated mean, error from (10), and $\max\{i : \max_{t \in \mathbb{Z}_+} s_i^\dagger(t) > 0\}$ for different values of n and m . Errors greater than 0.05 are boldfaced and highlighted in orange. Cell corresponding to s_i^\dagger such that $\max_{t \in \mathbb{Z}_+} s_i^\dagger(t) = 0$ are highlighted in blue.

m	n	d	Simulated mean: $\mathbb{E}[\bar{s}_i^\dagger]/n$					Error from (10): $\mathbb{E}[\bar{s}_i^\dagger - \bar{s}_i]/n$					$i : s_{\geq i} = 0$
			$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	
1	10^2	49	0.92	0.10	0.00	0.00	0.00	-0.02	—	0.00	0.00	0.00	3
2	10^2	7	0.92	0.52	0.02	0.00	0.00	-0.02	—	0.02	0.00	0.00	4
3	10^2	3	0.91	0.70	0.32	0.04	0.00	-0.03	-0.10	—	0.04	0.00	6
4	10^2	2	0.92	0.76	0.54	0.26	0.07	-0.03	-0.03	-0.16	—	0.07	8
1	10^3	161	0.97	0.04	0.00	0.00	0.00	0.00	—	0.00	0.00	0.00	3
2	10^3	12	0.97	0.66	0.01	0.00	0.00	0.00	-0.15	—	0.00	0.00	4
3	10^3	5	0.97	0.83	0.38	0.01	0.00	-0.01	-0.02	—	0.01	0.00	5
4	10^3	3	0.97	0.88	0.66	0.28	0.02	0.00	0.02	-0.05	—	0.02	7
5	10^3	2	0.97	0.91	0.79	0.61	0.36	0.00	0.03	0.03	0.02	—	9
1	10^4	493	0.99	0.01	0.00	0.00	0.00	0.00	—	0.00	0.00	0.00	3
2	10^4	22	0.99	0.75	0.00	0.00	0.00	0.00	-0.03	—	0.00	0.00	4
3	10^4	7	0.99	0.90	0.48	0.01	0.00	0.00	0.02	—	0.01	0.00	5
4	10^4	4	0.99	0.94	0.77	0.35	0.02	0.00	0.02	0.02	—	0.02	6
5	10^4	3	0.99	0.95	0.84	0.59	0.21	0.00	0.02	0.04	-0.04	—	8
1	10^5	1456	1.00	0.00	0.00	0.00	0.00	0.00	—	0.00	0.00	0.00	3
2	10^5	38	1.00	0.82	0.00	0.00	0.00	0.00	0.03	—	0.00	0.00	4
3	10^5	11	0.99	0.94	0.51	0.00	0.00	0.00	0.02	—	0.00	0.00	5
4	10^5	6	1.00	0.97	0.81	0.28	0.00	0.00	0.01	0.04	—	0.00	6
5	10^5	4	0.99	0.98	0.90	0.65	0.18	0.00	0.01	0.04	0.03	—	7

document the simulated mean $\{\mathbb{E}[\bar{s}_i^\dagger]/n\}_{i=1}^5$, the error from (10), i.e. $\{\mathbb{E}[\bar{s}_i^\dagger - \bar{s}_i]/n\}_{i=1}^5$, and the maximum queue lengths that were observed in the simulation, i.e. $\min\{i : s_i^\dagger(t) = 0 \forall t \in \mathbb{Z}_+\}$. Furthermore, to understand the transient behavior, we plot the evolution of $s^\dagger(t)$ with time for $n = 10^5$ and $m = (2, 4, 5)$ in Fig. 9 and 10. We now summarize the takeaways from Table 2, Fig. 9, and Fig. 10.

- Equation (10) closely approximates the stationary distribution even for n as small as 10^4 . Thus, one can directly employ (10) in practice to obtain close approximations of the stationary distribution for $n \geq 10^4$ and any $d \in \mathbb{Z}_+$.

Figure 9 Evolution of the load balancing CTMC with $n = 10^5$ and $d = 4$.

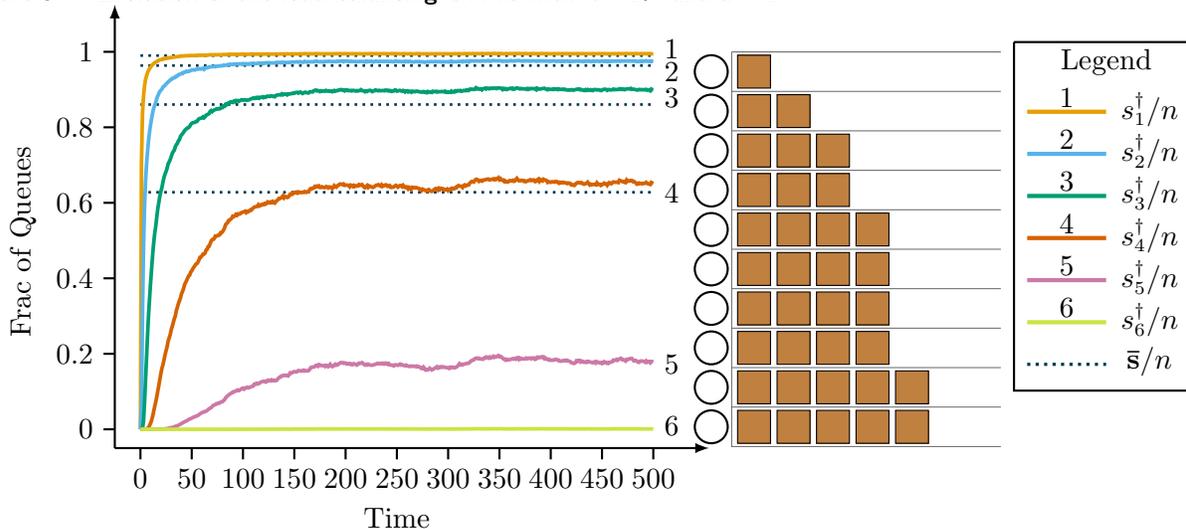
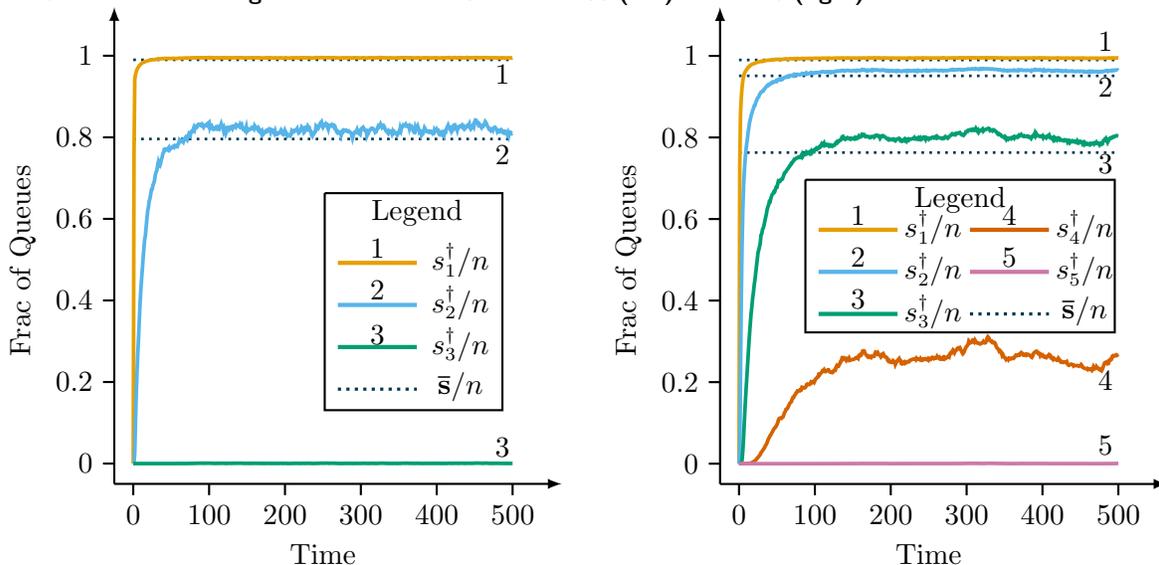


Figure 10 Load balancing CTMC with $n = 10^5$ and $d = 38$ (left) and $d = 6$ (right).



- No incoming customers are rejected due to a finite waiting space for $b \geq 8$. This suggests that the assumption $b = O(\log(n)^3)$ is not fundamental to the model. It is merely an artefact of the proof.
- Under steady state, $\mathbf{s}^\dagger(t)$ stays close to $\bar{\mathbf{s}}$ given by (10) for all $t \in \mathbb{R}_+$. In particular, $\mathbf{s}^\dagger(t) \approx \bar{\mathbf{s}}$ w.h.p.

7. Conclusion and Future Work

In this paper, we characterized the performance of Power-of- d choices routing algorithm for the sub-Halfin-Whitt regime in the load balancing model. We showed that if d grows polynomially with n , then the jobs experience a finite delay. On the other hand, if d grows only as Poly-Log(n), then the jobs experience infinite asymptotic delay. In particular, we characterized the delay for Power-of- d with $d \in [\Omega(\log(n)^3), n]$ and $\gamma \in (0, 0.5)$. Future work is to similarly understand the performance of Power-of- d choices for other many-server-heavy-traffic regimes.

8. Acknowledgement

We thank Dr. Debankur Mukherjee for insightful discussions that helped in proving the result. In addition, the illustration of the fixed point, as shown in Fig. 3 is inspired from a figure in [27]. Illustration of the asymptotic regime as a 2D graph in Fig. 1 is inspired by Dr. Lei Ying.

References

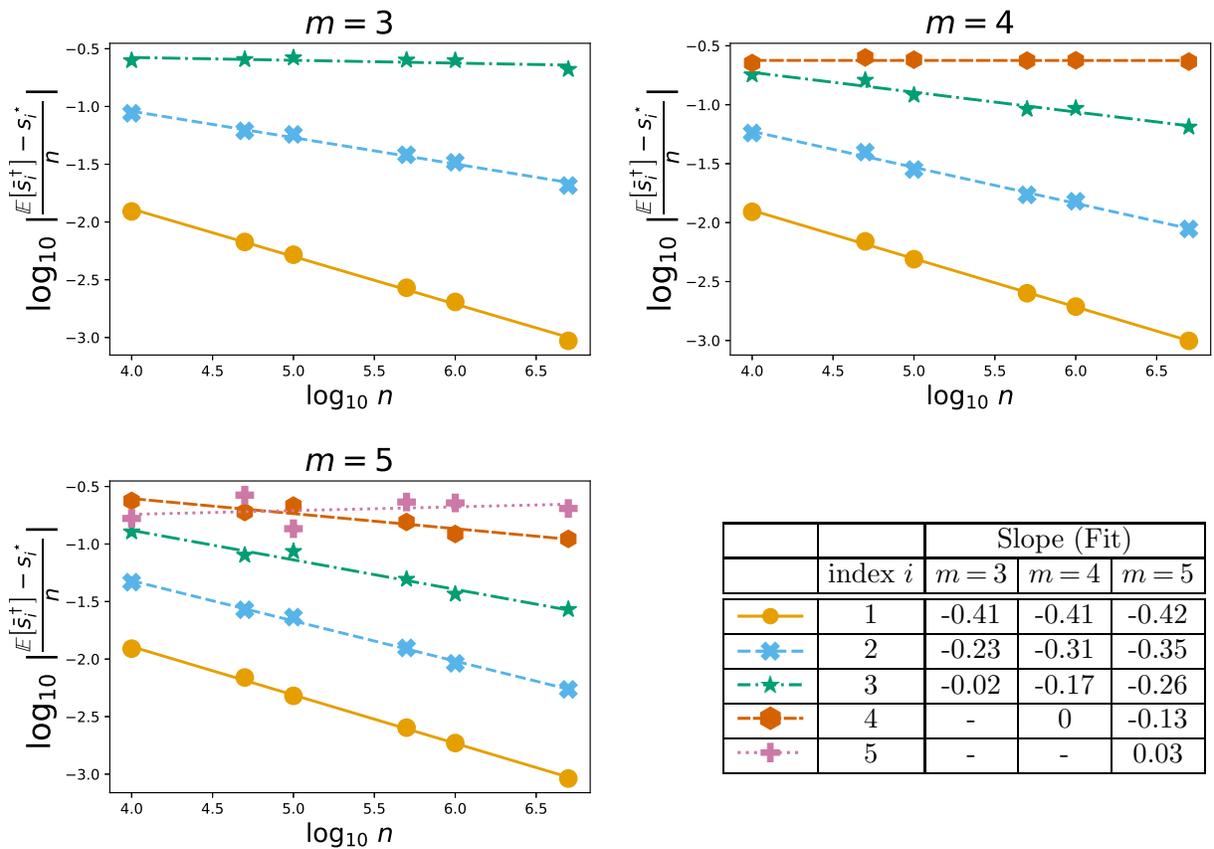
- [1] Banerjee S, Mukherjee D (2019) Join-the-shortest queue diffusion limit in halfin–whitt regime: Tail asymptotics and scaling of extrema. *The Annals of Applied Probability* 29(2):1262–1309.
- [2] Banerjee S, Mukherjee D (2020) Join-the-shortest queue diffusion limit in halfin–whitt regime: Sensitivity on the heavy-traffic parameter. *The Annals of Applied Probability* 30(1):80–144.
- [3] Bertsimas D, Gamarnik D, Tsitsiklis JN (2001) Performance of multiclass markovian queueing networks via piecewise linear lyapunov functions. *Annals of Applied Probability* 11(4):1384–1428.
- [4] Bhamidi S, Budhiraja A, Dewaskar M (2022) Near equilibrium fluctuations for supermarket models with growing choices. *The Annals of Applied Probability* 32(3):2083–2138.
- [5] Borkar VS (2009) *Stochastic approximation: a dynamical systems viewpoint*, volume 48 (Springer).
- [6] Braverman A (2020) Steady-state analysis of the join-the-shortest-queue model in the halfin–whitt regime. *Mathematics of Operations Research* 45(3):1069–1103.
- [7] Brightwell G, Fairthorne M, Luczak MJ (2018) The supermarket model with bounded queue lengths in equilibrium. *Journal of Statistical Physics* 173:1149–1194.
- [8] Brightwell G, Luczak M (2012) The supermarket model with arrival rate tending to one. arXiv:1201.5523.
- [9] der Boor MV, Borst SC, Van Leeuwen JS, Mukherjee D (2022) Scalable load balancing in networked systems: A survey of recent advances. *SIAM Review* 64(3):554–622.
- [10] Eryilmaz A, Srikant R (2012) Asymptotically tight steady-state queue length bounds implied by drift conditions. *Queueing Systems* 72(3):311–359.
- [11] Eschenfeldt P, Gamarnik D (2016) Supermarket queueing system in the heavy traffic regime. short queue dynamics.
- [12] Eschenfeldt P, Gamarnik D (2018) Join the shortest queue with many servers. the heavy-traffic asymptotics. *Mathematics of Operations Research* 43(3):867–886.
- [13] Foschini G, Salz J (1978) A basic dynamic routing problem and diffusion. *IEEE Transactions on Communications* 26(3):320–327.
- [14] Gupta V, Walton N (2019) Load balancing in the nondegenerate slowdown regime. *Operations Research* 67(1):281–294.
- [15] Halfin S, Whitt W (1981) Heavy-traffic limits for queues with many exponential servers. *Operations research* 29(3):567–588.
- [16] Haque SU, Khodadadian S, Maguluri ST (2023) Tight finite time bounds of two-time-scale linear stochastic approximation with markovian noise. *arXiv preprint arXiv:2401.00364* .
- [17] Hurtado-Lange D, Maguluri ST (2020) Load balancing system under join the shortest queue: Many-server-heavy-traffic asymptotics.
- [18] Hurtado-Lange D, Maguluri ST (2020) Transform methods for heavy-traffic analysis. *Stochastic Systems* 10(4):275–309.
- [19] Jonckheere M, Prabhu BJ (2018) Asymptotics of insensitive load balancing and blocking phases. *Queueing Systems* 88(3):243–278.
- [20] Liu X, Gong K, Ying L (2022) Large-system insensitivity of zero-waiting load balancing algorithms. *Abstract Proceedings of the 2022 ACM SIGMETRICS/IFIP PERFORMANCE Joint International Conference on Measurement and Modeling of Computer Systems*, 101–102, SIGMETRICS/PERFORMANCE '22 (New York, NY, USA: Association for Computing Machinery), ISBN 9781450391412, URL <http://dx.doi.org/10.1145/3489048.3526955>.
- [21] Liu X, Gong K, Ying L (2022) Steady-state analysis of load balancing with coxian-2 distributed service times. *Naval Research Logistics (NRL)* 69(1):57–75.
- [22] Liu X, Ying L (2020) Steady-state analysis of load-balancing algorithms in the sub-halfin–whitt regime. *Journal of Applied Probability* 57(2):578–596, URL <http://dx.doi.org/10.1017/jpr.2020.13>.

- [23] Liu X, Ying L (2022) Universal scaling of distributed queues under load balancing in the super-halfin-whitt regime. *IEEE/ACM Transactions on Networking* 30(1):190–201, URL <http://dx.doi.org/10.1109/TNET.2021.3105480>.
- [24] Maguluri ST, Srikant R, Ying L (2014) Heavy traffic optimal resource allocation algorithms for cloud computing clusters. *Performance Evaluation* 81:20–39.
- [25] Mitzenmacher M (1996) Load balancing and density dependent jump markov processes. *Proceedings of 37th Conference on Foundations of Computer Science*, 213–222 (New York, NY, USA: IEEE), URL <http://dx.doi.org/10.1109/SFCS.1996.548480>.
- [26] Mitzenmacher M (2001) The power of two choices in randomized load balancing. *IEEE Transactions on Parallel and Distributed Systems* 12(10):1094–1104.
- [27] Mukherjee D, Borst SC, Van Leeuwen JS, Whiting PA (2018) Universality of power-of-d load balancing in many-server systems. *Stochastic Systems* 8(4):265–292.
- [28] Vvedenskaya ND, Dobrushin RL, Karpelevich FI (1996) Queueing system with selection of the shortest of two queues: An asymptotic approach. *Problemy Peredachi Informatsii* 32(1):20–34.
- [29] Weng W, Wang W (2020) Achieving zero asymptotic queueing delay for parallel jobs. *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 4(3):1–36.
- [30] Ying L (2017) Stein’s method for mean field approximations in light and heavy traffic regimes. *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 1(1):1–27.
- [31] Zhao Z, Banerjee S, Mukherjee D (2021) Many-server asymptotics for join-the-shortest queue in the super-halfin-whitt scaling window.

Appendix A: Error from the Fixed Point

In Theorem 1, we show that the stochastic system concentrates around the fixed point of the corresponding dynamical system for large values of n . In this section, we want to understand how the error of the pre-limit system from this fixed point scales with n . We proceed as follows: fix a value of $m \in \{3, 4, 5\}$ and set d to be the smallest integer such that $n - n^{1-\gamma}d^{m-1} \leq 0$, motivated by our intuition for the choice of d given by (3). We then simulate this system for $n \in \{10^4, 5 \cdot 10^4, 10^5, 5 \cdot 10^5, 10^6, 5 \cdot 10^6\}$ and report the results in Fig. 11.

Figure 11 Log-log plot of the error of the simulated mean $\mathbb{E}[\bar{s}_i^\dagger]$ from the fixed point (s_i^*) of the ODE (2) for $m = 3$ (top left), $m = 4$ (top right), and $m = 5$ (bottom left).



Observe that the reported values of the slope in Fig. 11 is an increasing function of i , and so, the error $(s_i - s_i^*)$ scales differently as a function of n for $i \in [m]$. Such a qualitative behavior is consistent with our prescription for the pre-limit s_i in (10). In particular, the error from the fixed point in (10) is $d^{i-1}/\sqrt{n} \sim n^{-0.5+(i-1)\gamma/m}$, which is qualitatively consistent with the simulations as the exponent is an increasing function of i . However, the numerical values of the slope reported in Fig. 11 does not closely match with the prescription $-0.5 + (i - 1)\gamma/m$ as in (10). Nonetheless, the error of s_i compared to the fixed points is non-negligible (Fig. 11) for large values of i and reduces significantly (Table 2) by adding the correction term d^{i-1}/\sqrt{n} in (10) inspired by our theoretical results.

Note that, [4] shows that the scaling of the error is $o(1/\sqrt{n})$ for all $i \in [m - 1]$, which appears to be inconsistent with our simulation results. While we considered a realistic and feasible range of the values of n , the error might eventually scale as $o(1/\sqrt{n})$ as the value of n increases further. We considered the range of values of n such that the run time of the simulation is at most one day.

Appendix B: Preliminary Lemmas

B.1. Taylor Series Based Inequalities

In this section, we present a few inequalities based on Taylor's series expansion which will be useful later to bound some of the terms.

LEMMA 11. *Let $f(d)$ be such that $df(d) \rightarrow 0$ as $d \rightarrow \infty$. Then, there exists d_0 such that for all $d \geq d_0$, we have*

$$\left(1 - r \frac{\log d}{d} + f(d)\right)^d \leq \frac{2}{d^r}.$$

LEMMA 12. *Let $f(d)$ be such that $df(d) \rightarrow 0$ as $d \rightarrow \infty$ and $f(d) \geq 0$ for $d \geq d_1$ for some $d_1 \in \mathbb{R}_+$. Then, there exists $d_2 \in \mathbb{R}_+$ such that for all $d \geq d_2$, we have*

$$1 - df(d) \leq (1 - f(d))^d \leq 1 - df(d) + \frac{1}{2}d^2 f(d)^2.$$

LEMMA 13. *There exists $n_a \in \mathbb{Z}_+$ such that for all $n \geq n_a$, we have*

$$\left(\frac{n}{n + \sqrt{mn} \log n}\right)^{\sqrt{mn}(\log n)/2} \leq \frac{1}{n^{(m \log n)/4}}.$$

For the rest of the appendix, we consider n to be large enough such that all the results in this section holds true.

B.2. Iterative State Space Collapse

We start by formally defining the drift of a Lyapunov function. Let $q_{\mathbf{s}, \mathbf{s}'}$ be the rate at which the CTMC transitions from \mathbf{s} to \mathbf{s}' .

DEFINITION 1. Consider a Lyapunov function $V : \mathcal{S} \rightarrow \mathbb{R}$ and define the drift of V at state \mathbf{s} as

$$\Delta V(\mathbf{s}) = \sum_{\mathbf{s}' \in \mathcal{S}, \mathbf{s}' \neq \mathbf{s}} q_{\mathbf{s}, \mathbf{s}'} (V(\mathbf{s}') - V(\mathbf{s})).$$

State Space Collapse: If the drift of $V(\cdot)$ is such that $\Delta V(\mathbf{s}) \leq -\gamma$ when $V(\mathbf{s}) \geq B$ for some $\gamma, B \in \mathbb{R}_+$, then, one can obtain high probability tail bounds on $V(\bar{\mathbf{s}})$ [3] that depends on B, γ and the properties of the CTMC. Intuitively, every time the CTMC jumps to a state \mathbf{s} such that $V(\mathbf{s}) \geq B$, due to a strong drift ($\Delta V(\mathbf{s}) \leq -\gamma$), the CTMC will quickly jump back to a state such that $V(\mathbf{s}) \leq B$. Loosely speaking, the stationary probability $\mathbb{P}(V(\bar{\mathbf{s}}) \geq B + j)$ decreases exponentially in terms of j . This implies that the underlying CTMC $\{\mathbf{s}(t) : t \geq 0\}$ ‘‘collapses’’ to a subset of the state space $\{\mathbf{s} \in \mathcal{S} : V(\mathbf{s}) \leq B + j\}$ for a large enough $j \in \mathbb{R}_+$.

Iterative State Space Collapse: The main challenge in obtaining SSC is to show that the Lyapunov function $V(\cdot)$ exhibits negative drift when $V(\mathbf{s}) \geq B$. The authors in [21] ingeniously showed that if $\mathcal{E} \subseteq \mathcal{S}$ such that $\mathbb{P}(\bar{\mathbf{s}} \notin \mathcal{E}) \approx 0$, then, it suffices to show negative drift for $\mathbf{s} \in \mathcal{E} : V(\mathbf{s}) \geq B$. Thus, by exploiting the properties of the stationary distribution of $\{\mathbf{s}(t) : t \geq 0\}$, one needs to show negative drift for only a subset of the state space. We state this as a lemma below.

LEMMA 14. *Consider a Lyapunov function $V : \mathcal{S} \rightarrow \mathbb{R}$ such that $V(\mathbf{s}) \geq D$ for all $\mathbf{s} \in \mathcal{S}$ for some $D \in \mathbb{R}$. Consider*

$$\nu_{\max} := \max_{\mathbf{s}, \mathbf{s}' \in \mathcal{S}, q_{\mathbf{s}, \mathbf{s}'} > 0} |V(\mathbf{s}') - V(\mathbf{s})| < \infty,$$

and define

$$q_{\max} := \max_{\mathbf{s} \in \mathcal{S}} \sum_{\mathbf{s}' \in \mathcal{S} : V(\mathbf{s}) < V(\mathbf{s}')} q_{\mathbf{s}, \mathbf{s}'}.$$

Assume that there exists a set \mathcal{E} with $B > 0$, $\gamma > 0$, $\delta \geq 0$ such that the following conditions are satisfied.

- $\Delta V(\mathbf{s}) \leq -\gamma$ when $V(\mathbf{s}) \geq B$ and $\mathbf{s} \in \mathcal{E}$,
- $\Delta V(\mathbf{s}) \leq \delta$ when $V(\mathbf{s}) \geq B$ and $\mathbf{s} \notin \mathcal{E}$.

Then,

$$\mathbb{P}(V(\bar{\mathbf{s}}) \geq B + 2\nu_{\max}j) \leq \alpha^j + \beta \mathbb{P}(\bar{\mathbf{s}} \notin \mathcal{E}) \quad \forall j \in \mathbb{Z}_+, \quad (11)$$

with

$$\alpha = \frac{q_{\max} \nu_{\max}}{q_{\max} \nu_{\max} + \gamma} \quad \text{and} \quad \beta = \frac{\delta}{\gamma} + 1.$$

The above lemma is obtained by directly using [21, Lemma 10] with $V' = V - D$ as [21, Lemma 10] requires the Lyapunov function to be non-negative. It'll be helpful to note that proof of almost all the following lemmas follows a four-part template.

1. Define a Lyapunov function $V(\mathbf{s})$ depending on what is required to be proved.
2. Show that $\Delta V(\mathbf{s}) \leq -\sqrt{mn} \log n$ for all $\mathbf{s} \in \mathcal{E}$ such that $V(\mathbf{s}) \geq 0$, where $\mathbb{P}(\bar{\mathbf{s}} \in \mathcal{E}) \approx 1$. Lemma 11 and Lemma 12 will be useful in completing this step.
3. Apply Lemma 14 to obtain a high probability bound on $V(\bar{\mathbf{s}})$. In all the instances of application of Lemma 14, we use $B = 0$, $j = \sqrt{mn}(\log n)/2$, $\nu_{\max} = 1$, $q_{\max} = \delta = n$, and $\gamma = \sqrt{mn} \log n$. This also implies that $\beta \leq \sqrt{n}$ for large enough n . Also note that, Lemma 13 will be useful in simplifying the expression obtained. In particular, we have

$$\begin{aligned} \mathbb{P}(V(\bar{\mathbf{s}}) \geq \sqrt{mn} \log n) &\leq \left(\frac{n}{n + \sqrt{mn} \log n} \right)^{\sqrt{mn}(\log n)/2} + \sqrt{n} \mathbb{P}(\bar{\mathbf{s}} \notin \mathcal{E}) \\ &\leq \left(\frac{1}{n} \right)^{(m \log n)/4} + \sqrt{n} \mathbb{P}(\bar{\mathbf{s}} \notin \mathcal{E}). \end{aligned}$$

4. Translate the high probability bound on $V(\bar{\mathbf{s}})$ to the required bound by employing basic results in probability like union bound, law of total probability, etc.

Appendix C: Proof of Lemmas for Lower Bound

Proof of Lemma 1 (Base Case) To prove the lemma, we consider the following Lyapunov function:

$$V_1(\mathbf{s}) = n - \frac{2mn \log d}{d} - \frac{2n \log d}{d^2} - \sqrt{mn} \log n - s_1.$$

Now, we analyze the drift of $V_1(\mathbf{s})$ when $V_1(\mathbf{s}) \geq 0$. Thus, we have $s_1 \leq n - \frac{2mn \log d}{d} - \frac{2n \log d}{d^2} - \sqrt{mn} \log n$. Now, the drift is given as follows:

$$\begin{aligned} \Delta V_1(\mathbf{s}) &= s_1 - s_2 - \lambda \left(1 - \left(\frac{s_1}{n} \right)^d \right) \\ &\leq n - \frac{2mn \log d}{d} - \frac{2n \log d}{d^2} - \sqrt{mn} \log n - \lambda \left(1 - \left(1 - \frac{2m \log d}{d} \right)^d \right) \\ &\stackrel{(a)}{\leq} n - \frac{2mn \log d}{d} - \frac{2n \log d}{d^2} - \sqrt{mn} \log n - \lambda \left(1 - \frac{2}{d^{2m}} \right) \\ &\leq -\frac{2mn \log d}{d} - \frac{2n \log d}{d^2} - \sqrt{mn} \log n + \frac{2n}{d^{2m}} + n^{1-\gamma} \\ &\stackrel{(b)}{\leq} -\sqrt{mn} \log n, \end{aligned}$$

where (a) follows for all $n \geq n_0$ for some $n_0 \in \mathbb{Z}_+$ by Lemma 11. Next, (b) follows by the following observations: $\frac{2mn \log d}{d} \geq \frac{2mn \log d}{d^m} = n^{1-\gamma}$, and $\frac{2n}{d^{2m}} \leq \frac{2n}{d^2}$. Thus, we have $V_1(\mathbf{s}) \leq -\sqrt{mn} \log n$ for all $n \geq n_0$. By Lemma 14, we get

$$\mathbb{P}(V_1(\bar{\mathbf{s}}) \geq \sqrt{mn} \log n) \leq \left(\frac{n}{n + \sqrt{mn} \log n} \right)^{\sqrt{mn}(\log n)/2} \leq \frac{1}{n^{(m \log n)/4}},$$

where the last inequality follows by Lemma 13. This completes the proof. \square

Proof of Lemma 2 Define a family of functions $\{V_{ik}(\mathbf{s}) : i \in [k]\}$ for $k \geq 2$ as follows:

$$\begin{aligned} V_{ik}(\mathbf{s}) &= s_i - n + 3im \frac{n \log d}{d^{k-i+1}} \quad \forall i \in [k-1] \\ V_{kk}(\mathbf{s}) &= n - 3km \frac{n \log d}{d} - s_k. \end{aligned}$$

Now, using the above family of functions, we define the Lyapunov functions $\{L_{lk} : l \in [k-1]\}$ as follows:

$$\begin{aligned} L_{lk}^{(1)} &= V_{kk} - \sum_{j=l+1}^{k-1} V_{jk}, \quad L_{lk}^{(2)} = V_{lk} \quad \forall l \in [k-1] \\ L_{lk} &= \min_{j=l+1} \left\{ L_{lk}^{(1)}, L_{lk}^{(2)} \right\} \quad \forall l \in [k-1]. \end{aligned}$$

To prove the lemma, we make use of the following claim:

CLAIM 1. *There exists $n_{c1} \in \mathbb{Z}_+$ such that for all $n \geq n_{c1}$, we have*

$$\mathbb{P}(L_{1k}(\bar{\mathbf{s}}) \geq \sqrt{mn} \log n) \leq \left(\frac{1}{n}\right)^{(m \log n)/4 - 4(k-1)m - (k-1)}.$$

We defer the proof of the claim to Appendix E and continue with the proof of Lemma 2. We analyze the drift of $L_{0k}^{(1)}(\mathbf{s})$ resulting in a high probability upper bound on $L_{0k}^{(1)}(\mathbf{s})$ which will imply a high probability lower bound on \bar{s}_k . We start by analyzing the drift of $L_{0k}^{(1)}(\mathbf{s})$ when $L_{0k}^{(1)}(\mathbf{s}) \geq 0$ and $\mathbf{s} \in \mathcal{C}_{1,k}^{(1)} \cap \bigcap_{l=1}^{k-1} \mathcal{D}_l^{(1)}$ where

$$\mathcal{C}_{l,k}^{(1)} = \{L_{lk} \leq \sqrt{mn} \log n\} \quad \forall l \in [k-1] \quad (12a)$$

$$\mathcal{D}_l^{(1)} = \left\{ s_l \geq n - \frac{5mn \log d}{2d^{k-l}} \right\} \quad \forall l \in [k-1]. \quad (12b)$$

First, by using that $\mathbf{s} \in \mathcal{C}_{1,k}^{(1)}$, we get a useful upper bound on s_1 as follows:

$$\begin{aligned} L_{0k}^{(1)}(\mathbf{s}) \geq 0 &\Rightarrow V_{kk}(\mathbf{s}) - \sum_{j=2}^k V_{jk}(\mathbf{s}) \geq V_{1k}(\mathbf{s}) \\ &\Rightarrow V_{1k}(\mathbf{s}) \leq \sqrt{mn} \log n \quad \text{as } \mathbf{s} \in \mathcal{C}_{1,k}^{(1)} \\ &\Rightarrow s_1 \leq n - 3m \frac{n \log d}{d^k} + \sqrt{mn} \log n \leq n - 5m \frac{n \log d}{2d^k}, \end{aligned} \quad (13)$$

where the last inequality holds because $m \frac{n \log d}{2d^k} \geq m \frac{n \log d}{2d^m} = n^{1-\gamma}/4 \geq \sqrt{mn} \log n$ for all $n \geq n_1^{(1)}$ for some $n_1^{(1)} \in \mathbb{Z}_+$ independent of k . Next, we get a useful upper bound on s_k as follows:

$$\begin{aligned} L_{0k}^{(1)}(\mathbf{s}) \geq 0 &\Rightarrow V_{kk}(\mathbf{s}) - \sum_{j=1}^{k-1} V_{jk}(\mathbf{s}) \geq 0 \\ &\Rightarrow s_k \leq n - 3km \frac{n \log d}{d} - \sum_{j=1}^{k-1} V_{jk}(\mathbf{s}) \\ &\Rightarrow s_k \leq n - 3km \frac{n \log d}{d} + 2.5mn \log d \sum_{l=1}^{k-1} \frac{1}{d^l} \quad \text{as } \mathbf{s} \in \bigcap_{l=1}^{k-1} \mathcal{D}_l^{(1)} \\ &\Rightarrow s_k \leq n - 3m \frac{n \log d}{d}, \end{aligned} \quad (14)$$

where the last assertion follows by using the bound $1/d^l \leq 1/d$ for $l \in \{1, 2, \dots, k-1\}$. Now, the drift is given as follows:

$$\begin{aligned}
\Delta L_{0k}^{(1)}(\mathbf{s}) &= s_1 - s_{k+1} - \lambda \left(1 - \left(\frac{s_k}{n} \right)^d \right) \\
&\stackrel{(a)}{\leq} n - 5m \frac{n \log d}{2d^k} - \lambda \left(1 - \left(1 - 3m \frac{\log d}{d} \right)^d \right) \\
&\stackrel{(b)}{\leq} n - 5m \frac{n \log d}{2d^k} - \lambda \left(1 - \frac{2}{d^{3m}} \right) \\
&\leq -5m \frac{n \log d}{2d^k} + \frac{2n}{d^{3m}} + n^{1-\gamma} \\
&\stackrel{(c)}{\leq} -m \frac{n \log d}{2d^k} + \frac{2n}{d^{3m}} \\
&\stackrel{(d)}{\leq} -m \frac{n \log d}{4d^k} \leq -m \frac{n \log d}{4d^m} \\
&\leq -\frac{1}{8} n^{1-\gamma} \stackrel{(e)}{\leq} -\sqrt{mn} \log n,
\end{aligned}$$

where (a) follows by lower bounding s_{k+1} by zero and using the upper bounds on s_1 and s_k given by (13) and (14) respectively. Next, (b) follows by Lemma 11. Further, (c) follows as $2m \frac{n \log d}{d^k} \geq 2m \frac{n \log d}{d^m} = n^{1-\gamma}$. Now, (d) follows as $m \frac{n \log d}{4d^k} \geq m \frac{n \log d}{4d^m}$. Thus, there exists $n_1^{(2)} \in \mathbb{Z}_+$ independent of k such that for all $n \geq n_1^{(2)}$, we have $m \frac{n \log d}{4d^k} \geq 2n/d^{3m}$ as d increases with n . Lastly, (e) follows as there exists $n_1^{(3)} \in \mathbb{Z}_+$ such that for all $n \geq n_1^{(3)}$, we have $n^{1-\gamma}/8 \geq \sqrt{mn} \log n$. Thus, for all $n \geq \max_{k \in [3]} \{n_1^{(k)}\}$, we have $\Delta L_{0k}^{(1)}(\mathbf{s}) \leq -\sqrt{mn} \log n$ when $L_{0k}^{(1)}(\mathbf{s}) \geq 0$ and $\mathbf{s} \in \mathcal{C}_{1,k}^{(1)} \cap \bigcap_{l=1}^{k-1} \mathcal{D}_l^{(1)}$. Now, by using Lemma 14, we get a high probability upper bound on $L_{0k}^{(1)}(\bar{\mathbf{s}})$ as follows:

$$\begin{aligned}
&\mathbb{P} \left(L_{0k}^{(1)}(\bar{\mathbf{s}}) \geq \sqrt{mn} \log n \right) \\
&\leq \left(\frac{n}{n + \sqrt{mn} \log n} \right)^{(\sqrt{mn} \log n)/2} + \sqrt{n} \mathbb{P} \left(\bar{\mathbf{s}} \notin \mathcal{C}_{1,k}^{(1)} \cap \bigcap_{l=1}^{k-1} \mathcal{D}_l^{(1)} \right) \\
&\stackrel{(a)}{\leq} \left(\frac{1}{n} \right)^{(m \log n)/4} + \sqrt{n} \left(\mathbb{P} \left(\bar{\mathbf{s}} \notin \mathcal{C}_{1,k}^{(1)} \right) + \sum_{l=1}^{k-1} \mathbb{P} \left(\bar{\mathbf{s}} \notin \mathcal{D}_l^{(1)} \right) \right) \\
&\stackrel{(b)}{\leq} \left(\frac{1}{n} \right)^{(m \log n)/4} + \sqrt{n} \left(\frac{1}{n} \right)^{(m \log n)/4 - 4(k-1)m - (k-1)} + \sqrt{n}(k-1) \left(\frac{1}{n} \right)^{(m \log n)/4 - 4(k-1)m} \\
&\stackrel{(c)}{\leq} \left(\frac{1}{n} \right)^{(m \log n)/4 - 4(k-1)m - k} \leq \left(\frac{1}{n} \right)^{(m \log n)/4 - 4(k-1)m - m},
\end{aligned}$$

where (a) follows by Lemma 13. Next, (b) follows by bounding $\mathbb{P} \left(\bar{\mathbf{s}} \notin \mathcal{C}_{1,k}^{(1)} \right)$ using Claim 1 and $\mathbb{P} \left(\bar{\mathbf{s}} \notin \mathcal{D}_l^{(1)} \right)$ is bounded by using (IH) and noting that the lower order terms in (IH) are upper bounded by $mn \log d / (2d^{k-l})$. In particular, by (6) there exists $n_1^{(4)} \in \mathbb{Z}_+$ independent of k such that for all $n \geq n_1^{(4)}$, we have

$$m \frac{n \log d}{2d^{k-l}} \geq 16m^3 \frac{n \log(d)^2}{d^{k-l+1}} + 4md^{l-1} \sqrt{mn} \log n \quad \forall l \in [k-1]. \quad (15)$$

Lastly, (c) holds for all $n \geq n_1^{(5)} \in \mathbb{Z}_+$ for some $n \geq n_1^{(5)} \in \mathbb{Z}_+$. Now, we use the above probability bound to obtain the required result for the lemma as follows.

$$\begin{aligned}
\left\{ L_{0k}^{(1)}(\mathbf{s}) \geq \sqrt{mn} \log n \right\} &= \left\{ kn - 3km \frac{n \log d}{d} - 3mn \log d \sum_{l=1}^{k-1} \frac{l}{d^{k-l+1}} - \sum_{l=1}^k s_l \geq \sqrt{mn} \log n \right\} \\
&\supseteq \left\{ s_k \leq n - 3km \frac{n \log d}{d} - 3mn \log d \sum_{l=1}^{k-1} \frac{l}{d^{k-l+1}} - \sqrt{mn} \log n \right\} \\
&\supseteq \left\{ s_k \leq n - 6m^2 \frac{n \log d}{d} \right\},
\end{aligned}$$

where the last assertion follows as there exists $n_1^{(6)} \in \mathbb{Z}_+$ independent of k such that for all $n \geq n_1^{(6)}$, we have

$$\begin{aligned} 3mn \log d \sum_{l=1}^{k-1} \frac{l}{d^{k-l+1}} + \sqrt{mn} \log n &\stackrel{(a)}{\leq} 3mn \log d \left(\frac{m}{d^2} + \frac{m^2}{d^3} \right) + \sqrt{mn} \log n \\ &\stackrel{(b)}{\leq} 6m^2 \frac{n \log d}{d^2} + \sqrt{mn} \log n \stackrel{(c)}{\leq} 3m^2 \frac{n \log d}{d}, \end{aligned}$$

where (a) follows by using the bounds $k \leq m$ and $1/d^l \leq 1/d^3$ for all $l \in \{3, \dots, k-1\}$. Next, (b) follows as $m^2/d^3 \leq m/d^2$ for large enough n . Lastly, (c) follows as $\sqrt{mn} \log n \leq n^{1-\gamma} = 2mn \log d/d^m \leq 2mn \log d/d \leq 2m^2 n \log d/d$ and $6/d^2 \leq 1/d$ for large enough n . Thus, by defining $n_1 \triangleq \max_{k \in [6]} \{n_1^{(k)}, n_{c1}\}$, the proof is complete. \square

Proof of Lemma 3 We will prove this using induction. For some $j \in \{1, \dots, k-1\}$, the induction hypothesis is given as follows. There exists $n_2^{(1)} \in \mathbb{Z}_+$ such that for all $n \geq n_2^{(1)}$, we have

$$\mathbb{P} \left(\bar{s}_j \leq n - 3(2m+k-j)m \frac{n \log d}{d^{k-j+1}} - \sqrt{mn} \log n \right) \leq \left(\frac{1}{n} \right)^{\frac{1}{4} m \log n - 4(k-\frac{3}{4})m - (k-j)}. \quad (\text{IH2})$$

Note that, the above expression would directly imply the required result as shown later in (16). The base case ($j=k$) is satisfied for all $n \geq n_1$ by Lemma 2. Now, we show the induction step for j . Define the family of Lyapunov functions $\{W_l(\mathbf{s}) : l \in [k-1]\}$ for $k \geq 2$ as follows:

$$W_l(\mathbf{s}) = n - 3(2m+k-l)m \frac{n \log d}{d^{k-l+1}} - s_l.$$

We analyze the drift of $W_j(\mathbf{s})$ when $W_j(\mathbf{s}) \geq 0$ and $\mathbf{s} \in \mathcal{C}_{j+1}^{(2)} \cap \mathcal{D}_{j-1}^{(2)}$ where

$$\begin{aligned} \mathcal{C}_l^{(2)} &= \left\{ W_l(\mathbf{s}) \leq \sqrt{mn} \log n \right\} \quad \forall l \in [k-1] \\ \mathcal{D}_l^{(2)} &= \left\{ s_l \geq n - \frac{5mn \log d}{2d^{k-l}} \right\} \quad \forall l \in [k-1]. \end{aligned}$$

The drift is given as follows:

$$\begin{aligned} \Delta W_j(\mathbf{s}) &= s_j - s_{j+1} - \lambda \left(\left(\frac{s_{j-1}}{n} \right)^d - \left(\frac{s_j}{n} \right)^d \right) \\ &\stackrel{(a)}{\leq} 3(2m+k-j-1)m \frac{n \log d}{d^{k-j}} + \sqrt{mn} \log n \\ &\quad - \lambda \left(\left(1 - \frac{5m \log d}{2d^{k-j+1}} \right)^d - \left(1 - 3(2m+k-j)m \frac{\log d}{d^{k-j+1}} \right)^d \right) \\ &\stackrel{(b)}{\leq} 3(2m+k-j-1)m \frac{n \log d}{d^{k-j}} + \sqrt{mn} \log n \\ &\quad - \lambda \left(-\frac{5m \log d}{2d^{k-j}} + 3(2m+k-j)m \frac{\log d}{d^{k-j}} - 9(2m+k-j)^2 m^2 \frac{\log(d)^2}{2d^{2k-2j}} \right) \\ &\leq -m \frac{n \log d}{2d^{k-j}} + \sqrt{mn} \log n + 3(2m+k-j)m \frac{n^{1-\gamma} \log d}{d^{k-j}} + 9(2m+k-j)^2 m^2 \frac{n \log(d)^2}{2d^{2k-2j}} \\ &\stackrel{(c)}{\leq} -m \frac{n \log d}{4d^{k-j}} \stackrel{(d)}{\leq} -\sqrt{mn} \log n. \end{aligned}$$

where (a) follows by bounding s_{j-1}, s_j and s_{j+1} using $\mathbf{s} \in \mathcal{C}_{j+1}^{(2)} \cap \mathcal{D}_{j-1}^{(2)}$ and $W_j(\mathbf{s}) \geq 0$. Next, (b) follows by Lemma 12.

Further, (c) follows as there exists $n_2^{(2)} \in \mathbb{Z}_+$, independent of j and k such that for all $n \geq n_2^{(2)}$, we have

$$\begin{aligned} m \frac{n \log d}{12d^{k-j}} &\stackrel{(c_1)}{\geq} m \frac{n \log d}{12d^m} \geq \frac{1}{24} n^{1-\gamma} \stackrel{(c_2)}{\geq} \sqrt{mn} \log n \\ m \frac{n \log d}{12d^{k-j}} &\stackrel{(c_3)}{\geq} m \frac{n \log d}{12d^{k-j}} \times 108mn^{-\gamma} \stackrel{(c_4)}{\geq} 3(2m+k-j)m \frac{n^{1-\gamma} \log d}{d^{k-j}} \\ m \frac{n \log d}{12d^{k-j}} &\stackrel{(c_5)}{\geq} m \frac{n \log d}{12d^{k-j}} \times \frac{486m^3 \log d}{d} \stackrel{(c_6)}{\geq} 9(2m+k-j)^2 m^2 \frac{n \log(d)^2}{2d^{2k-2j}}, \end{aligned}$$

where (c_1) follows as $k - j \leq k \leq m$, and (c_2) follows as $\gamma < 0.5$. Next, (c_3) follows as $\gamma > 0$ and $m \leq \log n$, and (c_4) follows as $3(2m + k - j) \leq 9m$. Lastly, by (5), (c_5) follows as $m^3 \log d/d \leq \gamma^3 \log(n)^3(1 + o(1))/(d \log(d)^2) \rightarrow 0$ as $n \rightarrow \infty$, and (c_6) follows as $3(2m + k - j) \leq 9m$.

Further, (d) follows as there exists $n_2^{(3)} \in \mathbb{Z}_+$, independent of j and k such that for all $n \geq n_2^{(3)}$ we have $mn \log d/4d^m \geq \sqrt{mn} \log n$. Thus, for all $n \geq \max\{n_2^{(2)}, n_2^{(3)}\}$, we have $\Delta W_j(\mathbf{s}) \leq -\sqrt{mn} \log n$ when $W_j(\mathbf{s}) \geq 0$ and $\mathbf{s} \in \mathcal{C}_{j+1}^{(2)} \cap \mathcal{D}_{j-1}^{(2)}$. Now, by using Lemma 14, we get a high probability upper bound on \bar{s}_j as follows:

$$\begin{aligned} & \mathbb{P}\left(\bar{s}_j \leq n - 9m^2 \frac{n \log d}{d^{k-j+1}}\right) \\ & \stackrel{(a)}{\leq} \mathbb{P}\left(W_j(\bar{\mathbf{s}}) \geq \sqrt{mn} \log n\right) \\ & \stackrel{(b)}{\leq} \left(\frac{n}{n + \sqrt{mn} \log n}\right)^{(\sqrt{mn} \log n)/2} + \sqrt{n} \mathbb{P}\left(\bar{\mathbf{s}} \notin \mathcal{C}_{j+1}^{(2)} \cap \mathcal{D}_{j-1}^{(2)}\right) \\ & \stackrel{(c)}{\leq} \left(\frac{1}{n}\right) + \sqrt{n} \left(\mathbb{P}\left(\bar{\mathbf{s}} \notin \mathcal{C}_{j+1}^{(2)}\right) + \mathbb{P}\left(\bar{\mathbf{s}} \notin \mathcal{D}_{j-1}^{(2)}\right)\right) \\ & \stackrel{(d)}{\leq} \left(\frac{1}{n}\right)^{(m \log n)/4} + \sqrt{n} \left(\frac{1}{n}\right)^{(m \log n)/4 - 4(k-1)m - m - (k-j)} + \sqrt{n} \left(\frac{1}{n}\right)^{(m \log n)/4 - 4(k-1)m} \\ & \stackrel{(e)}{\leq} \left(\frac{1}{n}\right)^{(m \log n)/4 - 4(k-1)m - m - (k-j+1)}, \end{aligned}$$

where (a) follows for all $n \geq n_2^{(4)}$ for some $n_2^{(4)} \in \mathbb{Z}_+$ as

$$\begin{aligned} & n - 3(2m + k - j)m \frac{n \log d}{d^{k-j+1}} - \sqrt{mn} \log n \\ & \geq n - 9m^2 \frac{n \log d}{d^{k-j+1}} + 3m \frac{n \log d}{d^{k-j+1}} - \sqrt{mn} \log n \\ & \geq n - 9m^2 \frac{n \log d}{d^{k-j+1}} + 3m \frac{n \log d}{d^m} - \sqrt{mn} \log n \geq n - 9m^2 \frac{n \log d}{d^{k-j+1}}. \end{aligned} \quad (16)$$

Next, (b) follows by Lemma 14 and (c) follows by Lemma 13. Now, (d) follows by upper bounding $\mathbb{P}\left(\bar{\mathbf{s}} \notin \mathcal{C}_{j+1}^{(2)}\right)$ using (IH2). In addition, $\mathbb{P}\left(\bar{\mathbf{s}} \notin \mathcal{D}_{j-1}^{(2)}\right)$ is upper bounded by using (IH) and noting that the lower order terms in (IH) are upper bounded by $mn \log d/(2d^{k-j+1})$ using (15) for all $n \geq n_2^{(5)}$ for some $n_2^{(5)} \in \mathbb{Z}_+$. Lastly, (e) follows for all $n \geq n_2^{(6)}$ for some $n_2^{(6)} \in \mathbb{Z}_+$. Now, by considering $n_2 \triangleq n_2^{(1)} \geq \max_{k \in [6]} \{n_2^{(k)}, n_1\}$, the induction step is complete. Thus, we have

$$\mathbb{P}\left(\bar{s}_i \leq n - 9m^2 \frac{n \log d}{d^{k-l+1}}\right) \leq \left(\frac{1}{n}\right)^{(m \log n)/4 - 4(k-1)m - m - (k-l+1)} \leq \left(\frac{1}{n}\right)^{(m \log n)/4 - 4(k-0.5)m} \quad \forall l \in [k].$$

This completes the proof of the lemma. \square

Proof of Lemma 4 Define the family of functions $\{W_{ik}(\mathbf{s}) : i \in [k]\}$ as follows:

$$\begin{aligned} W_{ik}(\mathbf{s}) &= s_i - n + 2m \frac{n \log d}{d^{k-i+1}} + (2i+1)d^{i-1} \sqrt{mn} \log n + 3(i+1)m \frac{n \log d}{d^{k-i+2}} \quad \forall i \in \{1, \dots, k-1\} \\ W_{kk}(\mathbf{s}) &= n - 2m \frac{n \log d}{d} - 2kd^{k-1} \sqrt{mn} \log n - 10m^2 \frac{n \log d}{d^2} - s_k. \end{aligned}$$

Now, we use the above family of functions to define the Lyapunov functions $\{Z_{ik} : i \in [k]\}$ as follows:

$$\begin{aligned} Z_{ik}^{(1)} &= W_{kk} - \sum_{l=i+1}^{k-1} W_{lk}, \quad Z_{ik}^{(2)} = W_{ik} \quad \forall i \in [k-1] \\ Z_{ik} &= \min\left\{Z_{ik}^{(1)}, Z_{ik}^{(2)}\right\} \quad \forall i \in [k-1]. \end{aligned}$$

To prove the lemma, we make use of the following claim:

CLAIM 2. *There exists $n_{e2} \in \mathbb{Z}_+$ such that for all $n \geq n_{e2}$, we have*

$$\mathbb{P}\left(Z_{1k}(\bar{\mathbf{s}}) \geq \sqrt{mn} \log n\right) \leq \left(\frac{1}{n}\right)^{(m \log n)/4 - 4(k-0.5)m - (k-1)}.$$

We defer the proof of claim to Appendix E and continue with the proof of Lemma 4. We analyze the drift of $Z_{0k}^{(1)}(\mathbf{s})$ when $Z_{0k}^{(1)}(\mathbf{s}) \geq 0$ and $\mathbf{s} \in \mathcal{C}_{1,k}^{(3)} \cap \bigcap_{l=1}^{k-1} \mathcal{D}_l^{(3)}$ where

$$\mathcal{C}_{l,k}^{(3)} = \left\{ Z_{l,k}(\mathbf{s}) \leq \sqrt{mn} \log n \right\} \quad \forall l \in [k-1] \quad (17a)$$

$$\mathcal{D}_l^{(3)} = \left\{ s_l \geq n - \frac{9m^2 n \log d}{d^{k-l+1}} \right\} \quad \forall l \in [k-1]. \quad (17b)$$

We first obtain a useful upper bound on s_1 as follows.

$$\begin{aligned} Z_{0k}^{(1)}(\mathbf{s}) \geq 0 &\Rightarrow W_{kk}(\mathbf{s}) - \sum_{l=2}^{k-1} W_{lk}(\mathbf{s}) \geq W_{1k}(\mathbf{s}) \\ &\Rightarrow W_{1k}(\mathbf{s}) \leq \sqrt{mn} \log n \quad \text{as } \mathbf{s} \in \mathcal{C}_{1,k}^{(3)} \\ &\Rightarrow s_1 \leq n - 2m \frac{n \log d}{d^k} - 2\sqrt{mn} \log n - 6m \frac{n \log d}{d^{k+1}}. \end{aligned} \quad (18)$$

Next, we will get a useful upper bound on s_k as follows:

$$\begin{aligned} Z_{0k}^{(1)}(\mathbf{s}) \geq 0 &\Rightarrow s_k \leq n - 2m \frac{n \log d}{d} - 2kd^{k-1} \sqrt{mn} \log n - 10m^2 \frac{n \log d}{d^2} - \sum_{l=1}^{k-1} W_{lk}(\mathbf{s}) \\ &\stackrel{(*)}{\Rightarrow} s_k \leq n - 2m \frac{n \log d}{d} - 2kd^{k-1} \sqrt{mn} \log n - 10m^2 \frac{n \log d}{d^2} + 9m^2 n \log d \sum_{l=2}^k \frac{1}{d^l} \\ &\stackrel{(**)}{\Rightarrow} s_k \leq n - 2m \frac{n \log d}{d}, \end{aligned} \quad (19)$$

where $(*)$ follows as $\mathbf{s} \in \bigcap_{l=1}^{k-1} \mathcal{D}_l^{(3)}$ and $(**)$ follows as there exists $n_3^{(1)} \in \mathbb{Z}_+$ such that for all $n \geq n_3^{(1)}$ we have

$$9m^2 n \log d \sum_{l=2}^k \frac{1}{d^l} \leq 9m^2 n \log d \left(\frac{1}{d^2} + \frac{m}{d^3} \right) \leq 10m^2 \frac{n \log d}{d^2}.$$

where the last inequality follows by noting that $m/d \rightarrow 0$ as $n \rightarrow \infty$. Now, the drift of $Z_{0k}^{(1)}(\mathbf{s})$ is given as follows:

$$\begin{aligned} \Delta Z_{0k}^{(1)}(\mathbf{s}) &= s_1 - s_{k+1} - \lambda \left(1 - \left(\frac{s_k}{n} \right)^d \right) \\ &\stackrel{(a)}{\leq} n - 2m \frac{n \log d}{d^k} - 2\sqrt{mn} \log n - 6m \frac{n \log d}{d^{k+1}} - \lambda \left(1 - \left(1 - 2m \frac{\log d}{d} \right)^d \right) \\ &\stackrel{(b)}{\leq} n - 2m \frac{n \log d}{d^k} - 2\sqrt{mn} \log n - 6m \frac{n \log d}{d^{k+1}} - \lambda \left(1 - \frac{2}{d^{2m}} \right) \\ &\leq -2m \frac{n \log d}{d^k} - 2\sqrt{mn} \log n - 6m \frac{n \log d}{d^{k+1}} + n^{1-\gamma} + \frac{2n}{d^{2m}} \\ &\stackrel{(c)}{\leq} -2\sqrt{mn} \log n \leq -\sqrt{mn} \log n. \end{aligned}$$

where (a) follows as $s_{k+1} \geq 0$, s_1 is upper bounded as in (18), and s_k is upper bounded as in (19). Next, by Lemma 11, there exists $n_3^{(2)} \in \mathbb{Z}_+$ independent of k such that for all $n \geq n_3^{(2)}$, (b) follows. Lastly, (c) follows by noting that

$$\begin{aligned} 2m \frac{n \log d}{d^k} &\geq 2m \frac{n \log d}{d^m} = n^{1-\gamma}, \\ 6m \frac{n \log d}{d^{k+1}} &\geq 6m \frac{n \log d}{d^{m+1}} \geq 6m \frac{n \log d}{d^{2m}} \geq \frac{2n}{d^{2m}}, \end{aligned}$$

where the last set of inequalities follows as $m \geq 2$ and there exists $n_3^{(3)} \in \mathbb{Z}_+$ such that for all $n \geq n_3^{(3)}$, we have $\log d \geq 1$. Thus, for all $n \geq \max_{k \in [3]} \left\{ n_3^{(k)} \right\}$, we have $\Delta Z_{0k}^{(1)}(\mathbf{s}) \leq -\sqrt{mn} \log n$ when $Z_{0k}^{(1)}(\mathbf{s}) \geq 0$ and

$\mathbf{s} \in \mathcal{C}_{1k}^{(3)} \cap \bigcap_{l=1}^{k-1} \mathcal{D}_l^{(3)}$. Now, by using Lemma 14, we obtain a high probability upper bound on $Z_{0k}^{(1)}(\bar{\mathbf{s}})$ as follows:

$$\begin{aligned}
& \mathbb{P}(Z_{0k}(\bar{\mathbf{s}}) \geq \sqrt{mn} \log n) \\
& \leq \left(\frac{n}{n + \sqrt{mn} \log n} \right)^{(\sqrt{mn} \log n)/2} + \sqrt{n} \mathbb{P}\left(\bar{\mathbf{s}} \notin \mathcal{C}_{1k}^{(3)} \cap \bigcap_{l=1}^{k-1} \mathcal{D}_l^{(3)}\right) \\
& \stackrel{(a)}{\leq} \left(\frac{1}{n} \right)^{(m \log n)/4} + \sqrt{n} \left(\mathbb{P}(\bar{\mathbf{s}} \notin \mathcal{C}_{1k}^{(3)}) + \sum_{l=1}^{k-1} \mathbb{P}(\bar{\mathbf{s}} \notin \mathcal{D}_l^{(3)}) \right) \\
& \stackrel{(b)}{\leq} \left(\frac{1}{n} \right)^{(m \log n)/4} + \sqrt{n} \left(\frac{1}{n} \right)^{(m \log n)/4 - 4(k-0.5)m - (k-1)} + \sqrt{n}(k-1) \left(\frac{1}{n} \right)^{(m \log n)/4 - 4(k-1)m} \\
& \stackrel{(c)}{\leq} \left(\frac{1}{n} \right)^{(m \log n)/4 - 4(k-0.5)m - k} \leq \left(\frac{1}{n} \right)^{(m \log n)/4 - 4(k-0.5)m - m},
\end{aligned}$$

where (a) follows by Lemma 13. Next, (b) follows by upper bounding $\mathbb{P}(\bar{\mathbf{s}} \notin \mathcal{C}_{1k}^{(3)})$ by Claim 2, and upper bounding $\mathbb{P}(\bar{\mathbf{s}} \notin \mathcal{D}_l^{(3)})$ for all $n \geq n_2$ by Lemma 3. Lastly, (c) follows for all $n \geq n_3^{(4)}$ for some $n_3^{(4)} \in \mathbb{Z}_+$ independent of k . Now, to complete the proof, note that

$$\begin{aligned}
& \left\{ Z_{0k}(\bar{\mathbf{s}}) \geq \sqrt{mn} \log n \right\} \\
& = \left\{ kn - 2mn \log d \sum_{l=1}^k \frac{1}{d^l} - \sqrt{mn} \log n \sum_{l=1}^k (2l+1)d^{l-1} - 10m^2 \frac{n \log d}{d^2} - 3mn \log d \sum_{l=1}^{k-1} \frac{(l+1)}{d^{k-l+2}} - \sum_{l=1}^k \bar{s}_l \right. \\
& \quad \left. \geq \sqrt{mn} \log n \right\} \\
& \supseteq \left\{ \bar{s}_k \leq n - 2m \frac{n \log d}{d} - 3md^{k-1} \sqrt{mn} \log n - 12m^2 \frac{n \log d}{d^2} \right\},
\end{aligned}$$

where the last assertion follows as there exists $n_3^{(5)} \in \mathbb{Z}_+$ independent of k such that for all $n \geq n_3^{(5)}$, we have

$$\begin{aligned}
& 3mn \log d \sum_{l=1}^{k-1} \frac{(l+1)}{d^{k-l+2}} \leq 3m^3 \frac{n \log d}{d^3} \leq m^2 \frac{n \log d}{2d^2} \\
& 2mn \log d \sum_{l=2}^k \frac{1}{d} \leq 2mn \log d \left(\frac{1}{d^2} + \frac{m}{d^3} \right) \leq \frac{3mn \log d}{d^2} \stackrel{(*)}{\leq} 3m^2 \frac{n \log d}{2d^2} \\
& \sqrt{mn} \log n \sum_{l=1}^k (2l+1)d^{l-1} + \sqrt{mn} \log n \leq (2m+1)(d^{k-1} + md^{k-2} + 1) \sqrt{mn} \log n \\
& \leq 3md^{k-1} \sqrt{mn} \log n,
\end{aligned}$$

where we used the fact that $k \leq m$ and $m^2/d \rightarrow 0$ as $n \rightarrow \infty$. Note that, (*) follows as $m \geq 2$. Now, by defining $n_3 \triangleq \max_{k \in [5]} \{n_3^{(k)}, n_2, n_{c2}\}$, the proof is complete. \square

Proof of Lemma 5 We will prove this using induction. For some $j \in [k]$, the induction hypothesis is given as follows: there exists $n_4^{(1)} \in \mathbb{Z}_+$ such that for all $n \geq n_4^{(1)}$, we have

$$\begin{aligned}
& \mathbb{P}\left(\bar{s}_j \leq n - 2m \frac{n \log d}{d^{k-j+1}} - (4m-j)d^{j-1} \sqrt{mn} \log n - 16(k-j+1)m^2 \frac{n \log(d)^2}{d^{k-j+2}} - \sqrt{mn} \log n\right) \\
& \leq \left(\frac{1}{n} \right)^{(m \log n)/4 - 4(k-0.25)m - (k-j)}. \tag{IH3}
\end{aligned}$$

The base case ($j = k$) is satisfied for all $n \geq n_3$ by Lemma 4 as

$$(4m-k)d^{k-1} \sqrt{mn} \log n + 16m^2 \frac{n \log(d)^2}{d^2} + \sqrt{mn} \log n \geq 3md^{k-1} \sqrt{mn} \log n + 12m^2 \frac{n \log d}{d^2}.$$

Now, we show the induction step corresponding to $j \in [k]$. Consider the Lyapunov function

$$\tilde{W}_j(\mathbf{s}) = n - 2m \frac{n \log d}{d^{k-j+1}} - (4m-j)d^{j-1} \sqrt{mn} \log n - 16(k-j+1)m^2 \frac{n \log(d)^2}{d^{k-j+2}} - s_j.$$

We analyze the drift of $\tilde{W}_j(\mathbf{s})$ when $\tilde{W}_j(\mathbf{s}) \geq 0$ and $\mathbf{s} \in \mathcal{C}_{j+1}^{(4)} \cap \mathcal{D}_{j-1}^{(4)}$, where

$$\begin{aligned} \mathcal{C}_l^{(4)} &= \left\{ \tilde{W}_l \leq \sqrt{mn} \log n \right\} \quad \forall l \in [k] \\ \mathcal{D}_l^{(4)} &= \left\{ s_l \geq n - 9m^2 \frac{n \log d}{d^{k-l+1}} \right\} \quad \forall l \in [k]. \end{aligned}$$

Now, the drift is given as follows:

$$\begin{aligned} & \Delta \tilde{W}_j(\mathbf{s}) \\ &= s_j - s_{j+1} - \lambda \left(\left(\frac{s_{j-1}}{n} \right)^d - \left(\frac{s_j}{n} \right)^d \right) \\ &\stackrel{(a)}{\leq} 2m \frac{n \log d}{d^{k-j}} + (4m - j - 1)d^j \sqrt{mn} \log n + 16(k-j)m^2 \frac{n \log(d)^2}{d^{k-j+1}} + \sqrt{mn} \log n \\ &\quad - \lambda \left(1 - 9m^2 \frac{\log d}{d^{k-j+2}} \right)^d + \lambda \left(1 - 2m \frac{\log d}{d^{k-j+1}} - (4m-j)\sqrt{m}d^{j-1} \frac{\log n}{\sqrt{n}} - 16(k-j+1)m^2 \frac{\log(d)^2}{d^{k-j+2}} \right)^d \\ &\stackrel{(b)}{\leq} 2m \frac{n \log d}{d^{k-j}} + (4m - j - 1)d^j \sqrt{mn} \log n + 16(k-j)m^2 \frac{n \log(d)^2}{d^{k-j+1}} + \sqrt{mn} \log n \\ &\quad - \lambda \left(-9m^2 \frac{\log d}{d^{k-j+1}} + 2m \frac{\log d}{d^{k-j}} + (4m-j)\sqrt{m}d^j \frac{\log n}{\sqrt{n}} + 16(k-j+1)m^2 \frac{\log(d)^2}{d^{k-j+1}} \right) \\ &\quad + \frac{3\lambda}{2} \left(4m^2 \frac{\log(d)^2}{d^{2k-2j}} + (4m-j)^2 m d^{2j} \frac{\log(n)^2}{n} + 256(k-j+1)^2 m^4 \frac{\log(d)^4}{d^{2k-2j+2}} \right) \\ &\leq -d^j \sqrt{mn} \log n - 7m^2 \frac{n \log(d)^2}{d^{k-j+1}} + \sqrt{mn} \log n + 6m^2 \frac{n \log(d)^2}{d^{2k-2j}} + 24m^3 d^{2j} \log(n)^2 + 384m^6 \frac{n \log(d)^4}{d^{2k-2j+2}} \\ &\quad + 2m \frac{n^{1-\gamma} \log d}{d^{k-j}} + 4\sqrt{m}m d^j n^{0.5-\gamma} \log n + 16m^3 \frac{n^{1-\gamma} \log(d)^2}{d^{k-j+1}} \\ &\stackrel{(c)}{\leq} -d^j \sqrt{mn} \log n - m^2 \frac{n \log(d)^2}{d^{k-j+1}} + \sqrt{mn} \log n + 24m^3 d^{2j} \log(n)^2 + 384m^6 \frac{n \log(d)^4}{d^{2k-2j+2}} \\ &\quad + 2m \frac{n^{1-\gamma} \log d}{d^{k-j}} + 4\sqrt{m}m d^j n^{0.5-\gamma} \log n + 16m^3 \frac{n^{1-\gamma} \log(d)^2}{d^{k-j+1}} \\ &\stackrel{(d)}{\leq} -\frac{1}{2}d^j \sqrt{mn} \log n + \sqrt{mn} \log n \stackrel{(e)}{\leq} -\sqrt{mn} \log n, \end{aligned}$$

where (a) follows by the lower bound on s_{j-1} , s_{j+1} and upper bound on s_j . In particular, as $\mathbf{s} \in \mathcal{D}_{j-1}^{(4)}$, we get a lower bound on s_{j-1} , as $\tilde{W}_j(\mathbf{s}) \geq 0$, we get an upper bound on s_j , and as $\mathbf{s} \in \mathcal{C}_{j+1}^{(4)}$, we get a lower bound on s_{j+1} . Next, (b) follows by Lemma 12 and using the identity $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$. Now, (c) follows as $k-j \geq 1$ which implies that $k-j+1 \leq 2(k-j)$. Lastly, (d) follows as there exists $n_4^{(2)} \in \mathbb{Z}_+$ such that for all $n \geq n_4^{(2)}$, we have

$$\begin{aligned} \frac{1}{2}m^2 \frac{n \log(d)^2}{d^{k-j+1}} &\stackrel{(d_1)}{\geq} \frac{1}{2}m^2 \frac{n \log(d)^2}{d^{k-j+1}} \times \frac{768m^4 \log(d)^2}{d^{k-j+1}} = 384m^6 \frac{n \log(d)^4}{d^{2k-2j+2}} \\ \frac{1}{2}m^2 \frac{n \log(d)^2}{d^{k-j+1}} &\stackrel{(d_2)}{\geq} \frac{1}{4}m \frac{n^{1-\gamma} \log d}{d^{k-j+1-m}} \stackrel{(d_3)}{\geq} 3m \frac{n^{1-\gamma} \log d}{d^{k-j}} \stackrel{(d_4)}{\geq} 2m \frac{n^{1-\gamma} \log d}{d^{k-j}} + 16m^3 \frac{n^{1-\gamma} \log(d)^2}{d^{k-j+1}} \\ \frac{1}{4}d^j \sqrt{mn} \log n &\stackrel{(d_5)}{\geq} 24m^3 d^j \log(n)^2 (2mn^\gamma \log d) \stackrel{(d_6)}{\geq} 24m^3 d^{2j} \log(n)^2 \\ \frac{1}{4}d^j \sqrt{mn} \log n &\stackrel{(d_7)}{\geq} 4\sqrt{m}m d^j n^{0.5-\gamma} \log n, \end{aligned}$$

where (d₁) follows as $m^4 \log(d)^2 / d^{k-j+1} \leq m^4 \log(d)^2 / d^2 \leq \log(n)^4 / (d \log d)^2 \rightarrow 0$ as $n \rightarrow \infty$. Next, (d₂) follows as $2n^\gamma m \log d \geq d^m$, (d₃) follows as $m \geq 2$, and (d₄) follows as $m^2 \log d / d \leq \log(n)^2 / (d \log d) \rightarrow 0$. Now, (d₅) follows as $\gamma < 0.5$ and $m \leq \log n$, and (d₆) follows as $d^j \leq d^m \geq 2mn^\gamma \log d$. Lastly, (d₇) follows as $\gamma > 0$ and $m \leq \log n$. Further, (e) follows as there exists $n_4^{(3)} \in \mathbb{Z}_+$ such that for all $n \geq n_4^{(3)}$, we have

$$\frac{1}{2}d^j \sqrt{mn} \log n \geq \frac{1}{2}d \sqrt{mn} \log n \geq 2\sqrt{mn} \log n.$$

Thus, for all $n \geq \max\{n_4^{(2)}, n_4^{(3)}\}$, we have $\Delta \tilde{W}_j(\mathbf{s}) \leq -\sqrt{mn} \log n$ when $\tilde{W}_j(\mathbf{s}) \geq 0$ and $\mathbf{s} \in \mathcal{C}_{j+1}^{(4)} \cap \mathcal{D}_{j-1}^{(4)}$. Now, by using Lemma 14, we obtain a high probability upper bound on \bar{s}_j as follows:

$$\begin{aligned}
& \mathbb{P}\left(\bar{s}_j \leq n - 2m \frac{n \log d}{d^{k-j+1}} - (4m-j)d^{j-1}\sqrt{mn} \log n - 16(k-j+1)m^2 \frac{n \log(d)^2}{d^{k-j+2}} - \sqrt{mn} \log n\right) \\
&= \mathbb{P}\left(\tilde{W}_j(\bar{\mathbf{s}}) \geq \sqrt{mn} \log n\right) \\
&\stackrel{(a)}{\leq} \left(\frac{n}{n + \sqrt{mn} \log n}\right)^{(\sqrt{mn} \log n)/2} + \sqrt{n} \mathbb{P}\left(\bar{\mathbf{s}} \notin \mathcal{C}_{j+1}^{(4)} \cap \mathcal{D}_{j-1}^{(4)}\right) \\
&\stackrel{(b)}{\leq} \left(\frac{1}{n}\right)^{(m \log n)/4} + \sqrt{n} \left(\mathbb{P}\left(\bar{\mathbf{s}} \notin \mathcal{C}_{j+1}^{(4)}\right) + \mathbb{P}\left(\bar{\mathbf{s}} \notin \mathcal{D}_{j-1}^{(4)}\right)\right) \\
&\stackrel{(c)}{\leq} \left(\frac{1}{n}\right)^{(m \log n)/4} + \sqrt{n} \left(\frac{1}{n}\right)^{(m \log n)/4 - 4(k-0.25)m - (k-j)} + \sqrt{n} \left(\frac{1}{n}\right)^{(m \log n)/4 - 4(k-1)m} \\
&\stackrel{(d)}{\leq} \left(\frac{1}{n}\right)^{(m \log n)/4 - 4(k-0.25)m - (k-j+1)},
\end{aligned}$$

where (a) follows by Lemma 14 and (b) follows by Lemma 13. Next, (c) follows by upper bounding $\mathbb{P}\left(\bar{\mathbf{s}} \notin \mathcal{C}_{j+1}^{(4)}\right)$ using induction hypothesis (IH3) and upper bounding $\mathbb{P}\left(\bar{\mathbf{s}} \notin \mathcal{D}_{j-1}^{(4)}\right)$ for all $n \geq n_2$ using Lemma 3. Lastly, (d) follows for all $n \geq n_4^{(4)}$ for some $n_4^{(4)} \in \mathbb{Z}_+$. By considering $n_4 \triangleq n_4^{(1)} \geq \max_{k \in [4]} \{n_4^{(k)}, n_2, n_3\}$, the induction step is complete. To complete the proof, note that

$$\begin{aligned}
& \mathbb{P}\left(\bar{s}_j \leq n - 2m \frac{n \log d}{d^{k-j+1}} - 4md^{j-1}\sqrt{mn} \log n - 16m^3 \frac{n \log(d)^2}{d^{k-j+2}}\right) \\
&\stackrel{(*)}{\leq} \mathbb{P}\left(\bar{s}_j \leq n - 2m \frac{n \log d}{d^{k-j+1}} - (4m-j)d^{j-1}\sqrt{mn} \log n - 16(k-j+1)m^2 \frac{n \log(d)^2}{d^{k-j+2}} - \sqrt{mn} \log n\right) \\
&\leq \left(\frac{1}{n}\right)^{(m \log n)/4 - 4(k-0.25)m - (k-j+1)} \leq \left(\frac{1}{n}\right)^{(m \log n)/4 - 4km} \quad \forall j \in [k],
\end{aligned}$$

where (*) follows as $j \geq 1$ and $k \leq m$. This completes the proof of the lemma. \square

Appendix D: Proof for the Upper Bound

D.1. Proof of Lemmas for Upper Bound

We start by re-stating the expression of B_i below that was defined in (9) for convenience.

$$B_i = 18md^{i-1}\sqrt{mn} \log n + 36m^3 \frac{n \log(d)^2}{d^{m-i+2}} + \frac{n^{1-\gamma}}{d^{m-i}} \mathbb{1}\{m > 1\} \quad \forall i \in [m].$$

Now, we start by stating and proving a ‘master’ lemma that will help us in proving Lemma 6, Lemma 8, and Lemma 9.

LEMMA 15. Let $B_{m+2} \leq nb$ and $x \in \mathbb{R}_+$ and $\tilde{n}_0^{(1)} \in \mathbb{Z}_+$ be such that for all $n \geq \tilde{n}_0^{(1)}$, we have

$$\mathbb{P}\left(\sum_{i=m+2}^b \bar{s}_i \geq B_{m+2}\right) \leq \left(\frac{1}{n}\right)^{\frac{m \log n}{x}}. \quad (20)$$

Define

$$B'_m = (1 + (b-1)\mathbb{1}\{B_{m+2} \geq 2\}) \left(17md^{m-1}\sqrt{mn} \log n + 36m^3 \frac{n \log(d)^2}{d^2} + n^{1-\gamma} \mathbb{1}\{m > 1\}\right).$$

Then, there exists $\tilde{n}_0^{(2)} \in \mathbb{Z}_+$ such that for all $n \geq \tilde{n}_0^{(2)}$, we have

$$\mathbb{P}(\bar{s}_{m+1} \geq B'_m + \sqrt{mn} \log n) \leq \left(\frac{1}{n}\right)^{\frac{m \log n}{\max\{x, 5\}} - m - 2} \quad (21)$$

$$\mathbb{P}\left(\bar{s}_m \geq n - 2m \frac{n \log d}{d} + B'_m + \sqrt{mn} \log n\right) \leq \left(\frac{1}{n}\right)^{\frac{m \log n}{\max\{x, 5\}} - m - 2}. \quad (22)$$

Proof of Lemma 15 Consider the following functions:

$$L_{m+1}(\mathbf{s}) = \sum_{l=m+1}^b s_l - (1 + (b-1)\mathbb{1}\{B_{m+2} \geq 2\}) \times \\ \left(8md^{m-1}\sqrt{mn}\log n + 18m^3\frac{n\log(d)^2}{d^2} + n^{1-\gamma}\mathbb{1}\{m > 1\} \right) \\ L_l(\mathbf{s}) = n - 2m\frac{n\log d}{d^{m-l+1}} + 3ld^{l-1}\sqrt{mn}\log n + lm\frac{n\log d}{d^{m-l+2}} - s_l \quad \forall l \in [m].$$

Now, we define the Lyapunov function $U_j(\mathbf{s})$ in terms of $\{L_l(\mathbf{s}) : l \in [m+1]\}$ as follows:

$$U_j^{(1)}(\mathbf{s}) = L_{m+1}(\mathbf{s}) - \sum_{l=j+1}^m L_l(\mathbf{s}) \\ U_j^{(2)}(\mathbf{s}) = L_j(\mathbf{s}) \\ U_j(\mathbf{s}) = \min \left\{ U_j^{(1)}(\mathbf{s}), U_j^{(2)}(\mathbf{s}) \right\}.$$

We use induction on $j \in \{1, \dots, m\}$ to show the following claim:

CLAIM 3. *There exists $\tilde{n}_c \geq \tilde{n}_0^{(1)}$ such that for all $n \geq \tilde{n}_c$, we have*

$$\mathbb{P}(U_1(\bar{\mathbf{s}}) \geq \sqrt{mn}\log n) \leq \left(\frac{1}{n}\right)^{\frac{m\log n}{\max\{x, 5\}} - m}.$$

We use the high probability upper bound $\sum_{l=m+2}^b s_l \leq B_{m+2}$ to prove the above claim. We defer the details of the proof to Appendix F and continue with the proof of Lemma 15. Now, we analyze the drift of $U_0^{(1)}(\mathbf{s})$ when $U_0^{(1)}(\mathbf{s}) \geq 0$ and $\mathbf{s} \in \tilde{\mathcal{C}}_1^{(1)}$ where $\tilde{\mathcal{C}}_1^{(1)} = \{U_1(\mathbf{s}) \leq \sqrt{mn}\log n\}$. First, we obtain a useful lower bound on s_1 as follows:

$$U_0^{(1)}(\mathbf{s}) \geq 0 \Rightarrow L_{m+1}(\mathbf{s}) - \sum_{l=2}^m L_l(\mathbf{s}) \geq L_1(\mathbf{s}) \\ \stackrel{(*)}{\Rightarrow} L_1(\mathbf{s}) \leq \sqrt{mn}\log n \\ \Rightarrow s_1 \geq n - 2m\frac{n\log d}{d^m} + 2\sqrt{mn}\log n + m\frac{n\log d}{d^{m+1}}, \quad (23)$$

where $(*)$ follows as $\mathbf{s} \in \tilde{\mathcal{C}}_1^{(1)}$. Now, the drift is given as follows:

$$\Delta U_0^{(1)}(\mathbf{s}) = -s_1 + \lambda \left(1 - \left(\frac{s_b}{n}\right)^d \right) \\ \stackrel{(a)}{\leq} -n + 2m\frac{n\log d}{d^m} - 2\sqrt{mn}\log n - m\frac{n\log d}{d^{m+1}} + n - n^{1-\gamma} \\ \stackrel{(b)}{=} -2\sqrt{mn}\log n - m\frac{n\log d}{d^{m+1}} \\ \leq -\sqrt{mn}\log n.$$

where (a) follows by using the trivial bound $s_b \geq 0$ and lower bounding s_1 using (23). Next, (b) follows as m is assumed to be integer. In particular, we have $n^{1-\gamma} = 2m\frac{n\log d}{d^m}$. Thus, we have $\Delta U_0^{(1)}(\mathbf{s}) \leq -\sqrt{mn}\log n$ when $U_0^{(1)}(\mathbf{s}) \geq 0$ and $\mathbf{s} \in \tilde{\mathcal{C}}_1^{(1)}$. Now, using Lemma 14, we have

$$\mathbb{P}(U_0^{(1)}(\bar{\mathbf{s}}) \geq \sqrt{mn}\log n) \leq \left(\frac{n}{n + \sqrt{mn}\log n}\right)^{(\sqrt{mn}\log n)/2} + \sqrt{n}\mathbb{P}(\bar{\mathbf{s}} \notin \tilde{\mathcal{C}}_1^{(1)}) \\ \stackrel{(a)}{\leq} \left(\frac{1}{n}\right)^{(m\log n)/4} + \sqrt{n}\mathbb{P}(\bar{\mathbf{s}} \notin \tilde{\mathcal{C}}_1^{(1)}) \\ \stackrel{(b)}{\leq} \left(\frac{1}{n}\right)^{(m\log n)/4} + \left(\frac{1}{n}\right)^{\frac{m\log n}{\max\{x, 5\}} - m} \\ \stackrel{(c)}{\leq} \left(\frac{1}{n}\right)^{\frac{m\log n}{\max\{x, 5\}} - m - 1}, \quad (24)$$

where (a) follows by Lemma 13 and (b) follows by Claim 3 for all $n \geq \tilde{n}_c$. Lastly, (c) follows for all $n \geq \tilde{n}_0^{(2)}$ for some $\tilde{n}_0^{(2)} \in \mathbb{Z}_+$. Now, we will translate the above probability bound to the one required for the lemma. Define $\tilde{\mathcal{D}}_l^{(1)}$ as follows:

$$\tilde{\mathcal{D}}_l^{(1)} = \left\{ s_l \geq n - 2m \frac{n \log d}{d^{m-l+1}} - 4md^{l-1} \sqrt{mn} \log n - 16m^3 \frac{n \log(d)^2}{d^{m-l+2}} \right\} \quad \forall l \in [m].$$

We first prove the high probability upper bound on \bar{s}_{m+1} below.

$$\begin{aligned} \mathbb{P}(\bar{s}_{m+1} \geq B'_m) &\stackrel{(a)}{=} \mathbb{P}\left(\bar{\mathbf{s}} \in \{s_{m+1} \geq B'_m\} \cap \bigcap_{l=1}^m \tilde{\mathcal{D}}_l^{(1)}\right) + \mathbb{P}\left(\bar{\mathbf{s}} \in \{s_{m+1} \geq B'_m\} \cap \bigcup_{l=1}^m \tilde{\mathcal{D}}_l^{(1),c}\right) \\ &\stackrel{(b)}{\leq} \mathbb{P}\left(\bar{\mathbf{s}} \in \{s_{m+1} \geq B'_m\} \cap \bigcap_{l=1}^m \tilde{\mathcal{D}}_l^{(1)}\right) + \sum_{l=1}^m \mathbb{P}\left(\bar{\mathbf{s}} \notin \tilde{\mathcal{D}}_l^{(1)}\right) \\ &\stackrel{(c)}{\leq} \mathbb{P}\left(\bar{\mathbf{s}} \in \{s_{m+1} \geq B'_m\} \cap \bigcap_{l=1}^m \tilde{\mathcal{D}}_l^{(1)}\right) + \left(\frac{1}{n}\right)^{(m \log n)/5-1} \\ &\stackrel{(d)}{\leq} \mathbb{P}\left(U_0^{(1)}(\bar{\mathbf{s}}) \geq \sqrt{mn} \log n\right) + \left(\frac{1}{n}\right)^{(m \log n)/5-1} \\ &\stackrel{(e)}{\leq} \left(\frac{1}{n}\right)^{\frac{m \log n}{\max\{x,5\}} - m-1} + \left(\frac{1}{n}\right)^{(m \log n)/5-1} \\ &\leq \left(\frac{1}{n}\right)^{\frac{m \log n}{\max\{x,5\}} - m-2}, \end{aligned}$$

where (a) follows by the law of total probability, and (b) follows by the union bound. Next, (c) follows by Theorem 2. Now, (d) follows by noting the following:

$$\begin{aligned} \{s_{m+1} \geq B'_m\} \cap \bigcap_{l=1}^m \{\mathbf{s} \in \tilde{\mathcal{D}}_l^{(1)}\} &\stackrel{(*)}{\subseteq} \left\{L_{m+1}(\mathbf{s}) \geq 9md^{m-1} \sqrt{mn} \log n + 18m^3 \frac{n \log(d)^2}{d^2}\right\} \cap \bigcap_{l=1}^m \{\mathbf{s} \in \tilde{\mathcal{D}}_l^{(1)}\} \\ &\stackrel{(**)}{\subseteq} \left\{U_0^{(1)}(\mathbf{s}) \geq \sqrt{mn} \log n\right\}. \end{aligned} \quad (25)$$

where $(*)$ follows as $\mathbb{1}\{m > 1\}n^{1-\gamma} \geq 0$ and $(**)$ holds for all $n \geq \tilde{n}_0^{(3)}$ for some $\tilde{n}_0^{(3)} \in \mathbb{Z}_+$. In particular, note that $U_0^{(1)}(\mathbf{s}) = L_{m+1}(\mathbf{s}) - \sum_{l=1}^m L_l(\mathbf{s})$ and we can upper bound $\sum_{l=1}^m L_l(\mathbf{s})$ as follows:

$$\begin{aligned} \mathbf{s} \in \bigcap_{l=1}^m \tilde{\mathcal{D}}_l^{(1)} &\Rightarrow \sum_{l=1}^m L_l(\mathbf{s}) \leq 7m \sqrt{mn} \log n \sum_{l=1}^m d^{l-1} + 17m^3 n \log(d)^2 \sum_{l=1}^m \frac{1}{d^{m-l+2}} \\ &\leq 8md^{m-1} \sqrt{mn} \log n + 18m^2 \frac{n \log(d)^2}{d^2}. \end{aligned}$$

Lastly, (e) follows by (24). This completes the proof of (21). Now, we will prove (22). Similar to the upper bound on \bar{s}_{m+1} , we get

$$\begin{aligned} &\mathbb{P}\left(\bar{s}_m \geq n - 2m \frac{n \log d}{d} + B'_m\right) \\ &\leq \mathbb{P}\left(\bar{\mathbf{s}} \in \left\{s_m \geq n - 2m \frac{n \log d}{d} + B'_m\right\} \cap \bigcap_{l=1}^{m-1} \{\mathbf{s} \in \tilde{\mathcal{D}}_l^{(1)}\}\right) + \left(\frac{1}{n}\right)^{(m \log n)/5-1} \\ &\stackrel{(*)}{\leq} \mathbb{P}\left(U_0^{(1)}(\bar{\mathbf{s}}) \geq \sqrt{mn} \log n\right) + \left(\frac{1}{n}\right)^{(m \log n)/5-1} \\ &\leq \left(\frac{1}{n}\right)^{\frac{m \log n}{\max\{x,5\}} - m-1} + \left(\frac{1}{n}\right)^{(m \log n)/5-1} \\ &\leq \left(\frac{1}{n}\right)^{\frac{m \log n}{\max\{x,5\}} - m-2}. \end{aligned}$$

where $(*)$ follows by noting that $U_0^{(1)}(\mathbf{s}) = L_{m+1}(\mathbf{s}) - \sum_{l=1}^m L_l(\mathbf{s})$ and bounding the terms $\{L_l(\mathbf{s}) : l \in [m+1] \setminus \{m\}\}$. In particular, note that

$$L_{m+1}(\mathbf{s}) \geq -(1 + (b-1) \mathbb{1}\{B_{m+2} \geq 2\}) \left(8md^{m-1} \sqrt{mn} \log n + 18m^3 \frac{n \log(d)^2}{d^2} + n^{1-\gamma} \mathbb{1}\{m > 1\}\right)$$

Also, as $\mathbf{s} \in \bigcap_{l=1}^{m-1} \tilde{\mathcal{D}}_l^{(1)}$, there exists $\tilde{n}_0^{(5)} \in \mathbb{Z}_+$ such that for all $n \geq \tilde{n}_0^{(5)}$, we have

$$\begin{aligned} \sum_{l=1}^{m-1} L_l(\mathbf{s}) &\leq 7m\sqrt{mn} \log n \sum_{l=1}^{m-1} d^{l-1} + 17m^3 n \log(d)^2 \sum_{l=1}^{m-1} \frac{1}{d^{m-l+2}} \\ &\leq 8md^{m-2} \sqrt{mn} \log n + 18m^3 \frac{n \log(d)^2}{d^3} \\ &\leq md^{m-1} \sqrt{mn} \log n + 17m^3 \frac{n \log(d)^2}{d^2}. \end{aligned}$$

Where the last inequality follows for $d \geq 18/17$. Combining the bounds, we have

$$\begin{aligned} &L_{m+1}(\mathbf{s}) - \sum_{l=1}^m L_l(\mathbf{s}) \\ &\geq -(1 + (b-1) \mathbb{1}\{B_{m+2} \geq 2\}) \left(9md^{m-1} \sqrt{mn} \log n + 35m^3 \frac{n \log(d)^2}{d^2} + n^{1-\gamma} \mathbb{1}\{m > 1\} \right) - L_m(\mathbf{s}) \\ &\geq -(1 + (b-1) \mathbb{1}\{B_{m+2} \geq 2\}) \left(12md^{m-1} \sqrt{mn} \log n + 36m^3 \frac{n \log(d)^2}{d^2} + n^{1-\gamma} \mathbb{1}\{m > 1\} \right) + B'_m \\ &\geq \sqrt{mn} \log n, \end{aligned}$$

where the second last inequality follows by using the lower bound $s_m \geq n - 2mn \log d/d + B'_m$. Thus, by defining $\tilde{n}_0^{(2)} = \max_{k \in \{3,4,5\}} \{\tilde{n}_0^{(k)}, \tilde{n}_c\}$, the proof is complete. \square

Proof of Lemma 6 By applying Lemma 15 with $B_{m+2} = nb$, $x = 1$, $\tilde{n}_0^{(1)} = 1$ and noting that $B'_m + \sqrt{mn} \log n \leq bB_m$, for all $n \geq \tilde{n}_0^{(2)}$, we have

$$\mathbb{P}(\bar{s}_{m+1} \geq bB_m) \leq \left(\frac{1}{n}\right)^{(m \log n)/5 - m - 2} \leq \left(\frac{1}{n}\right)^{(m \log n)/6},$$

where the last inequality follows for all $n \geq \tilde{n}_1^{(1)}$ for some $\tilde{n}_1^{(1)} \in \mathbb{Z}_+$ and B_m is defined in (9). Thus, by defining $\tilde{n}_1 \triangleq \max\{\tilde{n}_0^{(2)}, \tilde{n}_1^{(1)}\}$, the proof is complete. \square

Proof of Lemma 7 Consider the Lyapunov function:

$$U_{m+2}(\mathbf{s}) = \sum_{l=m+2}^b s_l.$$

In steady-state, we have

$$0 = \mathbb{E}[\Delta U_{m+2}(\bar{\mathbf{s}})] = -\mathbb{E}[\bar{s}_{m+2}] + \mathbb{E}\left[\lambda \left(\left(\frac{\bar{s}_{m+1}}{n}\right)^d - \left(\frac{\bar{s}_b}{n}\right)^d \right)\right].$$

Thus, we have

$$\begin{aligned} &\mathbb{E}[\bar{s}_{m+2}] \\ &= \mathbb{E}\left[\lambda \left(\left(\frac{\bar{s}_{m+1}}{n}\right)^d - \left(\frac{\bar{s}_b}{n}\right)^d \right)\right] \\ &\stackrel{(a)}{\leq} n \mathbb{E}\left[\left(\frac{\bar{s}_{m+1}}{n}\right)^d\right] \\ &\stackrel{(b)}{=} n \mathbb{E}\left[\left(\frac{\bar{s}_{m+1}}{n}\right)^d \mid \bar{s}_{m+1} \geq bB_m\right] \mathbb{P}(\bar{s}_{m+1} \geq bB_m) + n \mathbb{E}\left[\left(\frac{\bar{s}_{m+1}}{n}\right)^d \mid \bar{s}_{m+1} \leq bB_m\right] \mathbb{P}(\bar{s}_{m+1} \leq bB_m) \\ &\stackrel{(c)}{\leq} n \left(\frac{bB_m}{n}\right)^d + n \mathbb{P}(\bar{s}_{m+1} \geq bB_m) \\ &\stackrel{(d)}{\leq} n \left(\frac{bB_m}{n}\right)^d + \left(\frac{1}{n}\right)^{(m \log n)/6 - 1} \\ &\stackrel{(e)}{\leq} \left(\frac{1}{n}\right)^{\log(n)^2} + \left(\frac{1}{n}\right)^{(m \log n)/6 - 1} \stackrel{(f)}{\leq} \left(\frac{1}{n}\right)^{(m \log n)/7}, \end{aligned}$$

where (a) follows as $\lambda \leq n$ and $s_b \geq 0$. Next, (b) follows by the law of total expectation. Further, (c) follows as $\mathbb{P}(\bar{s}_{m+1} \geq bB_m) \leq 1$ and $s_{m+1} \leq n$. Now, (d) follows for all $n \geq \tilde{n}_1$ by Lemma 6. Lastly, (e) follows for all

$n \geq \tilde{n}_2^{(1)}$ for some $\tilde{n}_2^{(1)} \in \mathbb{Z}_+$ as $bB_m = o(n)$ and $d \geq \Omega(\log(n)^3)$ and (f) follows for all $n \geq \tilde{n}_2^{(2)}$ for some $\tilde{n}_2^{(2)}$ as $m = o(\log n)$. Now, by Markov's inequality, we have

$$\mathbb{P}(\bar{s}_{m+2} \geq 1) \leq \mathbb{E}[\bar{s}_{m+2}] \leq \left(\frac{1}{n}\right)^{(m \log n)/7}.$$

Lastly, note that as $s_{k_1} \geq s_{k_2}$ for all $k_1 \leq k_2$, we have

$$\{\bar{s}_{m+2} \geq 1\} = \left\{ \sum_{l=m+2}^b \bar{s}_l \geq 1 \right\}.$$

Thus, by defining $\tilde{n}_2 \triangleq \max\{\tilde{n}_1, \tilde{n}_2^{(1)}, \tilde{n}_2^{(2)}\}$, the proof is complete. \square

Proof of Lemma 8 and Lemma 9 By using Lemma 7, for all $n \geq \tilde{n}_2$, we have

$$\mathbb{P}\left(\sum_{l=m+2}^b \bar{s}_l \geq 1\right) \leq \left(\frac{1}{n}\right)^{(m \log n)/7}.$$

Now, by using Lemma 15 with $B_{m+2} = 1$, $x = 7$, $\tilde{n}_0^{(1)} = \tilde{n}_2$ and noting that $B'_m + \sqrt{mn} \log n \leq B_m$ as $B_{m+2} = 1$, for all $n \geq \tilde{n}_0^{(2)}$, we have

$$\begin{aligned} \mathbb{P}(\bar{s}_{m+1} \geq B_m) &\leq \left(\frac{1}{n}\right)^{(m \log n)/7 - m - 2} \stackrel{(*)}{\leq} \left(\frac{1}{n}\right)^{(m \log n)/8} \\ \mathbb{P}\left(\bar{s}_m \geq n - 2m \frac{n \log d}{d} + B_m\right) &\leq \left(\frac{1}{n}\right)^{(m \log n)/7 - m - 2} \stackrel{(*)}{\leq} \left(\frac{1}{n}\right)^{(m \log n)/8}, \end{aligned}$$

where (*) follows for all $n \geq \tilde{n}_3^{(1)}$ for some $\tilde{n}_3^{(1)} \in \mathbb{Z}_+$. Thus, by defining $\tilde{n}_3 = \max\{\tilde{n}_0^{(2)}, \tilde{n}_3^{(1)}\}$, the proof is complete. \square

Proof of Lemma 10 Define the Lyapunov function:

$$\tilde{W}_j(\mathbf{s}) = s_j - n + 2m \frac{n \log d}{d^{m-j+1}} - B_j - 2(m-j)m \frac{n \log d}{d^{m-j+2}},$$

where $B_j = o(mn \log d / d^{m-j+1})$ as defined in (9). We analyze the drift of $\tilde{W}_j(\mathbf{s})$ when $\tilde{W}_j(\mathbf{s}) \geq 0$ and $\mathbf{s} \in \tilde{\mathcal{C}}_{j+1}^{(2)}$ where

$$\tilde{\mathcal{C}}_{j+1}^{(2)} = \left\{ \tilde{W}_{j+1}(\mathbf{s}) \leq \sqrt{mn} \log n \right\}.$$

Thus, we have

$$s_j \geq n - 2m \frac{n \log d}{d^{m-j+1}} + B_j + 2(m-j)m \frac{n \log d}{d^{m-j+2}} \quad (26a)$$

$$s_{j+1} \leq n - 2m \frac{n \log d}{d^{m-j}} + B_{j+1} + 2(m-j-1)m \frac{n \log d}{d^{m-j+1}} + \sqrt{mn} \log n. \quad (26b)$$

Now, the drift is given as follows:

$$\begin{aligned} \Delta \tilde{W}_j(\mathbf{s}) &= -s_j + s_{j+1} + \lambda \left(\left(\frac{s_{j-1}}{n}\right)^d - \left(\frac{s_j}{n}\right)^d \right) \\ &\stackrel{(a)}{\leq} 2m \frac{n \log d}{d^{m-j+1}} - B_j - 2(m-j)m \frac{n \log d}{d^{m-j+2}} \\ &\quad - 2m \frac{n \log d}{d^{m-j}} + B_{j+1} + 2(m-j-1)m \frac{n \log d}{d^{m-j+1}} + \sqrt{mn} \log n \\ &\quad + n \left(1 - \left(1 - 2m \frac{\log d}{d^{m-j+1}} + \frac{B_j}{n} + 2(m-j)m \frac{\log d}{d^{m-j+2}} \right)^d \right) \\ &\stackrel{(b)}{\leq} 2m \frac{n \log d}{d^{m-j+1}} - B_j - 2(m-j)m \frac{n \log d}{d^{m-j+2}} \\ &\quad - 2m \frac{n \log d}{d^{m-j}} + B_{j+1} + 2(m-j-1)m \frac{n \log d}{d^{m-j+1}} + \sqrt{mn} \log n \\ &\quad + n \left(2m \frac{\log d}{d^{m-j}} - \frac{B_{j+1}}{n} - 2(m-j)m \frac{\log d}{d^{m-j+1}} \right) \\ &\leq \sqrt{mn} \log n - B_j \\ &\leq \sqrt{mn} \log n - 18md^{j-1} \sqrt{mn} \log n \\ &\leq -\sqrt{mn} \log n, \end{aligned}$$

where (a) follows as $s_{j-1} \leq n$, and s_{j-1} and s_j is bounded as in (26) and (b) follows by Lemma 12 and noting that $dB_j = B_{j+1}$. Thus, we have $\Delta \tilde{W}_j(\mathbf{s}) \leq -\sqrt{mn} \log n$ when $\tilde{W}_j(\mathbf{s}) \geq 0$ and $\mathbf{s} \in \tilde{\mathcal{C}}_{j+1}^{(2)}$. Thus, by Lemma 14, we have

$$\begin{aligned} \mathbb{P}\left(\tilde{W}_j(\bar{\mathbf{s}}) \geq \sqrt{mn} \log n\right) &\leq \left(\frac{n}{n + \sqrt{mn} \log n}\right)^{(\sqrt{mn} \log n)/2} + \sqrt{n} \mathbb{P}\left(\bar{\mathbf{s}} \notin \tilde{\mathcal{C}}_{j+1}^{(2)}\right) \\ &\stackrel{(a)}{\leq} \left(\frac{1}{n}\right)^{(m \log n)/4} + \sqrt{n} \mathbb{P}\left(\bar{\mathbf{s}} \notin \tilde{\mathcal{C}}_{j+1}^{(2)}\right) \\ &\stackrel{(b)}{\leq} \left(\frac{1}{n}\right)^{(m \log n)/4} + \sqrt{n} \left(\frac{1}{n}\right)^{(m \log n)/8 - (m-j-1)} \\ &\stackrel{(c)}{\leq} \left(\frac{1}{n}\right)^{(m \log n)/8 - (m-j)}, \end{aligned}$$

where (a) follows by Lemma 13 and (b) follows by the induction hypothesis (IH2). Lastly, (c) follows for all $n \geq \tilde{n}_5$ for some $\tilde{n}_5 \in \mathbb{Z}_+$. Now, by setting $\tilde{n}_{IH} \geq \tilde{n}_5$, the induction step is complete. This completes the proof of the lemma. \square

Appendix E: Proof of Claims for Lower Bound

Proof of Claim 1 For $l \in \{1, \dots, k-1\}$, we consider the following induction hypothesis: There exists $n_{c1} \in \mathbb{Z}_+$ such that for all $n \geq n_{c1}$, we have

$$\mathbb{P}\left(L_{lk}(\bar{\mathbf{s}}) \geq \sqrt{mn} \log n\right) \leq \left(\frac{1}{n}\right)^{(m \log n)/4 - 4(k-1)m - (k-l)}. \quad (27)$$

Base Case: We analyze the drift of $L_{k-1,k}(\mathbf{s})$ when $L_{k-1,k}(\mathbf{s}) \geq 0$. Thus, we have

$$s_{k-1} \geq n - 3(k-1)m \frac{n \log d}{d^2} \quad (28a)$$

$$s_k \leq n - \frac{3kmn \log d}{d}. \quad (28b)$$

First, consider the case when $L_{k-1,k}^{(1)}(\mathbf{s}) \geq L_{k-1,k}^{(2)}(\mathbf{s})$. The drift is given as follows:

$$\begin{aligned} \Delta L_{k-1,k}(\mathbf{s}) &\leq \lambda \left(\left(\frac{s_{k-2}}{n}\right)^d - \left(\frac{s_{k-1}}{n}\right)^d \right) - s_{k-1} + s_k \\ &\stackrel{(a)}{\leq} \lambda \left(1 - \left(1 - 3(k-1)m \frac{\log d}{d^2}\right)^d \right) + 3(k-1)m \frac{n \log d}{d^2} - \frac{3kmn \log d}{d} \\ &\stackrel{(b)}{\leq} \lambda \left(3(k-1)m \frac{\log d}{d} \right) + 3(k-1)m \frac{n \log d}{d^2} - \frac{3kmn \log d}{d} \\ &\leq -\frac{3mn \log d}{d} + 3(k-1)m \frac{n \log d}{d^2} \\ &\stackrel{(c)}{\leq} -\frac{2mn \log d}{d} \stackrel{(d)}{\leq} -\sqrt{mn} \log n, \end{aligned}$$

where (a) follows as $s_{k-2} \leq n$, and we use the bounds on s_{k-1} and s_k given by (28). Next, (b) follows by Lemma 12. Now, (c) follows as there exists $n_{c1}^{(1)}$ such that for all $n \geq n_{c1}^{(1)}$, we have

$$\frac{mn \log d}{d} \stackrel{(c_1)}{\geq} 3m^2 \frac{n \log d}{d^2} \geq 3(k-1)m \frac{n \log d}{d^2},$$

where (c₁) follows as $m/d \rightarrow 0$ as $d \rightarrow \infty$. Lastly, (d) follows as

$$\frac{2mn \log d}{d} \geq \frac{2mn \log d}{d^m} = n^{1-\gamma} \geq \sqrt{mn} \log n.$$

where the last inequality follows as $\gamma \in (0, 0.5)$. Now, consider the case when $L_{k-1,k}^{(1)}(\mathbf{s}) \leq L_{k-1,k}^{(2)}(\mathbf{s})$. The drift is given as follows:

$$\begin{aligned} \Delta L_{k-1,k}(\mathbf{s}) &\leq s_k - s_{k+1} - \lambda \left(\left(\frac{s_{k-1}}{n} \right)^d - \left(\frac{s_k}{n} \right)^d \right) \\ &\stackrel{(a)}{\leq} n - \frac{3kmn \log d}{d} - \lambda \left(\left(1 - 3(k-1)m \frac{\log d}{d^2} \right)^d - \left(1 - \frac{3km \log d}{d} \right)^d \right) \\ &\stackrel{(b)}{\leq} n - \frac{3kmn \log d}{d} - \lambda \left(1 - 3(k-1)m \frac{\log d}{d} - \frac{2}{d^{3km}} \right) \\ &\leq -\frac{3mn \log d}{d} + \frac{2n}{d^{3km}} + n^{1-\gamma} \\ &\stackrel{(c)}{\leq} -\frac{mn \log d}{d} + \frac{2n}{d^{3km}} \stackrel{(d)}{\leq} -\sqrt{mn} \log n, \end{aligned}$$

where (a) follows as $s_{k-2} \leq n$, and we use the bounds on s_{k-1} and s_k given by (28). Next, (b) follows for all $n \geq n_{c1}^{(2)}$ for some $n_{c1}^{(2)} \in \mathbb{Z}_+$ by Lemma 12 and Lemma 11. Now, (c) follows by (29a). Lastly, (d) follows as there exists $n_{c1}^{(3)}$ such that for all $n \geq n_{c1}^{(3)}$ we have

$$\begin{aligned} \frac{mn \log d}{2d} &\geq \frac{2n}{d^{3k}} \geq \frac{2n}{d^{3km}} \\ \frac{mn \log d}{2d} &\geq \frac{mn \log d}{2d^m} \geq \frac{1}{4} n^{1-\gamma} \geq \sqrt{mn} \log n, \end{aligned}$$

where the last inequality follows as $\gamma < 0.5$. By combining the two cases, we get $\Delta L_{k-1,k}(\mathbf{s}) \leq -\sqrt{mn} \log n$ when $L_{k-1,k}(\mathbf{s}) \geq 0$. Thus, by Lemma 14, we have

$$\mathbb{P}(L_{k-1,k}(\bar{\mathbf{s}}) \geq \sqrt{mn} \log n) \leq \left(\frac{n}{n + \sqrt{mn} \log n} \right)^{(\sqrt{mn} \log n)/2} \leq \left(\frac{1}{n} \right)^{m \log n/4} \leq \left(\frac{1}{n} \right)^{m \log n/4 - 4(k-1)m-1},$$

where the last inequality follows by Lemma 13. By considering $n_{c1} \geq \max_{k \in [3]} \{n_{c1}^{(k)}\}$, the base case is complete.

Induction Step: We analyze the drift of $L_{i-1,k}(\mathbf{s})$ when $L_{i-1,k}(\mathbf{s}) \geq 0$ and $\mathbf{s} \in \mathcal{C}_{i,k}^{(1)} \cap \bigcap_{l=i}^{k-1} \mathcal{D}_l^{(1)}$ defined as in (12). Similar to the proof of Lemma 2 (Eq. (13) and (14)), there exists $n_{c1}^{(4)} \in \mathbb{Z}_+$, such that for all $n \geq n_{c1}^{(4)}$, we have

$$s_i \leq n - \frac{3imn \log d}{d^{k-i+1}} + \sqrt{mn} \log n \quad (29a)$$

$$s_k \leq n - 3m \frac{n \log d}{d}. \quad (29b)$$

We proceed by analyzing the drift for the case when $L_{i-1,k}^{(1)}(\mathbf{s}) \geq L_{i-1,k}^{(2)}(\mathbf{s})$.

$$\begin{aligned} \Delta L_{i-1,k}(\mathbf{s}) &\leq \lambda \left(\left(\frac{s_{i-2}}{n} \right)^d - \left(\frac{s_{i-1}}{n} \right)^d \right) - s_{i-1} + s_i \\ &\stackrel{(a)}{\leq} \lambda \left(1 - \left(1 - 3(i-1)m \frac{\log d}{d^{k-i+2}} \right)^d \right) + 3(i-1)m \frac{n \log d}{d^{k-i+2}} - \frac{3imn \log d}{d^{k-i+1}} + \sqrt{mn} \log n \\ &\stackrel{(b)}{\leq} \lambda \left(3(i-1)m \frac{\log d}{d^{k-i+1}} \right) + 3(i-1)m \frac{n \log d}{d^{k-i+2}} - \frac{3imn \log d}{d^{k-i+1}} + \sqrt{mn} \log n \\ &\leq -3m \frac{n \log d}{d^{k-i+1}} + 3(i-1)m \frac{n \log d}{d^{k-i+2}} + \sqrt{mn} \log n \\ &\stackrel{(c)}{\leq} -m \frac{n \log d}{d^{k-i+1}} \stackrel{(c)}{\leq} -\sqrt{mn} \log n, \end{aligned}$$

where (a) follows by lower bounding s_{i-1} using $L_{i-1,k}(\mathbf{s}) \geq 0$, upper bounding s_i using (29), and trivially upper bounding s_{i-2} by n . Next, (b) follows by Lemma 12. Lastly, (c) follows as there exists $n_{c1}^{(5)} \in \mathbb{Z}_+$ such that for all $n \geq n_{c1}^{(5)}$, we have

$$\begin{aligned} m \frac{n \log d}{d^{k-i+1}} &\stackrel{(c_1)}{\geq} 3m^2 \frac{n \log d}{d^{k-i+2}} \geq 3(i-1)m \frac{n \log d}{d^{k-i+2}} \\ m \frac{n \log d}{d^{k-i+1}} &\geq m \frac{n \log d}{d^m} \geq \frac{1}{2} n^{1-\gamma} \stackrel{(c_2)}{\geq} \sqrt{mn} \log n, \end{aligned}$$

where (c_1) follows as $m/d \leq \log n/d \rightarrow 0$ as $n \rightarrow \infty$ and (c_2) follows as $\gamma < 0.5$. Now, consider the case when $L_{i-1,k}^{(1)}(\mathbf{s}) \leq L_{i-1,k}^{(2)}(\mathbf{s})$. The drift is given as follows:

$$\begin{aligned} \Delta L_{i-1,k}(\mathbf{s}) &\stackrel{(a)}{\leq} s_i - s_{k+1} - \lambda \left(\left(\frac{s_{i-1}}{n} \right)^d - \left(\frac{s_k}{n} \right)^d \right) \\ &\stackrel{(b)}{\leq} n - \frac{3imn \log d}{d^{k-i+1}} + \sqrt{mn} \log n - \lambda \left(\left(1 - 3(i-1)m \frac{\log d}{d^{k-i+2}} \right)^d - \left(1 - 3m \frac{\log d}{d} \right)^d \right) \\ &\leq n - \frac{3imn \log d}{d^{k-i+1}} + \sqrt{mn} \log n - \lambda \left(1 - 3(i-1)m \frac{\log d}{d^{k-i+1}} - \frac{2}{d^{3m}} \right) \\ &\leq -3m \frac{n \log d}{d^{k-i+1}} + \sqrt{mn} \log n + \frac{2n}{d^{3m}} + n^{1-\gamma} \\ &\stackrel{(c)}{\leq} -m \frac{n \log d}{d^{k-i+1}} + \sqrt{mn} \log n + \frac{2n}{d^{3m}} \stackrel{(d)}{\leq} -\sqrt{mn} \log n, \end{aligned}$$

where (a) follows by lower bounding s_{i-1} using $L_{i-1,k}(\mathbf{s}) \geq 0$, upper bounding s_i and s_k using (29), and trivially lower bounding s_{k+1} by 0. Next, (b) follows for all $n \geq n_{c1}^{(6)}$ for some $n_{c1}^{(6)} \in \mathbb{Z}_+$ by Lemma 11 and Lemma 12. Further, (c) follows as

$$2m \frac{n \log d}{d^{k-i+1}} \geq 2m \frac{n \log d}{d^m} = n^{1-\gamma}.$$

Lastly, (d) follows as there exists $n_{c1}^{(7)}$ such that for all $n \geq n_{c1}^{(7)}$, we have

$$\begin{aligned} m \frac{n \log d}{2d^{k-i+1}} &\geq m \frac{n \log d}{2d^m} \geq \frac{2n}{d^{3m}} \\ m \frac{n \log d}{4d^{k-i+1}} &\geq m \frac{n \log d}{4d^m} \geq \frac{1}{8} n^{1-\gamma} \geq \sqrt{mn} \log n. \end{aligned}$$

By combining the two cases, we get $\Delta L_{i-1,k}(\mathbf{s}) \leq -\sqrt{mn} \log n$ when $L_{i-1,k}(\mathbf{s}) \geq 0$ and $\mathbf{s} \in \mathcal{C}_{i,k}^{(1)} \cap \bigcap_{l=i}^{k-1} \mathcal{D}_l^{(1)}$. Thus, by Lemma 14, we have

$$\begin{aligned} &\mathbb{P} \left(L_{i-1,k}(\bar{\mathbf{s}}) \geq \sqrt{mn} \log n \right) \\ &\leq \left(\frac{n}{n + \sqrt{mn} \log n} \right)^{(\sqrt{mn} \log n)/2} + \sqrt{n} \mathbb{P} \left(\bar{\mathbf{s}} \notin \mathcal{C}_{i,k}^{(1)} \cap \bigcap_{l=i}^{k-1} \mathcal{D}_l^{(1)} \right) \\ &\stackrel{(a)}{\leq} \left(\frac{1}{n} \right)^{(m \log n)/4} + \sqrt{n} \left(\mathbb{P} \left(\bar{\mathbf{s}} \notin \mathcal{C}_{i,k}^{(1)} \right) + \sum_{l=i}^{k-1} \mathbb{P} \left(\bar{\mathbf{s}} \notin \mathcal{D}_l^{(1)} \right) \right) \\ &\stackrel{(b)}{\leq} \left(\frac{1}{n} \right)^{(m \log n)/4} + \sqrt{n} \left(\frac{1}{n} \right)^{(m \log n)/4 - 4(k-1)m - (k-i)} + \sqrt{n} m \left(\frac{1}{n} \right)^{(m \log n)/4 - 4(k-1)m} \\ &\stackrel{(c)}{\leq} \left(\frac{1}{n} \right)^{m(\log n)/4 - 4(k-1)m - (k-i+1)}, \end{aligned}$$

where (a) follows by Lemma 13. Next, (b) follows by upper bounding $\mathbb{P} \left(\bar{\mathbf{s}} \notin \mathcal{C}_{i,k}^{(1)} \right)$ using the induction hypothesis (27). Also, similar to (15), $\mathbb{P} \left(\bar{\mathbf{s}} \notin \mathcal{D}_l^{(1)} \right)$ is upper bounded for all $n \geq n_{c1}^{(8)}$ for some $n_{c1}^{(8)} \in \mathbb{Z}_+$ by Theorem 2. Lastly, (c) follows for all $n \geq n_{c1}^{(9)}$ for some $n_{c1}^{(9)} \in \mathbb{Z}_+$. By fixing $n_{c1} \geq \max_{k \in [9]} \{n_{c1}^{(k)}\}$, the induction step is complete. This completes the proof of the claim. \square

Proof of Claim 2 For some $l \in [k-1]$, we consider the following induction hypothesis: There exists $n_{c2} \in \mathbb{Z}_+$, such that for all $n \geq n_{c2}$, we have

$$\mathbb{P} \left(Z_{lk}(\bar{\mathbf{s}}) \geq \sqrt{mn} \log n \right) \leq \left(\frac{1}{n} \right)^{m(\log n)/4 - 4(k-0.5)m - (k-l)}. \quad (30)$$

Base Case: We analyze the drift of $Z_{k-1,k}(\mathbf{s})$ when $Z_{k-1,k}(\mathbf{s}) \geq 0$. Thus, we have

$$s_{k-1} \geq n - 2m \frac{n \log d}{d^2} - (2k-1)d^{k-2} \sqrt{mn} \log n - 3km \frac{n \log d}{d^3} \quad (31a)$$

$$s_k \leq n - 2m \frac{n \log d}{d} - 2kd^{k-1} \sqrt{mn} \log n - 10m^2 \frac{n \log d}{d^2}. \quad (31b)$$

First, consider the case when $Z_{k-1,k}^{(1)}(\mathbf{s}) \geq Z_{k-1,k}^{(2)}(\mathbf{s})$. The drift is given as follows:

$$\begin{aligned}
 & \Delta Z_{k-1,k}(\mathbf{s}) \\
 & \leq \lambda \left(\left(\frac{s_{k-2}}{n} \right)^d - \left(\frac{s_{k-1}}{n} \right)^d \right) - s_{k-1} + s_k \\
 & \stackrel{(a)}{\leq} \lambda \left(1 - \left(1 - 2m \frac{\log d}{d^2} - (2k-1)d^{k-2} \frac{\sqrt{mn} \log n}{\sqrt{n}} - 3km \frac{\log d}{d^3} \right)^d \right) + 2m \frac{n \log d}{d^2} \\
 & \quad + (2k-1)d^{k-2} \sqrt{mn} \log n + 3km \frac{n \log d}{d^3} - 2m \frac{n \log d}{d} - 2kd^{k-1} \sqrt{mn} \log n - 10m^2 \frac{n \log d}{d^2} \\
 & \stackrel{(b)}{\leq} \lambda \left(2m \frac{\log d}{d} + (2k-1)d^{k-1} \frac{\sqrt{m} \log n}{\sqrt{n}} + 3km \frac{\log d}{d^2} \right) + (2k-1)d^{k-2} \sqrt{mn} \log n \\
 & \quad + 3km \frac{n \log d}{d^3} - 2m \frac{n \log d}{d} - 2kd^{k-1} \sqrt{mn} \log n - 8m^2 \frac{n \log d}{d^2} \\
 & \stackrel{(c)}{\leq} -d^{k-1} \sqrt{mn} \log n + (2k-1)d^{k-2} \sqrt{mn} \log n \\
 & \stackrel{(d)}{\leq} -\frac{1}{2}d^{k-1} \sqrt{mn} \log n \stackrel{(e)}{\leq} -\sqrt{mn} \log n,
 \end{aligned}$$

where (a) follows by upper bounding s_{k-2} by n and using the bounds on s_{k-1} and s_k given by (31). Next, (b) follows by Lemma 12. Now, (c) follows as

$$3km \frac{n \log d}{d^2} + 3km \frac{n \log d}{d^3} \leq 3m^2 \frac{n \log d}{d^2} + 3m^2 \frac{n \log d}{d^3} \leq 6m^2 \frac{n \log d}{d^2} \leq 8m^2 \frac{n \log d}{d^2}.$$

Lastly, (d) follows as there exists $n_{c2}^{(1)}$ such that for all $n \geq n_{c2}^{(1)}$, we have

$$\frac{1}{2}d^{k-1} \sqrt{mn} \log n \geq 2md^{k-2} \sqrt{mn} \log n \geq (2k-1)d^{k-2} \sqrt{mn} \log n,$$

where the first inequality follows as $m/d \leq \log n/d \rightarrow 0$ as $n \rightarrow \infty$. Lastly, (e) follows for all $n \geq n_{c2}^{(2)}$ for some $n_{c2}^{(2)} \in \mathbb{Z}_+$ as $k \geq 2$. Note that $k=1$ corresponds to the base case of (IH) which is proved in Lemma 1. Now, consider the case when $Z_{k-1,k}^{(1)}(\mathbf{s}) \leq Z_{k-1,k}^{(2)}(\mathbf{s})$. The drift is given as follows:

$$\begin{aligned}
 \Delta Z_{k-1,k}(\mathbf{s}) & \leq s_k - s_{k+1} - \lambda \left(\left(\frac{s_{k-1}}{n} \right)^d - \left(\frac{s_k}{n} \right)^d \right) \\
 & \stackrel{(a)}{\leq} n - 2m \frac{n \log d}{d} - 2kd^{k-1} \sqrt{mn} \log n - 10m^2 \frac{n \log d}{d^2} \\
 & \quad - \lambda \left(\left(1 - 2m \frac{\log d}{d^2} - (2k-1)d^{k-2} \frac{\sqrt{m} \log n}{\sqrt{n}} - 3km \frac{\log d}{d^3} \right)^d - \left(1 - 2m \frac{\log d}{d} \right)^d \right) \\
 & \stackrel{(b)}{\leq} n - 2m \frac{n \log d}{d} - 2kd^{k-1} \sqrt{mn} \log n - 10m^2 \frac{n \log d}{d^2} \\
 & \quad - \lambda \left(1 - 2m \frac{\log d}{d} - (2k-1)d^{k-1} \frac{\sqrt{m} \log n}{\sqrt{n}} - 3km \frac{\log d}{d^2} - \frac{2}{d^{2m}} \right) \\
 & \leq -d^{k-1} \sqrt{mn} \log n - 10m^2 \frac{n \log d}{d^2} + 3km \frac{n \log d}{d^2} + \frac{2n}{d^{2m}} + n^{1-\gamma} \\
 & \stackrel{(c)}{\leq} -d^{k-1} \sqrt{mn} \log n \leq -\sqrt{mn} \log n.
 \end{aligned}$$

where (a) follows by noting that $s_{k+1} \geq 0$ and using the bounds on s_{k-1} and s_k given by (31). Next, (b) follows for all $n \geq n_{c2}^{(3)}$ for some $n_{c2}^{(3)} \in \mathbb{Z}_+$ by Lemma 11 and Lemma 12. Lastly, (c) follows as

$$3km \frac{n \log d}{d^2} + \frac{2n}{d^{2m}} + n^{1-\gamma} \leq 5m^2 \frac{n \log d}{d^2} + n^{1-\gamma} \leq 5m^2 \frac{n \log d}{d^2} + 2m \frac{n \log d}{d^m} \leq 10m^2 \frac{n \log d}{d^2},$$

where the last inequality follows as $m \geq 2$. Thus, by the above two cases, we have $\Delta Z_{k-1,k}(\mathbf{s}) \leq -\sqrt{mn} \log n$ when $Z_{k-1,k}(\mathbf{s}) \geq 0$. Thus, by Lemma 14, we have

$$\mathbb{P}(Z_{k-1,k}(\bar{\mathbf{s}}) \geq \sqrt{mn} \log n) \leq \left(\frac{n}{n + \sqrt{mn} \log n} \right)^{(\sqrt{mn} \log n)/2} \leq \left(\frac{1}{n} \right)^{(m \log n)/4},$$

where the last inequality follows by Lemma 13. Thus, by considering $n_{c2} \geq \max_{k \in [3]} \{n_{c2}^{(k)}\}$, the base case is complete.

Induction Step: We analyze the drift of $Z_{i-1,k}(\mathbf{s})$ when $Z_{i-1,k}(\mathbf{s}) \geq 0$ and $\mathbf{s} \in \mathcal{C}_{i,k}^{(3)} \cap \bigcap_{l=i}^{k-1} \mathcal{D}_l^{(3)}$, where $\mathcal{C}_{i,k}^{(3)}$ and $\mathcal{D}_l^{(3)}$ are defined in (17). Similar to the proof of Lemma 4 (Eq. (18) and (19)), we can get the following bounds on s_i and s_k :

$$s_i \leq n - 2m \frac{n \log d}{d^{k-i+1}} - (2i+1)d^{i-1}\sqrt{mn} \log n - 3(i+1)m \frac{n \log d}{d^{k-i+2}} + \sqrt{mn} \log n, \quad (32)$$

$$s_k \leq n - 2m \frac{n \log d}{d}. \quad (33)$$

Now, we analyze the drift for the case when $Z_{i-1,k}^{(1)}(\mathbf{s}) \geq Z_{i-1,k}^{(2)}(\mathbf{s})$.

$$\begin{aligned} & \Delta Z_{i-1,k}(\mathbf{s}) \\ & \leq \lambda \left(\left(\frac{s_{i-2}}{n} \right)^d - \left(\frac{s_{i-1}}{n} \right)^d \right) - s_{i-1} + s_i \\ & \stackrel{(a)}{\leq} \lambda \left(1 - \left(1 - 2m \frac{\log d}{d^{k-i+2}} - (2i-1)d^{i-2} \frac{\sqrt{m} \log n}{\sqrt{n}} - 3im \frac{\log d}{d^{k-i+3}} \right)^d \right) + 2m \frac{n \log d}{d^{k-i+2}} \\ & \quad + (2i-1)d^{i-2}\sqrt{mn} \log n + 3im \frac{n \log d}{d^{k-i+3}} - 2m \frac{n \log d}{d^{k-i+1}} - (2i+1)d^{i-1}\sqrt{mn} \log n \\ & \quad - 3(i+1)m \frac{n \log d}{d^{k-i+2}} + \sqrt{mn} \log n \\ & \stackrel{(b)}{\leq} \lambda \left(2m \frac{\log d}{d^{k-i+1}} + (2i-1)d^{i-1} \frac{\sqrt{m} \log n}{\sqrt{n}} + 3im \frac{\log d}{d^{k-i+2}} \right) + 2m \frac{n \log d}{d^{k-i+2}} + (2i-1)d^{i-2}\sqrt{mn} \log n \\ & \quad + 3im \frac{n \log d}{d^{k-i+3}} - 2m \frac{n \log d}{d^{k-i+1}} - (2i+1)d^{i-1}\sqrt{mn} \log n - 3(i+1)m \frac{n \log d}{d^{k-i+2}} + \sqrt{mn} \log n \\ & \leq -2d^{i-1}\sqrt{mn} \log n - m \frac{n \log d}{d^{k-i+2}} + (2i-1)d^{i-2}\sqrt{mn} \log n + 3im \frac{n \log d}{d^{k-i+3}} + \sqrt{mn} \log n \\ & \stackrel{(c)}{\leq} -d^{i-1}\sqrt{mn} \log n \leq -\sqrt{mn} \log n, \end{aligned}$$

where (a) follows by noting that $s_{i-2} \leq n$, using the bound on s_i given by (32), and bounding s_{i-1} by using the fact that $Z_{i-1,k}(\mathbf{s}) \geq 0$. Next, (b) follows by Lemma 12. Lastly, (c) follows as there exists $n_{c2}^{(4)} \in \mathbb{Z}_+$ such that for all $n \geq n_{c2}^{(4)}$, we have

$$\begin{aligned} m \frac{n \log d}{d^{k-i+2}} & \stackrel{(c_1)}{\geq} 3m^2 \frac{n \log d}{d^{k-i+3}} \geq 3im \frac{n \log d}{d^{k-i+3}} \quad \forall i \leq m \\ d^{i-1}\sqrt{mn} \log n & \stackrel{(c_2)}{\geq} 2md^{i-2}\sqrt{mn} \log n \geq 2id^{i-2}\sqrt{mn} \log n \\ & \geq (2i-1)d^{i-2}\sqrt{mn} \log n + \sqrt{mn} \log n \quad \forall i \leq m, \end{aligned}$$

where (c₁) and (c₂) follows as $m/d \leq \log n/d \rightarrow 0$ as $n \rightarrow \infty$. Now, consider the case when $Z_{i-1,k}^{(1)}(\mathbf{s}) \leq Z_{i-1,k}^{(2)}(\mathbf{s})$. The drift is given as follows:

$$\begin{aligned} \Delta Z_{i-1,k}(\mathbf{s}) & \leq s_i - s_{k+1} - \lambda \left(\left(\frac{s_{i-1}}{n} \right)^d - \left(\frac{s_k}{n} \right)^d \right) \\ & \stackrel{(a)}{\leq} n - 2m \frac{n \log d}{d^{k-i+1}} - (2i+1)d^{i-1}\sqrt{mn} \log n - 3(i+1)m \frac{n \log d}{d^{k-i+2}} + \sqrt{mn} \log n \\ & \quad - \lambda \left(\left(1 - 2m \frac{\log d}{d^{k-i+2}} - (2i-1)d^{i-2} \frac{\sqrt{m} \log n}{\sqrt{n}} - 3im \frac{\log d}{d^{k-i+3}} \right)^d - \left(1 - 2m \frac{\log d}{d} \right)^d \right) \\ & \stackrel{(b)}{\leq} n - 2m \frac{n \log d}{d^{k-i+1}} - (2i+1)d^{i-1}\sqrt{mn} \log n - 3(i+1)m \frac{n \log d}{d^{k-i+2}} + \sqrt{mn} \log n \\ & \quad - \lambda \left(1 - 2m \frac{\log d}{d^{k-i+1}} - (2i-1)d^{i-1} \frac{\sqrt{m} \log n}{\sqrt{n}} - 3im \frac{\log d}{d^{k-i+2}} - \frac{2}{d^{2m}} \right) \\ & \leq -2d^{i-1}\sqrt{mn} \log n - 3m \frac{n \log d}{d^{k-i+2}} + \frac{2n}{d^{2m}} + n^{1-\gamma} + \sqrt{mn} \log n \\ & \stackrel{(c)}{\leq} -2d^{i-1}\sqrt{mn} \log n + \sqrt{mn} \log n \leq -\sqrt{mn} \log n, \end{aligned}$$

where (a) follows by using the bound on s_i and s_k given by (32) and (33), and bounding s_{i-1} by using the fact that $Z_{i-1,k}(\mathbf{s}) \geq 0$. Next, (b) follows for all $n \geq n_{c_2}^{(5)}$ for some $n_{c_2}^{(5)} \in \mathbb{Z}_+$ by Lemma 11 and Lemma 12. Lastly, (c) follows as there exists $n_{c_2}^{(6)}$ such that for all $n \geq n_{c_2}^{(6)}$, we have

$$3m \frac{n \log d}{d^{k-i+2}} \stackrel{(c_1)}{\geq} 3m \frac{n \log d}{d^m} \geq \frac{2n}{d^{2m}} + 2m \frac{n \log d}{d^m} \geq \frac{2n}{d^{2m}} + n^{1-\gamma},$$

where (c₁) follows as $i \geq 2$ and $k \leq m$. By the above two cases, we get $\Delta Z_{i-1,k}(\mathbf{s}) \leq -\sqrt{mn} \log n$ when $Z_{i-1,k}(\mathbf{s}) \geq 0$ and $\mathbf{s} \in \mathcal{C}_{i,k}^{(3)} \cap \bigcap_{l=i}^{k-1} \mathcal{D}_l^{(3)}$. Thus, by Lemma 14, we have

$$\begin{aligned} & \mathbb{P}(Z_{i-1,k}(\bar{\mathbf{s}}) \geq \sqrt{mn} \log n) \\ & \leq \left(\frac{n}{n + \sqrt{mn} \log n} \right)^{(m \log n)/2} + \sqrt{n} \mathbb{P}\left(\bar{\mathbf{s}} \notin \mathcal{C}_{i,k}^{(3)} \cap \bigcap_{l=i}^{k-1} \mathcal{D}_l^{(3)}\right) \\ & \stackrel{(a)}{\leq} \left(\frac{1}{n} \right)^{(m \log n)/4} + \sqrt{n} \left(\mathbb{P}(\bar{\mathbf{s}} \notin \mathcal{C}_{i,k}^{(3)}) + \sum_{l=i}^{k-1} \mathbb{P}(\bar{\mathbf{s}} \notin \mathcal{D}_l^{(3)}) \right) \\ & \stackrel{(b)}{\leq} \left(\frac{1}{n} \right)^{(m \log n)/4} + \sqrt{n} \left(\frac{1}{n} \right)^{(m \log n)/4 - 4(k-0.5)m - (k-i)} + \sqrt{nm} \left(\frac{1}{n} \right)^{(m \log n)/4 - 4(k-1)m} \\ & \stackrel{(c)}{\leq} \left(\frac{1}{n} \right)^{(m \log n)/4 - 4(k-0.5)m - (k-i+1)}, \end{aligned}$$

where (a) follows by Lemma 13. Next, (b) follows by upper bounding $\mathbb{P}(\bar{\mathbf{s}} \notin \mathcal{C}_{i,k}^{(3)})$ using the inducting hypothesis given by (30). Also, $\mathbb{P}(\bar{\mathbf{s}} \notin \mathcal{D}_l^{(3)})$ is upper bounded for all $n \geq n_2$ by Lemma 3. Lastly, (c) follows for all $n \geq n_{c_2}^{(7)}$ for some $n_{c_2}^{(7)} \in \mathbb{Z}_+$. By fixing $n_{c_2} \geq \max_{k \in [7]} \{n_{c_2}^{(k)}, n_2\}$, the induction step is complete. \square

Appendix F: Proof of Claims for Theorem 3

Proof of Claim 3 The proof is induction based. The induction hypothesis is as follows. There exists $\tilde{n}_c \in \mathbb{Z}_+$ such that for all $n \geq \tilde{n}_c$, we have

$$\mathbb{P}(U_j(\bar{\mathbf{s}}) \geq \sqrt{mn} \log n) \leq \left(\frac{1}{n} \right)^{\frac{m \log n}{\max\{x,5\}} - (m-j+1)}. \quad (34)$$

Base Case ($j = m$): We analyze the drift of $U_m(\mathbf{s})$ when $U_m(\mathbf{s}) \geq 0$ and $\mathbf{s} \in \tilde{\mathcal{D}}_{m-1} \cap \tilde{\mathcal{D}}_{m+2}$ where

$$\begin{aligned} \tilde{\mathcal{D}}_{m-1} &= \left\{ s_{m-1} \geq n - \left(2m \frac{n \log d}{d^2} + 4md^{m-2} \sqrt{mn} \log n + 16m^3 \frac{n \log(d)^2}{d^3} \right) \mathbb{1}\{m > 1\} \right\}. \\ \tilde{\mathcal{D}}_{m+2} &= \left\{ \sum_{l=m+2}^b s_l \leq B_{m+2} \right\}. \end{aligned} \quad (35)$$

As $U_m(\mathbf{s}) \geq 0$, we get the following bounds on s_m and s_{m+1} :

$$s_{m+1} \geq 8md^{m-1} \sqrt{mn} \log n + 18m^3 \frac{n \log(d)^2}{d^2} + n^{1-\gamma} \mathbb{1}\{m > 1\} - 2 \quad (36a)$$

$$s_m \leq n - 2m \frac{n \log d}{d} + 3md^{m-1} \sqrt{mn} \log n + m^2 \frac{n \log d}{d^2}, \quad (36b)$$

where (36a) follows by considering two cases. If $B_{m+2} \geq 2$, then we use the bound $\sum_{l=m+1}^b s_l \leq bs_{m+1}$ to obtain (36a). Else if $B_{m+2} < 2$, then we use the bound $\sum_{l=m+1}^b s_l \leq s_{m+1} + B_{m+2} \leq s_{m+1} + 2$ to obtain (36a). First, consider the case when $U_m^{(1)}(\mathbf{s}) \leq U_m^{(2)}(\mathbf{s})$. In this case, the drift is as follows:

$$\begin{aligned} & \Delta U_m(\mathbf{s}) \\ & \leq -s_{m+1} + \lambda \left(\left(\frac{s_m}{n} \right)^d - \left(\frac{s_b}{n} \right)^d \right) \\ & \stackrel{(a)}{\leq} -8md^{m-1} \sqrt{mn} \log n - 18m^3 \frac{n \log(d)^2}{d^2} + 2 + n \left(1 - 2m \frac{\log d}{d} + 3m \sqrt{md}^{m-1} \frac{\log n}{\sqrt{n}} + m^2 \frac{\log d}{d^2} \right)^d \\ & \stackrel{(b)}{\leq} -8md^{m-1} \sqrt{mn} \log n - 18m^3 \frac{n \log(d)^2}{d^2} + 2 + \frac{2n}{d^{2m}} \stackrel{(c)}{\leq} -\sqrt{mn} \log n, \end{aligned}$$

where (a) follows as $s_b \geq 0$, and s_m and s_{m+1} are bounded as in (36). Next, (b) follows by Lemma 11. Lastly, (c) follows as $18m^3 \frac{n \log(d)^2}{d^2} \geq \frac{2n}{d^{2m}}$ and $8md^{m-1} \sqrt{mn} \log n \geq \sqrt{mn} \log n$. Now, consider the case when $U_m^{(1)}(\mathbf{s}) \geq U_m^{(2)}(\mathbf{s})$. In this case, the drift is as follows:

$$\begin{aligned}
& \Delta U_m(\mathbf{s}) \\
& \leq s_m - s_{m+1} - \lambda \left(\left(\frac{s_{m-1}}{n} \right)^d - \left(\frac{s_m}{n} \right)^d \right) \\
& \stackrel{(a)}{\leq} n - 2m \frac{n \log d}{d} + 3md^{m-1} \sqrt{mn} \log n + m^2 \frac{n \log d}{d^2} - 8md^{m-1} \sqrt{mn} \log n - 18m^3 \frac{n \log(d)^2}{d^2} \\
& \quad - n^{1-\gamma} \mathbb{1}\{m > 1\} + 2 - \lambda \left(1 - \left(2m \frac{\log d}{d^2} + 4md^{m-2} \frac{\sqrt{m} \log n}{\sqrt{n}} + 16m^3 \frac{\log(d)^2}{d^3} \right) \mathbb{1}\{m > 1\} \right)^d \\
& \quad + \lambda \left(1 - 2m \frac{\log d}{d} + 3md^{m-1} \frac{\sqrt{m} \log n}{\sqrt{n}} + m^2 \frac{\log d}{d^2} \right)^d \\
& \stackrel{(b)}{\leq} n - 2m \frac{n \log d}{d} - 5md^{m-1} \sqrt{mn} \log n - 17m^3 \frac{n \log(d)^2}{d^2} - n^{1-\gamma} \mathbb{1}\{m > 1\} + 2 \\
& \quad - \lambda \left(1 - 2m \frac{\log d}{d} \mathbb{1}\{m > 1\} - 4md^{m-1} \frac{\sqrt{m} \log n}{\sqrt{n}} - 16m^3 \frac{\log(d)^2}{d^2} - \frac{2}{d^{2m}} \right) \\
& \stackrel{(c)}{\leq} -md^{m-1} \sqrt{mn} \log n - m^3 \frac{n \log(d)^2}{d^2} + \frac{2n}{d^{2m}} + 2 \\
& \stackrel{(d)}{\leq} -\sqrt{mn} \log n,
\end{aligned}$$

where (a) follows by substituting bounds on s_{m-1} , s_m , and s_{m+1} given by (35) and (36). Next, (b) follows by Lemma 12 and Lemma 11. Now, (c) follows as $2mn \log d/d \geq n^{1-\gamma}$. Lastly, (d) follows as $m^3 \frac{n \log(d)^2}{d^2} \geq \frac{2n}{d^{2m}}$ and $md^{m-1} \sqrt{mn} \log n \geq \sqrt{mn} \log n$. Thus, by the above two cases, we have $\Delta U_m(\mathbf{s}) \leq -\sqrt{mn} \log n$ when $U_m(\mathbf{s}) \geq 0$ and $\mathbf{s} \in \mathcal{D}_{m-1} \cap \tilde{\mathcal{D}}_{m+2}$. Combining the two cases and using Lemma 14, we get

$$\begin{aligned}
\mathbb{P}(U_m(\bar{\mathbf{s}}) \geq \sqrt{mn} \log n) & \leq \left(\frac{n}{n + \sqrt{mn} \log n} \right)^{(\sqrt{mn} \log n)/2} + \mathbb{P}(\bar{\mathbf{s}} \notin \tilde{\mathcal{D}}_{m-1} \cap \tilde{\mathcal{D}}_{m+2}) \\
& \stackrel{(a)}{\leq} \left(\frac{1}{n} \right)^{(m \log n)/4} + \sqrt{n} \mathbb{P}(\bar{\mathbf{s}} \notin \tilde{\mathcal{D}}_{m-1}) + \sqrt{n} \mathbb{P}(\bar{\mathbf{s}} \notin \tilde{\mathcal{D}}_{m+2}) \\
& \stackrel{(b)}{\leq} \left(\frac{1}{n} \right)^{(m \log n)/4} + \left(\frac{1}{n} \right)^{(m \log n)/5-0.5} + \sqrt{n} \mathbb{P}(\bar{\mathbf{s}} \notin \tilde{\mathcal{D}}_{m+2}) \\
& \stackrel{(c)}{\leq} \left(\frac{1}{n} \right)^{(m \log n)/4} + \left(\frac{1}{n} \right)^{(m \log n)/5-0.5} + \left(\frac{1}{n} \right)^{(m \log n)/x-0.5} \\
& \stackrel{(d)}{\leq} \left(\frac{1}{n} \right)^{\frac{m \log n}{\max\{x, 5\}}-1},
\end{aligned}$$

where (a) follows by Lemma 13. Next, if $m = 1$, then (b) follows trivially as $\mathbb{P}(\bar{\mathbf{s}} \notin \tilde{\mathcal{D}}_{m-1}) = 0$. Else if, $m > 1$, then (b) follows for all $n \geq n_{LB}$ by Theorem 2. Now, (c) follows by the high probability upper bound on $\sum_{l=m+2}^b \bar{s}_l$ assumed in the statement of the lemma. Lastly, (d) follows for all $n \geq \tilde{n}_c^{(1)}$ for some $\tilde{n}_c^{(1)} \in \mathbb{Z}_+$. By considering $\tilde{n}_c \geq \tilde{n}_c^{(1)}$, the base case is complete.

Induction Step: We will analyze the drift of $U_{j-1}(\mathbf{s})$ when $U_{j-1}(\mathbf{s}) \geq 0$ and $\mathbf{s} \in \tilde{\mathcal{C}}_j^{(1)} \cap \tilde{\mathcal{D}}_{j-2}^{(1)}$ where

$$\begin{aligned}
\tilde{\mathcal{C}}_j^{(1)} & = \{U_j(\mathbf{s}) \leq \sqrt{mn} \log n\} \\
\tilde{\mathcal{D}}_{j-2}^{(1)} & = \left\{ s_{j-2} \geq n - 2m \frac{n \log d}{d^{m-j+3}} - 4md^{j-3} \sqrt{mn} \log n - 16m^3 \frac{n \log(d)^2}{d^{m-j+4}} \right\}.
\end{aligned} \tag{37}$$

Now, we obtain a useful lower bound on s_j as follows:

$$\begin{aligned}
U_{j-1}(\mathbf{s}) \geq 0 & \Rightarrow L_{m+1}(\mathbf{s}) - \sum_{l=j+1}^m L_l(\mathbf{s}) \geq L_j(\mathbf{s}) \\
& \stackrel{(*)}{\Rightarrow} L_j(\mathbf{s}) \leq \sqrt{mn} \log n \\
& \Rightarrow s_j \geq n - 2m \frac{n \log d}{d^{m-j+1}} + 3jd^{j-1} \sqrt{mn} \log n + jm \frac{n \log d}{d^{m-j+2}} - \sqrt{mn} \log n,
\end{aligned} \tag{38}$$

where (*) follows as $\mathbf{s} \in \tilde{C}_j^{(1)}$. Next, we obtain a useful upper bound on s_{j-1} as follows:

$$U_{j-1}(\mathbf{s}) \geq 0 \Rightarrow s_{j-1} \leq n - 2m \frac{n \log d}{d^{m-j+2}} + 3(j-1)d^{j-2}\sqrt{mn} \log n + (j-1)m \frac{n \log d}{d^{m-j+3}}. \quad (39)$$

First, consider the case when $U_j^{(1)}(\mathbf{s}) \leq U_j^{(2)}(\mathbf{s})$. The drift is given as follows:

$$\begin{aligned} \Delta U_{j-1}(\mathbf{s}) &\leq -s_j + \lambda \left(\left(\frac{s_{j-1}}{n} \right)^d - \left(\frac{s_b}{n} \right)^d \right) \\ &\stackrel{(a)}{\leq} -n + 2m \frac{n \log d}{d^{m-j+1}} - 3jd^{j-1}\sqrt{mn} \log n - jm \frac{n \log d}{d^{m-j+2}} + \sqrt{mn} \log n \\ &\quad + n \left(1 - 2m \frac{\log d}{d^{m-j+2}} + 3(j-1)d^{j-2} \frac{\sqrt{m} \log n}{\sqrt{n}} + (j-1)m \frac{\log d}{d^{m-j+3}} \right)^d \\ &\stackrel{(b)}{\leq} -n + 2m \frac{n \log d}{d^{m-j+1}} - 3jd^{j-1}\sqrt{mn} \log n - jm \frac{n \log d}{d^{m-j+2}} + \sqrt{mn} \log n \\ &\quad + n \left(1 - 2m \frac{\log d}{d^{m-j+1}} + 3(j-1)d^{j-1} \frac{\sqrt{m} \log n}{\sqrt{n}} + (j-1)m \frac{\log d}{d^{m-j+2}} \right) \\ &\leq -3d^{j-1}\sqrt{mn} \log n - m \frac{n \log d}{d^{m-j+2}} + \sqrt{mn} \log n \\ &\leq -\sqrt{mn} \log n, \end{aligned}$$

where (a) follows as $s_b \geq 0$, $\lambda \leq n$, and s_{j-1} and s_j are bounded as in (39) and (38) respectively. Next, (b) follows by Lemma 12. Now, consider the case when $U_j^{(1)}(\mathbf{s}) \geq U_j^{(2)}(\mathbf{s})$. The drift is given as follows:

$$\begin{aligned} &\Delta U_{j-1}(\mathbf{s}) \\ &\leq s_{j-1} - s_j - \lambda \left(\left(\frac{s_{j-2}}{n} \right)^d - \left(\frac{s_{j-1}}{n} \right)^d \right) \\ &\stackrel{(a)}{\leq} -2m \frac{n \log d}{d^{m-j+2}} + 3(j-1)d^{j-2}\sqrt{mn} \log n + (j-1)m \frac{n \log d}{d^{m-j+3}} \\ &\quad + 2m \frac{n \log d}{d^{m-j+1}} - 3jd^{j-1}\sqrt{mn} \log n - jm \frac{n \log d}{d^{m-j+2}} + \sqrt{mn} \log n \\ &\quad - \lambda \left(1 - 2m \frac{\log d}{d^{m-j+3}} - 4md^{j-3} \frac{\sqrt{m} \log n}{\sqrt{n}} - 16m^3 \frac{\log(d)^2}{d^{m-j+4}} \right)^d \\ &\quad + \lambda \left(1 - 2m \frac{\log d}{d^{m-j+2}} + 3(j-1)d^{j-2} \frac{\sqrt{m} \log n}{\sqrt{n}} + (j-1)m \frac{\log d}{d^{m-j+3}} \right)^d \\ &\stackrel{(b)}{\leq} -2m \frac{n \log d}{d^{m-j+2}} + 3(j-1)d^{j-2}\sqrt{mn} \log n + (j-1)m \frac{n \log d}{d^{m-j+3}} \\ &\quad + 2m \frac{n \log d}{d^{m-j+1}} - 3jd^{j-1}\sqrt{mn} \log n - jm \frac{n \log d}{d^{m-j+2}} + \sqrt{mn} \log n \\ &\quad - \lambda \left(1 - 2m \frac{\log d}{d^{m-j+2}} - 4md^{j-2} \frac{\sqrt{m} \log n}{\sqrt{n}} - 16m^3 \frac{\log(d)^2}{d^{m-j+3}} \right) \\ &\quad + \lambda \left(1 - 2m \frac{\log d}{d^{m-j+1}} + 3(j-1)d^{j-1} \frac{\sqrt{m} \log n}{\sqrt{n}} + (j-1)m \frac{\log d}{d^{m-j+2}} \right) \\ &\quad + \frac{3\lambda}{2} \left(4m^2 \frac{\log(d)^2}{d^{2m-2j+2}} + 9m(j-1)^2 d^{2j-2} \frac{\log(n)^2}{n} + (j-1)^2 m^2 \frac{\log(d)^2}{d^{2m-2j+4}} \right) \\ &\leq -3d^{j-1}\sqrt{mn} \log n - m \frac{n \log d}{d^{m-j+2}} + (4m + 3(j-1))d^{j-2}\sqrt{mn} \log n + m(16m^2 + j-1) \frac{n \log(d)^2}{d^{m-j+3}} \\ &\quad + \sqrt{mn} \log n + 6m^2 \frac{n \log(d)^2}{d^{2m-2j+2}} + \frac{27m}{2}(j-1)^2 d^{2j-2} \log(n)^2 + \frac{3}{2}(j-1)^2 m^2 \frac{n \log(d)^2}{d^{2m-2j+4}} \\ &\stackrel{(c)}{\leq} -3d^{j-1}\sqrt{mn} \log n + (4m + 3(j-1))d^{j-2}\sqrt{mn} \log n + \sqrt{mn} \log n \\ &\stackrel{(d)}{\leq} -d^{j-1}\sqrt{mn} \log n \leq -\sqrt{mn} \log n, \end{aligned}$$

where (a) follows by using the bounds on s_{j-2} , s_{j-1} and s_j given by (37), (39) and (38) respectively. Next, (b) follows by Lemma 12. Now, (c) follows as there exists $\tilde{n}_c^{(2)} \in \mathbb{Z}_+$ independent of j such that for all $n \geq \tilde{n}_c^{(2)}$,

we have

$$\begin{aligned} \frac{m}{3} \frac{n \log d}{d^{m-j+2}} &\stackrel{(c_1)}{\geq} 17m^3 \frac{n \log(d)^2}{d^{m-j+3}} \geq m (16m^2 + (j-1)) \frac{n \log(d)^2}{d^{m-j+3}} \\ \frac{m}{3} \frac{n \log d}{d^{m-j+2}} &\stackrel{(c_2)}{\geq} 7m^2 \frac{n \log(d)^2}{d^{2m-2j+2}} \stackrel{(c_3)}{\geq} 6m^2 \frac{n \log(d)^2}{d^{2m-2j+2}} + \frac{3}{2}(j-1)^2 m^2 \frac{n \log(d)^2}{d^{2m-2j+4}} \\ \frac{m}{3} \frac{n \log d}{d^{m-j+2}} &\stackrel{(c_4)}{=} \frac{1}{12m \log d} d^{m+j-2} n^{1-2\gamma} \stackrel{(c_5)}{\geq} \frac{1}{12m \log d} d^{2j-2} n^{1-2\gamma} \stackrel{(c_6)}{\geq} \frac{27m}{2} (j-1)^2 d^{2j-2} \log(n)^2, \end{aligned}$$

where (c₁) follows as $m^2 \log d/d \leq \log(n)^2/(d \log d) \rightarrow 0$ as $n \rightarrow \infty$. Next, (c₂) follows as $m-j+2 < 2m-2j+2$ as $j \leq m-1$. Further, (c₃) follows as $(j-1)^2/d^2 \leq m^2/d^2 \rightarrow 0$ as $n \rightarrow \infty$. Lastly, (c₄) follows by noting that $2mn \log d/d^m = n^{1-\gamma}$ as m is assumed to be an integer, (c₅) follows as $m \geq j$, and (c₆) follows as $1-2\gamma > 0$. Next, (d) follows as there exists $\tilde{n}_c^{(3)} \in \mathbb{Z}_+$ independent of j such that for all $n \geq \tilde{n}_c^{(3)}$, we have

$$\begin{aligned} d^{j-1} \sqrt{mn} \log n &\geq \sqrt{mn} \log n \\ d^{j-1} \sqrt{mn} \log n &\stackrel{(*)}{\geq} 7m d^{j-2} \sqrt{mn} \log n \geq (4m + 3(j-1)) d^{j-2} \sqrt{mn} \log n, \end{aligned}$$

where (*) follows as $m/d \rightarrow 0$ as $n \rightarrow \infty$. By combining the above two cases, we get $\Delta U_{j-1}(\mathbf{s}) \leq -\sqrt{mn} \log n$ when $U_{j-1}(\mathbf{s}) \geq 0$ and $\mathbf{s} \in \tilde{\mathcal{C}}_j^{(1)} \cap \tilde{\mathcal{D}}_{j-2}^{(1)}$. Now, for all $n \geq \max\{\tilde{n}_c^{(2)}, \tilde{n}_c^{(3)}\}$, using Lemma 14, we get

$$\begin{aligned} \mathbb{P}(U_{j-1}(\bar{\mathbf{s}}) \geq \sqrt{mn} \log n) &\leq \left(\frac{n}{n + \sqrt{mn} \log n} \right)^{(\sqrt{mn} \log n)/2} + \sqrt{n} \mathbb{P}(\bar{\mathbf{s}} \notin \tilde{\mathcal{C}}_j^{(1)} \cap \tilde{\mathcal{D}}_{j-2}^{(1)}) \\ &\stackrel{(a)}{\leq} \left(\frac{1}{n} \right)^{(m \log n)/4} + \sqrt{n} \left(\mathbb{P}(\bar{\mathbf{s}} \notin \tilde{\mathcal{C}}_j^{(1)}) + \mathbb{P}(\bar{\mathbf{s}} \notin \tilde{\mathcal{D}}_{j-2}^{(1)}) \right) \\ &\stackrel{(b)}{\leq} \left(\frac{1}{n} \right)^{(m \log n)/4} + \left(\frac{1}{n} \right)^{\frac{m \log n}{\max\{x,5\}} - (m+1-j) - 0.5} + \sqrt{n} \mathbb{P}(\bar{\mathbf{s}} \notin \tilde{\mathcal{D}}_{j-2}^{(1)}) \\ &\stackrel{(c)}{\leq} \left(\frac{1}{n} \right)^{(m \log n)/4} + \left(\frac{1}{n} \right)^{\frac{m \log n}{\max\{x,5\}} - (m+1-j) - 0.5} + \left(\frac{1}{n} \right)^{(m \log n)/5 - 0.5} \\ &\stackrel{(d)}{\leq} \left(\frac{1}{n} \right)^{\frac{m \log n}{\max\{x,5\}} - (m+2-j)}, \end{aligned}$$

where (a) follows by Lemma 13, (b) follows by the induction hypothesis (34), and (c) follows by Theorem 2. Lastly, (d) follows for all $n \geq \tilde{n}_c^{(4)}$ for some $\tilde{n}_c^{(4)} \in \mathbb{Z}_+$. By considering $\tilde{n}_c \geq \max_{k \in \{2,3,4\}} \{\tilde{n}_c^{(k)}\}$, the induction step is complete. \square

Appendix G: Proof of Preliminary Lemmas

Proof of Lemma 11 Let d_0 be such that $(r \log d - df(d))/d > -1$ as $df(d) \rightarrow 0$. Now, we have

$$\begin{aligned} \log \left(d^r \left(1 - r \frac{\log d}{d} + f(d) \right)^d \right) &= d \log \left(1 - r \frac{\log d}{d} + f(d) \right) + r \log d \\ &\leq df(d) \rightarrow 0 \text{ as } d \rightarrow \infty, \end{aligned} \tag{40}$$

where the last inequality follows as $\log(1+x) \leq x$ for $x > -1$. Lastly, note that (40) implies that $\limsup_{d \rightarrow \infty} d^r \left(1 - r \frac{\log d}{d} + f(d) \right)^d \leq 1$ which completes the proof. \square

Proof of Lemma 12 Note that $1 - xd \leq (1-x)^d$ holds for all $x \leq 1$ by the Bernoulli's inequality. This completes the first part of the lemma. Now to prove the second part of the lemma, by Binomial series expansion, we have

$$\begin{aligned} (1-f(d))^d &= \sum_{k=0}^{\infty} \binom{d}{k} (-f(d))^k \\ &= 1 - df(d) + \frac{1}{2} d(d-1) f(d)^2 - \frac{1}{6} d(d-1)(d-2) f(d)^3 + \sum_{k=4}^{\infty} \binom{d}{k} (-f(d))^k \\ &\stackrel{(a)}{\leq} 1 - df(d) + \frac{1}{2} d^2 f(d)^2 - \frac{1}{6} d(d-1)(d-2) f(d)^3 + \frac{1}{6} d(d-1)(d-2) f(d)^3 \sum_{k=1}^{\infty} (df(d))^k \end{aligned}$$

$$\begin{aligned}
&= 1 - df(d) + \frac{1}{2}d^2 f(d)^2 - \frac{1}{6}d(d-1)(d-2)f(d)^3 \left(1 - \frac{df(d)}{1 - df(d)}\right) \\
&\stackrel{(b)}{\leq} 1 - df(d) + \frac{1}{2}d^2 f(d)^2 - \frac{1}{12}d(d-1)(d-2)f(d)^3 \\
&\stackrel{(c)}{\leq} 1 - df(d) + \frac{1}{2}d^2 f(d)^2,
\end{aligned}$$

where (a) follows as $f(d) \geq 0$ for all $d \geq d_1$, and we have $\binom{d}{k} \leq (d(d-1)(d-2)d^k/6)$ for all $k \geq 4$. Next, (b) follows for all $d \geq d_2$ for some $d_2 \geq d_1$ as $df(d) \rightarrow 0$. Lastly, (c) follows for all $d \geq d_1$ as $f(d) \geq 0$. This completes the proof. \square

Proof of Lemma 13 We have

$$\begin{aligned}
\left(\frac{n}{n + \sqrt{mn} \log n}\right)^{(\sqrt{mn} \log n)/2} &= \left(1 + \frac{\sqrt{m} \log n}{\sqrt{n}}\right)^{-(\sqrt{mn} \log n)/2} = e^{-m \log(n)^2/2 \left(\frac{\log(1 + \sqrt{m} \log n / \sqrt{n})}{\sqrt{m} \log n / \sqrt{n}}\right)} \\
&\leq e^{-m \log(n)^2/4} = \left(\frac{1}{n}\right)^{(m \log n)/4},
\end{aligned}$$

where the inequality holds due to the following. As $\sqrt{m} \log n / \sqrt{n} \rightarrow 0$, we have $\log(1 + \sqrt{m} \log n / \sqrt{n}) / (\sqrt{m} \log n / \sqrt{n}) \rightarrow 1$. Thus, there exists $n_a \in \mathbb{Z}_+$ such that for all $n \geq n_a$, we have $\log(1 + \sqrt{m} \log n / \sqrt{n}) / (\sqrt{m} \log n / \sqrt{n}) \geq 0.5$. This completes the proof. \square