Stability of the Greedy Algorithm on the Circle

Leonardo T. Rolla, Vladas Sidoravicius

Argentinian National Research Council, University of Buenos Aires

NYU-ECNU Institute of Mathematical Sciences at NYU Shanghai

Courant Institute of Mathematical Sciences, New York University

Abstract

We consider a single-server system with service stations in each point of the circle. Customers arrive after exponential times at uniformly-distributed locations. The server moves at finite speed and adopts a greedy routing mechanism. It was conjectured by Coffman and Gilbert in 1987 that the service rate exceeding the arrival rate is a sufficient condition for the system to be positive recurrent, for any value of the speed. In this paper we show that the conjecture holds true.

1 Introduction

In this paper we study a greedy single-server system on the unit-length circle $\mathbb{R}/\mathbb{Z}$. Customers arrive following a Poisson process with rate λ. Each arriving customer chooses a position on $\mathbb{R}/\mathbb{Z}$ uniformly at random and waits for service. If there are no customers in the system, the server stands still. Otherwise, the server chooses the nearest waiting customer and travels in that direction at speed $v > 0$, ignoring any new arrivals. Upon reaching the position of such customer, the server stays there until service completion, which takes a random time T that is independent of the past configurations and has expectation μ^{-1}.

The above system was introduced by Coffman and Gilbert in 1987 [CG87], and since then became a paradigm example of a routing mechanism that depends on the system state. This is the so-called *greedy server*, due to the simple strategy of targeting the nearest customer.

Continuous-space models provide natural approximations for systems with a large number of service stations embedded in a spacial structure, and their description is usually more transparent than the discrete-space formulation, mostly because the latter often is obscured by combinatorial aspects. However, systems with greedy routing strategies in the continuum are extremely sensitive to microscopic perturbations, and their rigorous study represents a mathematical challenge.

It was conjectured in [CG87] that the greedy server on the circle should be a stable system when $\lambda < \mu$ for any $v > 0$. Since then, a number of related models have been proposed and studied. Stability was verified under light-traffic assumptions, i.e., for λ and μ fixed and v large enough,

and for the greedy server on a discrete ring $\mathbb{Z}/n\mathbb{Z}$. However, these approximations were unable to identify and tackle the main difficulty of this system, which is is due to the interplay between the server's motion and the environment of waiting customers that surround it. This interplay is given by the interaction resulting from the choice of the next customer and the removal of those who have been served. In this paper we prove stability for the greedy server.

Definition. We say that t is a *regeneration time* if the system becomes empty at time t, i.e., if there is one customer at time $t-$ and no customers at time $t+$. Let $\tau_\varnothing := \inf\{t > 0 : t \text{ is a regeneration time}\}$. We say that the system is *recurrent* if, starting from the empty state $\varnothing$, there will be a.s. a regeneration time, i.e., $\mathbb{P}^\varnothing[\tau_\varnothing < \infty] = 1$. We say that the system is *stable*, or *positively recurrent*, if $\mathbb{E}^\varnothing[\tau_\varnothing] < \infty$.

Theorem 1.1. *Suppose that the distribution of the service time T is geometric, exponential, or deterministic. For any $\lambda < \mu$ and any $v > 0$, the greedy server on the circle is stable.*

Remark. In our approach, it is crucial that the arrivals are Poisson in space-time. There is a *dynamic* version of the greedy server, where *new arrivals are not ignored* while the server is traveling. This variation might be studied by similar arguments, but the dynamic mechanism introduces some extra complications that will not be considered here. A proof of stability for general service times having an exponential moment follows from the same approach as presented here, requiring a little extra work due to the lack of Markov property. We present the proof for exponentially distributed service times with $\mu = 1$. The cases of geometric or deterministic times only differ in notation.

For the proof of Theorem 1.1, we consider a representation for the conditional distribution of the set of waiting customers in terms of a stochastic evolution of profiles. In this framework, the server learns only the information that is necessary and sufficient to determine the next movement, and the positions of further waiting customers remain unknown. This approach was used in [FRS15] to show that the greedy server on the real line is transient, which is an important ingredient in our proof of stability.

In the remainder of this introduction, we review some known results on the greedy server and related models, discuss the problem of self-interaction, describe the approach based on a stochastic evolution of profiles, present a heuristic discussion in order to highlight the main ideas of the proof, and finally give a brief outline of the paper structure.

1.1 Previous results on the greedy server and related models

Stability was verified for the greedy server on $\mathbb{R}/\mathbb{Z}$ under light-traffic assumptions [KS96] and for the greedy server on a discrete ring $\mathbb{Z}/n\mathbb{Z}$ [FL96, FL98, MQ99], see below. It was also shown for several related models, including a class of non-greedy policies [KS94], a gated-greedy variant on convex spaces [AL94], and random non-greedy servers on general spaces [AF97]. See [RNFK11] and references therein for a recent review.

The *light-traffic* regime is given by

$$\lambda\left(\frac{1}{\mu} + \frac{1}{2v}\right) < 1.$$

This regime was studied in [KS96], particularly the limit $\lambda \to 0$ for which the first terms of Taylor expansions of some performance measurements were computed. A simple coupling argument works for proving stability under light-traffic assumption. In this case, $\frac{1}{2v}$ gives an upper bound for the travel time between two consecutive services, since $\frac{1}{2}$ is the largest distance within the unit-length circle. Adding this bound to the service time allows a comparison between the greedy server on $\mathbb{R}/\mathbb{Z}$ and a stable $M/G/1$ system, which proves that the former is stable.

This kind of argument could be pushed down to smaller values of v than the above inequality allows by obtaining a stochastic upper bound on the distance to the nearest customer better than $\frac{1}{2}$. But in any case, it may not be extended to general $v > 0$, because the presence of the server interferes severely with the conditional distribution of the locations of remaining customers.

On the other hand, stability under the general condition $\lambda < \mu$ is known to hold for the *polling server* on $\mathbb{R}/\mathbb{Z}$, i.e., the server whose strategy is to always travel in the same direction. In [KS92] this fact was proven using a decomposition of the set of waiting customers into a collection of Galton-Watson trees that turn out to be subcritical for $\lambda < \mu$. This decomposition provides a detailed description of the busy cycles (sequence of configurations observed between two consecutive regeneration times) and the stationary state, but if one only wants to prove stability, there is a simple and robust argument. Take $\varepsilon < 1 - \frac{\lambda}{\mu}$ and K such that

$$\frac{K}{\mu} + \frac{1}{v} < (1 - \varepsilon)\frac{K}{\lambda}.$$

The above inequality implies that, whenever the number of waiting customers N is larger than K, the time it will typically take to serve the first N customers, including service and travel time, is less than the time it will typically take for the next $N - \varepsilon N$ arrivals, resulting in a net decrease by εN on the number of waiting customers.

Simulations indicate that, under heavy traffic conditions, the greedy server dynamics resembles that of the polling server [CG87]. This suggests that a possible strategy for proving stability of the greedy server might be to adapt the above argument. In this case, the first step would be to understand its local behavior, and a natural approach is to consider a system on an infinite line. A model on $\mathbb{Z}$ was studied in [KM97], where it is shown that the server is eventually going to move in a fixed random direction.

Some direct attempts also include the study of a greedy server model on the finite ring $\mathbb{Z}/n\mathbb{Z}$, which was shown to be stable in [FL96, FL98, MQ99]. Each of these references provides different arguments under a variety of general assumptions.

Yet discrete models have not been able to grasp the microscopic nature of the greedy mechanism in continuous space, neither on $\mathbb{Z}$ nor on $\mathbb{Z}/n\mathbb{Z}$, and there are major obstacles in extrapolating any approach based on a discrete approximation. This difficulty is due to the *self-interaction* of the server's path at the *microscopic level*, which takes place because the server's trajectory influences the set of waiting customers and at the same time is determined by the latter.

1.2 Self-interacting motions

The main difficulty in studying greedy server systems in continuous spaces is due to the interplay between the server's motion and the environment of waiting customers that surround it. This

interplay is given by the interaction resulting from the greedy choice of the next customer and the removal of those who have been served. The server's path is *self-repelling*, since the removal of already served customers makes it less likely for the greedy server to take the next step back into the recently visited regions.

In some well-known examples of self-repelling motions, the self-interaction comes from an explicit prescription of the distribution of the next step in terms of the past occupation times. For the excited random walks [BW03], perturbed Brownian motions [CPY98, CD99, Dav96, Dav99, PW97], and excited Brownian motions [RS11], whenever there is a drift, it is pushing the motion in a certain fixed direction. For the random walk avoiding its past convex hull [ABV03, Zer05] and the prudent walk [BFV10, BM10], there is a growing forbidden region containing the previous trajectory, which strongly pushes the motion outwards.

For our greedy server, and also for the true self-avoiding walk [Tót95, Tót99], the true self-repelling motion [TW98], and the Brownian motion with repulsion [MT08], there is a mixture of information, and "self-repulsion" does not immediately imply "repulsion towards ∞", since the particle is allowed to cross its past path, receiving contradictory signals from its left and right-hand sides. In fact, some of the latter models are recurrent and some are transient.

It was clear since these models were introduced that they could not be treated via standard methods and tools. A lot remains to be understood even in dimension $d = 1$, and, despite the existence of a few disconnected techniques that have proved useful in particular situations, this rich field of study lacks a systematic basis.[1]

The greedy server model has two particular features. Unlike most of the above models, here there is no direct prescription of how the past trajectory influences the future in terms of occupation times. Moreover, this evolution is time-inhomogeneous in the sense that customers keep accumulating, which yields an increasing bias towards least recently explored regions with decreasing traveled inter-distances after each service.

1.3   Stochastic evolution of profiles

To address the issues mentioned in the above subsection, we consider a representation of the customers environment which reflects its randomness as perceived by the server.

More precisely, we only want to learn the information that is necessary and sufficient to determine the next movement, and the positions of further waiting customers should remain unknown. Each time the server has to scan the system state to determine the position of the next target, we acquire exactly two pieces of information: the presence of a customer at that position and the absence of any other customer at smaller distances, as illustrated in Figure 1.1.

The arrivals are represented by a space-time Poisson Point Process $\nu \subseteq (\mathbb{R}/\mathbb{Z}) \times \mathbb{R}$, and in this approach one is ignoring the points of ν that have not yet influenced the server's trajectory. One can think of this scheme as re-sampling the set of waiting customers at each departure time, according to the appropriate conditional distribution. The latter is given in terms of the space-time region where the configuration ν has not been revealed. In this setting, the state of the system is given by

---

[1]Except for the family of universality classes given by the Schramm-Löwner Evolutions [Sch00], which include 2-dimensional loop-erased random walk [LSW04a] and several other models [LSW04b, Smi01, Smi10].

the positions of the server and the current customer, plus the profile corresponding to the boundary of this region where ν is unknown. The knowledge of this triplet determines the distribution of its future without the need of any further information from the past, yielding a Markovian evolution. See Figure 1.1.

1.4   Heuristics

If the server is busy most of the time, the system must be stable, since on average the service time is smaller than the inter-arrival time. The fundamental problem in showing stability is therefore the possibility that the server spend a long time zigzagging on regions with low density of customers due to a trapping configuration produced by the stochastic dynamics.

For the analogous model on the real line, this cannot be the case: the server may zigzag for a finite period of time, but it is bound to eventually choose a direction and head that way [FRS15].

On the same grounds, since the greedy routing mechanism is *local*, this can neither be the case on the circle – at least *until the server realizes that it is not operating on the infinite line*.

Suppose we are given a configuration where the circle is crowded of waiting customers, and, from this point on, our goal is to alleviate this situation. We would like to say that, with high probability, after a short time the server will choose a direction and then cope with its workload as the polling server would.
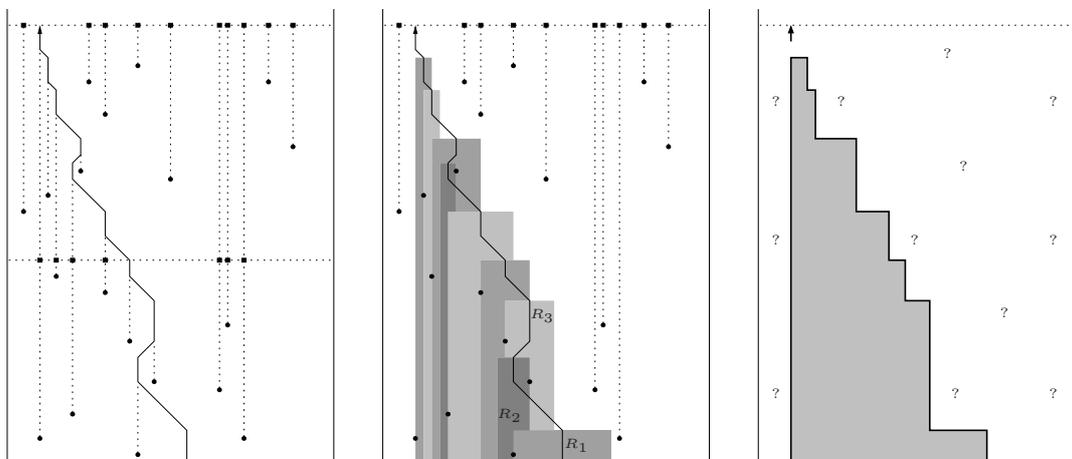

Figure 1.1: On the left, space-time evolution of the greedy server system: Points of ν correspond to costumer arrivals and are represented by round points. The continuous path represents the server's motion. For two different times we depicted the set of waiting customers with squared points. In the center, for the same evolution, each of the contiguous areas in different tones of gray correspond to regions where ν is revealed at the moment when the server needs to know the new target customer. Each of these regions contains the point corresponding to such customer at their lateral boundary, and no other points. On the right, the profile corresponds to the union of all the regions where ν has been revealed up to the time t corresponding to the dotted line. The configuration ν outside this region has been erased, since the server's path up to time t gave no information about this configuration. The bold arrow indicates the positions of both the server and the current customer.

There are two situations where the server may feel that it is on the circle rather than on the line. First, if it arrives at a given point x for the second time after performing a whole turn on the circle, it will encounter an environment that has been affected by its previous visit. This is not an actual problem, because if it happens it will imply that all the customers which were initially present will have then been served, and typically the server will have served more customers than new ones will have arrived.

The second difference is what poses a real issue. The server has a tendency to go into regions that have been *least recently visited*, since in these regions the average interdistance between customers is smaller, and they have bigger chance to attract the server via its greedy mechanism. This is indeed how transience is proved on $\mathbb{R}$. Let us call the *age* of a point in space the measurement in time units of how recently it was visited by the server in the past. On the line, the age is minimal at the server's position, and *increases as we go further away from the server.* The new regions encountered thus become older and older, and the server surrenders to the fact that the cleared regions it is leaving behind cannot compete with the old regions ahead.

However, this is not true on the circle: the age profile cannot increase indefinitely. This gives rise to the possibility of the following tricky scenario. Imagine that on a tiny region around some point x the system is much older than on any close neighborhood. When the server enters this region, it will take a very long time to finish with all the waiting customers. After finishing with all these customers tightly packed in space, there will no longer be a strong difference between the ages ahead and behind the server, who may end up going back to the region that has just been cleared, invalidating the argument.

We deal with this difficulty by making two key observations. First, the age of the points on the circle is monotone in some sense: there is only one local minimum, located at the server's position, and one local maximum x, and the age increases as we move from the server towards x. Second, if the above scenario effectively happens and the server changes direction, the new configuration may become worse in terms of the number of waiting customers, but will be better in the sense that this sharp peak in the age profile has been flattened. In order to say that the new configuration is "better" in this situation, we need to quantify "badness," taking into account a trade-off between diminishing the overall workload and leveling this singular region with excessively high concentration. This is achieved by considering a Lyapunov functional that combines the total number of customers and the maximum local density.

1.5   Outline of the paper

This paper is divided as follows. In Section 2, we give some definitions and notation used throughout the text, and describe the process evolution. In Section 3, we introduce the stochastic evolution of profiles. In Section 4.1, we define an observable B that will serve as a Lyapunov functional, along with a stopping time $\mathcal{T}$ so that $B_{\mathcal{T}}$ has a downwards drift, and finally prove Theorem 1.1. The proof of downwards drift is given in Section 4.2 by showing that the greedy server behaves most of the time like a polling server via a coupling with a system on the infinite line, which is done in Section 4.3. The latter is studied with a block construction in Section 4.4 and a renewal argument in Section 4.5.

2 Setup and notation

The symbol $\preccurlyeq$ means *stochastic domination* between random elements taking value on the same partially ordered space. Define $a \wedge b = \min\{a, b\}$, $a \vee b = \max\{a, b\}$, and $[a]^+ = a \vee 0$. The *indicator* that $x \in J$ is denoted by $\mathbb{1}_J(x)$, and the indicator that the system is in a given state at time s is denoted by $\mathbb{1}_{\text{state}}(s)$. The complement of a set J is denoted by J^c when the space where we take the complement is clear.

We consider the circle as the quotient space $\mathbb{R}/\mathbb{Z}$, and for $x, y \in \mathbb{R}$ we write $x \cong y$ if $x - y \in \mathbb{Z}$. Moreover, we identify classes of $\mathbb{R}/\mathbb{Z}$ with their representants on $\mathbb{R}$, and refer to the points or their representants without distinction unless mentioned otherwise. We denote *arcs* on the circle by $[x, y] \subseteq \mathbb{R}/\mathbb{Z}$, given by the projection of $[x, y] \subseteq \mathbb{R}$ for any pair of representants $x, y \in \mathbb{R}$ such that $x \leqslant y < x + 1$. Analogously for open or closed arcs. We define the *clockwise distance* $\vec{d}$ by $\vec{d}(x, y) = y - x \in [0, 1)$. In particular, $(x, y] = (y, x]^c$ and $\vec{d}(x, y) = 1 - \vec{d}(y, x)$. The distance on $\mathbb{R}/\mathbb{Z}$ is given by $d(x, y) = \vec{d}(x, y) \wedge \vec{d}(y, x)$. We say that f is *increasing* on $[x, y] \subseteq \mathbb{R}/\mathbb{Z}$ if f is increasing on any lifting $[x, y] \subseteq \mathbb{R}$ with $x \leqslant y < x + 1$; analogously for f nonincreasing, nondecreasing, etc.

Evolution of the greedy server system The state of the system at time t is described by the triplet $(\mathscr{C}_t, \mathcal{S}_t, \mathcal{C}_t)$. Here $\mathscr{C}_t$ denotes the set of customers present at the system, $\mathcal{S}_t$ denotes the position of the server, $\mathcal{C}_t \in \mathscr{C}_t$ denotes the position of the customer being served or targeted by the server, and $\mathcal{C}_t = \mathscr{C}_t = \emptyset$ when the system is empty. The process $(\mathscr{C}_t, \mathcal{S}_t, \mathcal{C}_t)_{t \geqslant 0}$ is a strong Markov process, whose stochastic evolution we describe now.

At all times, $t \mapsto \mathcal{S}_t$ is continuous and

$$\frac{\mathrm{d}\mathcal{S}_t}{\mathrm{d}t} = \mathcal{V}_t := \begin{cases} 0, & \mathcal{S}_t = \mathcal{C}_t \text{ or } \mathcal{C}_t = \emptyset, \\ v, & \mathcal{S}_t \neq \mathcal{C}_t, \ \vec{d}(\mathcal{S}_t, \mathcal{C}_t) < \vec{d}(\mathcal{C}_t, \mathcal{S}_t), \\ -v, & \mathcal{S}_t \neq \mathcal{C}_t, \ \vec{d}(\mathcal{S}_t, \mathcal{C}_t) \geqslant \vec{d}(\mathcal{C}_t, \mathcal{S}_t), \end{cases} \tag{2.1}$$

in the sense of right derivative.

There are three different regimes: *moving* when $\mathcal{S}_t \neq \mathcal{C}_t \in \mathbb{R}/\mathbb{Z}$, *serving* when $\mathcal{S}_t = \mathcal{C}_t$, and *idle* when $\mathcal{C}_t = \emptyset$. While the system is *idle*, $\mathcal{S}$ and $\mathcal{C}$ remain unchanged until an arrival happens. While the server is *moving*, the evolution of $\mathcal{S}$ obeys (2.1), and $\mathcal{C}$ remains constant. This regime lasts until *service starts*, i.e., until the time s when $\mathcal{S}_s = \mathcal{C}_s$. During *service*, the evolution of $\mathcal{S}$ is again given by (2.1), $\mathcal{C}$ also remains constant, and *service finishes* according to an exponential clock of rate 1.

The moments when *service finishes* will be called *departure times*. At departure times t, the new regime will be either *moving* or *idle*. First, the current customer is removed from the system: $\mathscr{C}_t = \mathscr{C}_{t-} \setminus \{\mathcal{C}_{t-}\}$. Then $\mathcal{C}_t$ is chosen as the nearest waiting customer, if any:

$$\mathcal{C}_t = \arg\min\{d(\mathcal{S}_t, x) : x \in \mathscr{C}_t\}, \quad \text{or} \quad \mathcal{C}_t = \emptyset \text{ if } \mathscr{C}_t = \emptyset. \tag{2.2}$$

During the whole evolution, *arrivals* happen at rate λ. An *arrival* consists of adding to $\mathscr{C}$ a new point z chosen uniformly at random on $\mathbb{R}/\mathbb{Z}$, i.e., $\mathscr{C}_t = \mathscr{C}_{t-} \cup \{z\}$. If $\mathcal{C}_{t-} \neq \emptyset$, i.e., the server was moving or serving, this is the only change. If $\mathcal{C}_{t-} = \emptyset$, i.e., the system was idle, then also $\mathcal{C}$ is updated by $\mathcal{C}_t = z$ and a.s. the new regime is moving.

3   The process viewed from the server

The process $(\mathscr{C}_t, \mathcal{S}_t, \mathcal{C}_t)_{t \geqslant 0}$ may be constructed from two point processes: a Poisson Point Process $\nu \subseteq (\mathbb{R}/\mathbb{Z}) \times \mathbb{R}_+$ with intensity $\lambda \cdot \mathrm{d}x\mathrm{d}t$, each point corresponding to the arrival of a new customer at position x at time t, and the Poisson Point Process $\mathscr{T} \subseteq \mathbb{R}_+$ corresponding to possible departure times (each mark $t \in \mathscr{T}$ effectively corresponds to a departure time if a customer was being served up to time $t-$ and is ignored if the server was idle or moving). For u and w denoting functions on $\mathbb{R}/\mathbb{Z}$ or constants, let

$$\Gamma_u^w = \{(x,s) : x \in \mathbb{R}/\mathbb{Z}, u(x) < s \leqslant w(x)\} \subseteq (\mathbb{R}/\mathbb{Z}) \times \mathbb{R}_+.$$

The σ-algebra $\mathcal{F}_t = \sigma(\nu_t, \mathscr{T}_t)$, where $\nu_t = \nu \cap \Gamma_0^t$ and $\mathscr{T}_t = \mathscr{T} \cap [0,t]$, contains all the information about arrivals and departures up to time t, and consequently about $(\mathscr{C}_s, \mathcal{S}_s, \mathcal{C}_s)_{s \in [0,t]}$.

The process $(\mathcal{S}_t, \mathcal{C}_t)_{t \geqslant 0}$ is not Markovian. Indeed, the conditional distribution of $(\mathcal{S}_s, \mathcal{C}_s)_{s \geqslant t}$ given $\mathcal{F}_t$ depends on both $(\mathcal{S}_t, \mathcal{C}_t)$ and $\mathscr{C}_t$. Yet the only interaction between $(\mathcal{S}, \mathcal{C})$ and $\mathscr{C}$ is given by (2.2). Namely, at each *departure time* t, $\mathscr{C}_t$ is *queried* about the *nearest* waiting customer $\mathcal{C}_t$, if any. The position of $\mathcal{C}_t$ reveals that $\mathscr{C}_t \cap [\mathcal{S}_t - z, \mathcal{S}_t + z] = \{\mathcal{C}_t\}$, where $z = d(\mathcal{S}_t, \mathcal{C}_t) < \frac{1}{2}$, and on the other hand it gives no information about the complementary set $\mathscr{C}_t \cap [\mathcal{S}_t - z, \mathcal{S}_t + z]^c$ of waiting customers.

In the sequel we discuss the conditional distribution of $\mathscr{C}_t$ given $(\mathcal{S}_s, \mathcal{C}_s)_{s \in [0,t]}$, the role played by this conditional law, and the evolution of this law itself.

Markovianity without $\mathscr{C}_t$ By the Markov property of $(\mathscr{C}_t, \mathcal{S}_t, \mathcal{C}_t)_{t \geqslant 0}$ with respect to $\{\mathcal{F}_t\}_{t \geqslant 0}$ we have that the conditional law of $(\mathcal{S}_s, \mathcal{C}_s)_{s \geqslant t}$ satisfies

$$\mathscr{L}\left[(\mathcal{S}_s, \mathcal{C}_s)_{s \geqslant t} \big| \mathcal{F}_t\right] = \mathscr{L}\left[(\mathcal{S}_s, \mathcal{C}_s)_{s \geqslant t} \big| (\mathscr{C}_t, \mathcal{S}_t, \mathcal{C}_t)\right].$$

Let

$$\mathcal{G}_t = \sigma\left((\mathcal{S}_s, \mathcal{C}_s)_{s \in [0,t]}\right) \subseteq \mathcal{F}_t.$$

In the sequel we consider the triple $\left(\mathscr{L}(\mathscr{C}_t | \mathcal{G}_t), \mathcal{S}_t, \mathcal{C}_t\right)$ and study its evolution.

By the observations in the previous paragraph, the evolution $(\mathcal{S}_s, \mathcal{C}_s)_{s \in [0,t]}$ gives information about $\nu \cap (\mathbb{R}/\mathbb{Z}) \times (0,t]$ in a very precise way. At each departure time s, the new $\mathcal{C}_s$ is chosen as the point of $\mathscr{C}_s$ that is closest to $\mathcal{S}_s$. At these times, $\mathcal{C}_s$ is given by $\mathcal{S}_s \pm z$, where z is the smallest distance for which there is a point $(\mathcal{S}_s \pm z, s')$ with $s' \in (0,s]$ in ν, not considering the points that correspond to customers who have already left the system. This reveals a rectangle $[\mathcal{S}_s - z, \mathcal{S}_s + z] \times (0,s]$ where ν has no more points that will participate in the construction of $(\mathscr{C}_r)_{r > s}$, and the law of $\mathscr{C}_s$ outside $[\mathcal{S}_s - z, \mathcal{S}_s + z]$ is not affected. For times r between s and the next departure time, $\mathscr{C}_r$ is given by the union of $\mathscr{C}_s$ and the Poisson arrivals corresponding to $\nu \cap (\mathbb{R}/\mathbb{Z}) \times (s,r]$. For the times s when the system is in the idle state, the revealed rectangle is the whole $\mathbb{R}/\mathbb{Z} \times (0,s]$.

Iterating this argument, by time t the configuration ν has been revealed on the region given by the union of such rectangles. Since all these rectangles have their base on $t = 0$, their union is of the form $\Gamma_0^{w_t}$, where $w_t(x)$ denotes the maximal height among all the rectangles whose base contains the point x. In other words, the value of $w_t(x)$ is the most recent among: the departure times $s \in (0,t]$ such that $x \in [\mathcal{S}_s - z, \mathcal{S}_s + z]$; and the times $s \in (0,t]$ when the system was idle. The set of waiting customers $\mathscr{C}_t \setminus \{\mathcal{C}_t\}$ is thus determined by the configuration ν on the complementary

region $\Gamma^t_{w_t}$. Therefore, the conditional distribution of $\mathscr{C}_t \setminus \{\mathcal{C}_t\}$ given $\mathcal{G}_t$ is that of an inhomogeneous Poisson process on $\mathbb{R}/\mathbb{Z}$, with local intensity at each point x given by

$$\lambda \left[t - w_t(x) \right] \mathrm{d}x.$$

In summary,

$$\mathscr{L} \left[(\mathcal{S}_s, \mathcal{C}_s)_{s \geqslant t} \middle| \mathcal{G}_t \right] = \mathscr{L} \left[(\mathcal{S}_s, \mathcal{C}_s)_{s \geqslant t} \middle| (w_t, \mathcal{S}_t, \mathcal{C}_t) \right].$$

Since the evolving region $(\Gamma^{w_t}_0)_{t \geqslant 0}$ is increased at departure times t by adding a rectangle to $\Gamma^{w_{t-}}_0$, this rectangle being in turn determined by $\mathcal{S}_t$ and $\mathcal{C}_t$, we have

$$\mathscr{L} \left[(w_s, \mathcal{S}_s, \mathcal{C}_s)_{s \geqslant t} \middle| (w_s, \mathcal{S}_s, \mathcal{C}_s)_{s \in [0,t]} \right] = \mathscr{L} \left[(w_s, \mathcal{S}_s, \mathcal{C}_s)_{s \geqslant t} \middle| (w_t, \mathcal{S}_t, \mathcal{C}_t) \right];$$

i.e., $(w_t, \mathcal{S}_t, \mathcal{C}_t)$ is a Markov process with respect to its natural filtration. In our framework, we shall consider

$$u_t = w_t - t \leqslant 0$$

instead of w, so that $(u_t, \mathcal{S}_t, \mathcal{C}_t)_{t \geqslant 0}$ is a time-homogeneous strong Markov process.

Evolution of $(u_t, \mathcal{S}_t, \mathcal{C}_t)$ The law of the evolution $(u_t, \mathcal{S}_t, \mathcal{C}_t)_{t \geqslant 0}$ is given as follows. As before, the system may be in one of three regimes, determined by $(\mathcal{S}_t, \mathcal{C}_t)$.

While *moving* or *serving*, the evolution of $\mathcal{S}$ and $\mathcal{C}$ are given by the same rules as in the previous section: $\mathcal{C}$ remains constant, $\mathcal{S}$ satisfies (2.1), and in the serving regime *service finishes* at rate 1. We no longer have $\mathscr{C}$ to account for the whole set of waiting customers. Instead of randomly adding new customers at rate λ, this information is now encoded in the function $u(x)$, with the rule

$$\frac{\mathrm{d}u_t(x)}{\mathrm{d}t} = -1 \quad \forall\, x \in \mathbb{R}/\mathbb{Z}, \tag{3.1}$$

which rather accounts for the time period when new customers have been arriving to the system at location $\mathrm{d}x$.

At *departure times*, instead of choosing the nearest point in $\mathscr{C}_t$ as in (2.2), we take what would be nearest point in a realization of a Poisson Point Process on $\mathbb{R}/\mathbb{Z}$ with intensity $-u_t(x)\mathrm{d}x$. More precisely, at the departure times the system goes through an instantaneous random transition, which may lead to either a *moving* or an *idle* state, as we describe below. Let $0 < E < \infty$ and $0 < U < 1$ denote exponential and uniform random variables, independent of each other and of the construction up to time $t-$. The meaning of E is that the measure of the interval that needs to be explored before finding a point is exponentially distributed, and U is important in deciding the position of such point in the boundary of this explored interval. The total intensity of waiting customers potentially present in the system is given by

$$A(u) = \int_{\mathbb{R}/\mathbb{Z}} -\lambda u(x)\mathrm{d}x.$$

If $E \geqslant A(u_{t-})$, take

$$\mathcal{C}_t = \emptyset,$$

and the system becomes *idle*. Otherwise, let $0 < z < \frac{1}{2}$ be the unique number such that $\int_{\mathcal{S}_t-z}^{\mathcal{S}_t+z}(-\lambda u_{t-})\mathrm{d}x = E$, let $a = -u_{t-}(\mathcal{S}_t - z)$, $b = -u_{t-}(\mathcal{S}_t + z)$, choose

$$
\mathcal{C}_t = \begin{cases} \mathcal{S}_t - z, & U \in (0, \frac{a}{a+b}), \\ \mathcal{S}_t + z, & U \in [\frac{a}{a+b}, 1), \end{cases}
\tag{3.2}
$$

and the new regime is *moving*. Finally, take

$$
u_t(x) = \begin{cases} u_{t-}(x) \cdot \mathbb{1}_{[\mathcal{S}_t-z,\mathcal{S}_t+z]^c}(x), & E < A(u_{t-}), \\ 0, & E \geqslant A(u_{t-}). \end{cases}
\tag{3.3}
$$

The *idle* regime $\mathcal{C} = \emptyset$ can only be achieved together with $u \equiv 0$ on $\mathbb{R}/\mathbb{Z}$. While the system is idle, the state $(u, \mathcal{S}, \mathcal{C})$ remains unchanged until the first customer arrival, which happens according to an exponential clock of rate λ. The *arrival* consists of letting $\mathcal{C}_t = z$, where z is chosen uniformly on $\mathbb{R}/\mathbb{Z}$. Immediately after an arrival, a.s. the new regime is *moving*.

Framework   A piecewise continuous, upper semi-continuous function $u(x) \leqslant 0$ on $\mathbb{R}/\mathbb{Z}$ is called a *potential*. Note that the evolution described above can start from any given potential u and points $\mathcal{S}, \mathcal{C}$ such that $\mathcal{C} \neq \emptyset$ if $u \not\equiv 0$. For shortness, the triplet $(u, \mathcal{S}, \mathcal{C})$ will be denoted by $\mathcal{U}$. Let $\mathbb{P}^{\mathcal{U}}$ denote the law of $(\mathcal{U}_t)_{t\geqslant 0}$ starting from $\mathcal{U}$ at $t = 0$.

We say that u is a *proper potential* if there exist $x_{\min}, x_{\max} \in \mathbb{R}/\mathbb{Z}$ such that u is nondecreasing on the arc $[x_{\min}, x_{\max}]$ and nonincreasing on the arc $[x_{\max}, x_{\min}]$, or if u is monotone on any arc not containing $x_{\max}$; see Figure 3.1. Given a proper potential u, we say that $\mathcal{U} = (u, \mathcal{S}, \mathcal{C})$ is a *proper state* if either $u(\mathcal{C}) = u(x_{\max})$, or $\mathcal{C} = \emptyset$ and $u \equiv 0$.

Proposition 3.1. *Starting from a proper state $\mathcal{U}$, $\mathbb{P}^{\mathcal{U}}$-a.s. the process $(\mathcal{U}_t)_{t\geqslant 0}$ remains in proper states for every $t > 0$.*

Proof. When the system is idle, $\mathcal{U}$ is a proper state by definition. At departure times, according to (3.2)-(3.3) the position of $\mathcal{C}$ changes and u is updated by increasing its value to 0 on an arc
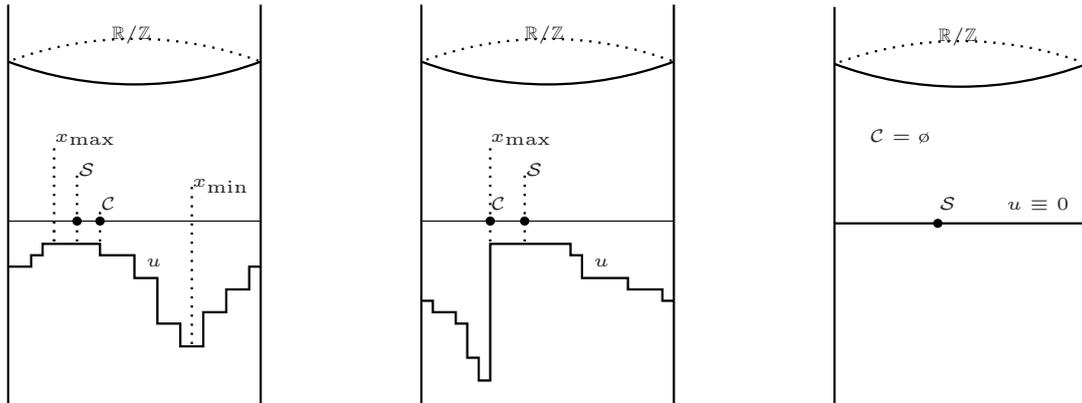

Figure 3.1: Three examples of *proper states*.

containing both the new and the old $\mathcal{C}$. This transformation preserves the condition of $\mathcal{U}$ being a proper state. When the system state is either moving or serving, u evolves according to (3.1), that is, the subtraction of a constant, which also preserves the proper state condition. $\qquad\square$

Remark. Although $\mathscr{C}_t$ is not determined by $(\mathcal{S}_t, \mathcal{C}_t)$, we have that $\mathscr{C}_t = \emptyset$ if and only if $\mathcal{C}_t = \emptyset$, and it is thus sufficient to consider the process $(\mathcal{U}_t)_{t \geqslant 0}$ in the study of positive recurrence, defined on page 2. This is the approach used henceforth.

4 Proof of stability

The goal of this this section is to prove Theorem 1.1. In Section 4.1, we define a Lyapunov functional B and a stopping time $\mathcal{T}$. We then state Proposition 4.1 about the downwards drift of B at time $\mathcal{T}$ and use it to prove Theorem 1.1. In Section 4.2, we prove Proposition 4.1 making use of Proposition 4.3, which states that the total time that the server spends traveling before time $\mathcal{T}$ is stochastically bounded. In Section 4.3, we prove Proposition 4.3 by coupling the system on the circle with another one on the real line. In Section 4.4, we show that the latter has positive probability of being transient (and in fact ballistic) using a block construction. In Section 4.5, we conclude with a renewal argument relying on the uniform estimates obtained in the block construction.

We spell some formulae for later reference.

$$\eta = 1 - \lambda, \qquad \Psi = 2\eta^{-1}, \qquad \varepsilon = \frac{\eta\lambda}{8}, \qquad \delta = \frac{\varepsilon}{2\Psi}. \tag{4.1}$$

The reason for these definitions will become clear as they are used in the proof.

4.1 Lyapunov functional and stopping times

Given a proper potential u, let

$$N = N(u) = \sup_{x \in \mathbb{R}/\mathbb{Z}} -u(x),$$

$$B = B(u) = A(u) + 4\varepsilon N(u).$$

Notice that the evolution of u is given by (3.1) when the state is moving or serving, at departure times it jumps upwards according to (3.3), and it remains constant when the state is idle. It thus follows that

$$u_{t+s} \geqslant u_t - s \qquad \forall\, s, t \geqslant 0. \tag{4.2}$$

Since $\lambda + 4\varepsilon < 1$, it follows from (4.2) that

$$B(u_{t+s}) \leqslant B(u_t) + s \qquad \forall\, s, t \geqslant 0. \tag{4.3}$$

Let B_* denote a finite number that will be fixed later. We claim that, for any proper state $\mathcal{U}$ with $B(u) \leqslant B_*$,

$$\mathbb{P}^{\mathcal{U}}\left[\tau_{\emptyset} < \frac{1}{2v} + 1\right] \geqslant \left(1 - e^{-1}\right) \exp\left(-B_* - 1 - \frac{1}{2v}\right) > 0. \tag{4.4}$$

To see why the claim is true, consider the event that the server travels towards the nearest customer $\mathcal{C}$, then finishes service within $T < 1$ time unit, and at this departure time the next state given by (3.2) and (3.3) is idle. When these events hold, since the distance $d = d(\mathcal{S}, \mathcal{C})$ is at most $\frac{1}{2}$, this departure time happens at $t' = \frac{d}{v} + T < \frac{1}{2v} + 1$, implying that $\tau_\emptyset < \frac{1}{2v} + 1$. The first term on the right-hand side corresponds to the probability that $T < 1$. The second term is a lower bound for the the conditional probability that $\mathcal{C}_{t'} = \emptyset$ given t', since the latter is given by $e^{-A(u_{t'})} \geqslant e^{-B(u_{t'})}$ which by (4.3) is bounded by $e^{-B(u)-t'}$, proving the claim.

By (4.3) and (4.4), the proof of Theorem 1.1 reduces to showing that

$$\sup\left\{\mathbb{E}^{\mathcal{U}}\left[\tau_{\{B \leqslant B_*\}}\right] : \mathcal{U} \text{ proper state}, B(u) < \bar{B}\right\} < \infty \quad \forall \bar{B} < \infty, \tag{4.5}$$

where $\tau_{\{B \leqslant B_*\}} = \inf\{t : B(u_t) \leqslant B_*\}$.

Let $\mathcal{U}$ be a proper state such that $B(u) > B_*$. In the proof of (4.5), we study the behavior of $B(u_t)$ at a particular stopping time $\mathcal{T}$ that is defined below.

Define the sets

$$\begin{aligned} U &= \left\{x \in \mathbb{R}/\mathbb{Z} : u(x) < -\tfrac{N}{2}\right\}, \\ G_t &= \left\{x \in \mathbb{R}/\mathbb{Z} : u_t(x) > -t\right\}. \end{aligned} \tag{4.6}$$

Since u is a proper potential, U must be either $\mathbb{R}/\mathbb{Z}$ or an open arc. Notice that $G_0 = \emptyset$ and by (4.2) we have that G_t is nondecreasing in t. By (3.1) and (3.3), it may only increase at departure times t, by adding a closed arc containing $\mathcal{S}_t$ and $\mathcal{C}_t$. Thus G_t is always either $\emptyset$, or all $\mathbb{R}/\mathbb{Z}$, or a closed arc containing $\mathcal{S}_t$.

We define the following stopping times:

$$\begin{aligned} \mathcal{T}^+ &= \mathcal{T}^+(u) = \Psi B(u), \\ \mathcal{T}_\circ &= \mathcal{T}_\circ(u) = \inf\{t : G_t \supseteq \mathbb{R}/\mathbb{Z}\}, \\ \mathcal{T}_\curlyvee &= \mathcal{T}_\curlyvee(u) = \inf\{t : G_t \supseteq U\}, \\ \mathcal{T} &= \mathcal{T}(u) = \mathcal{T}_\circ(u) \wedge \mathcal{T}_\curlyvee(u) \wedge \mathcal{T}^+(u). \end{aligned} \tag{4.7}$$

It follows from (4.3) that

$$B(u_\mathcal{T}) \leqslant B(u) + \mathcal{T} \leqslant B(u) + \mathcal{T}^+ = (\Psi + 1)B(u). \tag{4.8}$$

A few comments are in order. Normally, $\mathcal{T}$ is attained because the condition for $\mathcal{T}_\curlyvee$ is attained. The deterministic time $\mathcal{T}^+$ is a *safety caution*: it bounds the possible damage that is caused when this condition is not attained in due time. The presence of $\mathcal{T}_\circ$ in the definition of $\mathcal{T}$ is innocuous from a formal point of view, since $\mathcal{T}_\circ \geqslant \mathcal{T}_\curlyvee$. We write it to indicate that $\mathcal{T}_\curlyvee$ may be attained in two conceptually different situations: either because U is "crossed" by $(G_s)_{s \geqslant 0}$, or because U is partly taken by $(G_s)_{s \geqslant 0}$ from one direction and then from the other, in which case the whole circle $\mathbb{R}/\mathbb{Z}$ is taken. See Figure 4.1.

Proposition 4.1 (Downwards drift). *For any proper state $\mathcal{U}$,*

$$\mathbb{P}^{\mathcal{U}}\left(B(u_\mathcal{T}) \leqslant (1-\varepsilon)B(u)\right) \geqslant 1 - \rho, \tag{4.9}$$

where $\rho = \rho(B(u))$ satisfies $\rho(B) \to 0$ as $B \to \infty$, and ε is defined in (4.1).

Writing $D(\cdot) = \log \frac{B(\cdot)}{B_*}$, (4.8) and (4.9) imply that

$$\mathbb{P}^{\mathcal{U}}\left(D(u_{\mathcal{T}}) \leqslant D(u) - \varepsilon\right) \geqslant 1 - \rho, \; D(u_{\mathcal{T}}) \leqslant D(u) + \Psi, \; \mathcal{T} \leqslant \mathcal{T}^+ = \Psi B_* e^D. \qquad (4.10)$$

We are going to use the following fact, whose proof is omitted.

Lemma 4.2. *Let $(Y_n)_{n\in\mathbb{N}}$ be i.i.d. Bernoulli random variables with*

$$\mathbb{P}(Y_1 = \Psi) = 1 - \mathbb{P}(Y_1 = -\varepsilon) = \rho.$$

Write $\mathbb{P}^s$ for the law of $(S_n)_{n\in\mathbb{N}}$ given by $S_n = s + Y_1 + \cdots + Y_n$, and define $\sigma = \inf\{n : S_n \leqslant 0\}$. Then there exists $\rho_* > 0$ such that $\mathbb{E}^s[\sigma e^{\Psi\sigma}] < \infty$ for any $\rho \leqslant \rho_*$ and $s < \infty$.

Proof of Theorem 1.1. We need to show (4.5). First we use Proposition 4.1 to fix the value of B_* with the property that $\rho(B) \leqslant \rho_*$ for any $B > B_*$.

Let $\mathcal{U}$ be a proper state such that $B_* < B(u) < \bar{B}$. We start with $D_0 = D(u_0) > 0$. Consider the stopping time $\mathcal{T}_1 = \mathcal{T}(u_0)$ defined by (4.7) and define $D_1 = D(u_{\mathcal{T}_1})$. For the shifted process $(\mathcal{U}_{\mathcal{T}_1+t})_{t\geqslant 0}$, consider the stopping time $\mathcal{T}_2 = \mathcal{T}(u_{\mathcal{T}_1})$ and write $D_2 = D(u_{\mathcal{T}_1+\mathcal{T}_2})$. Analogously, once $\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_n$ have been constructed, consider, for the shifted process $(\mathcal{U}_{\mathcal{T}_1+\mathcal{T}_2+\cdots+\mathcal{T}_n+t})_{t\geqslant 0}$, the stopping time $\mathcal{T}_{n+1} = \mathcal{T}(u_{\mathcal{T}_1+\mathcal{T}_2+\cdots+\mathcal{T}_n})$, and write $D_{n+1} = D(u_{\mathcal{T}_1+\mathcal{T}_2+\cdots+\mathcal{T}_n+\mathcal{T}_{n+1}})$. Let $\gamma = \inf\{n : D_n \leqslant 0\}$.

Taking $s = D_0$, it follows from (4.10) that

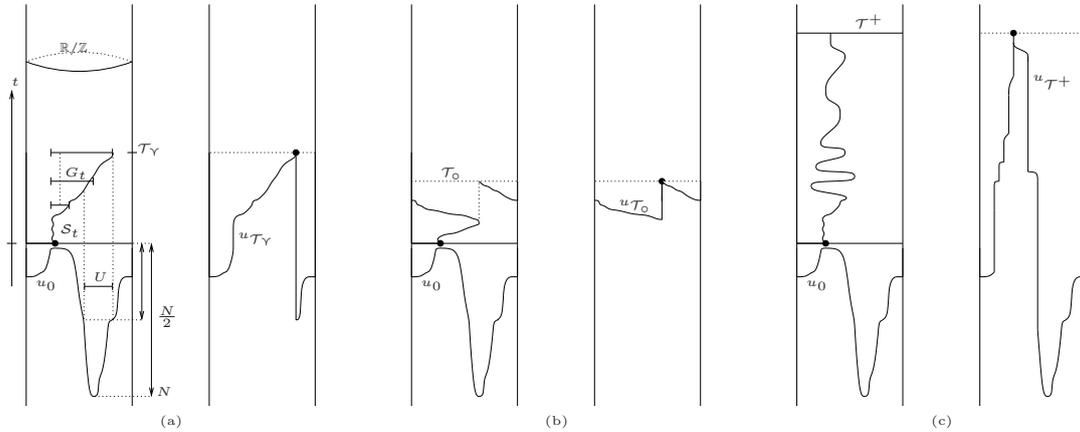$$(D_{n\wedge\gamma})_{n=0,1,2,\ldots} \preccurlyeq ([S_{n\wedge\sigma}]^+)_{n=0,1,2,\ldots},$$

Figure 4.1: Evolution of G_t starting from a given potential u_0, with the new potential $u_{\mathcal{T}}$ at the stopping time $\mathcal{T}$. From left to right there are three pairs of graphs, each one embedded in the space-time $(\mathbb{R}/\mathbb{Z}) \times \mathbb{R}$. In (a) we depict a typical example when $\mathcal{T} = \mathcal{T}_\Upsilon$. In (b) we show an instance where $\mathcal{T} = \mathcal{T}_\circ$. Finally, in (c) there is an example where the server remains confined for a long time, preventing the condition for $\mathcal{T}_\Upsilon$ to be attained up to time $\mathcal{T} = \mathcal{T}^+$. Each pair represents the system evolution and the resulting potential: the graph on the left shows the parametrized curves $(\mathcal{S}_t, t)_{t\in[0,\mathcal{T}]}$ and $(x, u_0(x))_{x\in\mathbb{R}/\mathbb{Z}}$, and on the right there is $(x, u_{\mathcal{T}}(x) + \mathcal{T})_{x\in\mathbb{R}/\mathbb{Z}}$ together with the point $(\mathcal{S}_{\mathcal{T}}, \mathcal{T})$.

whence $\gamma \preccurlyeq \sigma$. Therefore we get

$$\frac{\tau_{\{B \leqslant B_*\}}}{\Psi B_*} \leqslant \sum_{n=1}^{\gamma} \frac{\mathcal{T}_n}{\Psi B_*} \leqslant \sum_{n=1}^{\gamma} e^{D_{n-1}} \leqslant \gamma \exp\left[\max_{0 \leqslant n < \gamma} D_n\right] \leqslant \gamma e^{D_0 + \Psi \gamma},$$

whence by Lemma 4.2

$$\frac{1}{\Psi B_*} \mathbb{E}^{\mathcal{U}}\left[\tau_{\{B \leqslant B_*\}}\right] \leqslant \mathbb{E}^{\mathcal{U}}\left[\gamma e^{D_0 + \Psi \gamma}\right] \leqslant \mathbb{E}^s\left[\sigma e^{s + \Psi \sigma}\right] \leqslant \mathbb{E}^{\bar{s}}\left[\sigma e^{\bar{s} + \Psi \sigma}\right] < \infty,$$

where $\bar{s} = \log \frac{\bar{B}}{B_*}$. $\square$

4.2 Downwards drift

Write $A = A(u_0)$, $N = N(u_0)$, $B = B(u_0)$, $A' = A(u_{\mathcal{T}})$, $N' = N(u_{\mathcal{T}})$, $B' = B(u_{\mathcal{T}})$. We decompose time in three parts:

$$\mathcal{T} = \mathcal{M} + \mathcal{S} + \mathcal{I}, \tag{4.11}$$

where

$$\mathcal{M} = \int_0^{\mathcal{T}} \mathbb{1}_{\text{moving}}(s)\mathrm{d}s, \quad \mathcal{S} = \int_0^{\mathcal{T}} \mathbb{1}_{\text{serving}}(s)\mathrm{d}s, \quad \mathcal{I} = \int_0^{\mathcal{T}} \mathbb{1}_{\text{idle}}(s)\mathrm{d}s.$$

By definition of $\mathcal{T}_\circ$, the system cannot be idle for any $t < \mathcal{T}$, thus $\mathcal{I} = 0$. For each $t > 0$, let $\mathcal{N}_t$ denote the *number of departures times* in $(0, t]$. Fix $\mathcal{N} = \mathcal{N}_{\mathcal{T}}$, the number of customers served up to time $\mathcal{T}$. The total time spent with services during $(0, \mathcal{T}]$ is given by

$$\mathcal{S} = \sum_{n=1}^{\mathcal{N}} T_n + \beta T_{\mathcal{N}+1}$$

for some $0 \leqslant \beta < 1$, where $(T_n)_{n \in \mathbb{N}}$ are i.i.d. exponential random variables.

Writing $A_t = A(u_t)$, it follows from (3.1) that $\frac{\mathrm{d}A_t}{\mathrm{d}t} = \lambda$ for Lebesgue-a.e. $t < \mathcal{T}$. Moreover, $(A_t)_t$ jumps downwards at departure times, and (3.3) reads as

$$A_t = \left[A_{t-} - E\right]^+.$$

Since $A_t > 0$ for all $t < \mathcal{T}$, A' satisfies

$$A' = A + \lambda \mathcal{T} - \left(\sum_{n=1}^{\mathcal{N}-1} E_n + \beta' E_{\mathcal{N}}\right),$$

where $0 < \beta' \leqslant 1$ and $(E_n)_{n \in \mathbb{N}}$ are i.i.d. exponential random variables.

We now present the last ingredient, which is proved in the next subsection.

Proposition 4.3 (Polling behavior)**.** *The distribution of $\mathcal{M}$ under $\mathbb{P}^{\mathcal{U}}$ is tight:*

$$\mathbb{P}^{\mathcal{U}}\left\{\mathcal{M} > t\right\} \underset{t \to \infty}{\longrightarrow} 0$$

uniformly over all proper states $\mathcal{U}$.

Proof of Proposition 4.1. It follows from Donsker's invariance principle and from Proposition 4.3 that

$$\sum_{n=1}^{\mathcal{N}+1} T_n < \mathcal{N} + \delta B, \tag{4.12}$$

$$\sum_{n=1}^{\mathcal{N}-1} E_n > \mathcal{N} - \delta B, \tag{4.13}$$

$$\mathscr{M} < \delta \Psi B, \tag{4.14}$$

hold with high probability as $B \to \infty$, uniformly in $\mathcal{U}$.

Assume that (4.12), (4.13), and (4.14) happen. Putting these altogether yields

$$
\begin{aligned}
0 \leqslant A' = {}& A + \lambda \mathcal{T} - \sum_{n=1}^{\mathcal{N}-1} E_n - \beta' E_{\mathcal{N}} \\
\leqslant {}& A + \lambda \mathcal{T} - \mathcal{N} + \delta B && \text{by (4.13)} \\
\leqslant {}& A + \lambda \mathcal{T} - \mathscr{S} + 2\delta B && \text{by (4.12)} \\
= {}& A + \lambda \mathcal{T} - \mathcal{T} + \mathscr{M} + 2\delta B && \text{by (4.11)} \\
\leqslant {}& A - \eta \mathcal{T} + \delta \Psi B + 2\delta B && \text{by (4.14)} \\
\leqslant {}& A - \eta \mathcal{T} + 2\delta \Psi B && \text{since } \Psi > 2 && (4.15) \\
< {}& 2B - \eta \mathcal{T} = \eta(\mathcal{T}^+ - \mathcal{T}). && \text{since } A < B, \ 2\delta \Psi < 1
\end{aligned}
$$

By the last inequality, we have $\mathcal{T} < \mathcal{T}^+$. It then follows from the definition of $\mathcal{T}$ that $U \subseteq G_{\mathcal{T}}$, whence $u_{\mathcal{T}}(x) > -\mathcal{T}$ for $x \in U$. But by (4.2) and the definition of U, we have $u_{\mathcal{T}}(x) \geqslant u_0(x) - \mathcal{T} \geqslant -N/2 - \mathcal{T}$ for $x \in U^c$. Therefore,

$$N' \leqslant N/2 + \mathcal{T}.$$

Combining this and (4.15):

$$
\begin{aligned}
B' - B \leqslant {}& (-\eta \mathcal{T} + 2\delta \Psi B) + 4(\mathcal{T} - N/2)\varepsilon \\
\leqslant {}& -(2\varepsilon - 2\delta \Psi)B - (\eta - 4\varepsilon)\mathcal{T} && \text{since } N > B \\
\leqslant {}& -(2\varepsilon - 2\delta \Psi)B && \text{since } 4\varepsilon < \eta \\
= {}& -\varepsilon B. && \qquad\square
\end{aligned}
$$

4.3 Polling behavior

In this section we prove Proposition 4.3 via a coupling with the greedy server on the real line. The latter eventually moves towards one of the two directions and spends little time going backwards, which was shown in [FRS15]. We consider a periodic extension of the initial potential u on the circle, and approximate it by another potential with less oscillations, for which we can generalize this result.

The rules described in Section 3 may also be used to construct evolutions on the real line, which we will couple with the greedy server on the circle.

Let us start with an informal description of how these systems are coupled.

First, define an evolution on $\mathbb{R}$ by a lifting from $\mathbb{R}/\mathbb{Z}$: extend the potential periodically and replace both the server and the currently served customer by infinitely many replicas at unit interdistances.

If all the replicas evolve using the same randomness, this system remains periodic for all times. Moreover, one can recover the original system by projecting back from the line back onto the circle, so it is basically the same system.

Now remove all the replicas and consider the system with a single server. For short times, this server evolves just like in the system with all the replicas. In fact, this remains true until the first time when the server needs to query for the presence of customers in a region that has already been queried by another replica. So the systems will uncouple at time $\mathcal{T}_{[1]}$ defined below, or $\mathcal{T}_\circ$ for the system on $\mathbb{R}/\mathbb{Z}$.

Finally, this system with a single server on $\mathbb{R}$ can be coupled with another similar system also on $\mathbb{R}$, which at $t = 0$ starts with the same positions for both the server and served customer, but a slightly different potential. If the same randomness is used for both systems, the servers will evolve together until the first time when they query for the presence of customers in a region where the potential was initially different. This time of uncoupling is given by $\mathcal{T}_U$ defined below. It typically occurs before $\mathcal{T}_{[1]}$, in which case it will correspond to $\mathcal{T}_\gamma$ on the circle.

In summary, we can couple the system on the circle with one on the line, and the latter with another one having different initial potential, this double coupling lasting until $\mathcal{T}_U \wedge T_{[1]}$. Figure 4.2 shows an example where $\mathcal{T}_U$ is attained first, and Figure 4.1(b) shows an example on the circle where $\mathcal{T}_{[1]}$ would be attained first.

We now make the above description precise.

Coupling with the greedy server on $\mathbb{R}$ A *potential* is a piecewise continuous, upper semi-continuous function $\bar{u}(x) \leqslant 0$ on $\mathbb{R}$ with $\int_{\mathbb{R}} -\bar{u}\mathrm{d}x = \infty$. The evolution of $(\bar{\mathcal{U}}_t)_{t \geqslant 0}$ is defined in the same way as on the circle, i.e., satisfying (2.1),(3.1),(3.2),(3.3).

Let $\mathcal{U}$ be a proper state on the circle and $\bar{u}$ the *periodic extension* of u on $\mathbb{R}$. Without loss of generality, in the sequel we assume that $\mathcal{S} = 0$. Take $\bar{\mathcal{S}} = 0$ and let $\bar{\mathcal{C}}$ be the only representant of $\mathcal{C}$ in $[-\frac{1}{2}, \frac{1}{2})$.

We define
$$H_t = \left\{ x \in \mathbb{R} : \bar{u}_t(x) > -t \right\}.$$

By the same arguments as for the greedy server on $\mathbb{R}/\mathbb{Z}$, H_t is nondecreasing in t; it is empty until the first departure time, after which it consists of a closed interval containing both $\bar{\mathcal{S}}_t$ and $\bar{\mathcal{C}}_t$.

Define the stopping time
$$\mathcal{T}_{[1]} = \inf \left\{ t : |H_t| \geqslant 1 \right\}.$$

For each $t < \mathcal{T}_{[1]}$, we define the map π_t that takes each point $x \in [L(t), L(t) + 1) \subseteq \mathbb{R}$ to its projection $x \in \mathbb{R}/\mathbb{Z}$, where $L(t)$ is chosen as follows. If $H_t = \emptyset$, we take $L(t) = -\frac{1}{2}$; otherwise if $H_t \neq \emptyset$, we take $L(t) = \inf H_t$. For a function $w : \mathbb{R} \to \mathbb{R}$ define $\pi_t w = w \circ \pi_t^{-1}$.

Recall from (4.6) that the set U is either the whole circle or an open arc not containing $\mathcal{S} = 0$. Let
$$\bar{U} = \left\{ x \in \mathbb{R} : \bar{u}(x) < -\tfrac{N(u)}{2} \right\}$$

and take
$$l = \inf \left(\bar{U} \cap [-1, 0] \right), \quad r = \sup \left(\bar{U} \cap [0, 1] \right).$$

Finally, consider another initial state $\tilde{\mathcal{U}}$ given by $\tilde{\mathcal{S}} = \bar{\mathcal{S}}$, $\tilde{\mathcal{C}} = \bar{\mathcal{C}}$, and

$$\tilde{u}(x) = \begin{cases} \bar{u}(x), & x \in (l, r), \\ -\frac{N}{2}, & \text{otherwise.} \end{cases} \tag{4.16}$$

Define the evolution $(\tilde{\mathcal{U}}_t)_{t \geqslant 0}$ again by the same rules as for $\bar{\mathcal{U}}$, and consider the stopping time

$$\mathcal{T}_U = \inf \left\{ t : H_t \not\subseteq (l, r) \right\}.$$

Lemma 4.4 (Coupling). *The evolutions $(\bar{\mathcal{U}}_t)_{t \geqslant 0}$ and $(\tilde{\mathcal{U}}_t)_{t \geqslant 0}$ on the line and $(\mathcal{U}_t)_{t \geqslant 0}$ on the circle may be constructed on the same probability space, satisfying*

$$\mathcal{T}_\circ = \mathcal{T}_{[1]}, \qquad\qquad\qquad \mathcal{T}_\curlyvee = \mathcal{T}_U \wedge T_{[1]},$$
$$\mathcal{U}_t = \pi_t(\bar{\mathcal{U}}_t) \text{ for all } t < \mathcal{T}_{[1]}, \qquad\qquad\qquad \bar{\mathcal{S}}_t = \tilde{\mathcal{S}}_t \text{ for all } t < \mathcal{T}_U.$$

Proof. The coupling given by Lemma 4.4 is illustrated in Figure 4.2. The evolution of $(\mathcal{U}_t)_t$ can be constructed using an i.i.d. sequence $(E_n, U_n, T_n)_n$, where E_n and U_n are the exponential uniform used as input for (3.2) and (3.3) at each departure time t_n, and T_n are the service times prior to the n-th departure time. (When the system enters the idle state, another clock will be needed to determine the next arrival time, but this state cannot be achieved before $\mathcal{T}_\circ$.)

The coupling is simple: we use the same sequence $(E_n, U_n, T_n)_n$ to build $(\bar{\mathcal{U}}_t)_t$ and $(\tilde{\mathcal{U}}_t)_t$. It remains to check that this coupling a.s. satisfies the identities stated in the lemma; the details are left to the reader. □

Strong transience   For the evolution $(\tilde{\mathcal{U}}_t)_t$, the total distance traveled by the server between times t and t' is denoted by

$$V_t^{t'}(\tilde{\mathcal{S}}_\cdot) := \int_t^{t'} |\tilde{\mathcal{V}}_s| \mathrm{d}s = v \int_t^{t'} \mathbb{1}_{\text{moving}}(s) \mathrm{d}s.$$

We say that $(\tilde{\mathcal{S}}_t)_{t \geqslant 0}$ is *transient* if, for each $M > 0$, $\sup \left\{ t : \tilde{\mathcal{S}}_t \in [-M, M] \right\} < \infty$. If moreover

$$\left| \tilde{\mathcal{S}}_t - \tilde{\mathcal{S}}_0 \right| \geqslant \frac{1}{3} V_0^t(\tilde{\mathcal{S}}_\cdot) \qquad \text{for all } t > 0, \tag{4.17}$$
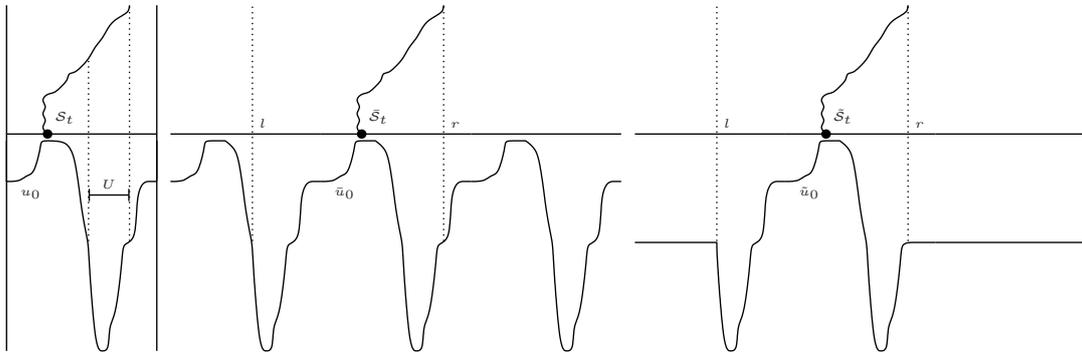

Figure 4.2: Illustration of how $\mathcal{U}$, $\bar{\mathcal{U}}$, and $\tilde{\mathcal{U}}$ evolve together until time $\mathcal{T}_\curlyvee$.

we say that $(\tilde{\mathcal{S}}_t)_{t \geqslant 0}$ is *strongly transient*. The latter means that the total displacement must increase linearly with the traveled distance.

For $0 \leqslant \alpha \leqslant 1$, we say that $\tilde{\mathcal{U}}$ is *α-unimodal* if $\tilde{u}$ attains its maximum on $\tilde{\mathcal{C}}$, and

$$\tilde{u}(x) \leqslant \alpha \cdot \inf_{y \in [\tilde{\mathcal{C}}, x]} \tilde{u}(y), \ \forall \ x > \tilde{\mathcal{C}}, \qquad \tilde{u}(x) \leqslant \alpha \cdot \inf_{y \in [x, \tilde{\mathcal{C}}]} \tilde{u}(y), \ \forall \ x < \tilde{\mathcal{C}}. \qquad (4.18)$$

Notice that with $\alpha = 1$ this is equivalent to $\tilde{u}$ being nondecreasing on $(-\infty, \tilde{\mathcal{C}}]$ and nonincreasing on $[\tilde{\mathcal{C}}, \infty)$, and for $\alpha < 1$ and $\tilde{u} \leqslant 0$, the condition is weaker.

The result below is a consequence of Proposition 1 in [FRS15], written in our notations.

Proposition 4.5. *Given any $\tilde{\mathcal{U}}$ that is α-unimodal with $\alpha = 1$, $(\tilde{\mathcal{S}}_t)_{t \geqslant 0}$ is a.s. transient.*

In order to prove Proposition 4.3, we shall obtain a slightly stronger result:

Proposition 4.6. *Given any $\tilde{\mathcal{U}}$ that is α-unimodal with $\alpha = \frac{1}{2}$, there exists a random time $\mathcal{T}_Z$ satisfying $\mathbb{P}^{\tilde{\mathcal{U}}}(\mathcal{T}_Z < \infty) = 1$, and such that $(\tilde{\mathcal{S}}_{\mathcal{T}_Z + t})_{t \geqslant 0}$ is strongly transient. Moreover, the number of departure times $\tilde{\mathcal{N}}_{\mathcal{T}_Z}$ before $\mathcal{T}_Z$ is tight:*

$$\mathbb{P}^{\tilde{\mathcal{U}}} \left\{ \tilde{\mathcal{N}}_{\mathcal{T}_Z} > k \right\} \underset{k \to \infty}{\longrightarrow} 0$$

uniformly over all α-unimodal $\tilde{\mathcal{U}}$.

Proposition 4.6 is proved in Sections 4.4 and 4.5 by adapting the multi-scale construction of [FRS15] to the case of α-unimodal initial states.

Proof of Proposition 4.3. We first observe that $\tilde{\mathcal{U}}$, with $\tilde{u}$ defined by (4.16), is α-unimodal for $\alpha = \frac{1}{2}$. By definition of $\mathscr{M}$ and V,

$$\mathscr{M} = \frac{1}{v} \, V_0^{\mathcal{T}}(\mathcal{S}.) \leqslant \frac{1}{v} \, V_0^{\mathcal{T}_Y}(\mathcal{S}.)$$

and by Lemma 4.4,

$$V_0^{\mathcal{T}_Y}(\mathcal{S}.) = V_0^{\mathcal{T}_{[1]} \wedge \mathcal{T}_U}(\bar{\mathcal{S}}.) = V_0^{\mathcal{T}_{[1]} \wedge \mathcal{T}_U}(\tilde{\mathcal{S}}.) \leqslant V_0^{\mathcal{T}_U}(\tilde{\mathcal{S}}.).$$

By definition of $\mathcal{T}_U$, we have that $\tilde{\mathcal{S}}_t \in [l, r] \subseteq [-1, 1]$ for all $t < \mathcal{T}_U$. The distance traveled by the server between consecutive departure times is thus bounded by 2, and therefore

$$V_0^{\mathcal{T}_Z \wedge \mathcal{T}_U}(\tilde{\mathcal{S}}.) \leqslant 2(\tilde{\mathcal{N}}_{\mathcal{T}_Z \wedge \mathcal{T}_U} + 1) \leqslant 2(\tilde{\mathcal{N}}_{\mathcal{T}_Z} + 1).$$

In case $\mathcal{T}_U \leqslant \mathcal{T}_Z$, this upper bound for $V_0^{\mathcal{T}_U}(\tilde{\mathcal{S}}.)$ is good enough. So consider the case $\mathcal{T}_Z < \mathcal{T}_U$ and write

$$V_0^{\mathcal{T}_U}(\tilde{\mathcal{S}}.) = V_0^{\mathcal{T}_Z}(\tilde{\mathcal{S}}.) + V_{\mathcal{T}_Z}^{\mathcal{T}_U}(\tilde{\mathcal{S}}.).$$

By (4.17) and the definition of $\mathcal{T}_Z$,

$$V_{\mathcal{T}_Z}^{\mathcal{T}_U}(\tilde{\mathcal{S}}.) \leqslant 6.$$

Summarizing,

$$\mathscr{M} \leqslant \frac{1}{v} \left(8 + 2\tilde{\mathcal{N}}_{\mathcal{T}_Z} \right)$$

and the result then follows from Proposition 4.6. $\qquad \square$

4.4 Multi-scale estimates

In the remainder of this section we give a short but self-contained proof of Proposition 4.6 using a block argument. The reader will find a similar construction, with a more extended explanation of the main ideas, in [FRS15]. Since only $\tilde{\mathcal{U}}$ is concerned, we shorten notation and write $\mathcal{U}$ instead. Each time C or c appears, it denotes a different constant that is positive, finite, and depends only on v.

Let $\mathcal{A}_t = \sigma((\mathcal{U}_s)_{s\in[0,t]})$ denote the natural filtration for $(\mathcal{U}_t)_{t\geqslant 0}$. We construct a sequence of stopping times $0 = L_0 < L_1 < \cdots$ and define the corresponding events of success $A_j \in \mathcal{A}_{L_{j+1}}$ in terms of $\mathcal{U}_{L_j}$. The construction will have the following properties. For some sequence p_j and any $\mathcal{U}$ that is α-unimodal,

$$\mathbb{P}^{\mathcal{U}}(A_j|\mathcal{A}_{L_j}) = \mathbb{P}^{\mathcal{U}}(A_j|\mathcal{U}_{L_j}) \geqslant p_j \quad \text{on } A_0 \cap \cdots \cap A_{j-1}, \quad \text{and} \quad \prod_j p_j > 0. \qquad (4.19)$$

The event $\cap_{j=0}^{\infty} A_j$ implies strong transience of $(\mathcal{S}_t)_{t\geqslant 0}$. Almost surely, for each $j = 0, 1, 2, \ldots$, the state of $\mathcal{U}_{L_j}$ is *serving*.

We assume without loss of generality that the state of $\mathcal{U}_0$ is serving, that $\lambda = 1$, and that $\mathcal{S}_0 = 0$. Take $\sigma = \operatorname{sgn} \mathcal{S}_{L_1}$ to indicate the direction in which subsequent blocks are supposed to grow. Let $Z_j = \sigma \mathcal{S}_{L_j}$, $N_j = L_j - u(\mathcal{S}_{L_j})$, $Q_j = \mathcal{N}_{L_{j+1}} - \mathcal{N}_{L_j}$, $X_j = Z_{j+1} - Z_j$, $M_j = L_{j+1} - L_j$.

The triggering step $j = 0$ is defined as follows. We always take $Q_0 = 1$, and the first step consists of finishing with the customer that is being served at time L_0, then traveling towards the nearest customer at position σZ_1, and L_1 is the stopping time attained as soon as the server reaches this position. The event A_0 means *success* at the Step $j = 0$, and is defined by the following conditions: $X_0^- \leqslant X_0 \leqslant X_0^+$ and $M_0^- \leqslant M_0 \leqslant M_0^+$, where $X_0^- = \frac{9}{N_1}$, $X_0^+ = 36$, $M_0^- = 1$, and $M_0^+ = 2 + \frac{36}{v}$. If there is no success, we declare Step 0 to have *failed* and stop. In the sequel we assume without loss of generality that $\sigma = +1$.

For $j \geqslant 1$, suppose that Steps $0, 1, 2, \ldots, j-1$ have been successful and start from u_{L_j}, and take $\ell_j = \lceil 54 j^{1/4} \rceil$, $D_j = \frac{1}{36}\ell_j$. Let $s_{j,0} = S_{L_j}$ and $s_{j,1}, s_{j,2}, \ldots, s_{j,\ell_j}, s_{j,\ell_j+1}$ denote the positions of the next $\ell_j + 1$ customers. Step j may be successful, which is denoted by the event A_j, in two situations: first, if $s_{j,n} > s_{j,n-1}$ for $n = 1, \ldots, \ell_j$, in which case we take $Q_j = \ell_j$; second, if there is only one $\tilde{n} \in \{1, \ldots, \ell_j + 1\}$ such that $\tilde{n} \neq \ell_j + 1$ and $s_{j,\tilde{n}} < s_{j,\tilde{n}-1}$, in which case we take $Q_j = \ell_j + 1$. If none of these two happen, we declare Step j to have *failed* and stop. Otherwise, in either of the above two cases we say that Step j is *successful* if moreover

$$\begin{cases} Q_j^- \leqslant Q_j \leqslant Q_j^+, \quad X_j^- \leqslant X_j \leqslant X_j^+, \quad M_j^- \leqslant M_j \leqslant M_j^+, \\ V_{L_j}^{L_{j+1}}(\mathcal{S}_\cdot) \leqslant X_j + \frac{4D_j}{N_j}, \quad Z_{j-1} < \mathcal{S}_t < Z_{j+1} \text{ for } L_j \leqslant t < L_{j+1}, \end{cases}$$

where $Q_j^- = \ell_j, Q_j^+ = \ell_j + 1, X_j^- = \frac{\ell_j - 1}{3N_{j+1}}, X_j^+ = \frac{3\ell_j}{N_j}, M_j^- = \frac{1}{2}Q_j^-, M_j^+ = 2Q_j^+ + \frac{3X_j^+}{v}$. Here the time L_{j+1} is given by the instant when the server reaches the last customer, located at Z_{j+1}, and the next block starts with this customer being served.

We now estimate the probability of success $\mathbb{P}^{\mathcal{U}}(A_j|\mathcal{U}_{L_j})$ on $A_0 \cap \cdots \cap A_{j-1}$ by considering a number of events that imply A_j.

First notice that $A_0 \cap \cdots \cap A_{j-1}$ implies that $M_j^- \geqslant Cj^{1/4}, N_j = -u(\sigma Z_j) + L_j \geqslant L_j \geqslant M_0^- + \cdots + M_{j-1}^- \geqslant Cj^{5/4}, X_j^+ \leqslant Cj^{-1}, X_{j-1}^- \leqslant Cj^{-1}$, and thus $M_j^+ \leqslant Cj^{1/4}$. Moreover, the condition of $\mathcal{U}_t$ being α-unimodal is preserved for all t. (Indeed, while the system state is serving or moving, by (3.1) the potential $u \leqslant 0$ changes by the subtraction of a constant that increases with time, and this preserves (4.18). At departure times, the state is updated by (3.2)-(3.3), which changes the position of $\mathcal{C}$ and increases the value of u to 0 on an interval that contains both the old and new $\mathcal{C}$. This increases the inf in (4.18), so this inequality still holds for x outside of such interval, whereas for x in such interval both $u(x)$ and the inf become 0.) Thus, for $j \geqslant 1$, the event A_{j-1} implies the following conditions on $\mathcal{U}_{L_j}$:

$$\begin{cases} u_{L_j}(x) = u_0(x) - L_j \leqslant -\frac{N_j}{2} & \text{for } x > Z_j, \\ u_{L_j}(x) \geqslant -M_{j-1} & \text{for } Z_j - X_{j-1}^- < x < Z_j. \end{cases} \tag{4.20}$$

Let $T_0, T_1, \ldots, T_{\ell_j}$ denote the service times of the customer being served at time L_0 and the following ℓ_j customers. Let $E_1, U_1, E_2, U_2, \ldots, E_{\ell_j}, U_{\ell_j}, E_{\ell_j+1}, U_{\ell_j+1}$ be the exponential and uniform random variables used for determining the positions of $s_{j,1}, s_{j,\ell_j+1}$ via (3.2).

For $j = 0$, consider the event that $1 < T_0 < 2$, $36 < E_1 < 72$, and that U_1 lies on the largest interval among $(0, \frac{a}{a+b})$ and $[\frac{a}{a+b}, 1)$. The probability that these conditions are satisfied is at least $p_0 = \frac{1}{2}(e^{-36} - e^{-72})(e^{-1} - e^{-2}) > 0$. The requirement for U_1 implies that $u_0(S_{L_1}) \leqslant u_0(-S_{L_1})$. Hence, by α-unimodality of u_0, the occurrence of the above events imply

$$36 < \int_{-z}^{+z} [-u_0(x) + T_0] \mathrm{d}x \leqslant \int_{-z}^{+z} \max_{[-z,+z]} (T_0 - u_0) \mathrm{d}x \leqslant -4X_0[u_0(S_{L_1}) + T_0] \leqslant 4X_0 N_1$$

and

$$72 > \int_{-z}^{+z} [-u_0(x) + T_0] \mathrm{d}x \geqslant \int_{-z}^{+z} T_0 \mathrm{d}x \geqslant 2z,$$

which in turn imply the bounds on X_0 and M_0, therefore $\mathbb{P}^u(A_0) \geqslant p_0 > 0$.

For $j \geqslant 1$, we observe that the positions $s_{j,1}, s_{j,2}, \ldots s_{j,\ell_j+1}$ can be sampled via a Poisson Point Process on the region

$$R = \left\{ (x,t) \in \mathbb{R}^2 : Z_j - X_{j-1}^- \leqslant x \leqslant Z_j + X_j^+, u_{L_j}(x) < t \leqslant M_j^+ \right\},$$

unless the elapsed time M_j exceeds M_j^+ or the exploration for customers leaves the interval $[Z_j - X_{j-1}^-, Z_j + X_j^+]$, which is ruled out a posteriori in case Step j is successful. This region R can be decomposed in a disjoint union $R_1 \cup R_2$, where $R_1 = \left\{ (x,t) \in \mathbb{R}^2 : Z_j \leqslant x \leqslant Z_j + X_j^+, u_{L_j}(x) < t \leqslant 0 \right\}$ and $R_2 \subseteq [Z_j - X_{j-1}^-, Z_j] \times [-M_{j-1}^+, M_j^+] \cup [Z_j, Z_j + X_j^+] \times [0, M_j^+]$.

The inequalities in (4.20) and those preceding it imply that

$$|R_2| \leqslant (X_{j-1}^- + X_j^+)(M_{j-1}^+ + M_j^+) \leqslant Cj^{-3/4}.$$

The probability that there are two or more points in R_2 is thus bounded by $Cj^{-3/2}$.

Define $\beta(x) = \int_{Z_j}^x [-u_{L_j}(z)] \mathrm{d}z$, $x \geqslant Z_j$, and write $(x_1, t_1), (x_2, t_2), \ldots$ the set of points found in R_1, labeled by ordering $x_1 < x_2 < x_3 < \cdots$. Writing $x_0 = Z_j$, we have that $(\beta(x_n) - \beta(x_{n-1}))_{n=1,\ldots,\ell_j}$

are i.i.d. exponential random variables. With positive probability, and tending to 1 exponentially fast in ℓ_j as $j \to \infty$, both events

$$\frac{2}{3}(\ell_j - 1) \leqslant \beta(x_{\ell_j - 1}) \leqslant \beta(x_{\ell_j}) \leqslant \frac{3}{2}\ell_j$$

and

$$\beta(x_n) - \beta(x_{n-1}) \leqslant D_j \quad \text{for } n = 1, 2, \ldots, \ell_j$$

occur. But since $\mathcal{U}_{L_j}$ is α-unimodal we have

$$\frac{1}{2}N_j \leqslant \frac{\beta(x_n) - \beta(x_{n-1})}{x_n - x_{n-1}} \leqslant 2N_{j+1}.$$

The above inequalities imply that $X_j^- \leqslant x_{\ell_j - 1} - Z_j \leqslant x_{\ell_j} - Z_j \leqslant X_j^+$ and $x_n - x_{n-1} \leqslant \frac{2D_j}{N_j} = \frac{\ell_j}{18N_j} \leqslant \frac{X_{j-1}^-}{3}$. On the event that there are no points in R_2, we have $V_{L_j}^{L_j+1} = X_j = x_{\ell_j} - Z_j$. On the event that there is one point in R_2, we have $x_{\ell_j-1} - Z_j \leqslant X_j \leqslant x_{\ell_j} - Z_j$ and $V_{L_j}^{L_j+1} \leqslant X_j + 4\frac{D_j}{N_j}$. This finishes the bounds on Q_j, X_j, $V_{L_j}^{L_j+1}$, and $(\mathcal{S}_t)_{t \in [L_j, L_{j+1})}$.

It remains to control M_j, which is given by the sum $T_0 + \cdots + T_{Q_j-1} + v^{-1}V_{L_j}^{L_j+1}(\mathcal{S}.)$ of service times plus traveling time. The latter is non-negative and bounded by $2X_j/v$, which is bounded by $2X_j^+/v$. Therefore the inequality $M_j^- \leqslant M_j \leqslant M_j^+$ holds whenever $Q_j/2 < T_0 + \cdots + T_{Q_j-1} < 2Q_j$, which in turn occur with exponentially high probability in ℓ_j.

Finally, we use the bound on $V_{L_j}^{L_j+1}(\mathcal{S}.)$ to prove strong transience. We add the requirement that $V_{L_j}^{L_j+1}(\mathcal{S}.) = X_j$ for $j = 1$. This changes the lower bound on probability of A_1, but it remains positive. Notice that the same equality is true for $j = 0$ by construction. Now one can decompose $V_0^t(\mathcal{S}.)$ in distances traveled in each of the two possible directions and again decompose these distances in the contribution from each block, and combine the bounds on X_{j-1} with $\frac{4D_j}{N_j} \leqslant \frac{2}{3}X_{j-1}^-$ to get $V_0^t(\mathcal{S}.) \leqslant \frac{5}{3}\mathcal{S}_t$.

4.5 Renewal argument

Having (4.19) in hands, we finally prove tightness of $\mathcal{N}_{\mathcal{T}_z}$. We need that the probability of success in each block j not only be bounded from below by some p_j but actually equal to p_j. We introduce an artificial coin toss to provide this last ingredient.

Enlarge the underlying probability space to add an independent sequence of i.i.d. uniform variables $\tilde{U}_j$. For each j, define the event $\tilde{A}_j \subseteq A_j$ by

$$\tilde{A}_j = A_j \cap \left[\tilde{U}_{j+1} < \frac{p_j}{\mathbb{P}(A_j | \mathcal{U}_{L_j})} \right].$$

In words, we add an extra coin toss in order to have an exact equality instead of an upper bound:

$$\mathbb{P}^{\mathcal{U}}(\tilde{A}_j | \mathcal{U}_{L_j}) = p_j \text{ on } \tilde{A}_0 \cap \cdots \cap \tilde{A}_{j-1}.$$

Notice that $J^0 = \min\{j : \tilde{A}_{j-1} \text{ does not occur}\} \in \{1, 2, 3, \ldots\} \cup \{\infty\}$ is a stopping time with respect to $\{\tilde{\mathcal{A}}_j\}_{j=0,1,2,\ldots}$, where $\tilde{\mathcal{A}}_j = \sigma(\mathcal{A}_{L_j}, \tilde{U}_0, \tilde{U}_1, \ldots, \tilde{U}_{j-1}, \tilde{U}_j)$. The distribution of J^0 is given

by $\mathbb{P}(J^0 > k+1) = p_0 p_1 \cdots p_k$. If $J^0 = \infty$, we have success for all j and $(\mathcal{S}_t)_{t \geqslant 0}$ is strongly transient, and we can take $\mathcal{T}_Z = 0$.

If otherwise, $J^0 < \infty$, we have $\mathcal{N}_{L_{J^0}} \leqslant Q^+(J^0)$, where $Q^+(j) = Q_0^+ + Q_1^+ + \cdots + Q_j^+$. In this case we can apply a time shift of L_{J^0} and define $(\mathcal{U}_t^1)_{t \geqslant 0}$ by $\mathcal{U}_t^1 = \mathcal{U}_{t+L_{J^0}}$. For this evolution $(\mathcal{U}_t^1)_{t \geqslant 0}$ we can define the stopping times $L_0^1, L_1^1, L_2^1, \ldots$, the events $\tilde{A}_0^1, \tilde{A}_1^1, \tilde{A}_2^1, \ldots$, and the step of first failure J^1.

By the strong Markov property, the conditional distribution of $(\mathcal{U}_t^1)_{t \geqslant 0}$ given that $J^0 < \infty$ is given by $\mathbb{P}^{\mathcal{U}_{L_{J^0}}}$, and since $\mathcal{U}_{L_{J^0}}$ is α-unimodal, the conditional distribution of J^1 given that $J^0 < \infty$ is the same: $\mathbb{P}(J^1 > k+1 | J^0 < \infty) = p_0 p_1 \cdots p_k$.

Again, if $J^1 = \infty$, $(\mathcal{S}_{t+L_{J^0}})_{t \geqslant 0}$ is strongly transient and we take $\mathcal{T}_Z = L_{J^0}$. Otherwise, $\mathcal{N}_{L_{J^0} + L_{J^1}^1} \leqslant Q^+(J^0) + Q^+(J^1)$. Analogously we can construct $\mathcal{U}^2, J^2, \mathcal{U}^3, J^3, \ldots$ until at some step $K+1$ we get $J^{K+1} = \infty$. We then take $\mathcal{T}_Z = L_{J^0} + L_{J^1}^1 + \cdots + L_{J^K}^K$, and we have that $(\mathcal{S}_{t+\mathcal{T}_Z})_{t \geqslant 0}$ is strongly transient. As before, $\mathcal{N}_{\mathcal{T}_Z} \leqslant Q^+(J^0) + Q^+(J^1) + \cdots + Q^+(J^K)$. But the distribution of the latter upper bound does not depend on $\mathcal{U}$, and therefore $\mathcal{N}_{\mathcal{T}_Z}$ is tight.

Acknowledgments

References

[ABV03]　O. ANGEL, I. BENJAMINI, AND B. VIRÁG, *Random walks that avoid their past convex hull*, Electron. Comm. Probab., 8 (2003), pp. 6–16.

[AF97]　E. ALTMAN AND S. FOSS, *Polling on a space with general arrival and service time distribution*, Oper. Res. Lett., 20 (1997), pp. 187–194.

[AL94]　E. ALTMAN AND H. LEVY, *Queueing in space*, Adv. in Appl. Probab., 26 (1994), pp. 1095–1116.

[BFV10]　V. BEFFARA, S. FRIEDLI, AND Y. VELENIK, *Scaling limit of the prudent walk*, Electron. Commun. Probab., 15 (2010), pp. 44–58.

[BM10]　M. BOUSQUET-MÉLOU, *Families of prudent self-avoiding walks*, J. Combin. Theory Ser. A, 117 (2010), pp. 313–344.

[BW03]　I. BENJAMINI AND D. B. WILSON, *Excited random walk*, Electron. Comm. Probab., 8 (2003), pp. 86–92.

[CD99]　L. CHAUMONT AND R. A. DONEY, *Pathwise uniqueness for perturbed versions of Brownian motion and reflected Brownian motion*, Probab. Theory Related Fields, 113 (1999), pp. 519–534.

[CG87]　J. E. G. COFFMAN AND E. N. GILBERT, *Polling and greedy servers on a line*, Queueing Systems Theory Appl., 2 (1987), pp. 115–145.

[CPY98]　P. CARMONA, F. PETIT, AND M. YOR, *Beta variables as times spent in* $[0, \infty[$ *by certain perturbed Brownian motions*, J. London Math. Soc. (2), 58 (1998), pp. 239–256.

[Dav96]　B. DAVIS, *Weak limits of perturbed random walks and the equation* $Y_t = B_t + \alpha \sup\{Y_s : s \leq t\} + \beta \inf\{Y_s : s \leq t\}$, Ann. Probab., 24 (1996), pp. 2007–2023.

[Dav99] ——, *Brownian motion and random walk perturbed at extrema*, Probab. Theory Related Fields, 113 (1999), pp. 501–518.

[FL96] S. Foss and G. Last, *Stability of polling systems with exhaustive service policies and state-dependent routing*, Ann. Appl. Probab., 6 (1996), pp. 116–137.

[FL98] ——, *On the stability of greedy polling systems with general service policies*, Probab. Engrg. Inform. Sci., 12 (1998), pp. 49–68.

[FRS15] S. Foss, L. T. Rolla, and V. Sidoravicius, *Greedy walk on the real line*, Ann. Probab., 43 (2015), pp. 1399–1418.

[KM97] I. A. Kurkova and M. V. Menshikov, *Greedy algorithm, $\mathbf{Z}^1$ case*, Markov Process. Related Fields, 3 (1997), pp. 243–259.

[KS92] D. P. Kroese and V. Schmidt, *A continuous polling system with general service times*, Ann. Appl. Probab., 2 (1992), pp. 906–927.

[KS94] ——, *Single-server queues with spatially distributed arrivals*, Queueing Systems Theory Appl., 17 (1994), pp. 317–345.

[KS96] ——, *Light-traffic analysis for queues with spatially distributed arrivals*, Math. Oper. Res., 21 (1996), pp. 135–157.

[LSW04a] G. F. Lawler, O. Schramm, and W. Werner, *Conformal invariance of planar loop-erased random walks and uniform spanning trees*, Ann. Probab., 32 (2004), pp. 939–995.

[LSW04b] ——, *On the scaling limit of planar self-avoiding walk*, in Fractal geometry and applications: a jubilee of Benoît Mandelbrot, Part 2, vol. 72 of Proc. Sympos. Pure Math., Amer. Math. Soc., Providence, RI, 2004, pp. 339–364.

[MQ99] R. Meester and C. Quant, *Stability and weakly convergent approximations of queueing systems on a circle*. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.12.7937, 1999.

[MT08] T. Mountford and P. Tarrès, *An asymptotic result for Brownian polymers*, Ann. Inst. Henri Poincaré Probab. Stat., 44 (2008), pp. 29–46.

[PW97] M. Perman and W. Werner, *Perturbed Brownian motions*, Probab. Theory Related Fields, 108 (1997), pp. 357–383.

[RNFK11] L. Rojas-Nandayapa, S. Foss, and D. P. Kroese, *Stability and performance of greedy server systems: A review and open problems*, Queueing Syst., 68 (2011), pp. 221–227.

[RS11] O. Raimond and B. Schapira, *Excited Brownian motions*, ALEA Lat. Am. J. Probab. Math. Stat., 8 (2011), pp. 19–41.

[Sch00] O. Schramm, *Scaling limits of loop-erased random walks and uniform spanning trees*, Israel J. Math., 118 (2000), pp. 221–288.

[Smi01] S. Smirnov, *Critical percolation in the plane: conformal invariance, Cardy's formula, scaling limits*, C. R. Acad. Sci. Paris Sér. I Math., 333 (2001), pp. 239–244.

[Smi10] ——, *Conformal invariance in random cluster models. I. Holomorphic fermions in the Ising model*, Ann. of Math. (2), 172 (2010), pp. 1435–1467.

[Tót95] B. Tóth, *The "true" self-avoiding walk with bond repulsion on $\mathbf{Z}$: limit theorems*, Ann. Probab., 23 (1995), pp. 1523–1556.

[Tót99] ——, *Self-interacting random motions—a survey*, in Random walks (Budapest, 1998), vol. 9 of Bolyai Soc. Math. Stud., János Bolyai Math. Soc., Budapest, 1999, pp. 349–384. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.40.8361.

[TW98] B. Tóth and W. Werner, *The true self-repelling motion*, Probab. Theory Related Fields, 111 (1998), pp. 375–452.

[Zer05] M. P. W. Zerner, *On the speed of a planar random walk avoiding its past convex hull*, Ann. Inst. H. Poincaré Probab. Statist., 41 (2005), pp. 887–900.