Large Tandem Queueing Networks With Blocking

James B. Martin[1]

University of Cambridge

Abstract

Systems consisting of many queues in series have been considered by Glynn and Whitt (1991) and Baccelli, Borovkov and Mairesse (2000). We extend their results to apply to situations where the queues have finite capacity and so various types of "blocking" can occur. The models correspond to max-plus type recursions, of simple form but in infinitely many dimensions; they are related to "percolation" problems of finding paths of maximum weight through a two-dimensional lattice with random weights at the vertices. Topics treated include: laws of large numbers for the speed of customers progressing through the system; stationary behaviour for systems with external arrival processes; a functional central limit theorem describing the behaviour of the "front of the wave" progressing through a system which starts empty; stochastic orderings for waiting times of customers at successive queues. Several open problems are noted.

1 Introduction

In this paper we study the asymptotics of a system consisting of many queues in series. Such systems have been studied by Glynn and Whitt [10] and by Baccelli, Borovkov and Mairesse [1] in the case where the queues have unlimited waiting room. We extend some of the results of those papers to situations where the queues have finite capacity, and so *blocking* occurs.

The underlying model consists of a half-infinite series of queues, labelled by $k \in \mathbb{Z}_+$. On leaving queue k, a customer proceeds to queue $k+1$. All queues behave identically (except in some respects the first queue, labelled 0, which will always have infinite capacity). No overtaking occurs between customers, and so the customers are served in the same order at every queue – they are labelled by $n \in \mathbb{Z}_+$ or by $n \in \mathbb{Z}$, according to the precise model, and retain their label as they move between queues. Customer n precedes customer $n+1$ at each queue. We write $S(n,k)$ for the service time of customer n at queue k, which will be non-negative and will in general be random.

[1]Correspondence address: Jesus College, Jesus Lane, Cambridge CB5 8BL, UK. Email: James.Martin@ccsr.cam.ac.uk

Keywords: tandem queueing network, last-passage percolation, hydrodynamic limit, superadditivity, greedy lattice animal.

We consider a fairly general form of blocking, along the lines of the framework of Cheng and Yao [5], in which constraints are possible both on the total capacity of a queue and on the number of already-served customers who can be contained in a queue when their progress to the next queue is blocked. This framework includes various well-known types of blocking such as *blocking before service*, *blocking after service* and various types of *Kanban* blocking.

We will mainly consider two variants of the model. In the first case, queues $k \geq 1$ are taken to be empty at time 0, while queue 0 contains an infinite supply of customers. This corresponds to the situation considered in [10]. In the second case, we consider a system with an external arrival process of customers arriving at queue 0; here we can address the issue of the stationary behaviour of the network (and the first question is whether the system has a well-defined non-trivial behaviour at all).

Let $D(n,k)$ be the time that customer n finishes service at queue k. Our main interest will be in the asymptotic behaviour of $D(n,k)$ as k becomes large or as n and k become large together.

In Section 2 we describe the models in detail. We derive recurrence relations for the variables $\{D(n,k)\}$ in terms of the service times $\{S(n,k)\}$; these recurrences represent the "local behaviour" of the system, and determine which kind of blocking occurs. The boundary conditions applied then determine which of the above-mentioned variants of the model is considered. We develop a representation for $D(n,k)$ as a "path of greatest weight" through a graph corresponding to a square lattice; the weight at point (n,k) corresponds to the service time $S(n,k)$. This graphical representation (which can be regarded as a "last-passage percolation" model) will be used throughout the paper.

In Section 3 we obtain a functional central limit theorem for the collection $\{D(n,k), n \geq 0\}$ as $k \to \infty$, which applies to systems empty at time 0 (either with a saturated initial queue or with a general one-sided arrival process). The limit describes the form of the "front of the wave" of customers progressing through the series of queues. Glynn and Whitt proved this result in [10] for the case of queues with infinite capacity. We show that the same limit holds in the presence of blocking; thus we may say that, over time, the first few customers meet sufficiently rarely that the blocking does not affect the form of the front of the wave.

In Section 4, we look at various stochastic ordering properties among the inter-service times of successive customers at the same queue, and of the same customer at successive queues, for the model with a saturated initial queue at time 0. For queues with infinite capacity, we relax the conditions needed to derive the results of Section 5 of [10] (we do not require the service times to be independent). We find that these results carry over partially, but not completely, to cases of queues with finite capacity.

The most substantial section is Section 5, in which we prove various *hydrodynamic limits* or *directional laws of large numbers* for the behaviour of $D(n,k)$ as n and k grow large together, under the assumption that the service times $\{S(n,k)\}$ are i.i.d.

We first treat the model with saturated initial queue at time 0; we derive a *growth function* γ such that

$$\frac{D(\lfloor xk \rfloor, k)}{k} \to \gamma(x) \text{ as } k \to \infty$$

for all $x \geq 0$ (Theorem 5.1). (This corresponds to a "shape theorem" for the associated percolation model). For queues with infinite capacity, this was done by Srinivasan [19] for the case where the service times have exponential distribution; by Glynn and Whitt [10] for general distributions with an exponential tail; and by Baccelli *et al.* [1] for distributions satisfying a much weaker moment condition. We follow [1] in applying the theory of *greedy lattice animals* (introduced in [6] and [9] — the precise results we use here are from [14]) to provide a condition on the distribution of the service times $\{S(n,k)\}$ which implies finiteness of the growth rates γ. A related hydrodynamic limit was proved by Seppäläinen in [18], Proposition 10.1 (though no conditions for the finiteness of the growth function γ are given there).

We then extend these hydrodynamic limits to systems with external arrival processes (Theorems 5.2 and 5.3), under the same condition on $\{S(n,k)\}$. A key result is that if the arrival process is ergodic and has expected inter-arrival time greater than some critical value (which depends on the distribution of $\{S(n,k)\}$), then the system has a "stable" behaviour over time, and a law of large numbers can be given for the speed of progress of a typical customer through the system. In the case of queues with infinite capacity, this critical value is equal to $\mathbb{E}S(n,k)$; for queues with finite capacity, the critical value is larger but still finite; thus the system still allows a strictly positive *throughput* (though in general the inefficiency caused by the blocking reduces the maximal throughput attainable). In earlier work, Baccelli and Konstantopoulous [2] considered a framework which includes systems of queues in tandem with blocking; in the case where the service-time distribution has an exponential tail, they used large deviations methods to give a positive lower bound (which is uniform in the size of the system) for the maximal throughput.

As well as establishing results to cover queues with finite capacity, we are able to relax slightly the moment conditions used for the infinite capacity case in [1], particularly for the case of systems with an external arrival process (Theorems 5.2, 5.3); we need a condition slightly weaker than $\mathbb{E}S^{2+\epsilon} < \infty$ (which is close to the condition $\mathbb{E}S^2 < \infty$ known to be necessary) rather than needing $\mathbb{E}S^{3+\epsilon} < \infty$.

The results concerning the hydrodynamic limits are stated in Section 5.1, and are followed by discussion and interpretations in terms of system behaviour and of interaction of customers. Several open problems are also noted. The proofs of the results are given in Sections 5.2-5.4. Tools used include a concentration of measure result derived from a theorem of Talagrand [20], and a truncation technique similar to that used in [9] and [1]. Superadditivity properties are central to the arguments.

2 Preliminaries

2.1 Basic Recurrences

The type of blocking which occurs is determined by two integer parameters b and c, with $1 \leq c \leq \infty$ and $0 \leq b \leq c$.

The parameter c is the capacity of each queue. Customer n may not move from queue k to queue $k+1$ until queue $k+1$ has fewer than c customers; equivalently, until customer $n-c$ has moved from queue $k+1$ to queue $k+2$. Queue 0, however, has infinite capacity.

The parameter b is the maximum number of already-served customers who are allowed to remain at a queue. A customer will not start service at a queue if completion of the service would lead to this maximum being exceeded. Thus if $b=0$, a customer cannot start service at queue k until there is an empty space at queue $k+1$ for him to move into on completion of service (i.e. until queue $k+1$ has fewer than c customers); if $b \geq 1$, customer n cannot start service at queue k until customer $n-b$ has moved from queue k to queue $k+1$. In both cases, an equivalent constraint is to say that customer n cannot start service at queue k until customer $n-b-c$ has left queue $k+1$.

Let $D(n,k)$ be the time at which customer n completes service at queue k. (Then $D(n,k)-S(n,k)$ is the time at which customer n starts service at queue k). Also let $U(n,k)$ be the time at which customer n enters queue k. The descriptions above give us the following recurrence relations for the "local" behaviour of the system:

$$U(n,k) = \max\{D(n,k-1), U(n-c,k+1)\} \tag{2.1}$$

$$D(n,k) = S(n,k) + \max\{U(n,k), D(n-1,k), U(n-b-c,k+2)\}. \tag{2.2}$$

Since $S(n,k)$ is assumed to be non-negative, it follows from (2.2) that $D(n,k)$ is non-decreasing in n; using this, we can expand (2.1) and (2.2) recursively to give recurrences for the variables $\{D(n,k)\}$ alone:

$$D(n,k) = S(n,k) + \sup_{(\tilde{n},\tilde{k})\in\mathcal{R}} D(n-\tilde{n}, k-\tilde{k}) \tag{2.3}$$

where $\mathcal{R} = \mathcal{R}^{(b,c)}$ is a subset of $\mathbb{Z}^2$ which depends on the parameters b and c:

$$\begin{aligned}
\mathcal{R}^{(0,1)} &= \{(0,1),(1,-1)\} && \\
\mathcal{R}^{(0,c)} &= \{(0,1),(1,0),(c,-1)\} && \text{for } 2 \leq c < \infty \\
\mathcal{R}^{(b,c)} &= \{(0,1),(1,0)\} \cup \{(b+c,-1),(b+2c,-2),(b+3c,-3),\dots\} && \text{for } 1 \leq b \leq c < \infty \\
\mathcal{R}^{(b,\infty)} &= \{(0,1),(1,0)\}. &&
\end{aligned} \tag{2.4}$$

The case $b=0$ is known as *blocking before service* or *communication blocking* (although the meaning of this last term is not consistent in the literature). The case $c=1$ is then related to the *totally asymmetric simple exclusion process*, while the case $c>1$ is related to the *totally asymmetric c-exclusion process*, studied for example by Seppäläinen in [18]. The case $b=1$ is known as *blocking after service* or *manufacturing blocking*, while the cases $b \geq 2$ correspond to various types of *Kanban blocking*.

In the special case $c=\infty$ no blocking occurs, and the value of b is irrelevant. This is the case of infinite buffers, studied for models like ours in [10] and [1]. In this case the recurrence (2.3) is the well-known *Lindley's equation*. Note that for $c=\infty$ the recurrence relation is symmetric between n and k (which does not hold in the cases with blocking).

Note: An apparently more general framework would be to include another parameter $a > 1$, with $a + b \geq c$, as a constraint on the number of customers allowed at a queue which have not yet completed their service. Such a setting is described by Cheng and Yao in [5]. In fact, this does not add any generality in our case because of the assumption that all the queues are identical; the behaviour in the case $a = a_0$, $b = b_0$, $c = c_0$ is equivalent to that in the case $a = c_0$, $b = a_0 + b_0 - c_0$, $c = c_0$ in which the constraint given by a on unserved jobs plays no role (being weaker than the overall capacity constraint given by c).

2.2 Path Representations

Let (n', k') and $(n, k) \in \mathbb{Z}^2$, with $n' \leq n$, $k' \leq k$. A sequence $\{(n_1, k_1), \ldots, (n_r, k_r)\}$ of points in $\mathbb{Z}^2$ is said to be a *feasible path* from (n', k') to (n, k) if:

(i) $(n_1, k_1) = (n', k')$ and $(n_r, k_r) = (n, k)$;

(ii) for $2 \leq i \leq r$, $(n_i - n_{i-1}, k_i - k_{i-1}) \in \mathcal{R}$;

(iii) for $2 \leq i \leq r$, $k_i \geq k_1$.

We then write $\mathcal{P}\big((n', k'), (n, k)\big)$ for the set of feasible paths from (n', k') to (n, k) – this set again depends on the parameters b and c.

Property (ii) says that the vectors in the set $\mathcal{R}$ form the set of "permissible steps" for the path. Restriction (iii) above is useful to us because of the particular boundary conditions with which we will be working. In other situations (for example in the case of a model with a doubly infinite sequence of queues) it might be more natural to remove this condition.

Figure 2.1 shows examples of feasible paths for various values of the blocking parameters b and c. (Throughout the paper we will draw n increasing to the right, and k increasing upwards).

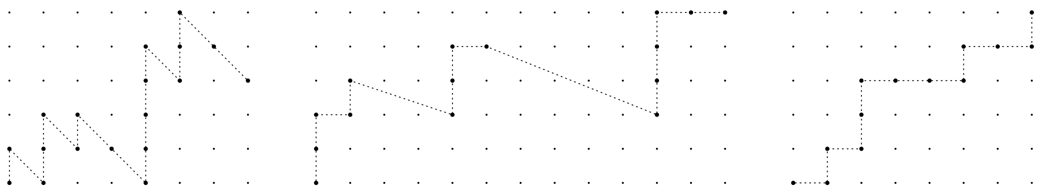

Figure 2.1: Examples of feasible paths: (i) $b = 0$, $c = 1$; (ii) $b = 1$, $c = 2$; (iii) $c = \infty$.

For a finite subset ξ of $\mathbb{Z}^2$, we define the *weight* $S(\xi)$ of ξ by

$$S(\xi) = \sum_{(n,k)\in\xi} S(n, k). \tag{2.5}$$

Any feasible path π is identified with the set of points which occur on the path, and the weight $S(\pi)$ of π is defined just as at (2.5). Then, for $n' \leq n$, $k' \leq k$, we define the *maximal weight* of a feasible path from (n',k') to (n,k) by

$$T\big((n',k'),(n,k)\big) = \max_{\pi\in\mathcal{P}\left((n',k'),(n,k)\right)} S(\pi). \tag{2.6}$$

For convenience of notation, we also set

$$T(n,k) = T\big((0,0),(n,k)\big)$$

for $n \geq 0$, $k \geq 0$.

We occasionally add superscripts (b,c) to the quantities $\mathcal{P}$ and T if we wish to emphasise their dependence on the type of blocking. The following lemma will provide a useful comparison between the behaviour of the system for different values of the blocking parameters. The larger the capacities b and c, the smaller will be the weight of a maximum path length — this will correspond to smaller queueing delay.

Lemma 2.1 *Let $b' \leq b$ and $c' \leq c$. Let $n_1 \leq n_2$ and $k_1 \leq k_2$.*

(i) If $\pi \in \mathcal{P}^{(b,c)}\big((n_1,k_1),(n_2,k_2)\big)$, then there exists $\pi' \in \mathcal{P}^{(b',c')}\big((n_1,k_1),(n_2,k_2)\big)$ such that $\pi \subset \pi'$.

(ii) $T^{(b,c)}\big((n_1,k_1),(n_2,k_2)\big) \leq T^{(b',c')}\big((n_1,k_1),(n_2,k_2)\big)$.

Proof: From the forms in (2.4), one can show that for all $(\tilde{n},\tilde{k}) \in \mathcal{R}^{(b,c)}$, there exist $(\tilde{n}_i,\tilde{k}_i) \in \mathcal{R}^{(b',c')}$, $1 \leq i \leq r$ (for some $r \geq 1$) such that

$$\tilde{n} = \sum_{i=1}^{r} \tilde{n}_i, \text{ and } \tilde{k} = \sum_{i=1}^{r} \tilde{k}_i,$$

and such that, for all $1 \leq j \leq r$,

$$\sum_{i=1}^{j} \tilde{k}_i \geq \min\{0,\tilde{k}_i\}.$$

Using this, we can make any (b,c)-feasible path into a (b',c')-feasibile path by adding intermediate points to it as necessary, and part (i) follows. Part (ii) then follows from the definition (2.6). $\square$

2.3 Boundary Conditions

The recurrences (2.3) give the "local" behaviour of the system.

In order to complete the model, we specify boundary conditions; we give a set $\Lambda \subset \mathbb{Z}^2$ and fix the values $D(n,k)$, $(n,k) \notin \Lambda$. For suitable Λ, the values $D(n,k)$, $(n,k) \in \Lambda$ can then be derived using the recurrence relations (2.3) together with the values $S(n,k)$, $(n,k) \in \Lambda$.

2.3.1 Saturated initial queue: quadrant model

Let $\Lambda = \{(n,k) : n \geq 0, k \geq 0\}$, and let $D(n,k) = 0$ for all $(n,k) \notin \Lambda$.

This represents a model of a semi-infinite sequence of queues. Each queue is empty at time 0 except queue 0, which contains infinitely many customers at time 0. The first customer (customer 0) begins service at queue 0 at time 0.

In the case $c = \infty$, this corresponds to the model considered in [10], and also to one of the models studied in [1], where it is called the *quadrant* model.

By iteration of the recurrence relations in (2.3), and by the definition of the set of feasible paths in Section 2.2, we have that, for $n \geq 0$, $k \geq 0$, $D(n,k)$ may be represented as the maximum weight of a feasible path from $(0,0)$ to (n,k), so that in this case

$$D(n,k) = T(n,k). \tag{2.7}$$

2.3.2 Two-sided arrival process: half-plane model

We now consider a two-sided sequence of interarrival times $\{A(n), n \in \mathbb{Z}\}$. Customer 0 arrives at time 0, and the interarrival time between customers n and $n+1$ is $A(n)$, for all $n \in \mathbb{Z}$.

To represent this model, we let $\Lambda = \{(n,k) : k \geq 0\}$, and let $D(n,k) = 0$ for $k < -1$. For $n \geq 0$, let $D(n,-1) = \sum_{i=0}^{n-1} A(i)$, and for $n < 0$, let $D(n,-1) = -\sum_{i=n}^{-1} A(i)$.

From the forms (2.4), note also that (for all b, c) the only member $(\tilde{n}, \tilde{k})$ of $\mathcal{R}$ such that $\tilde{k} > 0$ is the point (0,1) (that is, the only upward step that a feasible path can make is a single step directly upwards).

The minimal solution to the recurrence relations for $\{D(n,k)\}$ given by (2.3) is then

$$\begin{aligned} D(n,k) &= \sup_{n' \leq n} \left\{ T\big((n',0),(n,k)\big) + D(n',-1) \right\}. \\ &= D(n,-1) + \sup_{n' \leq n} \left\{ T\big((n',0),(n,k)\big) - \sum_{i=n'}^{n-1} A(i) \right\}. \end{aligned} \tag{2.8}$$

This is the maximum, over all points $(n',0)$ on the n-axis, of the sum of the arrival time $D(n',-1)$ of customer n' and the maximum weight of a path from $(n',0)$ to (n,k). (We exclude $n' > n$ since then no such feasible path exists).

A model of this kind is treated for the case $c = \infty$ in [1], where it is called the *half-plane* model.

Note that here one first needs to verify that the $D(n,k)$ are well-defined — the supremum in (2.8) could be infinite, since infinitely many values of n' are possible.

The set-up can be used to model the "stable" state of a system, when such a state exists. If the distribution of $\{A(n),\ S(n,k),\ k \geq 0,\ n \in \mathbb{Z}\}$ is stationary in n, then so is the distribution of the collection $\{D(n,k+1) - D(n,k),\ D(n,k) - D(n-1,k),\ k \geq 0, n \in \mathbb{Z}\}$ derived from (2.8); for example, $D(0,k)$ represents the time taken by a typical customer between arrival in the system and the end of service at queue k.

3 Functional central limit theorems

In this section we obtain a functional central limit theorem (FCLT) which describes the behaviour of the "front of the wave" progressing through a system which is empty at time 0.

For all the types of blocking described in Section 2.1, we obtain the same result as was obtained by Glynn and Whitt in [10] for the case of queues with infinite buffers; thus we may say that, over time, the first few customers meet sufficiently rarely that the various types of blocking do not affect the form of the front of the wave.

The result is stated in terms of the variables $\{T(n,k)\}$, and so, using (2.7), applies directly to the model of Section 2.3.1; it can also cover the case of a one-sided arrival process at queue 0 – see the comment after the proof.

Let $D[0,\infty)$ be the space of real-valued functions on the interval $[0,\infty)$ which are right-continuous with left limits, endowed with the Skorohod topology (see for example [21]), and let $D[0,\infty)^{\mathbb{Z}_+}$ be the associated product space, endowed with the product topology.

Let $\beta,\theta > 0$, and, for $n,k \geq 0$, define the random variables $\tilde{S}_{nk}$ and $\tilde{T}_{nk}$ taking values in $D[0,\infty)$ by

$$\tilde{S}_{nk}(t) = k^{-\theta}\left[\sum_{i=0}^{\lfloor kt \rfloor} S(n,i) - \beta kt\right] \tag{3.1}$$

and

$$\tilde{T}_{nk}(t) = k^{-\theta}\left[T(n,\lfloor kt \rfloor) - \beta kt\right], \tag{3.2}$$

for $t \geq 0$. Then define $\tilde{\boldsymbol{S}}_k$ and $\tilde{\boldsymbol{T}}_k$, taking values in $D[0,\infty)^{\mathbb{Z}_+}$, by

$$\tilde{\boldsymbol{S}}_k = \left(\tilde{S}_{0k}, \tilde{S}_{1k}, \tilde{S}_{2k}, \ldots\right)$$

and

$$\tilde{\boldsymbol{T}}_k = \left(\tilde{T}_{0k}, \tilde{T}_{1k}, \tilde{T}_{2k}, \ldots\right).$$

To obtain an FCLT for $\tilde{\boldsymbol{T}}_k$, we do not require that the variables $S(n,k)$ be i.i.d., but only that $\tilde{\boldsymbol{S}}_k$ itself obeys an FCLT. If in fact $S(n,k)$ are i.i.d. with finite mean and variance, then the appropriate scaling in (3.1) and (3.2) to provide a non-trivial limit is given by $\theta = 1/2$ and $\beta = \mathbb{E}S$.

Theorem 3.1 *If $\tilde{\boldsymbol{S}}_k \Rightarrow \tilde{\boldsymbol{S}}$ in $D[0,\infty)^{\mathbb{Z}_+}$ as $k \to \infty$, where $\tilde{\boldsymbol{S}}$ has continuous sample paths with probability 1, then $\tilde{\boldsymbol{T}}_k \Rightarrow \tilde{\boldsymbol{T}}$ in $D[0,\infty)^{\mathbb{Z}_+}$ as $k \to \infty$, where $\tilde{\boldsymbol{T}} = f(\tilde{\boldsymbol{S}})$ and where $f : D[0,\infty)^{\mathbb{Z}_+} \to D[0,\infty)^{\mathbb{Z}_+}$ is defined by*

$$f_0(\mathbf{x})(t) = x_0(t)$$

and, for $n \geq 1$,

$$f_n(\mathbf{x})(t) = \sup_{0 \leq s \leq t} \{f_{n-1}(\mathbf{x})(s) + x_n(t) - x_n(s)\}$$
$$= x_n(t) - \inf_{0 \leq s \leq t} \{x_n(s) - f_{n-1}(\mathbf{x})(s)\}.$$

Proof: Glynn and Whitt show in Theorem 3.1 of [10] that the result holds in the case $c = \infty$. We adapt their proof slightly to show that the same conclusion holds in the case $b = 0$, $c = 1$. The result for other values of b and c then follows from these two cases and from the monotonicity property in Lemma 2.1(ii) (though one could also extend the proof below to cover these cases directly).

So let $b = 0$, $c = 1$. By the form of the recurrence in (2.3) and (2.4), we have that

$$T(0,k) = \sum_{i=0}^{k} S(0,i) \tag{3.3}$$

for $k \geq 0$, and, for $n > 0$, $k \geq 0$,

$$T(n,k) = \max_{0 \leq l \leq k} \left\{ \sum_{i=l}^{k} S(n,i) + T(n-1, l+1) \right\}. \tag{3.4}$$

We will show by induction on n that $\tilde{T}_{nk} \Rightarrow f_n(\tilde{\boldsymbol{S}})$ in $D[0,\infty)$ as $k \to \infty$, for all n. Since we use the product topology on $D[0,\infty)^{\mathbb{Z}_+}$, this gives the desired result.

Note first that, since $\tilde{\boldsymbol{S}}$ has a.s. continuous sample paths, so does $f(\tilde{\boldsymbol{S}})$ (since if $\mathbf{x}$ is continuous then, by induction, each of the component paths of $f(\mathbf{x})$ can be shown to be continuous, and we use the product topology).

Note also that f is continuous as a function from $D[0,\infty)^{\mathbb{Z}_+}$ to $D[0,\infty)^{\mathbb{Z}_+}$ (again one can show by induction that each of the component functions is continuous, e.g. by Section 6 of [21]). By the continuous mapping theorem (e.g. [4], Theorem 5.1) we then have that $f(\tilde{\boldsymbol{S}}_k) \Rightarrow f(\tilde{\boldsymbol{S}})$ as $k \to \infty$.

Now, from (3.3), $\tilde{T}_{0k} = \tilde{S}_{0k} = f_0(\tilde{\boldsymbol{S}}_k)$, so that $\tilde{T}_{0k} \to f_0(\tilde{\boldsymbol{S}})$ as $k \to \infty$, as required.

So let $n \geq 1$, and suppose that $\tilde{T}_{(n-1)k} \Rightarrow f_{n-1}(\tilde{\boldsymbol{S}})$. Using (3.4), we have

$$\begin{aligned}
\tilde{T}_{nk}(t) &= k^{-\theta} \left[\max_{0 \leq l \leq \lfloor kt \rfloor} \left\{ \sum_{i=l}^{\lfloor kt \rfloor} S(n,i) + T(n-1, l+1) \right\} - \beta kt \right] \\
&= k^{-\theta} \sup_{0 \leq s \leq t} \left\{ \sum_{i=\lfloor ks \rfloor}^{\lfloor kt \rfloor} S(n,i) - \beta k(t-s) + T(n-1, \lfloor ks \rfloor + 1) - \beta ks \right\} \\
&= \sup_{0 \leq s \leq t} \left\{ \tilde{S}_{nk}(t) - \tilde{S}_{nk}(s - 1/k) + \tilde{T}_{(n-1)k}(s + 1/k) + k^{-\theta}\beta \right\}.
\end{aligned} \tag{3.5}$$

Since $\tilde{S}_{nk}$ and $\tilde{T}_{(n-1)k}$ converge weakly to $\tilde{\boldsymbol{S}}$ and $f_{n-1}(\tilde{\boldsymbol{S}})$ respectively, and since both these limits have paths which are a.s. continuous, the function of t given by the RHS of (3.5) has the weak limit

$$\sup_{0 \leq s \leq t} \left\{ \tilde{S}_n(t) - \tilde{S}_n(s) + f_{n-1}(\tilde{\boldsymbol{S}})(s) \right\}.$$

From the definition of f, we then have that $\tilde{T}_{nk} \Rightarrow f_n(\tilde{\boldsymbol{S}})$, as desired. $\square$

The further results of Section 3 of [10] now apply also. In particular, if, for some j, $\{S(n,k), n \geq 0, k \geq j\}$ are i.i.d. with finite variance (here the exclusion of finitely many queues $k < j$ would allow us for example to cover the case of a general one-sided arrival process at queue 0) then the limiting scaled interdeparture process $(\tilde{T}_1 - \tilde{T}_0, \tilde{T}_2 - \tilde{T}_1, \tilde{T}_3 - \tilde{T}_2, \dots)$ in $D[0,\infty)^{\mathbb{Z}_+}$ may be characterised as a reflected Brownian motion on the infinite-dimensional non-negative orthant, with zero drift and with covariance matrix $\boldsymbol{\Sigma}$ given by $\Sigma_{ij} = 2I\{i=j\} - I\{|i-j|=1\}$. This process has received a lot of attention recently; as well as arising in connection with queueing and percolation problems as here, it also appears in the context of random matrix theory. See for example [3], [11], [15], [17] for a variety of viewpoints.

4 Stochastic orderings for inter-service times

In this section we again work with the variables $\{T(n,k)\}$, which correspond to the service completion times $\{D(n,k)\}$ in the model of Section 2.3.1.

For $n, k \geq 0$, we define $W(n,k)$, the *waiting time* of customer n between completing service at queue k and at queue $k+1$, by

$$W(n,k) = T(n,k+1) - T(n,k).$$

In this section, we establish stochastic ordering properties of the random variables $W(n,k)$ as n and k vary.

In the case $c = \infty$, the conclusion of the theorem is the same as that of Theorem 5.1 of Glynn and Whitt, but the assumptions are weaker in that we do not require the $\{S(n,k)\}$ to be independent. The underlying coupling used is the same as in [10], but we give a graphical argument which simplifies the proof and avoids the need for the induction used there. For queues with finite capacity, we find that the methods extend to the cases of blocking before service ($b=0$) and blocking after service ($b=1$) but not to larger values of b.

Theorem 4.1 *Suppose that $c = \infty$ or that $b \leq 1$.*

(i) If the collection $\{S(n,k), n \geq 0, k \geq 0\}$ is stationary in n, then

$$\{W(n,k), n \geq 0, k \geq 0\} \leq_{st} \{W(n+1,k), n \geq 0, k \geq 0\}.$$

(ii) If the collection $\{S(n,k), n \geq 0, k \geq 0\}$ is stationary in k, then

$$\{W(n,k), n \geq 0, k \geq 0\} \geq_{st} \{W(n,k+1), n \geq 0, k \geq 0\}.$$

Proof: We prove part (i); the proof of part (ii) is very similar.

We will produce a second identically distributed copy $\{\bar{W}(n,k)\}$ of $\{W(n,k)\}$ such that $\bar{W}(n,k) \leq W(n+1,k)$ for all n, k.

Let $\bar{S}(n,k) = S(n+1,k)$ for all $n, k \geq 0$, and define $\{\bar{T}(n,k)\}$ and $\{\bar{W}(n,k)\}$ just as $\{T(n,k)\}$ and $\{W(n,k)\}$ but using the array $\{\bar{S}(n,k)\}$ in place of the array $\{S(n,k)\}$.

Now $T(n,k)$ is the weight of a maximal path from $(0,0)$ to (n,k) in the graph determined by the array $\{S(n,k)\}$. Correspondingly, $\bar{T}(n,k)$ is the weight of a maximal path from $(0,0)$ to (n,k) in the graph determined by $\{\bar{S}(n,k)\}$ — or, equivalently, the weight of a maximal path from $(1,0)$ to $(n+1,k)$ in the graph determined by $\{S(n,k)\}$.

We therefore wish to show that

$$\bar{T}(n,k+1) - \bar{T}(n,k) \leq T(n+1,k+1) - T(n+1,k),$$

which we can rewrite as

$$\bar{T}(n,k+1) + T(n+1,k) \leq \bar{T}(n,k) + T(n+1,k+1). \tag{4.1}$$

The LHS of (4.1) consists of the weights of maximal paths (in the graph determined by the array $\{S(n,k)\}$), from $(1,0)$ to $(n+1,k+1)$ and from $(0,0)$ to $(n+1,k)$.

First take the case $c=\infty$. Since the only feasible path steps are single steps in the direction of increasing n or in the direction of increasing k, these paths must meet at some point.

Now alter the two paths by exchanging the portions which occur after the last point at which they meet. This yields feasible paths from $(0,0)$ to $(n+1,k+1)$ and from $(1,0)$ to $(n+1,k)$ but does not change the combined weight of the two paths. See Figure 4.1 for an example.

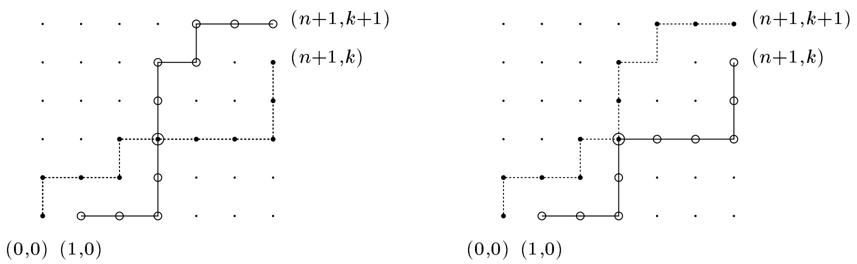

Figure 4.1:

The RHS of (4.1) consists of the combined weights of *maximal* paths from $(0,0)$ to $(n+1,k+1)$ and from $(1,0)$ to $(n+1,k)$, and so is at least as great as the combined weight of the two new paths. But this combined weight was the LHS of (4.1); so the inequality (4.1) holds as required.

This completes the proof for the case $c=\infty$. The case $b=0, c=1$ is very similar; the only feasible steps are of the form $(0,1)$ and $(1,-1)$, and again the paths must share a point.

For other b, c, we need to modify the proof; the paths from $(0,0)$ to $(n+1,k)$ and from $(1,0)$ to $(n+1,k+1)$ must still cross, as before, but now it is possible for them to cross without sharing a point. It will suffice, though, to show that if there are feasible *steps* from (n_1',k_1') to (n_1,k_1) and from (n_2',k_2') to (n_2,k_2), and if the lines drawn from (n_1',k_1') to (n_1,k_1) and from (n_2',k_2') to (n_2,k_2) intersect, then there are feasible *paths* from (n_1',k_1') to (n_2,k_2) and from (n_2',k_2') to (n_1,k_1) — if this is true then we can exchange the portions of the paths which occur after the last crossing point, as before. We can replace *steps* by *paths* if necessary, since our purpose is to obtain the inequality (4.1); it does not matter if we add extra points which were not present on the original paths, since this can only increase the value of the RHS.

First take the case $b = 0$, $c \geq 2$. Here we have steps of the form $(0, 1)$, $(1, 0)$ and $(c, -1)$. The only way for the paths to cross without sharing a point is for a vertical step $(1, 0)$ to cross a diagonal step $(c, -1)$. This may be replaced by two or more horizontal steps $(0, 1)$ as in Figure 4.2.

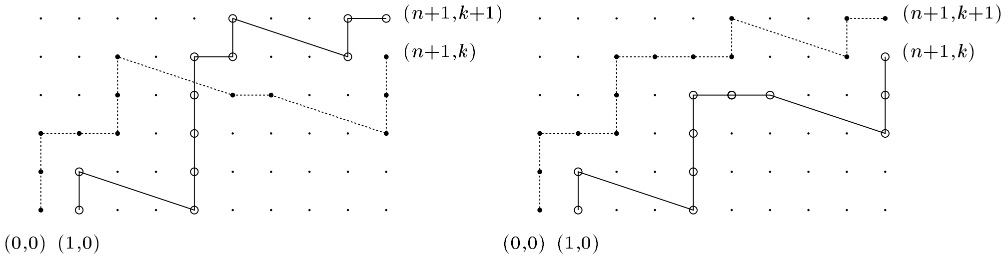

Figure 4.2:

Now consider $b = 1$; the feasible steps are $(0, 1)$, $(1, 0)$ and, for any $r \geq 1$, $(1+rc, -r)$. The situation is more complicated now that steps which decrease k by more than 1 are allowed. A diagonal step $(1 + rc, -r)$ may be crossed by a vertical step $(0, 1)$, by a horizontal step $(1, 0)$, or by another diagonal step $(1+r'c, -r')$. One can verify that each possible type of crossing may be removed in the way desired; however, the enumeration of the various cases is extensive and we give just an indication of how one can proceed. The easiest method is perhaps to break up each diagonal step $(1 + rc, -r)$ into a succession of r "substeps" $(c, -1)$ followed by a final substep $(1, 0)$; this creates r "intermediate points". Every possible crossing of the diagonal step then involves either a meeting at one of these intermediate points, or the crossing of a substep $(c, -1)$ by a vertical step $(0, 1)$ (resolvable in a similar way to that in Figure 4.2). $\square$

However, this resolution is no longer possible for all types of crossing if $b \geq 2$, and the coupling no longer works. For example, let $b = 2$, $c = 2$ and consider one path with a step $(0, 2) \to (6, 0)$ and another path with a pair of steps $(3, 0) \to (3, 1) \to (3, 2)$. This crossing can't be removed, since no feasible path which contains $(3, 1)$ can contain either $(0, 2)$ or $(6, 0)$. For these larger values of b, the results of Theorem 4.1 do not hold in general. However, we have neither a proof nor a counterexample under stronger hypotheses, for example that the service times $S(n, k)$ are i.i.d.

For $n, k \geq 0$, we can also define $\Delta(n, k)$, the *interservice time* between customers n and $n + 1$ at queue k by

$$\Delta(n, k) = T(n + 1, k) - T(n, k).$$

In the case $c = \infty$, the symmetry between n and k of the recurrence relations defining $T(n, k)$ implies that an analogue of Theorem 4.1 holds:

Theorem 4.2 *Let $c = \infty$. Then*

(i) If the collection $\{S(n,k), n \geq 0, k \geq 0\}$ is stationary in k, then

$$\{\Delta(n,k), n \geq 0, k \geq 0\} \leq_{st} \{\Delta(n,k+1), n \geq 0, k \geq 0\}.$$

(ii) If the collection $\{S(n,k), n \geq 0, k \geq 0\}$ is stationary in n, then

$$\{\Delta(n,k), n \geq 0, k \geq 0\} \geq_{st} \{\Delta(n+1,k), n \geq 0, k \geq 0\}.$$

However, these results fail to hold in general for cases involving blocking. The coupling argument breaks down in this case (for all b) because, once steps which decrease k are allowed, one can for example construct paths from $(0,0)$ to $(n,k+1)$ and from $(0,1)$ to $(n+1,k+1)$ which do not meet or cross, as in Figure 4.3, which illustrates the case $b = 0$, $c = 2$.

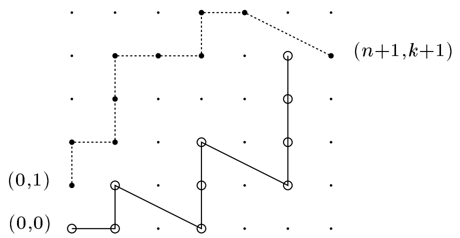

Figure 4.3:

5 Hydrodynamic limits

5.1 Main results and discussion

In this section we give various "hydrodynamic limits" or "directional laws of large numbers" for the time taken by customers to progress through the system, under various boundary conditions. In particular we will be able to look at the long-term behaviour of systems with external arrival processes of customers arriving at the first queue. The proofs of the results will be given in Sections 5.2-5.4.

We will assume that the service times $\{S(n,k)\}$ are i.i.d., and, to prove that the limiting growth rates are finite, we will need a condition on the tail of their common distribution, namely that

$$\int_0^\infty \mathbb{P}(S \geq s)^{1/2} ds < \infty. \tag{5.1}$$

This condition is used in [14] to treat the "greedy lattice animals" model, which was introduced in [6] and [9]. The finiteness of the growth rates in our current model can be deduced from the finiteness of the growth rate of a "greedy lattice animal", derived in [14] under condition (5.1). The consequences of (5.1) which we use are given in Lemma

5.5. Condition (5.1) is stronger than the condition that $\mathbb{E}S^2 < \infty$, but marginally weaker than the condition

$$\mathbb{E}S^2(\log^+ S)^{2+\epsilon} < \infty \text{ for some } \epsilon > 0 \tag{5.2}$$

needed to apply the results of [6].

The first result gives hydrodynamic limits for the weight of the heaviest path in a given direction in the plane; that is, laws of large numbers for the variables $T(n,k)$ (and hence for $D(n,k)$ in the situation of Section 2.3.1) when n grows proportionally to k:

Theorem 5.1 *Suppose that the service times $\{S(n,k)\}$ are i.i.d. with common distribution satisfying (5.1), and with $\mathbb{E}S = \beta$. Then there exists a function γ on $[0,\infty)$ such that, for all $x \geq 0$,*

$$\frac{T\big(\lfloor xK \rfloor, K\big)}{K} \to \gamma(x) \textit{ as } K \to \infty, \textit{ a.s. and in } \mathcal{L}^1, \tag{5.3}$$

and such that

$$\mathbb{E}T(n,k) \leq k\gamma(n/k) + \beta \textit{ for all } n, k \geq 0. \tag{5.4}$$

The function γ has the following properties:

(i) $\gamma(0) = \beta$;

(ii) $\gamma(x+y) - \gamma(x) \geq \beta y$ *for all* $x \geq 0$, $y > 0$;

(iii) γ *is concave;*

(iv) The right derivative $\gamma'(x)$ exists for all $x > 0$, and $\lim_{x\to\infty} \gamma'(x)$ exists and is in $[\beta, \infty)$.

We then extend these hydrodynamic limits to provide convergence results for the situation of a two-sided external arrival process to queue 0 (Section 2.3.2). If the arrival process is ergodic and the expected interarrival time between successive customers is sufficiently large, then the system is "stable" and we can obtain a law of large numbers for the amount of time taken by a typical customer to progress through a large number of queues.

To derive the convergence, we use the fact that for any given (n,k), there are only countably many possible directions for a feasible path from the n-axis to the point (n,k) We use a concentration of measure result (Lemmas 5.6 and 5.7) derived from a theorem of Talagrand in order to control the likelihood of significant deviations from the asymptotic growth rates in each of these possible directions; (these growth rates are given by the function γ of Theorem 5.1).

Note that we do not require the arrival process $\{A(n)\}$ to be independent of the service times $\{S(n,k)\}$. The arrival process could for example be affected in some way by the state of the system, provided the ergodicity of $\{A(n)\}$ is preserved.

Theorem 5.2 *Suppose that $\{S(n,k)\}$ are i.i.d. with common distribution satisfying (5.1). Suppose also that the sequence $\{A(n), n \in \mathbb{Z}\}$ of interarrival times at queue 0 is ergodic with $\mathbb{E}A(n) = \alpha$ and $\alpha > \lim_{x\to\infty} \gamma'(x)$, where the function γ is given by Theorem 5.1. Then for all $K > 0$,*

$$\sup_{n\geq 0} \Big\{ T\big((-n,0),(0,K)\big) - \sum_{i=1}^{n} A(-i) \Big\} < \infty \tag{5.5}$$

with probability 1, and

$$\frac{1}{K} \sup_{n\geq 0} \Big\{ T\big((-n,0),(0,K)\big) - \sum_{i=1}^{n} A(-i) \Big\} \to \sup_{x\geq 0} \{\gamma(x) - \alpha x\} \tag{5.6}$$

a.s. as $K \to \infty$. Thus, for the model of Section 2.3.2, $D(0,K) < \infty$ a.s. for all K, (so, similarly, $D(n,K) < \infty$ a.s. for all n, K) and

$$\frac{D(0,K)}{K} \to \sup_{x\geq 0} \{\gamma(x) - \alpha x\}$$

a.s. as $K \to \infty$.

Note that, unlike that in Theorem 5.1, the convergence given in Theorem 5.2 is now only a.s. In the case of queues without blocking, the next theorem shows that the $\mathcal{L}_1$ convergence also holds, under an additional condition on the arrival process; the proof is done by comparing the LHS of (5.6) to a process with a superadditive property. The result was shown in [1] under the additional condition that $\mathbb{E}S^{3+\epsilon} < \infty$ for some $\epsilon > 0$.

For a discussion of when the condition (5.7) holds in general, see for example [7]. Certainly it holds if the interarrival times $A(n)$ are i.i.d.; a sufficient condition is that they are strongly mixing with mixing coefficients whose sum is finite.

Theorem 5.3 *Consider now only the case $c = \infty$. If the conditions of Theorem 5.2 hold, and if in addition the arrival process $\{A(n)\}$ satisfies*

$$\mathbb{E} \sup_{n\geq 0} \left[\eta n - \sum_{i=1}^{n} A(-i) \right] < \infty \tag{5.7}$$

for some η with $\lim_{x\to\infty} \gamma(x)/x < \eta < \alpha$, then the convergence in (5.6) occurs also in $\mathcal{L}_1$.

For a discussion of when the condition (5.7) holds in general, see for example [7]. Certainly it holds if the interarrival times $A(n)$ are i.i.d.; a sufficient condition is that they are strongly mixing with mixing coefficients whose sum is finite.

For $c < \infty$, on the other hand, the question of the $\mathcal{L}_1$ convergence in (5.6) seems harder to treat. We explain during the proof (Section 5.4) why the superadditive argument that we use to prove Theorem 5.3 breaks down in the case $c < \infty$. On the other hand, the alternative argument used for the case $c = \infty$ in [1] makes use of the fact that, as n grows with K held fixed, the set of points that can occur on a path from $(-n,0)$ to $(0,K)$ grows only proportionally to n; in the case $c < \infty$, this no longer holds — the size of the set now grows proportionally to n^2.

Indeed, we should probably first ask: when is $\mathbb{E}D(0,0) < \infty$? That is, when is the *stationary expected waiting time* at the first queue finite? At the moment we need rather restrictive conditions on $\{S(n,k)\}$ and $\{A(n)\}$ in order to show this.

We now discuss various issues concerning the growth function γ given by Theorem 5.1 and the various hydrodynamic limits.

The asymptotic slope $\lim_{x\to\infty}\gamma'(x)$ given by Theorem 5.1(iv) can be interpreted as the reciprocal of the *maximal throughput* of the system of queues. That is, if the expected interarrival time in an ergodic arrival process is $\alpha > \gamma'(x)$, then (by Theorem 5.2) the system has a stable behaviour, under which, for example, the expectation of the total number of customers served at queue 0 (or queue k for any $k \in \mathbb{Z}^+$) during a typical unit time interval is α^{-1}; this can be made as close as desired to $(\lim_{x\to\infty}\gamma'(x))^{-1}$. On the other hand, if $\alpha < \lim_{x\to\infty}\gamma'(x)$, then one can adapt (5.3) to show that the supremum in (5.5) is almost surely infinite, and the system has no stable behaviour. In the case of infinite buffers, the symmetry between k and n implies that $\lim_{x\to\infty}\gamma'(x) = \lim_{x\to 0}\gamma(x)$, and it is shown in Proposition 5.8 of [1] that the function γ is continuous at $x = 0$, so that $\lim_{x\to\infty}\gamma'(x) = \gamma(0) = \mathbb{E}S$.

To our knowledge, the only cases in which the function γ is known explicitly are those where the distribution of S is either exponential or geometric, and either there is no blocking (see for example [1] or [16] for references and discussion) or there is blocking before service and the queues have capacity 1 (the formulas here can be obtained by a simple transformation of those in the case without blocking). In other cases, there are various interesting questions concerning the form of γ. For example:

(i) *When is γ strictly concave?* One can show that if S has bounded support and achieves its supremum with sufficiently large probability, then there is an interval on which γ is linear with slope equal to this supremum. Is this the only way in which γ can fail to be strictly concave?

This question has implications for the asymptotic form of the of the maximising paths in situations such as that considered in Theorem 5.2. For example, let $n^*(K)$ be the n which maximises the LHS of (5.5). Then one can easily extend the proof of the theorem to show that

$$\liminf_{K\to\infty} -\frac{n^*(K)}{K} \geq \sup\{x : \gamma'(x) \geq \alpha\} \tag{5.8}$$

and that

$$\limsup_{K\to\infty} -\frac{n^*(K)}{K} \leq \inf\{x : \gamma'(x) \leq \alpha\}. \tag{5.9}$$

If γ is strictly concave, then the RHS of (5.8) is the same as that of (5.9), say equal to $x^* = x^*(\alpha)$. The interpretation for the model of Section 2.3.2 is then as follows: for large n, the first time that customer m experiences a delay caused (indirectly) by customer $m - n$ occurs when customer m is around queue n/x^*; and, in general, if n and k are large with $k > n/x^*$, then those service times of customer $m - n$ which affect the amount of time taken by customer m to reach queue k tend to be those experienced around queue $k - n/x^*$.

(ii) *Is γ always continuously differentiable?* That is, do its left and right derivatives (which exist by concavity) coincide for all $x > 0$?

In the case without blocking, the smoothness of γ would imply that, for all $\alpha > \beta$, there is a law of a (two-sided) ergodic external arrival process with expected interarrival time α at queue 0 which is a "fixed point" for the system, in the sense that the departure process from queue 0 (which then becomes the arrival process for queue 1) has the same law – see [13].

The case of blocking before service with exponential service times is related to the Markovian *totally asymmetric c-exclusion process*, studied for example in [18]. There the smoothness of γ would imply that for all ρ, $0 \leq \rho \leq c$, there exists an ergodic equilibrium distribution under which the average number of customers per queue is ρ. This is known for $c = 1$ and $c = \infty$ (where these equilibrium distributions have product form) but not for larger finite values of c. (See also [8]).

(iii) *Can (5.1) be weakened?* It's known that the condition $\mathbb{E}S^2 < \infty$ is necessary for the finiteness of γ, and for the convergence in (5.6); (when $c < \infty$, it is also necessary for the finiteness in (5.5)). The sufficient condition (5.1) is not much stronger, but it would nevertheless be more satisfying to complete the characterisation by determining whether γ is finite in cases where the variance is finite but where (5.1) does not hold; apparently the behaviour is not yet known for any such example. See for example [14] for further discussion.

5.2 Preliminary Lemmas

Regard $\mathbb{Z}^2$ as a graph in the normal way, so that (n, k) and (n', k') are adjacent if $|n - n'| + |k - k'| = 1$. Then for $m \in \mathbb{N}$, let $\mathcal{A}(m)$ be the set of connected subsets of $\mathbb{Z}^2$ of size m which contain the origin (so-called *lattice animals of size m*).

The following lemma collects some basic properties of the feasible paths defined in Section 2.2 which are true for all b, c. They follow easily from the forms of the possible path steps given by the sets $\mathcal{R}$ in (2.4); (or indeed just from the form for $b = 0$, $c = 1$ and then the monotonicity result Lemma 2.1). The results given are a lot less than best possible, but we keep the form as simple as possible for later convenience.

Lemma 5.4 *Let $\pi \in \mathcal{P}\big((n_1, k_1), (n_2, k_2)\big)$, where $n_1 \leq n_2$ and $k_1 \leq k_2$. Let $R = |n_1| + |n_2| + |k_1| + |k_2|$, and assume that $R > 0$. Then:*

(i) $|\pi| \leq 3R$.

(ii) If $(n, k) \in \pi$, then $\max\big(|n|, |k|\big) \leq R$.

(iii) There exists a $\xi \in \mathcal{A}(4R)$ such that $\pi \subset \xi$.

We next define truncated versions of the random variables $\{S(n, k)\}$, in order to be able to work with bounded random variables when we apply the concentration of measure results which we will use. The truncation method is similar to that used in [9] or [1].

For $(n,k)\in\mathbb{Z}^2$, let

$$\hat{S}(n,k)=\min\left[S(n,k),\left(\max|n|,|k|\right)^{1/4}\right], \tag{5.10}$$

and for a finite subset ξ of $\mathbb{Z}^2$, let

$$\hat{S}(\xi)=\sum_{(n,k)\in\xi}\hat{S}(n,k).$$

Analogously to (2.6), we also write

$$\hat{T}\big((n_1,k_1),(n_2,k_2)\big)=\max_{\pi\in\mathcal{P}\left((n_1,k_1),(n_2,k_2)\right)}\hat{S}_\pi$$

for $n_1\le n_2$, $k_1\le k_2$, and put $\hat{T}(n,k)=\hat{T}\big((0,0),(n,k)\big)$.

The following result then summarises the consequences of the condition (5.1) which we will use:

Lemma 5.5 *If $\{S(n,k)\}$ are i.i.d. with common distribution satisfying (5.1), then*

$$\limsup_{m\to\infty}\frac{1}{m}\max_{\xi\in\mathcal{A}(m)}S(\xi)<\infty \ \textit{a.s.}, \tag{5.11}$$

and

$$\frac{1}{m}\max_{\xi\in\mathcal{A}(m)}\{S(\xi)-\hat{S}(\xi)\}\to 0 \ \textit{a.s. as } m\to\infty. \tag{5.12}$$

Proof: (5.11) is given by Theorem 3.3 of [14]. For (5.12), Lemma 4.1 of [14] shows that, for all $y>0$,

$$\limsup_{m\to\infty}\frac{1}{m}\max_{\xi\in\mathcal{A}(m)}\sum_{(n,k)\in\xi}[S(n,k)-y]_+\le C\int_y^\infty\mathbb{P}(S\ge s)^{1/2}ds,$$

for some constant C. If (5.1) holds, then for any $\epsilon>0$, we can take y such that $C\int_y^\infty\mathbb{P}(S\ge s)^{1/2}ds\le\epsilon$. If $\max\{|n|,|k|\}\ge y^4$, then

$$S(n,k)-\hat{S}(n,k)\le[S(n,k)-y]_+.$$

So

$$\begin{aligned}
&\limsup_{m\to\infty}\frac{1}{m}\max_{\xi\in\mathcal{A}(m)}\{S(\xi)-\hat{S}(\xi)\}\\
&\quad\le\limsup_{m\to\infty}\frac{1}{m}\max_{\xi\in\mathcal{A}(m)}\Bigg\{\sum_{\substack{(n,k)\in\xi\\ \max\{|n|,|k|\}\ge y^4}}[S(n,k)-y]_+ +\sum_{\substack{(n,k)\in\xi\\ \max\{|n|,|k|\}<y^4}}[S(n,k)-\hat{S}(n,k)]\Bigg\}\\
&\quad\le\limsup_{m\to\infty}\frac{1}{m}\max_{\xi\in\mathcal{A}(m)}\sum_{(n,k)\in\xi}[S(n,k)-y]_+ +\sum_{|n|<y^4,|k|<y^4}S(n,k)\times\limsup_{m\to\infty}\frac{1}{m}\\
&\quad\le\epsilon+0.
\end{aligned}$$

Since this holds for all $\epsilon>0$, we have (5.12). □

Next we give the “concentration of measure” result which we need, followed by a particular application of it to our path model.

Lemma 5.6 *Let $L > 0$, and let $X_i, 1 \le i \le N$ be independent random variables, such that*

$$\mathbb{P}(0 \le X_i \le L) = 1$$

for each i. Let $\mathcal{C}$ be a set of subsets of $\{1, 2, \dots, N\}$, such that

$$\max_{C \in \mathcal{C}} |C| \le r,$$

and let

$$Z = \max_{C \in \mathcal{C}} \sum_{i \in C} X_i.$$

Then, for all $u > 0$,

$$\mathbb{P}\big(|Z - \mathbb{E}Z| \ge u\big) \le \exp\left(-\frac{u^2}{16rL^2} + 64\right).$$

Proof: This follows easily from Theorem 8.1.1 of [20]; see for example Lemma 5.1 of [14].
□

Lemma 5.7 *Let $n_1 \le n_2$, $k_1 \le k_2$. Let $R = |n_1| + |n_2| + |k_1| + |k_2|$, and assume that $R > 0$. Let $u > 0$. Then*

$$\mathbb{P}\Big(\big|\hat{T}\big((n_1, k_1), (n_2, k_2)\big) - \mathbb{E}\hat{T}\left((n_1, k_1), (n_2, k_2)\right)\big| \ge u\Big) \le \exp\left(-\frac{u^2}{48R^{3/2}} + 64\right).$$

Proof: From Lemma 5.4(i), any path from (n_1, k_1) to (n_2, k_2) has at most $3R$ points, and, from Lemma 5.4(ii) and the definition (5.10) of the truncated variables, any point (n, k) on such a path has truncated weight $\hat{S}(n, k) \le R^{1/4}$. Hence we can apply Lemma 5.6 with $\mathcal{C} = \mathcal{P}\big((n_1, k_1), (n_2, k_2)\big)$, $L = R^{1/4}$ and $r = 3R$. □

5.3 Proof of Theorem 5.1

First take the case of rational x. Let $x = n/k$, where $n \ge 0$, $k > 0$. For $0 \le l < m$, define

$$X(l, m) = T\big((ln, lk), (mn, mk)\big) - S(ln, lk).$$

From the definition (2.6) of the quantities $\{T(n, k)\}$ and the i.i.d. property for the collection $\{S(n, k)\}$, one has the following properties:

Non-negativity:	$X(l, m) \ge 0, 0 \le l < m$;
Superadditivity:	$X(l, r) \ge X(l, m) + X(m, r), 0 \le l < m < r$;
Stationarity:	$\{X(j + l, j + m), 0 \le l < m\}$ has the same distribution for all $j \ge 0$.
Independence:	$X(l, m)$ and $X(r, s)$ are independent whenever $0 \le l < m \le r < s$.

Under condition (5.1), it also follows from Lemma 5.4(iii) and from (5.11) that

$$\begin{aligned}\limsup_{m \to \infty} \frac{1}{m} X(0, m) &\le \limsup_{m \to \infty} \frac{1}{m} \max_{\xi \in \mathcal{A}(3mn + nk + 1)} S(\xi) \\ &< \infty\end{aligned}$$

with probability 1. Then, by a superadditive ergodic theorem, there exists $L < \infty$ such that

$$\frac{1}{m} X(0, m) \to L \text{ a.s. and in } \mathcal{L}_1$$

as $m \to \infty$, and

$$\mathbb{E} X_{0,1} \leq L.$$

This yields (5.3) for all rational x, and (5.4) for all n, k.

The extension of (5.3) to all nonnegative real x, along with the proof of properties (ii) and (iii), can then be done just as in the proof of Theorem 6.3 of Glynn and Whitt [10].

For property (i), note that

$$\begin{aligned}
\gamma(0) &= \lim_{k \to \infty} \frac{1}{k} T(0, k) \\
&= \lim_{k \to \infty} \frac{1}{k} \sum_{i=0}^{\infty} S(i, k) \\
&= \mathbb{E} S
\end{aligned}$$

by the law of large numbers.

For part (iv), the concavity of γ implies that, for $x > 0$, γ is continuous at x and has a right derivative $\gamma'(x)$, which is decreasing in x. On the other hand, from (ii), $\gamma'(x)$ is bounded below by β. Hence the limit $\lim_{x \to \infty} \gamma'(x)$ exists in $[\beta, \infty)$. □

5.4 Proof of Theorems 5.2 and 5.3

We first need a lemma concerning the deviations of arrival times from their expectations:

Lemma 5.8 *Let $\{A(n),\, n \geq 1\}$ be stationary and ergodic, with $\mathbb{E} A(n) = \alpha$. Let*

$$U_N = \frac{1}{N} \max_{1 \leq r \leq N} \left| \sum_{n=1}^{r} A(n) - r\alpha \right|.$$

Then $U_N \to 0$ a.s. as $N \to \infty$.

Proof: Take $\epsilon > 0$; by the ergodic theorem, there exists with probability 1 an R_ϵ such that

$$\frac{1}{r} \left| \sum_{n=1}^{r} A(n) - r\alpha \right| < \epsilon \text{ for all } r > R_\epsilon.$$

Then also

$$\frac{1}{N} \left| \sum_{n=1}^{r} A(n) - r\alpha \right| < \epsilon \text{ whenever } r > R_\epsilon \text{ and } r \leq N.$$

But since there are only finitely many $r \leq R_\epsilon$, there also exists with probability 1 an $N_\epsilon \geq R_\epsilon$ such that

$$\frac{1}{N}\left|\sum_{n=1}^{r} A(n) - r\alpha\right| < \epsilon \text{ whenever } r \leq R_\epsilon \text{ and } N \geq N_\epsilon.$$

Then if $N \geq N_\epsilon$, $U_N < \epsilon$, so the a.s. convergence holds as desired. $\square$

Proof of Theorem 5.2: Let η and m be such that

$$\lim_{x\to\infty} \gamma'(x) < \alpha - m < \eta < \alpha. \tag{5.13}$$

Then

$$\lim_{x\to\infty} \frac{1}{x}\big[\gamma(x) - \alpha x\big] < -m,$$

so we can choose ξ such that, for all $x > \xi$,

$$\gamma(x) - \alpha x \leq -mx, \tag{5.14}$$

and such that

$$\sup_x\{\gamma(x) - \alpha x\} = \sup_{x\leq\xi}\{\gamma(x) - \alpha x\}.$$

For convenience of notation, define

$$M(n, K) = T\big((-n, 0), (0, K)\big) \tag{5.15}$$

and

$$\hat{M}(n, K) = \hat{T}\big((-n, 0), (0, K)\big). \tag{5.16}$$

To show (5.5) and (5.6), it will suffice to show that

$$\sup_{n>\xi K}\Big\{M(n, K) - \sum_{i=1}^{n} A(-i)\Big\} < \infty \text{ a.s.} \tag{5.17}$$

for all K, that

$$\left[\frac{1}{K}\sup_{n>\xi K}\Big\{M(n, K) - \sum_{i=1}^{n} A(-i)\Big\}\right]_+ \to 0 \text{ a.s.} \tag{5.18}$$

as $K \to \infty$, and that

$$\frac{1}{K}\max_{n\leq\xi K}\Big\{M(n, K) - \sum_{i=1}^{n} A(-i)\Big\} \to \sup_{x\leq\xi}\{\gamma(x) - \alpha x\} \text{ a.s.} \tag{5.19}$$

as $K \to \infty$.

We treat the situations $n > \xi K$ and $n \leq \xi K$ separately.

Step 1: $n > \xi K$; proof of (5.17) and (5.18):

By the ergodic theorem, we have that, with probability 1, $\sum_{i=0}^{n-1} A(-i) - \eta n$ will be positive for all sufficiently large n; so it suffices to show that

$$\sup_{n>\xi K} \left\{M(n,K) - \eta n\right\} < \infty \text{ a.s.} \tag{5.20}$$

for each K, and that

$$\left[\frac{1}{K}\sup_{n>\xi K} \left\{M(n,K) - \eta n\right\}\right]_+ \to 0 \text{ a.s.} \tag{5.21}$$

as $K \to \infty$.

By (5.4) and then (5.14), we have, for any n, K,

$$\begin{aligned}
\mathbb{E}\hat{M}(n,K) &\leq \mathbb{E}M(n,K) \\
&= \mathbb{E}T(n,K) \\
&\leq K\gamma(n/K) + \beta \\
&\leq K\left[\gamma\left(\frac{n}{K}\right) - \alpha\frac{n}{K}\right] + \alpha n + \beta \\
&\leq K\left[-m\frac{n}{K}\right] + \alpha n + \beta \\
&= \eta n - 2m'n + \beta,
\end{aligned}$$

where $m' = (m + \eta - \alpha)/2 > 0$. We can then decompose the LHS of (5.17) and (5.18) by writing

$$\begin{aligned}
&\sup_{n>\xi K} \{M(n,K) - \eta n\} \\
&\quad = \sup_{n>\xi K} \left\{[M(n,K) - \hat{M}(n,K)] + [\hat{M}(n,K) - \mathbb{E}\hat{M}(n,K)] + [\mathbb{E}\hat{M}(n,K) - \eta n]\right\} \\
&\quad \leq \sup_{n>\xi K} \left\{[M(n,K) - \hat{M}(n,K)] + [\hat{M}(n,K) - \mathbb{E}\hat{M}(n,K)] - 2m'n + \beta\right\} \\
&\quad \leq \sup_{n>\xi K} \left\{M(n,K) - \hat{M}(n,K) - m'n\right\} \\
&\qquad\qquad + \sup_{n>\xi K} \left\{\hat{M}(n,K) - \mathbb{E}\hat{M}(n,K) - m'n\right\} + \beta.
\end{aligned} \tag{5.22}$$

We can now prove (5.20) and (5.21) by showing that, with probability 1, both the suprema on the RHS of (5.22) are finite for all K and are negative for all sufficiently large K.

For the first term, (5.12) together with Lemma 5.4(iii) implies that

$$\frac{1}{n}\max_{K<\xi^{-1}n} \{M(n,K) - \hat{M}(n,K)\} \to 0 \text{ a.s.}$$

as $n \to \infty$. Thus

$$\begin{aligned}
\mathbb{P}&\left(\exists \text{ infinitely many } (n,K) : n > \xi K \text{ and } M(n,K) - \hat{M}(n,K) \geq m'n\right) \\
&\qquad = \mathbb{P}\left(\exists \text{ infinitely many } n : \max_{K<\xi^{-1}n} \{M(n,K) - \hat{M}(n,K)\} \geq m'n\right) \\
&\qquad = 0,
\end{aligned}$$

so, with probability 1, the first supremum on the RHS of (5.22) is always finite and eventually negative.

For the second term on the RHS of (5.22), using Lemma 5.7 gives

$$\mathbb{P}(\hat{M}(n,K) - \mathbb{E}\hat{M}(n,K) \geq m'n) \leq \exp\left(-\frac{(m'n)^2}{48(n+K)^{3/2}} + 64\right).$$

Thus

$$\begin{aligned}
\sum_{n,K:n>\xi K} & \mathbb{P}(\hat{M}(n,K) - \mathbb{E}\hat{M}(n,K) - m'n \geq 0) \\
&\leq \sum_{n=1}^{\infty} \sum_{K=0}^{\lfloor \xi^{-1} n \rfloor} \exp\left(-\frac{(m'n)^2}{48(n+K)^{3/2}} + 64\right) \\
&\leq \sum_{n=1}^{\infty} (1+\xi^{-1}n) \exp\left(-\frac{(m'n)^2}{48(1+\xi^{-1})^{3/2} n^{3/2}} + 64\right) \\
&= \sum_{n=1}^{\infty} c_1 n \exp(-c_2 n^{1/2}) \text{ for some } c_1, c_2 > 0 \\
&< \infty.
\end{aligned}$$

It then follows from Borel-Cantelli that, with probability 1, the second supremum on the RHS of (5.22) is eventually negative. This completes the proof of (5.17) and (5.18).

Step 2: $n \leq \xi K$; proof of (5.19):
Using (5.12) and Lemma 5.4(iii), we have that

$$\frac{1}{K} \max_{0 \leq n \leq \xi K} \left| M(n,K) - \hat{M}(n,K) \right| \to 0 \text{ a.s.}$$

as $K \to \infty$. Also, from Lemma 5.8, we have that

$$\frac{1}{K} \max_{0 \leq n \leq \xi K} \left| \sum_{i=1}^{n} A(-i) - n\alpha \right| \to 0 \text{ a.s.}$$

as $K \to \infty$. Hence to prove (5.19), it suffices to show that

$$\frac{1}{K} \max_{0 \leq n \leq \xi K} \left\{ \hat{M}(n,K) - n\alpha \right\} \to \sup_{0 \leq x \leq \xi} \left\{ \gamma(x) + \alpha x \right\} \text{ a.s.} \tag{5.23}$$

as $K \to \infty$.

Now for all $0 \leq x \leq \xi$,

$$\begin{aligned}
\left| \frac{1}{K} \mathbb{E} M\big(\lfloor xK \rfloor, K\big) - \gamma(x) \right| &= \left| \frac{1}{K} \mathbb{E} T\Big(\big(-\lfloor xK \rfloor, 0\big), \big(0, K\big)\Big) - \gamma(x) \right| \\
&= \left| \frac{1}{K} \mathbb{E} T\big(\lfloor xK \rfloor, K\big) - \gamma(x) \right| \\
&\to 0
\end{aligned}$$

as $K \to \infty$, by stationarity and the $\mathcal{L}_1$ convergence in (5.3); then, using (5.12) again and dominated convergence, we have that

$$\left| \frac{1}{K} \mathbb{E} \hat{M}\big(\lfloor xK \rfloor, K\big) - \gamma(x) \right| \to 0$$

as $K \to \infty$ also. This gives

$$\liminf_{K\to\infty} \frac{1}{K} \max_{0\le n\le \xi K} \left\{\hat{M}(n,K) - n\alpha\right\} \ge \sup_{0\le x\le \xi} \{\gamma(x) - \alpha x\}.$$

On the other hand, using (5.4),

$$\begin{aligned}\frac{1}{K}\left[\mathbb{E}\hat{M}(n,K) - n\alpha\right] &= \frac{1}{K}\left[\mathbb{E}\hat{T}(n,K) - n\alpha\right] \\ &\le \frac{1}{K}\left[\mathbb{E}T(n,K) - n\alpha\right] \\ &\le \gamma\left(\frac{n}{K}\right) + \frac{\beta}{K} - \frac{n}{K}\alpha,\end{aligned}$$

for all $0 \le n \le \xi K$, giving

$$\limsup_{K\to\infty} \frac{1}{K} \max_{0\le n\le \xi K} \left\{\mathbb{E}\hat{M}(n,K) - n\alpha\right\} \le \sup_{0\le x\le \xi} \{\gamma(x) - \alpha x\}.$$

Combining these two gives

$$\frac{1}{K} \max_{0\le n\le \xi K} \left\{\mathbb{E}\hat{M}(n,K) - n\alpha\right\} \to \sup_{0\le x\le \xi} \{\gamma(x) - \alpha x\} \tag{5.24}$$

as $K \to \infty$.

The concentration inequality in Lemma 5.7 gives, for any $\epsilon > 0$,

$$\begin{aligned}&\mathbb{P}\left(\left|\frac{1}{K} \max_{0\le n\le \xi K} \left\{\hat{M}(n,K) - n\alpha\right\} - \frac{1}{K} \max_{0\le n\le \xi K} \left\{\mathbb{E}\hat{M}(n,K) - n\alpha\right\}\right| > \epsilon\right) \\ &\qquad\le \sum_{n=0}^{\lfloor \xi K\rfloor} \mathbb{P}\left(\left|\hat{M}(n,K) - \mathbb{E}\hat{M}(n,K)\right| > \epsilon K\right) \\ &\qquad\le (\xi K + 1)\exp\left(-\frac{(\epsilon K)^2}{48[(1+\xi)K]^{3/2}} + 64\right) \\ &\qquad\le (\xi K + 1)\exp\left(-\frac{\epsilon^2 K^{1/2}}{48(1+\xi)^{3/2}} + 64\right).\end{aligned}$$

The sum of the RHS over all $K \ge 1$ is finite; since this holds for all $\epsilon > 0$, combining it with (5.24) gives the desired convergence in (5.23). This completes the proof of (5.19) and hence of the theorem. □

Proof of Theorem 5.3: Take η from (5.7), and define ξ as in (5.13) and (5.14). Define $M(n,K)$ and $\hat{M}(n,K)$ as at (5.15) and (5.16).

It suffices to demonstrate that, in the case $c < \infty$, convergence in (5.18) and (5.19) holds in $\mathcal{L}_1$ as well as a.s.

The $\mathcal{L}_1$ convergence in (5.19) can be shown using the Dominated Convergence Theorem, since the LHS of (5.19) is non-negative and no larger than

$$\frac{1}{K}T\Big(\big(-\lfloor \xi K\rfloor, 0\big), (0, K)\Big),$$

which, by (5.3) and stationarity, converges in $\mathcal{L}_1$ to $\gamma(\xi)$ as $K \to \infty$.

Using (5.7), the $\mathcal{L}_1$ convergence in (5.18) will follow if we show that

$$\left[\frac{1}{K}\sup_{n>\xi K}\{M(n,K)-\eta n\}\right]_+ \to 0 \tag{5.25}$$

in $\mathcal{L}_1$ as $K\to\infty$. This will be proved by exploiting the fact that the process

$$\Big\{\sup_{n\geq\xi(K)}\{M(n,K)-\eta n\}, K\geq 0\Big\}$$

has the same marginal distributions as a process with a suitable superadditive property.

Extend the i.i.d. collection $\{S(n,k)\}$ to include all negative values of k. Then for $i\geq 1$, let

$$Q_{0,i}=\sup_{n>\xi i}\left\{T\big((-n,-i+1),(0,0)\big)-\eta n\right\}, \tag{5.26}$$

and let N_i be the (smallest, say) value of n attaining the supremum on the RHS; note that N_i is a.s. finite by the same argument used to prove (5.17). Define also $N_0=0$.

Then for $1\leq i<j$, let

$$Q_{i,j}=\sup_{n>\xi(j-i)}\left\{T\big((-N_i-n,-j+1),(-N_i,-i)\big)-\eta n\right\}. \tag{5.27}$$

For $Q_{0,i}$ we have maximised a quantity over a set of paths between "level $-i+1$" and "level 0" Then $Q_{i,j}$ takes a similar maximum over paths between levels $-j+1$ and $-i$ which are constrained to finish immediately below the maximising path from level $-i+1$ to level 0. This gives us the superadditive property we need; formally, for all $1\leq i<j$,

$$\begin{aligned}Q_{0,j}&=\sup_{n>\xi j}\left\{T\big((-n,-j+1),(0,0)\big)-\eta n\right\}\\&\geq\sup_{n'>\xi(j-i)}\Big\{T\big((-N_i-n,-j+1),(-N_i,-i)\big)\\&\qquad\qquad\qquad+T\big((-N_i,-i+1),(0,0)\big)-\eta(N_i+n')\Big\}\end{aligned}$$

(since $N_i>\xi i$ and since there is a feasible path step $(-N_i,-i)\to(-N_i,-i+1)$)

$$\begin{aligned}&=\sup_{n'>\xi(j-i)}\Big\{T\big((-N_i-n',-j+1),(-N_i,-i)\big)-\eta n'\Big\}\\&\qquad\qquad\qquad+T\big((-N_i,-i+1),(0,0)\big)-\eta N_i\\&=Q_{0,i}+Q_{i,j}.\end{aligned}$$

So far, this is all true for any value of c. However, in the case $c=\infty$ in particular, there are no allowable path steps which decrease k or n; thus all the paths involved in the definition of $Q_{i,j}$ visit only points (n,k) with $n\leq N_i$ and $-j<k\leq -i$.

From this observation and from the form of the definitions (5.26) and (5.27), we can in fact write

$$Q_{m,m+1}=f_i\big(\{S(-N_m+n,-m+k),n\leq 0,-i<k\leq 0\}\big)$$

for all $m \geq 0$, $i > 0$, where the function f_i is independent of m.

Now, for all $m \geq 1$, N_m depends only on $\{S(n,k), n \leq 0, -m < k \leq 0\}$, so, since the entire array $S(n,k)$ is i.i.d., we have that $\{S(-N_m + n, -m + k), n \leq 0, k \leq 0\}$ is independent of N_m and of $\{S(n,k), n \geq 0, -m < k \leq 0\}$, and is itself an i.i.d. array with the same common distribution as the original array $\{S(n,k)\}$.

Hence, for all $m \geq 1$, the collection $\{Q_{i,j}, m \leq i < j\}$ is independent of the collection $\{Q_{i,j}, 0 \leq i < j \leq m\}$, and the collection $\{Q_{m,m+1}, i \geq 1\}$ has the same distribution for each $m \geq 0$. (Note that although the superadditivity still holds for the cases $c < \infty$, these independence and stationarity properties both fail in general). It then follows from a superadditive version of the subadditive ergodic theorem proved by Liggett in [12] that there exists a constant $q \leq \infty$ such that

$$\frac{1}{K} Q_{0,K} \to q \text{ a.s. as } K \to \infty,$$

and, if $q < \infty$,

$$\frac{1}{K} Q_{0,K} \to q \text{ in } \mathcal{L}_1 \text{ as } K \to \infty. \tag{5.28}$$

Observe now that for each K, $\sup_{n > \xi(K-1)}\{M(n, K-1) - \eta n\}$ has the same distribution as $Q_{0,K}$; we have merely translated the set of paths considered by a distance $(K-1)$ in the second coordinate. Since the $\{S(n,k)\}$ are i.i.d., this makes no difference to the distribution for any given K, (though joint distributions over several values of K are generally affected).

Using (5.21), we have

$$\begin{aligned}
\limsup_{K\to\infty} \frac{1}{K} \sup_{n > \xi(K-1)} \{M(n, K-1) - \eta n\} &= \limsup_{K\to\infty} \frac{1}{K} \sup_{n > \xi K} \{M(n,K) - \eta n\} \\
&\leq 0 \text{ a.s.}
\end{aligned}$$

Then for any $\epsilon > 0$,

$$\begin{aligned}
\lim_{K\to\infty} \mathbb{P}\left(\frac{1}{K} Q_{0,K} \geq \epsilon\right) &= \lim_{K\to\infty} \mathbb{P}\left(\frac{1}{K} \sup_{n \geq \xi(K-1)} \{M(n, K-1) - \eta n\} \geq \epsilon\right) \\
&= 0.
\end{aligned}$$

Hence $q \leq 0$, and so, from (5.28),

$$\mathbb{E}\left[\frac{1}{K} Q_{0,K} - q\right]_+ \to 0 \text{ as } K \to \infty.$$

Thus also

$$\mathbb{E}\left[\frac{1}{K} \sup_{n \geq \xi(K-1)} \{M(n, K-1) - \eta n\} - q\right]_+ \to 0 \text{ as } K \to \infty,$$

which yields (5.25) as required. $\square$

Acknowledgements

I'm very grateful to François Baccelli and Jean Mairesse for many valuable conversations during the course of this work. I thank the referee for a very thorough reading and for helpful comments.

References

[1] Baccelli, F., Borovkov, A. and Mairesse, J., (2000) Asymptotic Results on Infinite Tandem Queueing Networks. *Probab. Theory Related Fields* **118**, 365–405.

[2] Baccelli, F. and Konstantopoulos, P., (1992) Estimates of Cycle Times in Stochastic Petri Nets. In *Applied Stochastic Analysis, Proc. US-French Workshop*, no. 177 in Lecture Notes in Control and Information Sciences, pages 186–209. Springer-Verlag.

[3] Baryshnikov, Y., (2001) GUEs and Queues. *Probab. Theory Related Fields* **119**, 256–274.

[4] Billingsley, P., (1968) *Convergence of Probability Measures.* John Wiley & Sons.

[5] Cheng, D. W. and Yao, D. D., (1993) Tandem Queues with General Blocking: a Unified Model and Comparison Results. *J. Discrete Event Dyn. Syst. Th. Appl.* **2**, 207–234.

[6] Cox, J. T., Gandolfi, A., Griffin, P. S. and Kesten, H., (1993) Greedy Lattice Animals I: Upper Bounds. *Ann. Appl. Prob.* **3**, 1151–1169.

[7] Daley, D. J. and Rolski, T., (1992) Finiteness of Waiting-Time Moments in General Stationary Single-Server Queues. *Ann. Appl. Prob.* **2**, 987–1008.

[8] Ekhaus, M. and Gray, L., (1994) Convergence to Equilibrium and a Strong Law for the Motion of Restricted Interfaces. Unpublished Manuscript.

[9] Gandolfi, A. and Kesten, H., (1994) Greedy Lattice Animals II: Linear Growth. *Ann. Appl. Prob.* **4**, 76–107.

[10] Glynn, P. W. and Whitt, W., (1991) Departures from Many Queues in Series. *Ann. Appl. Prob.* **1**, 546–572.

[11] Gravner, G., Tracy, C. A. and Widom, H., (2001) Limit Theorems for Height Fluctuations in a Class of Discrete Space and Time Growth Models. *J. Stat. Phys.* **102**, 1085–1132.

[12] Liggett, T. M., (1985) An Improved Subadditive Ergodic Theorem. *Ann. Prob.* **13**, 1279–1285.

[13] Mairesse, J. and Prabhakar, B., (1999) On the Existence of Fixed Points for the $./GI/1/\infty$ Queue. LIAFA research report.

[14] Martin, J. B., (2001) Linear Growth for Greedy Lattice Animals. *Stoch. Proc. Appl.* To appear. Currently available from `<http://www.statslab.cam.ac.uk/~jbm/papers.html>`.

[15] O'Connell, N. and Yor, M. Brownian Analogues of Burke's Theorem. *Stoch. Proc. Appl.* To appear.

[16] O'Connell, N. M., (1999) Directed Percolation and Tandem Queues. DIAS Technical Report, DIAS-STP-99-12.

[17] Seppäläinen, T., (1997) A Scaling Limit for Queues in Series. *Ann. Appl. Prob.* **7**, 855–872.

[18] Seppäläinen, T., (1999) Existence of Hydrodynamics for the Totally Asymmetric Simple K-Exclusion Process. *Ann. Prob.* **27**, 361–415.

[19] Srinivasan, R., (1993) Queues in Series via Interacting Particle Systems. *Math. Oper. Res.* **18**, 39–50.

[20] Talagrand, M., (1995) Concentration of Measure and Isoperimetric Inequalities in Product Spaces. *Inst. Hautes Etudes Sci. Publ. Math.* **81**, 73–205.

[21] Whitt, W., (1980) Some Useful Functions for Functional Limit Theorems. *Math. Oper. Res.* **5**, 67–85.