J. Appl. Prob. **55,** 968–997 (2018)
*doi:*10.1017/jpr.2018.60

APPROXIMATION OF EXCESSIVE BACKLOG PROBABILITIES OF TWO TANDEM QUEUES

ALI DEVIN SEZER,* *Middle East Technical University*

Abstract

Let X be the constrained random walk on $\mathbb{Z}_+^2$ having increments $(1, 0)$, $(-1, 1)$, and $(0, -1)$ with respective probabilities λ, μ_1, and μ_2 representing the lengths of two tandem queues. We assume that X is stable and $\mu_1 \neq \mu_2$. Let τ_n be the first time when the sum of the components of X equals n. Let Y be the constrained random walk on $\mathbb{Z} \times \mathbb{Z}_+$ having increments $(-1, 0)$, $(1, 1)$, and $(0, -1)$ with probabilities λ, μ_1, and μ_2. Let τ be the first time that the components of Y are equal to each other. We prove that $P_{n-x_n(1),x_n(2)}(\tau < \infty)$ approximates $p_n(x_n)$ with relative error *exponentially decaying* in n for $x_n = \lfloor nx \rfloor$, $x \in \mathbb{R}_+^2$, $0 < x(1) + x(2) < 1$, $x(1) > 0$. An affine transformation moving the origin to the point $(n, 0)$ and letting $n \to \infty$ connect the X and Y processes. We use a linear combination of basis functions constructed from single and conjugate points on a characteristic surface associated with X to derive a simple expression for $\mathbb{P}_y(\tau < \infty)$ in terms of the utilization rates of the nodes. The proof that the relative error decays exponentially in n uses a sequence of subsolutions of a related Hamilton–Jacobi–Bellman equation on a manifold consisting of three copies of $\mathbb{R}_+^2$ glued to each other along the constraining boundaries. We indicate how the ideas of the paper can be generalized to more general processes and other exit boundaries.

Keywords: Large deviation; constrained random walk; buffer overflow; queueing system; exit time; harmonic system

2010 Mathematics Subject Classification: Primary 60G50
Secondary 60G40; 60F10; 60J45

1. Introduction and definitions

Let X be a random walk with independent and identically distributed (i.i.d.) increments $\{I_1, I_2, I_3, \ldots\}$, constrained to remain in $\mathbb{Z}_+^2$, i.e.

$$X_0 = x \in \mathbb{Z}_+^2, \qquad X_{k+1} \doteq X_k + \pi(X_k, I_k), \qquad k = 1, 2, 3, \ldots,$$

$$I_k \in \{(1, 0), (-1, 1), (0, -1)\}, \qquad \pi(x, v) \doteq \begin{cases} v & \text{if } x + v \in \mathbb{Z}_+^2, \\ 0 & \text{otherwise,} \end{cases}$$

$$\mathbb{P}(I_k = (1, 0)) = \lambda, \qquad \mathbb{P}(I_k = (-1, 1)) = \mu_1, \qquad \mathbb{P}(I_k = (0, -1)) = \mu_2.$$

Let $\partial_i \doteq \{x \in \mathbb{Z}^2 : x(i) = 0\}$, $i = 1, 2$, denote the constraining boundaries of the process and let $\sigma_i \doteq \inf\{k : X_k \in \partial_i\}$, $i = 1, 2$, denote the first time that X hits these boundaries. The components of X represent the number of customers at jump times of a Jackson network consisting of two tandem queues.

Received 23 June 2016; revision received 26 June 2018.
* Postal address: Institute of Applied Mathematics, Middle East Technical University, Ankara, Turkey.
Email address: devin.sezer@gmail.com

We assume that X is stable, i.e. $\lambda < \mu_1, \mu_2$. We also assume that $\mu_1 \neq \mu_2$; see Subsection 7.1 for a discussion on the $\mu_1 = \mu_2$ case. Define

$$A_n = \{x \in \mathbb{Z}_+^2 \colon x(1) + x(2) \le n\}$$

and its boundary

$$\partial A_n = \{x \in \mathbb{Z}_+^2 \colon x(1) + x(2) = n\}.$$

Let τ_n be the first time that X hits ∂A_n, i.e.

$$\tau_n \doteq \inf\{k \ge 0 \colon X_k \in \partial A_n\}, \qquad n \ge 0. \tag{1}$$

The set A_n models a system-wide shared buffer of size n. One can change A_n to model other buffer structures, e.g. $\{x \in \mathbb{Z}_+^2 \colon x(i) \le n,\ i = 1, 2\}$, but for the purposes of this paper we will confine ourselves to A_n as defined above. For comments on generalizations, see Section 7. Our aim is to develop an approximation formula for the probability $p_n(x) \doteq \mathbb{P}_x(\tau_n < \tau_0)$, i.e. the probability that, starting from an initial state $x \in A_n$, the total number of customers in the system reaches n before the system empties. If we measure time in the number of independent cycles that restart each time X hits 0, p_n is the probability that the current cycle finishes successfully (i.e. without a buffer overflow). Knowledge of how to compute p_n can be useful in computing other performance measures, e.g. the expected number of cycles before failure or the expected cost of operating the system modeled by X. These make p_n a natural and useful measure of reliability. Despite the considerable attention that the analysis of probabilities such as p_n has received over the years (see the literature review below), to the best of the author's knowledge, sharp approximation results do not exist in the current literature; the difficulty arises from the multi-dimensionality of the problem and the constraining boundaries of the process. For this reason, we focus on one of the simplest two-dimensional constrained random walks: the tandem walk above. We expect the basic ideas of the paper to apply to more general models than the two-dimensional tandem walk, we comment on these in Section 7; Subsection 7.5 contains connections to the analysis of random perturbations of stable dynamical systems.

Since X is a Markov process, p_n, as a function of the initial point x, satisfies a system of linear equations; see (12). Since the number of unknowns grows as n^2, the resource requirements for an exact or numerical solution for this system becomes nontrivial for even moderate n (e.g. for $n = 1000$, a system will have half a million unknowns). For larger networks (for d nodes, the number of unknowns grows as n^d) this point is more pronounced. Therefore, more efficient approximation techniques are of interest. We develop the following approximation equation for p_n. Let Y be the random walk on $\mathbb{Z} \times \mathbb{Z}_+$ having increments $(-1, 0)$, $(1, 1)$, and $(0, -1)$ with respective probabilities λ, μ_1, and μ_2, constrained to be positive only on its second component. Let τ be the first time that the components of Y equal each other (the relation between X and Y is explained in the paragraphs below). In Section 3 we derive the following explicit formula:

$$\mathbb{P}_y(\tau < \infty) = W(y) \doteq \left(\rho_2^{y(1)-y(2)} - \frac{\mu_2 - \lambda}{\mu_2 - \mu_1}\rho_2^{y(1)-y(2)}\rho_1^{y(2)}\right) + \frac{\mu_2 - \lambda}{\mu_2 - \mu_1}\rho_1^{y(1)}, \tag{2}$$

$\rho_i \doteq \lambda/\mu_i$, $y \in \mathbb{Z}_+^2$, $y(1) > y(2)$. Fix $x \in \{x \in \mathbb{R}_+^2 \colon 0 < x(1) + x(2) < 1\}$ and define $x_n = \lfloor nx \rfloor$. In Section 4 we show that $W(n - x_n(1), x_n(2))$ approximates $p_n(x_n)$ with relative error *exponentially vanishing* in n (see Proposition 8). In the following paragraphs we summarize the analysis behind the results stated above.

When X is stable (as we have assumed) and when the initial position $X_0 = x$ is away from the exit boundary ∂A_n, the event $\{\tau_n < \tau_0\}$ rarely happens and its probability p_n decays

https://doi.org/10.1017/jpr.2018.60 Published online by Cambridge University Press

exponentially with buffer size n. The classical approximation technique to approximate small probabilities such as p_n is large deviations (LD) analysis, which identifies the exponential decay rate of p_n. To put our approach into context, we summarize the ideas involved in LD analysis. Note that p_n itself decays to 0, which is trivial. To obtain a nontrivial limit transform p_n to $V_n \doteq -(1/n)\log p_n$, using convex duality we can write the $-\log$ of an expectation as an optimization problem involving the relative entropy function (see [12]) and, thus, V_n can be interpreted as the value function of a discrete-time stochastic optimal control problem. The LD analysis consists of the law of large numbers limit analysis of this control problem; the limit problem is a deterministic optimal control problem whose value function satisfies a first-order Hamilton–Jacobi–Bellman (HJB) equation (see (38) in Section 4). Thus, LD analysis amounts to the computation of the limit of a *convex transformation* of the problem.

We use another (*affine*) transformation of X for the limit analysis. The proposed transformation is very simple: *observe* X *from the exit boundary.* For the two tandem walk, the most natural vantage point on the exit boundary ∂A_n turns out to be the corner $(n, 0)$. Therefore, we transform the process thus

$$Y^n \doteq T_n(X), \qquad T_n \colon \mathbb{R}^2 \to \mathbb{R}^2, \qquad T_n(x) \doteq y, \qquad y(2) = x(2), \qquad y(1) = n - x(1);$$

see Figure 1. The transformation T_n is affine and its inverse is equal to itself. The process Y^n, i.e. the process X as observed from the corner $(n, 0)$ is a constrained process on the domain $\Omega_Y^n \doteq (n - \mathbb{Z}_+) \times \mathbb{Z}_+$. The transformation T_n maps the set A_n to $B_n \subset \Omega_Y^n$, $B_n \doteq T_n(A_n)$, the corner $(n, 0)$ to the origin of Ω_Y^n, the exit boundary ∂A_n to $\partial B_n \doteq \{y \in \Omega_Y^n,\ y(1) = y(2)\}$, and finally the constraining boundary $\{x \in \mathbb{Z}_+^2,\ x(1) = 0\}$ to $\{y \in \mathbb{Z}_+^2 \colon y(1) = n\}$. As $n \to \infty$, the last boundary vanishes and Y^n converges to the limit process Y on the domain $\Omega_Y \doteq \mathbb{Z} \times \mathbb{Z}_+$ and the set B_n to $B \doteq \{y \in \Omega_Y,\ y(1) \geq y(2)\}$. The exit boundary for the limit problem is $\partial B = \{y \in \Omega_Y,\ y(1) = y(2)\}$, and the limit stopping time

$$\tau \doteq \inf\{k \geq 0 \colon Y_k \in \partial B\}$$

is the first time Y hits ∂B.

The stability of X *and the vanishing of the boundary constraint on* ∂_1 *implies that* Y *is* unstable/transient, i.e. *with probability* 1 *it wanders off to* ∞. Therefore, in our formulation, the limit process is an *unstable* constrained random walk in the same space and time scale as the original process but with fewer constraints, and the limit problem amounts to whether this unstable process ever hits the fixed boundary ∂B.

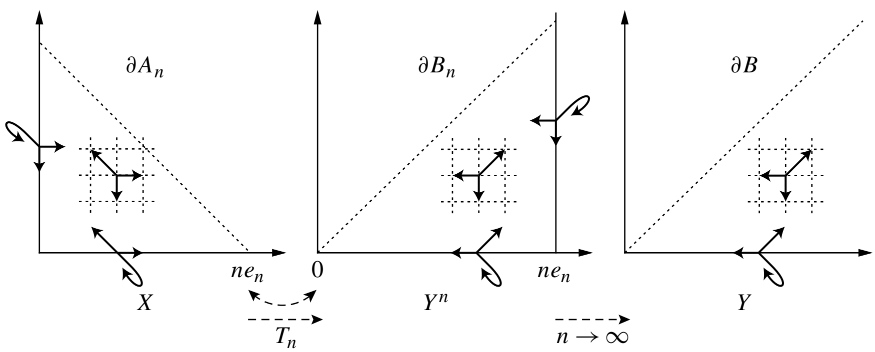

FIGURE 1: The transformation T_n.

https://doi.org/10.1017/jpr.2018.60 Published online by Cambridge University Press

Fix an initial point $y \in B$ in the new coordinates; our first convergence result is Proposition 1, i.e.

$$p_n = \mathbb{P}_{x_n}(\tau_n < \tau_0) \to \mathbb{P}_y(\tau < \infty), \tag{3}$$

where $x_n = T_n(y)$. The proof uses the law of large numbers and LD lower bounds to show that the difference between the two sides of (3) vanishes with n. With (3) we see that the limit problem in our formulation is to compute the hitting probability of the unstable Y to the boundary ∂B.

The convergence statement (3) involves a fixed initial condition for the process Y. In classical LD analysis, one specifies the initial point in scaled coordinates as follows: $x_n = \lfloor nx \rfloor \in A_n$ for $x \in \mathbb{R}^d_+$. Then the initial condition for the Y^n process is $y_n = T_n(x_n)$ (i.e. we do not fix the y coordinate but the scaled x coordinate). When x_n is defined in this way, (3) becomes a trivial statement since both sides decay to 0. For this reason, Section 4 is devoted to the study of the relative error

$$\frac{|\mathbb{P}_{x_n}(\tau_n < \tau_0) - \mathbb{P}_{y_n}(\tau < \infty)|}{\mathbb{P}_{x_n}(\tau_n < \tau_0)}; \tag{4}$$

in Proposition 8 we see that this error converges exponentially to 0 for the case of the two-dimensional tandem walk (i.e. the process X in Figure 1). The proof relies on showing that the probability of the intersection of the events $\{\tau_n < \tau_0\}$ and $\{\tau < \infty\}$ dominates the probabilities of both as $n \to \infty$. For this we calculate bounds in Proposition 10 on the LD decay rates of the probability of the differences between these events using a sequence of subsolutions of an HJB equation *on a manifold*; the manifold consists of three copies of $\mathbb{R}^2_+$, the zeroth copy glued to the first along ∂_1, and the first to the second along ∂_2, where $\partial_i = \{x \in \mathbb{R}^2_+ : x(i) = 0\}$. Extension of this argument to more complex processes and domains remains for future work.

The convergence results (3) and (4) reduce the problem of calculating $\mathbb{P}_x(\tau_n < \tau_0)$ to that of $\mathbb{P}_y(\tau < \infty)$. This constitutes the first step of our analysis and we expect it to apply more generally; see Subsection 7.4.

In Section 3 we apply the *principle of superposition* of classical linear analysis to the computation of $\mathbb{P}_y(\tau < \infty)$. The key for its application is to construct the right class of efficiently computable basis functions to be superposed. The construction of our basis functions is as follows. The distribution of the increments of Y is used to define the *characteristic polynomial* $\boldsymbol{p} : \mathbb{C}^2 \to \mathbb{C}$. We can represent $\boldsymbol{p}$ both as a rational function and as a polynomial. We call the 1-level set of $\boldsymbol{p}$, the *characteristic surface* of Y and denote it by $\mathcal{H}$; see (17). The 1-level $\mathcal{H}$ is, more precisely, a one-dimensional complex affine algebraic variety of degree 3. Each point on the characteristic surface $\mathcal{H}$ defines a log-linear function (see Proposition 2) that satisfies the interior harmonicity condition of Y (i.e. defines a harmonic function of the completely unconstrained version of Y); similarly, each boundary of the state space of Y has an associated characteristic polynomial and surface. We can write $\boldsymbol{p}$ as a second-order polynomial in each of its arguments; this implies that most points on $\mathcal{H}$ come in conjugate pairs. The keystone of the approach developed in Section 3 is the following observation: log-linear functions defined by two conjugate points on $\mathcal{H}$ can be linearly combined to obtain nontrivial functions which satisfy the corresponding boundary harmonicity condition (as well as the interior one); see Figure 2 and Proposition 4. In Subsection 3.3 we present a graph representation of the constructed Y-harmonic functions.

There is a direct connection between our computations and the Balayage operator in [46]; we point out this connection in Subsection 3.4, Remark 2. In Section 5 we provide a numerical example. In the conclusion (Section 7) we discuss several directions for future research.

Among these is the application of our approach to constrained diffusion processes and the associated elliptic equations with Neumann boundary conditions (Subsection 7.2).

There is a wide literature on the approximation of p_n and similar quantities. Glasserman and Kou [23] and Ignatiouk-Robert [27] computed the large deviation limit of $p_n(x)$ for $x = (1, 0)$ as

$$\lim_{n\to\infty} -\frac{1}{n} \log p_n((1,0)) = \min(-\log \rho_1, -\log \rho_2),$$

where $\rho_i = \lambda/\mu_i$. Since p_n is a small probability, i.e. the probability of a rare event, a natural idea is to use importance sampling (IS) to approximate it via simulation. To the best of the author's knowledge, Parekh and Walrand [45] were the first to study the optimal IS simulation of the two tandem walk model for the boundary ∂A_n; it was observed that static changes of measure implied by optimal large deviation sample paths may not lead to optimal IS changes of measure because of the constraining boundaries of the process. The authors introduced boundary layers to the problem and allowed the change of measure to depend on whether the process is in these layers. Glasserman and Kou [23] observed that a simple change of measure implied by LD analysis (exchange the arrival rate with the smaller of the service rates) can perform poorly for the exit boundary $\partial A_n = \{x : x(1) + x(2) = n\}$ for a range of parameter values. An asymptotically optimal change of measure for this boundary was developed by Dupius *et al.* [18] using a subsolution of a limit HJB equation; similar to the heuristic constructions in [45], the change of measure developed in [18] is dynamic, i.e. it depends on the position of the process X; see [16], [17], [52], and [54] for studies in higher dimensions, more general dynamics, and different exit boundaries using the subsolution approach. Let $\boldsymbol{\tau}_0 \doteq \inf\{k > 0 : X_k = 0\}$, i.e. $\boldsymbol{\tau}_0$ is the first *return* time to the origin. McDonald [39] proposed an alternative approximation approach to probabilities of the type $\mathbb{P}_0(\tau_n < \boldsymbol{\tau}_0)$ for a class of models under a number of assumptions; the approximation idea in [39] is to replace τ_0 with $\tau_\triangle$, and the initial position $0 \in \mathbb{Z}_+^d$ with a random initial point on $\triangle$ with distribution $\pi_\triangle$, where $\triangle$ are the constraining boundaries corresponding to a set of 'nonsuper-stable' nodes, $\tau_\triangle$ is the first nonzero time when one of these nodes becomes empty, and $\pi_\triangle$ is the stationary measure of the underlying process conditioned on $\triangle$; McDonald [39] and its approach are further reviewed in Section 6. Literature on the analysis and simulation of rare events of constrained random walks in particular, and on the analysis of constrained random walks in general, is vast; see, e.g. [1], [3]–[6], [8]–[11], [13], [14], [16], [17], [20], [21], [25], [28]–[30], [32], [33], [35]–[45], [47], [53], [54], [57], see also [2, Chapter VI], [26, Chapter 11], and [49] for further references. In Section 6 we review a number of the works listed above in relation to the results and the techniques in this paper.

2. Derivation of the limit problem

In this section we derive the limit problem resulting from the affine transformation T_n. The derivation is simple enough and we therefore state it for a more general setup: for the purposes of the present section we assume X to be the embedded random walk of a d-dimensional stable Jackson network; let, as before, I_k denote the unconstrained i.i.d. increments of X. Define

$$\mathcal{I}_1 \in \mathbb{R}^{d\times d}, \qquad \mathcal{I}_1(j,k) = 0, \quad j \neq k, \qquad \mathcal{I}_1(j,j) = 1, \quad j \neq 1, \qquad \mathcal{I}_1(1,1) = -1,$$

where $\mathcal{I}_1$ is the identity operator on $\mathbb{R}^d$ except that its first diagonal term is -1 rather than 1. The affine change of the coordinate map is

$$T_n = ne_1 + \mathcal{I}_1,$$

https://doi.org/10.1017/jpr.2018.60 Published online by Cambridge University Press

where $e_1 \doteq (1, 0, 0, \ldots, 0) \in \mathbb{R}^d$. Define the sequence of transformed increments

$$J_k \doteq \mathcal{I}_1(I_k).$$

The domain of the limit Y process is $\Omega_Y = \mathbb{Z} \times \mathbb{Z}_+^{d-1}$ and the limit process has dynamics

$$Y_{k+1} = Y_k + \pi_1(Y_k, J_k),$$

where

$$\pi_1(x, v) \doteq \begin{cases} v & \text{if } x + v \in \Omega_Y, \\ 0 & \text{otherwise.} \end{cases}$$

Let $A_n = \{x \in \mathbb{Z}_+^d : x(1) + x(2) + \cdots + x(d) \le n\}$ and τ_n be the first time that X hits $\partial A_n = \{x \in \mathbb{Z}_+^d : x(1) + x(2) + \cdots + x(d) = n\}$. The limit exit boundary is $\partial B = \{y \in \Omega_Y, y(1) \ge \sum_{i=2}^d y(i)\}$ and τ is the first time that Y hits ∂B. Set $\partial_1 = \{z \in \mathbb{Z}^d : z(1) = 0\}$ with σ_1 being the first time that X hits ∂_1.

Denote by $\mathcal{X}$ the law of large numbers limit of X, i.e. the deterministic function satisfying

$$\lim_n \mathbb{P}_{x_n}\left(\sup_{k \le t_0 n}\left|\frac{X_k}{n} - \mathcal{X}_{k/n}\right| > \delta\right) = 0 \quad \text{for any } \delta > 0,\ t_0 > 0, \tag{5}$$

where $x_n \in \mathbb{Z}_+^d$ is a sequence of initial positions satisfying $x_n/n \to \chi \in \mathbb{R}_+^d$; see, e.g. [48, Proposition 9.5] or [7, Theorem 7.23]. The limit process starts from $\mathcal{X}_0 = \chi$, is piecewise affine, and takes values in $\mathbb{R}_+^d$; then $s_t \doteq \sum_{i=1}^d \mathcal{X}_t(i)$ starts from $\sum_i \chi(i)$ and is also piecewise linear and continuous (and therefore differentiable except for a finite number of points) with values in $\mathbb{R}_+$. The stability and bounded i.i.d. increments of X imply that s is strictly decreasing and

$$c_1 > -\dot{s} > c_0 > 0 \quad \text{for two constants } c_1 \text{ and } c_0. \tag{6}$$

These imply that $\mathcal{X}$ goes in finite time t_1 to $0 \in \mathbb{R}_+^d$ and remains there afterward.

Fix an initial point $y \in \Omega_Y$ for the process Y and set $x_n = T_n(y)$; it follows from the definition of T_n that

$$\frac{x_n}{n} \to e_1 \doteq (1, 0, 0, \ldots, 0) \in \mathbb{R}^d. \tag{7}$$

Proposition 1. *Let y and x_n be as above. Then*

$$\lim_{n\to\infty} \mathbb{P}_{x_n}(\tau_n < \tau_0) = \mathbb{P}_y(\tau < \infty).$$

Proof. Note that $x_n \in A_n$ for $n > y(1)$. Define

$$M_k = \max_{l \le k} Y_l(1), \qquad M_k^X = \min_{l \le k} X_l(1).$$

The process M is increasing and M_τ is the greatest that the first component of Y becomes before hitting ∂B (if this happens in finite time). The monotone convergence theorem implies that

$$\mathbb{P}_y(\tau < \infty) = \lim_{n \nearrow \infty} \mathbb{P}_y(\tau < \infty, M_\tau < n).$$

Thus,

$$\mathbb{P}_y(\tau < \infty) = \mathbb{P}_y(\tau < \infty, M_\tau < n) + P_y(\tau < \infty, M_\tau \ge n) \tag{8}$$

and the second term goes to 0 with n. Decomposing $\mathbb{P}_{x_n}(\tau_n < \tau_0)$ similarly using M^X yields

$$\mathbb{P}_{x_n}(\tau_n < \tau_0) = \mathbb{P}_{x_n}(\tau_n < \tau_0, M_{\tau_n}^X > 0) + \mathbb{P}_{x_n}(\tau_n < \tau_0, M_{\tau_n}^X = 0).$$

https://doi.org/10.1017/jpr.2018.60 Published online by Cambridge University Press

On the set $\{M^X_{\tau_n} > 0\}$, the process X cannot reach the boundary ∂_1 before τ_n; therefore, over this set:

- the events $\{\tau_n < \tau_0\}$ and $\{\tau < \infty\}$ coincide (recall that X and Y are defined on the same probability space);
- the distribution of $(T_n(X), n - M^X)$ is the same as that of (Y, M) up to time τ_n.

Therefore, this is equal to

$$\mathbb{P}_y(\tau < \infty, M_\tau < n) + \mathbb{P}_{x_n}(\tau_n < \tau_0, M^X_{\tau_n} = 0).$$

The first term on the right-hand side is equal to the first term on the right-hand side of (8). We know that the second term in (8) goes to 0 with n. Then to complete our proof, it suffices to show that

$$\lim_n \mathbb{P}_{x_n}(\tau_n < \tau_0, M^X_{\tau_n} = 0) = 0, \tag{9}$$

where $M^X_{\tau_n} = 0$ means that X has hit ∂_1 before τ_n. Then the last probability is equal to

$$\mathbb{P}_{x_n}(\sigma_1 < \tau_n < \tau_0), \tag{10}$$

which, we will now argue, goes to 0 (σ_1 is the first time X hits ∂_1); (7) implies $\mathcal{X}_0 = e_1$. Define $t^1 \doteq \inf\{t \colon \mathcal{X}_t(1) = 0\}$ and $t^0 \doteq \inf\{t \colon \mathcal{X}_t = 0 \in \mathbb{R}^d\}$. By definition, $t^1 \le t^0 < \infty$. Now choose t_0 in (5) to be equal to t^0, define $\mathcal{C}_n \doteq \{\sup_{k \le t^0 n} \in |X_k/n - \mathcal{X}_{k/n}| > \delta\}$, and partition (10) with $\mathcal{C}_n$, i.e.

$$\mathbb{P}_{x_n}(\sigma_1 < \tau_n < \tau_0) = \mathbb{P}_{x_n}(\{\sigma_1 < \tau_n < \tau_0\} \cap \mathcal{C}_n) + \mathbb{P}_{x_n}(\{\sigma_1 < \tau_n < \tau_0\} \cap \mathcal{C}_n^c). \tag{11}$$

The first of these goes to 0 by (5). The event in the second term behaves as follows: X remains at most $n\delta$ distance away from $n\mathcal{X}$ until its nt^0th step, hits ∂_1 then ∂A_n and then 0. These and (6) imply that, for large enough n, any sample path lying in this event can hit ∂A_n only after time nt^0. Thus, the second probability on the right-hand side of (11) is bounded above by

$$\mathbb{P}_{x_n}(\{nt^0 < \tau_n < \tau_0\} \cap \mathcal{C}_n^c).$$

The Markov property of X, $\{\sigma_1 < \tau_n < \tau_0\} \subset \{\tau_n < \tau_0\}$, and (5) imply that the last probability is less than

$$\sum_{x \colon |x| \le n\delta} \mathbb{P}_x(\tau_n < \tau_0)\mathbb{P}_{x_n}(X_{nt^0} = x).$$

For $|x| \le n\delta$, the probability $\mathbb{P}_x(\tau_n < \tau_0)$ decays exponentially in n (see [23, Theorem 2.3]); then, the above sum goes to 0. This establishes (9) and completes the proof. □

3. Analysis of the limit problem

In this section and the rest of the paper we will focus on the two tandem queue process and its limit defined in Section 1. The analysis in the previous section suggests that we approximate

$$\mathbb{P}_x(\tau_n < \tau_0)$$

with $\mathbb{P}_{T_n(x)}(\tau < \infty) = \mathbb{E}_{T_n(x)}[\mathbf{1}_{\{\tau < \infty\}}]$.

https://doi.org/10.1017/jpr.2018.60 Published online by Cambridge University Press

The goal of this section is to develop a framework in which we will derive the following explicit formula for $\mathbb{P}_y(\tau < \infty)$:

$$\mathbb{P}_y(\tau < \infty) = \left(\rho_2^{y(1)-y(2)} - \frac{\mu_2 - \lambda}{\mu_2 - \mu_1}\rho_2^{y(1)-y(2)}\rho_1^{y(2)}\right) + \frac{\mu_2 - \lambda}{\mu_2 - \mu_1}\rho_1^{y(1)},$$

$y \in \mathbb{Z}_+^2$, $y(1) \geq y(2)$ (recall that we have assumed $\mu_1 \neq \mu_2$; for the $\mu_1 = \mu_2$ case, see Subsection 7.1); the proof of this equation is the final result (Proposition 7) of this section.

It follows from the Markov property of Y that $y \mapsto \mathbb{P}_y(\tau < \infty)$ is a harmonic function of Y (or Y-harmonic), i.e. it satisfies

$$V(y) = \mathbb{E}_y[V(Y_1)] = \sum_{v \in \mathcal{V}} V(y + \pi_1(y, v))p(v), \qquad y \in B, \tag{12}$$

where

$$\mathcal{V} \doteq \{(-1, 0), (1, 1), (0, -1)\},$$
$$\pi_1(x, v) \doteq \begin{cases} v & \text{if } x + v \in \mathbb{Z} \times \mathbb{Z}_+, \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

Then $\mathbb{P}_y(\tau < \infty) = 1$ for $y \in \partial B$ implies that $y \mapsto \mathbb{P}_y(\tau < \infty)$ also satisfies the boundary condition

$$V|_{\partial B} = 1. \tag{14}$$

A Y-harmonic function h is said to be ∂B-determined if it is of the form

$$h(y) = \mathbb{E}[f(Y_\tau)\mathbf{1}_{\{\tau<\infty\}}], \quad y \in \mathbb{Z} \times \mathbb{Z}_+, \qquad y(1) \geq y(2).$$

By definition, $y \mapsto \mathbb{P}_y(\tau < \infty)$ is ∂B-determined. Then $y \mapsto \mathbb{P}_y(\tau < \infty)$ is the unique ∂B-determined solution of (12) and (14).

Let Z denote the ordinary unconstrained random walk on $\mathbb{Z}^2$ with the same increments as Y. The unconstrained version of (12) is

$$V(z) = \mathbb{E}_z[V(Z_1)] = \sum_{v \in \mathcal{V}} V(z + v)p(v), \qquad z \in \mathbb{Z}^2. \tag{15}$$

A function is said to be a harmonic function of the unconstrained random walk Z if it satisfies (15).

Our approach to solving (12) and (14) (and, hence, obtaining a formula for $\mathbb{P}_y(\tau < \infty)$) is as follows.

1. Construct a class $\mathcal{F}_Y$ of 'simple' harmonic functions for the process Y (a class of solutions to (12)), i.e.
   (a) construct a class $\mathcal{F}_Z$ of harmonic functions for the unconstrained process Z of the form $z \mapsto \beta^{z(1)-z(2)}\alpha^{z(2)}$, $(\beta, \alpha) \in \mathbb{C}^2$;
   (b) use linear combinations of elements of $\mathcal{F}_Z$ to find solutions to (12).
2. Represent the boundary condition (14) by linear combinations of the boundary values of the ∂B-determined members of the class $\mathcal{F}_Y$.

The definition of the class $\mathcal{F}_Z$ is as in (20) and that of $\mathcal{F}_Y$ is as in (29).

https://doi.org/10.1017/jpr.2018.60 Published online by Cambridge University Press

We first remark on uniqueness. We have assumed that X is stable. This implies that $Y_{\tau\wedge k}$, $k = 1, 2, 3, \ldots$, is unstable and, therefore, the Martin boundary of this process has points at ∞. Then, one cannot expect all harmonic functions of Y to be ∂B-determined and the system in (12) and (14) will therefore not have a unique solution. In particular, the constant function $j(y) = 1$ solves this system but, as we will see below, j is not ∂B-determined. Hence, once we obtain a solution to (12) and (14) that we believe to be equal to $\mathbb{P}_y(\tau < \infty)$, we will then have to prove that it is ∂B-determined.

3.1. Characteristic polynomial and surface

Denote

$$\boldsymbol{p}(\beta, \alpha) \doteq \sum_{v\in\mathcal{V}} p(v)\beta^{v(1)-v(2)}\alpha^{v(2)} = \lambda\frac{1}{\beta} + \mu_1\alpha + \mu_2\frac{\beta}{\alpha}, \qquad (\beta, \alpha) \in \mathbb{C}^2, \tag{16}$$

the interior *characteristic polynomial* of the process Y,

$$\boldsymbol{p}(\beta, \alpha) = 1$$

the interior *characteristic equation* of Y, and

$$\mathcal{H} \doteq \{(\beta, \alpha)\colon \boldsymbol{p}(\beta, \alpha) = 1\} \tag{17}$$

the interior *characteristic surface* of Y. We borrow the adjective ‘characteristic’ from the classical theory of linear ordinary differential equations; the development below parallels that theory. Note that $\boldsymbol{p}$ is a rational function not a polynomial, but it obviously becomes polynomial in α (respectively, β) when multiplied by β (respectively, α) or a polynomial in β and α when multiplied by $\beta\alpha$; these polynomial representations are useful when we solve $\boldsymbol{p}(\beta, \alpha) = 1$, but the rational representation is simpler. For this reason, we use the rational representation whenever possible and switch to the polynomial representations when needed.

In Figure 2 we present a representation of the real section of the characteristic surface of the walk for $\lambda = 0.1$, $\mu_1 = 0.5$, and $\mu_2 = 0.4$. The characteristic surface $\mathcal{H}$ is an affine algebraic curve of degree 3; see [24, Definition 8.1, p. 32]. The characteristic equation $\boldsymbol{p} = 1$ becomes a quadratic equation in α when it is multiplied by α; the discriminant of this quadratic equation is

$$\Delta(\beta) = \left(\frac{\lambda}{\beta} - 1\right)^2 - 4\mu_1\mu_2\beta.$$

Therefore, for $\beta \in \mathbb{C}$, $\Delta(\beta) \neq 0$, and $\beta \neq 0$, points on $\mathcal{H}$ come in conjugate pairs (β, α_1) and (β, α_2) satisfying

$$\alpha_i = \frac{1}{\alpha_{3-i}}\frac{\mu_2\beta}{\mu_1}, \qquad i \in \{1, 2\}. \tag{18}$$

These conjugate pairs are central to the construction of Y-harmonic functions in Subsection 3.2.2 below.

Any point on $\mathcal{H}$ defines a harmonic function of Z as we now show.

Proposition 2. *For any $(\beta, \alpha) \in \mathcal{H}$, $z \mapsto \beta^{z(1)-z(2)}\alpha^{z(2)}$, $z \in \mathbb{Z}^2$, is a harmonic function of Z; in particular, it satisfies (12) for $y \in \mathbb{Z}^2$, $y(1), y(2) > 0$.*

Proof. Condition Z on its first step and use $\boldsymbol{p}(\beta, \alpha) = 1$. □

https://doi.org/10.1017/jpr.2018.60 Published online by Cambridge University Press

For $(\beta, \alpha) \in \mathbb{C}^2$, define

$$[(\beta, \alpha), \cdot]\colon \mathbb{Z}^2 \mapsto \mathbb{C}, \qquad [(\beta, \alpha), z] \doteq \beta^{z(1)-z(2)}\alpha^{z(2)}. \tag{19}$$

The last proposition yields the class of harmonic functions

$$\mathcal{F}_Z \doteq \{[(\beta, \alpha), \cdot], (\beta, \alpha) \in \mathcal{H}\} \quad \text{for } Z. \tag{20}$$

3.2. log-linear harmonic functions of Y

Define $B^o \doteq \{y \in \mathbb{Z}_+^2,\ y(1) > y(2)\}$. We write (12) separately for the boundary ∂_2 and the interior $B^o - \partial_2$, i.e.

$$V(y) = \sum_{v \in \mathcal{V}} V(y+v)p(v), \qquad y \in B^o - \partial_2, \tag{21}$$

$$V(y) = V(y)\mu_2 + \sum_{v \in \mathcal{V},\, v(2) \neq -1} V(y+v)p(v), \qquad y \in \partial_2 \cap B^o. \tag{22}$$

Any $g \in \mathcal{F}_Z$ satisfies (21) (since (21) is the restriction of (15) to $B^o - \partial_2$); (21) is linear and so any finite linear combination of members of $\mathcal{F}_Z$ continues to satisfy (21). In the next two subsections we will show that appropriate linear combinations of members of $\mathcal{F}_Z$ will also satisfy the boundary condition (22), and define harmonic functions of the constrained process Y.

3.2.1. *A Y-harmonic function defined by a single point on $\mathcal{H}$.* Recall that members of $\mathcal{F}_Z$ are of the form $[(\beta, \alpha), \cdot]\colon z \to \beta^{z(1)-z(2)}\alpha^{z(2)}$ and $(\beta, \alpha) \in \mathcal{H}$; these define harmonic functions for Z and they therefore satisfy (21). The simplest way to construct a Y-harmonic function is to look for $[(\beta, \alpha), \cdot]$ which satisfies (12), i.e. which satisfies (21) and (22) at the same time. Substituting $[(\beta, \alpha), \cdot]$ into (22), we see that it is a solution to (22) if and only if $(\beta, \alpha) \in \mathcal{H}$ also satisfies

$$\boldsymbol{p}_2(\beta, \alpha) = 1, \tag{23}$$

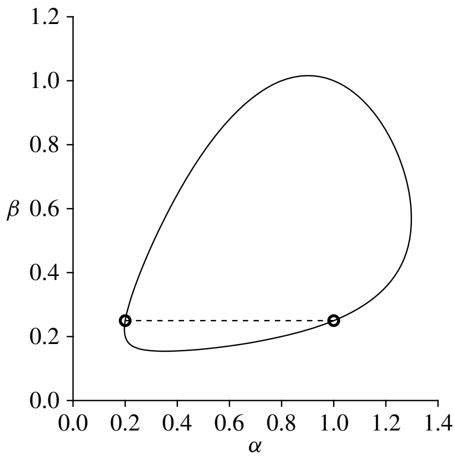

FIGURE 2: The real section of the characteristic surface $\mathcal{H}$ for $\lambda = 0.1$, $\mu_1 = 0.5$, and $\mu_2 = 0.4$; the end points of the dashed line are an example of a pair of conjugate points (β, α_1) and (β, α_2); together they define the Y-harmonic function $h_\beta(y) = \beta^{y(1)-y(2)}(C(\beta, \alpha_2)\alpha_1^{y(2)} - C(\beta, \alpha_1)\alpha_2^{y(2)})$; see Proposition 4. Each horizontal line intersecting the curve $\mathcal{H}$ twice gives a pair of conjugate points defining a Y-harmonic function.

https://doi.org/10.1017/jpr.2018.60 Published online by Cambridge University Press

where
$$\boldsymbol{p}_2(\beta,\alpha) \doteq \sum_{v\in\mathcal{V},\, v(2)\neq -1} p(v)\beta^{v(1)-v(2)}\alpha^{v(2)} + \mu_2 = \lambda\frac{1}{\beta} + \mu_1\alpha + \mu_2,$$
noting that
$$\boldsymbol{p}_2(\beta,\alpha) = \boldsymbol{p}(\beta,\alpha) - \mu_2\left(\frac{\beta}{\alpha} - 1\right). \tag{24}$$

We call (23) 'the characteristic equation of Y on ∂_2' and $\boldsymbol{p}_2$ its characteristic polynomial on the same boundary. Define the boundary characteristic surface of Y for ∂_2 as
$$\mathcal{H}_2 \doteq \{(\beta,\alpha)\in\mathbb{C}^2 \colon \boldsymbol{p}_2(\beta,\alpha) = 1\}.$$

For $[(\beta,\alpha),\cdot]$ to be Y-harmonic, (β,α) must lie on
$$\mathcal{H}\cap\mathcal{H}_2 = \{(0,0),(1,1),(\rho_1,\rho_1)\}\subset\mathbb{C}^2;$$
the last of these points yields our first nontrivial Y-harmonic function as we now show.

Proposition 3. *The function*
$$[(\rho_1,\rho_1),\cdot]\colon y\mapsto \rho_1^{y(1)-y(2)}\rho_1^{y(2)} = \rho_1^{y(1)} \quad \textit{is } Y\textit{-harmonic.}$$

Proof. The fact that $[(\rho_1,\rho_1),\cdot]$ satisfies (21) follows from the Markov property of Y and $(\rho_1,\rho_1)\in\mathcal{H}$, and that $[(\rho_1,\rho_1),\cdot]$ satisfies (22) follows from the Markov property of Y and $(\rho_1,\rho_1)\in\mathcal{H}_2$. □

3.2.2. *Y-harmonic functions via conjugate points.* Define the boundary operator D_2 acting on functions on $\mathbb{Z}^2$ and giving functions on ∂_2, i.e.
$$D_2V = g, \qquad V\colon\mathbb{Z}^2\to\mathbb{C},$$
$$g(y,0) \doteq (\mu_2 + \lambda V(y-1,0) + \mu_1 V(y+1,1)) - V(y,0), \qquad y\in\mathbb{Z};$$
D_2 is the difference between the left- and the right-hand sides of (22) and indicates how much V deviates from being Y-harmonic along the boundary ∂_2 as we now demonstrate.

Lemma 1. *It holds that $D_2V = 0$ if and only if V is Y-harmonic on ∂_2.*

The proof follows from the definitions involved. For $(\beta,\alpha)\in\mathbb{C}^2$ and $\beta,\alpha\neq 0$,
$$[D_2([(\beta,\alpha),\cdot])](y,0) = (\boldsymbol{p}_2(\beta,\alpha) - 1)\beta^{y},$$
where the left-hand side denotes the value of the function $D_2([(\beta,\alpha),\cdot])$ at $(y,0)$, $y\in\mathbb{Z}$. By definition, $\boldsymbol{p}(\beta,\alpha) = 1$ for $(\beta,\alpha)\in\mathcal{H}$; this and the last display and (24) imply that
$$[D_2([(\beta,\alpha),\cdot])](y,0) = \mu_2\left(1 - \frac{\beta}{\alpha}\right)\beta^{y} \tag{25}$$
if $(\beta,\alpha)\in\mathcal{H}$. One can write the function $(y,0)\mapsto\beta^{y}$ as $[(\beta,\alpha),\cdot]|_{\partial_2} = [(\beta,1),\cdot]|_{\partial_2}$; in addition, define
$$C(\beta,\alpha) \doteq \mu_2\left(1 - \frac{\beta}{\alpha}\right). \tag{26}$$
With these, (25) can be written as
$$D_2([(\beta,\alpha),\cdot]) = C(\beta,\alpha)[(\beta,1),\cdot]|_{\partial_2}. \tag{27}$$

https://doi.org/10.1017/jpr.2018.60 Published online by Cambridge University Press

The key observation here is that $D_2([(\beta, \alpha), \cdot])$ is a constant multiple of $[(\beta, 1), \cdot]|_{\partial_2}$. This and the linearity of D_2 imply that for $\alpha_1 \neq \alpha_2$,

$$(\beta, \alpha_1), (\beta, \alpha_2) \in \mathcal{H},$$

i.e. when (β, α_1) and (β, α_2) are conjugate points on $\mathcal{H}$, $[(\beta, \alpha_1), \cdot]$ and $[(\beta, \alpha_2), \cdot]$ can be linearly combined to cancel out each other's value under D_2. In the next proposition, we use these conjugate pairs and the above argument to obtain new Y-harmonic functions.

Proposition 4. *Assume that $\beta \in \mathbb{C}$, and $\beta \neq 0$ satisfies $\Delta(\beta) \neq 0$. Then*

$$h_\beta \doteq C(\beta, \alpha_2)[(\beta, \alpha_1), \cdot] - C(\beta, \alpha_1)[(\beta, \alpha_2), \cdot] \quad \textit{is } Y\textit{-harmonic.} \tag{28}$$

Proof. By assumption, (β, α_1) and (β, α_2) are both on $\mathcal{H}$ and, therefore, $[(\beta, \alpha_1), \cdot]$ and $[(\beta, \alpha_2), \cdot]$ are harmonic functions of Z. In particular, they both satisfy (21). Then their linear combination h_β also satisfies (21), since (21) is linear in V. It remains to show that h_β is also a solution to (22). Note that $\beta \neq 0$ implies $\alpha_1, \alpha_2 \neq 0, 1$. Then (27) implies that

$$\begin{aligned} D_2(h_\beta) &= C(\beta, \alpha_2) D_2([(\beta, \alpha_1), \cdot]) - C(\beta, \alpha_1) D_2([\beta, \alpha_2, \cdot]) \\ &= C(\beta, \alpha_2) C(\beta, \alpha_1)[(\beta, 1), \cdot]|_{\partial_2} - C(\beta, \alpha_1) C(\beta, \alpha_2)[(\beta, 1), \cdot]|_{\partial_2} \\ &= 0 \end{aligned}$$

and Lemma 1 implies that h_β satisfies (22). □

The function $y \mapsto \mathbb{P}_y(\tau < \infty)$ takes value 1 on ∂B. For this reason, the conjugate pair on $\mathcal{H}$ most relevant to the computation of $\mathbb{P}_y(\tau < \infty)$ consists of $(\rho_2, 1)$ and (ρ_2, ρ_1); we illustrate this pair in Figure 2. Then h_{ρ_2}, the Y-harmonic function defined by this pair, is equal to

$$h_{\rho_2}(y) = C(\rho_2, \rho_1)[(\rho_2, 1), y] - C(\rho_2, 1)[(\rho_2, \rho_1), y],$$

which, by definitions (19) and (26), is equal to

$$(\mu_2 - \mu_1)\rho_2^{y(1)-y(2)} - (\mu_2 - \lambda)\rho_2^{y(1)-y(2)}\rho_1^{y(2)}.$$

Note that the first term in the definition (2) of W is equal to $(1/(\mu_2 - \mu_1))h_{\rho_2}$.

With Proposition 4, we define our basic class of harmonic functions of Y as

$$\mathcal{F}_Y \doteq \{h_\beta, \beta \neq 0, \Delta(\beta) \neq 0\}. \tag{29}$$

Members of $\mathcal{F}_Y$ consist of linear combinations of log-linear functions; with a slight abuse of language, we will also refer to such functions as log-linear.

Remark 1. For the purposes of computing $\mathbb{P}_y(\tau < \infty)$ for the tandem network case treated here, a single member of $\mathcal{H}_Y$ will suffice, i.e. h_{ρ_2}; see Proposition 7. But $\mathcal{H}_Y$ is a family of simple-to-compute Y-harmonic functions and they can be used to approximate other expectations or even $\mathbb{P}_y(\tau < \infty)$ when the underlying network is not tandem; see Remark 3.

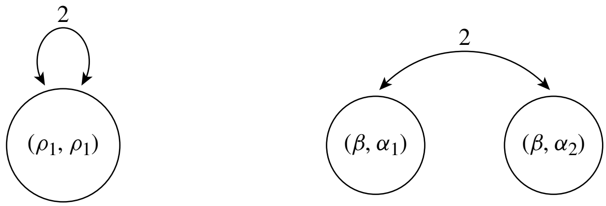

FIGURE 3: A graph representation of the Y-harmonic functions constructed in Propositions 3 and 4.

3.3. Graph representation of log-linear harmonic functions of *Y*

In Figure 3 we present a graph representation of the harmonic functions developed in the last subsection. Each node in this figure represents a member of $\mathcal{F}_Z$. The edges represent the boundary conditions; in this case there is only one ((22) of ∂_2) and the edge label '2' refers to ∂_2. A self-connected vertex represents a member of $\mathcal{F}_Z$ that also satisfies the ∂_2 boundary condition (22), i.e. $y \mapsto [(\rho_1, \rho_1), y] = \rho_1^{y(1)}$ from Proposition 3; the graph on the left represents exactly this function. The '2' labeled edge on the right represents the conjugacy relation (18) between α_1 and α_2, which allows these functions to be linearly combined to satisfy the harmonicity condition of Y on ∂_2.

We call the graphs shown in Figure 3, and the system of characteristic equations they represent, a *harmonic system*. One can also define harmonic systems for d-dimensional constrained random walks (see [55, Section 5]); these systems and their solutions play a key role in the generalization of the analysis of this section to higher dimensions.

3.4. ∂B-determined harmonic functions of *Y*

In Subsections 3.2.1 and 3.2.2 we constructed classes of Y-harmonic functions. For the purpose of computing $\mathbb{P}_y(\tau < \infty)$, $y \in B$, we need ∂B-determined Y-harmonic functions. In Proposition 5 we derive simple conditions that allow us to check whether a member of $\mathcal{F}_Y$ is ∂B-determined. In this regard, the following fact are useful.

Lemma 2. *Define*

$$\zeta_n \doteq \inf\{k : Y_k(1) = Y_k(2) + n\}.$$

For $y \in \mathbb{Z}_+^2$, $0 \le y(1) - y(2) \le n$,

$$\mathbb{P}_y(\zeta_n \wedge \zeta_0 = \infty) = 0.$$

Proof. The proof follows from the fact that when in $C = \{y \in \mathbb{Z}_+^2,\ y(2) \le y(1) \le y(2)+n\}$, the process Y hits $\partial C = \{y \in \mathbb{Z}_+^2 : y(1) - y(2) = n \text{ or } (1) = y(2)\}$ in at most n steps with probability greater than λ^n. For a detailed version of this argument, we refer the reader to the proof of [55, Proposition 2.2]. □

Proposition 5. *Let* α_1, α_2, *and* β *be as in Proposition 4. If*

$$|\beta| < 1, \qquad |\alpha_1|, |\alpha_2| \le 1 \tag{30}$$

then h_β *of (28) is* ∂B*-determined.*

Proof. By Proposition 4, h_β is Y-harmonic; (30) and its definition (28) imply that h_β is also bounded on B^o. Then $M_k = h_\beta(Y_{\tau\wedge\zeta_n\wedge k})$ is a bounded martingale. This, Proposition 2, and the optional sampling theorem imply that

$$h_\beta(y) = \mathbb{E}_y[h_\beta(Y_\tau)\mathbf{1}_{\{\tau<\zeta_n\}}] + \mathbb{E}_y[h_\beta(Y_{\zeta_n})\mathbf{1}_{\{\zeta_n\le\tau\}}], \qquad y \in B^o, \tag{31}$$

https://doi.org/10.1017/jpr.2018.60 Published online by Cambridge University Press

with $Y_{\zeta_n}(1) = n$ for $\tau > \zeta_n$. This and (30) imply that

$$\lim_{n\to\infty} \mathbb{E}_y[h_\beta(Y_{\zeta_n})\mathbf{1}_{\{\zeta_n\le\tau\}}] \le \lim_{n\to\infty} \beta^n = 0.$$

Thus, $\lim_n \zeta_n = \infty$ and letting $n \to \infty$ in (31) yields

$$h_\beta(y) = \mathbb{E}_y[h_\beta(Y_\tau)\mathbf{1}_{\{\tau<\infty\}}],$$

i.e. h_β is ∂B-determined. □

In addition, we have the following result.

Proposition 6. *The Y-harmonic function $[(\rho_1, \rho_1), \cdot]$ of Proposition 3 is ∂B-determined.*

Proof. The proof is identical to that of Proposition 5 and follows from $0 \le [(\rho_1, \rho_1), y] \le 1$ for $y \in B$ and the Y-harmonicity of $[(\rho_1, \rho_1), \cdot]$. □

Proposition 5 rests on condition (30); we refer the reader to [55, Section 4], in particular to Proposition 4.13 in which the conditions for (30) to hold in the context of general two-node Jackson networks were derived. For the purpose of computing $y \mapsto \mathbb{P}_y(\tau < \infty)$, we need only to consider the point (ρ_1, ρ_1) and the conjugate pair $(\rho_2, 1)$ and (ρ_2, ρ_1); it is trivial to check the conditions in (30) for these points. This brings us to the main result of this section.

Proposition 7. *Under the stability assumption $\lambda < \mu_1, \mu_2$, h_{ρ_2} is ∂B-determined and we have*

$$\mathbb{P}_y(\tau < \infty) = W(y) = \frac{1}{C(\rho_2, \rho_1)} h_{\rho_2}(y) + \frac{C(\rho_2, 1)}{C(\rho_2, \rho_1)}[(\rho_1, \rho_1), y], \qquad y \in B.$$

Definitions (19) and (26) yield the following expanded formula for W:

$$\begin{aligned} W(y) &= \frac{1}{C(\rho_2, \rho_1)} h_{\rho_2}(y) + \frac{C(\rho_2, 1)}{C(\rho_2, \rho_1)}[(\rho_1, \rho_1), y] \\ &= \left(\rho_2^{y(1)-y(2)} + \frac{\mu_2 - \lambda}{\mu_1 - \mu_2}\rho_2^{y(1)-y(2)}\rho_1^{y(2)}\right) + \frac{\mu_2 - \lambda}{\mu_2 - \mu_1}\rho_1^{y(1)}, \end{aligned}$$

which is the one given in (2).

Proof. The conjugate points on $\mathcal{H}$ for $\beta = \rho_2$ are $(\rho_2, 1)$ and (ρ_2, ρ_1); the stability assumption $\lambda < \mu_1, \mu_2$ implies that both these points satisfy (30). It follows from Propositions 4 and 5 that h_{ρ_2} is a ∂B-determined Y-harmonic function; similarly, it follows from Propositions 3 and 6 that $[(\rho_1, \rho_1), \cdot]$ is a ∂B-determined Y-harmonic function. It follows that their linear combination W is also ∂B-determined and Y-harmonic, i.e.

$$W(y) = \mathbb{E}_y[\mathbf{1}_{\{\tau<\infty\}} W(Y_\tau)].$$

However, $W(y) = 1$ on ∂B; therefore, this is equal to

$$\mathbb{P}_y(\tau < \infty).$$ □

Remark 2. The Balayage operator $\mathcal{T}$ (see [46, p. 25]) for the set ∂B is the operator mapping a function f on ∂B to the Y-harmonic function g on B, defined as:

$$\mathcal{T} : f \to g, \qquad g(x) = \mathbb{E}_x[f(X_\tau)\mathbf{1}_{\{\tau<\infty\}}].$$

https://doi.org/10.1017/jpr.2018.60 Published online by Cambridge University Press

Therefore, by definition, a Y-harmonic function h is ∂B-determined if and only if it is the image of some function under the Balayage operator $\mathcal{T}$. Computing $\mathbb{P}_y(\tau < \infty)$ amounts to computing the image of the constant function j on ∂B under the Balayage operator. Propositions 3–6 provide us with a collection of *basis functions* for which the Balayage operator $\mathcal{T}$ is very simple to compute; these functions play the same role in the current problem as exponential functions play in the solution of linear ordinary differential equations or the trigonometric functions in the solution of the heat and the Laplace equations. We now write Proposition 5 more explicitly. Suppose that α_1, α_2, and β are as in Proposition 5; recall that

$$h_\beta(y) = \beta^{y(1)-y(2)}(C(\beta,\alpha_2)\alpha_1^{y(2)} - C(\beta,\alpha_1)\alpha_2^{y(2)}), \qquad y \in \mathbb{Z}^2.$$

Then, from Proposition 5 we see that

$$\mathbb{E}_y[h_\beta(Y_\tau)\mathbf{1}_{\{\tau<\infty\}}] = h_\beta(y), \tag{32}$$

i.e. $\mathcal{T}(h_\beta|_{\partial B}) = h_\beta$.

Remark 3. We are interested in the computation of $\mathbb{P}_y(\tau < \infty) = \mathbb{E}_y[\mathbf{1}_{\tau<\infty}]$. More generally we may be interested in computing $g(y) = \mathbb{E}_y[f(Y_\tau)\mathbf{1}_{\{\tau<\infty\}}]$ for some function f. To approximate this expectation, one can proceed as follows. First, approximate f with a finite superposition of the form

$$f^* = \sum_{i=1}^{K} w_i f_i|_{\partial B},$$

where $w_i \in \mathbb{C}$ and $f_i \in \mathcal{F}_{\mathcal{Y}}$, i.e. a Y-harmonic function of the form

$$f_i = C(\beta_i,\alpha_i^*)[(\beta_i,\alpha_i),\cdot] - C(\beta_,\alpha_i)[(\beta_i,\alpha_i^*),\cdot],$$

and $|\beta_i|$, $|\alpha_i|$, $|\alpha_i^*| < 1$; then, by (32),

$$\mathbb{E}_y[f^*(Y_\tau)\mathbf{1}_{\{\tau<\infty\}}] = \sum_{i=1}^{K} w_i f_i(y)$$

would lead to an approximation of $\mathbb{E}_y[f(Y_\tau)\mathbf{1}_{\{\tau<\infty\}}]$ for $y \in B$. The error generated by this approximation is bounded by $\max_{y\in\partial B}|f^*(y) - f(y)|$.

4. Convergence–initial condition set for X

In the convergence argument of Section 2 we used an initial point for the Y process. The goal of this section is to provide a convergence argument starting from an initial position specified for the X process as $X(0) = \lfloor nx \rfloor$ for a fixed $x \in \mathbb{R}_+^2$ with $x(1) + x(2) < 1$, as is performed in LD analysis. We will show that the relative error

$$\frac{|\mathbb{P}_{x_n}(\tau_n < \tau_0) - \mathbb{P}_{T(x_n)}(\tau < \infty)|}{\mathbb{P}_{x_n}(\tau_n < \tau_0)}$$

decays *exponentially* in n; see Proposition 8 below.

For the current analysis, we will also use the limit process Y expressed in the original coordinates of the X process, which is $\bar{X} \doteq T_n(Y)$. Process $\bar{X}$ is the same process as X except that it is constrained only at the boundary ∂_2, i.e.

$$X_{k+1} = X_k + \pi(X_k, I_k), \qquad \bar{X}_{k+1} = \bar{X}_k + \pi_1(\bar{X}_k, I_k),$$

where π_1 is as in (13). We will assume that X and $\bar{X}$ start from the same initial position $X_0 = \bar{X}_0$ and whenever we specify an initial position below it is for both processes.

https://doi.org/10.1017/jpr.2018.60 Published online by Cambridge University Press

As before, $\tau_n = \inf\{k \geq 0\colon X_1(k) + X_2(k) = \partial A_n\}$ and $\tau = \inf\{k \geq 0\colon Y_k \in \partial B\}$; define $\partial \bar{A}_n \doteq \{x \in \mathbb{Z} \times \mathbb{Z}_+\colon x(1) + x(2) = n\}$. By definition, $\bar{X}$ hits $\partial \bar{A}_n$ exactly when Y hits ∂B; therefore,

$$\tau = \bar{\tau}_n \doteq \inf\{k \geq 0\colon \bar{X}_k \in \partial \bar{A}_n\} \tag{33}$$

and $\mathbb{P}_{T(x_n)}(\tau < \infty) = \mathbb{P}_{x_n}(\bar{\tau}_n < \infty)$.

Proposition 8. *For $x \in \mathbb{R}^2_+$, $0 < x(1) + x(2) < 1$ (and $x(1) > 0$ if $\rho_1 > \rho_2$ and $x(2) \leq 1 - log(\rho_1)/\log(\rho_2)$), set $x_n \doteq \lfloor nx \rfloor$. Then*

$$\frac{|\mathbb{P}_{x_n}(\tau_n < \tau_0) - \mathbb{P}_{T(x_n)}(\tau < \infty)|}{\mathbb{P}_{x_n}(\tau_n < \tau_0)} = \frac{|\mathbb{P}_{x_n}(\tau_n < \tau_0) - \mathbb{P}_{x_n}(\bar{\tau}_n < \infty)|}{\mathbb{P}_{x_n}(\tau_n < \tau_0)}$$

decays exponentially in n.

The proof will require several supporting results on $\sigma_1 = \inf\{k \geq 0\colon X_k \in \partial_1\}$,

$$\sigma_{1,2} \doteq \inf\{k\colon k \geq \sigma_1, X_k \in \partial_2\}, \quad \text{and} \quad \bar{\sigma}_{1,2} \doteq \inf\{k\colon k \geq \sigma_1, \bar{X}_k(1) = -\bar{X}_k(2)\}.$$

Proposition 9. *It holds that*

$$X_k(1) + X_k(2) = \bar{X}_k(1) + \bar{X}_k(2) \quad \textit{for } k \leq \sigma_{1,2}. \tag{34}$$

Proof. We see that

$$X_k = \bar{X}_k \quad \text{for } k \leq \sigma_1 \tag{35}$$

implies (34) for $k \leq \sigma_1$. If $\sigma_1 = \sigma_{1,2}$ then we are done. Otherwise, $X_{\sigma_1}(2) = \bar{X}_{\sigma_1}(2) > 0$ and $X_k(2) > 0$ for $\sigma_1 < k < \sigma_{1,2}$; let $\sigma_1 = \nu_1 < \nu_2 < \cdots < \nu_K < \sigma_{1,2}$ be the times when X hits ∂_1 before hitting ∂_2. The definitions of $\bar{X}$ and X imply that these are the only times when the increments of X and $\bar{X}$ differ: $X_{\nu_j+1} - X_{\nu_j} = 0$ and $\bar{X}_{\nu_j+1} - \bar{X}(\nu_j) = (-1, 1)$ if $I_{\nu_j} = (-1, 1)$; otherwise both differences are equal to I_{ν_j}. This and (35) imply that

$$X_k - \bar{X}_k = \varsigma_k \cdot (-1, 1) \quad \text{for } k \leq \sigma_{1,2}, \tag{36}$$

where $\varsigma_k \doteq \sum_{j=1}^{K} \mathbf{1}_{\{\nu_j \leq k\}} \mathbf{1}_{\{I_{\nu_j} = (-1,1)\}}$ and '$\cdot$' denotes scalar multiplication. Summing the components of both sides of (36) yields (34). □

Define

$$\Gamma_n \doteq \{\sigma_1 < \sigma_{1,2} < \tau_n < \tau_0\}.$$

Then Γ_n is one particular way for $\{\tau_n < \tau_0\}$ to occur. In the next proposition, we obtain an upper bound on its probability in terms of

$$\gamma \doteq -(\log(\rho_1) \vee \log(\rho_2)).$$

Proposition 10. *For any $\epsilon > 0$, there is $N > 0$ such that if $n > N$,*

$$\mathbb{P}_{x_n}(\Gamma_n) \leq \mathrm{e}^{-n(\gamma - \epsilon)},$$

where $x_n = \lfloor nx \rfloor$ and $x \in \mathbb{R}^2_+$, $x(1) + x(2) < 1$.

Proof. The proof will use the following definitions. Let $v_0 = (0, 1)$, $v_1 = (-1, 1)$, $v_2 = (0, -1)$, $p_X(v_0) = \lambda$, $p_X(v_1) = \mu_1$, $p_X(v_2) = \mu_2$, and

$$H_a(q) \doteq -\log\bigg(\sum_{i\in\{0,1,2\}-a} p_X(v_i)\mathrm{e}^{-\langle v_i, q\rangle} + \sum_{\{i\in a\}} p_X(v_i)\bigg), \qquad a \subset \{1, 2\}, \tag{37}$$

where $\langle\cdot,\cdot\rangle$ denotes the standard inner product in $\mathbb{R}^2$. For $x \in \mathbb{R}^2_+$, set

$$\boldsymbol{b}(x) \doteq \{i : x(i) = 0\}.$$

We will write H rather than $H_\varnothing$.

We now address the gradient operator on smooth functions on $\mathbb{R}^2$ with ∇. The authors of [18] and [51] used a smooth subsolution of

$$H_{\boldsymbol{b}(x)}(\nabla V(x)) = 0 \tag{38}$$

to find a lower bound on the decay rate of the second moment of IS estimators for the probability $\mathbb{P}_{x_n}(\tau_n < \tau_0)$ (note that V is said to be a subsolution of (38) if $H_{\boldsymbol{b}(x)}(\nabla V(x)) \geq 0$). The event Γ_n consists of three stages: the process first hits ∂_1, then ∂_2, and then finally hits ∂A_n without hitting 0. To handle this, we use a function $(s, x) \to V(s, x)$ with two variables; for the x variable we will substitute the scaled position of the X process, and the discrete variable $s \in \{0, 1, 2\}$ is for keeping track of which of the above three stages the process is in; V is a subsolution in the x variable and continuous in (s, x) (when (s, x) is thought of as a point on the manifold $\mathcal{M}$ consisting of three copies of $\mathbb{R}^2_+$ (one for each stage); the zeroth glued to the first along ∂_1 and the first to the second along ∂_2) and therefore one can think of V as three subsolutions (one for each stage) glued together along the boundaries of the state space of X where transitions between the stages occur. We will call a function $(s, x) \to V(s, x)$ with the above properties a subsolution of (38) on the manifold $\mathcal{M}$.

Define

$$\tilde{V}_i^\varepsilon(x) \doteq \langle \boldsymbol{r}_i, x\rangle + \gamma - (3 - i)\varepsilon, \qquad \tilde{V}^{\varepsilon,j} \doteq \bigwedge_{i=0}^{j} \tilde{V}_i^\varepsilon, \tag{39}$$

where

$$\boldsymbol{r}_0 \doteq (0, 0), \qquad \boldsymbol{r}_1 = -\gamma(1, 0), \qquad \boldsymbol{r}_2 \doteq -\gamma(1, 1).$$

The subsolution for stage j is a smoothed version of $\tilde{V}^{\varepsilon,j}$. As in [18] and [51], we will need to vary ε with n in the convergence argument; for this reason, ε will appear as the third parameter of the constructed subsolution. The details are as follows.

The subsolution for the zeroth stage is $\tilde{V}^{0,\varepsilon}$: $V(0, x, \varepsilon) \doteq \gamma - 3\varepsilon$, $\nabla V(0, \cdot) = 0$, and it trivially satisfies (38) and is therefore a subsolution.

Define the smoothing kernel

$$\eta_\delta(x) \doteq \frac{1}{\delta^2 M}\eta\left(\frac{x}{\delta}\right), \qquad \eta(x) \doteq \mathbf{1}_{\{|x|\leq 1\}}(|x|^2 - 1), \qquad M \doteq \int_{\mathbb{R}^2} \eta(x)\,\mathrm{d}x$$

To construct the subsolution for the first and the second stages, we will mollify $\tilde{V}^{j,\varepsilon}$, $j = 1, 2$, with η, i.e.

$$V(j, x, \varepsilon) \doteq \int_{\mathbb{R}^2} \tilde{V}^{j,\varepsilon}(y)\eta_{c_2\varepsilon}(x - y)\,\mathrm{d}y,$$

https://doi.org/10.1017/jpr.2018.60 Published online by Cambridge University Press

and c_2 is chosen so that

$$V(1,x,\varepsilon)=\begin{cases}V(2,x,\varepsilon) & \text{for } x\in\partial_2, & (40)\\ V(0,x,\varepsilon) & \text{for } x\in\partial_1, & (41)\end{cases}$$

(this is possible since $V(j,\varepsilon,x)\to\tilde V^{j,\varepsilon}$ as $c_2\to 0$ and all the involved functions are affine; see [51, p. 38] for details on how to compute c_2 explicitly). The fact that $V(j,\cdot,\varepsilon)$, $j=1,2$, are subsolutions follows from the concavity of H_a and the choices of the gradients $\boldsymbol{r}_i$; for details we refer the reader to [51, Lemma 2.3.2]. A direct computation yields

$$\left|\frac{\partial^2 V(j,\cdot,\varepsilon)}{\partial x_i\,\partial x_j}\right|\le\frac{c_3}{\varepsilon},\qquad j=1,2, \tag{42}$$

for a constant $c_3>0$ (again, see the proof of [51, Lemma 2.3.2] for more details of this computation).

The construction above implies that

$$V(2,x,\varepsilon)<0,\qquad x\in\{x\colon x(1)+x(2)=1\}, \tag{43}$$

completing the proof. □

Now on to the proof of Proposition 10.

Proof of Proposition 10. We see that $V(0,\cdot,\varepsilon)$ maps to a constant and, thus,

$$\langle\nabla W(x),v_i\rangle=W(x+v_i)-W(x)\quad\text{if } W=V(0,\cdot,\varepsilon). \tag{44}$$

For $W=V(j,\cdot,\varepsilon)$, $j=1,2$, Taylor's formula and (42) yield

$$\left|\left\langle\nabla W(x),\frac{1}{n}v_i\right\rangle-\left(W\left(x+\frac{1}{n}v_i\right)-W(x)\right)\right|\le\frac{c_3}{n\varepsilon}. \tag{45}$$

We will allow ε to depend on n so that $\varepsilon_n\to 0$ and $n\varepsilon_n\to\infty$. Define $S_k=\mathbf{1}_{\{\sigma_1<k\}}+\mathbf{1}_{\{\sigma_{1,2}<k\}}$, $M_0\doteq 1$, and

$$M_{k+1}\doteq M_k\exp\left(-n\left(V\left(S_{k+1},\frac{X_{k+1}}{n},\varepsilon_n\right)-V\left(S_k,\frac{X_k}{n},\varepsilon_n\right)\right)-\mathbf{1}_{\{n>\sigma_1\}}\frac{c_3}{n\varepsilon_n}\right).$$

The fact that $V(j,\cdot,\varepsilon_n)$, $j=0,1,2$, are subsolutions of (38), together with (40), (41), (44), and (45) imply that M is a supermartingale. Equations (44) and (45) allow us to replace the gradients in (37) and (38) with finite differences, and (40) and (41) preserve the supermartingale property of M as S passes from 0 to 1 and from 1 to 2. This and $M\ge 0$ imply (see [19, Theorem 7.6]) that

$$\mathbb{E}_{x_n}\left[\prod_{k=1}^{\tau_{0,n}}\exp\left(-n\left(V\left(S_{k+1},\frac{X_{k+1}}{n},\varepsilon_n\right)-V\left(S_k,\frac{X_k}{n},\varepsilon_n\right)\right)-\mathbf{1}_{\{n>\sigma_1\}}\frac{c_3}{n\varepsilon_n}\right)\right]\le 1,$$

where $\tau_{0,n}\doteq\tau_n\wedge\tau_0$. Restricting the expectation on the left to $\mathbf{1}_{\Gamma_n}$ and replacing $\mathbf{1}_{\{n>\sigma_1\}}$ with 1 to make the expectation smaller, we have

$$\mathbb{E}_{x_n}\left[\mathbf{1}_{\Gamma_n}\exp\left(-\frac{c_3}{n\varepsilon_n}\tau_{0,n}\right)\exp\left(-n\sum_{k=1}^{\tau_{0,n}}V\left(S_{k+1},\frac{X_{k+1}}{n},\varepsilon_n\right)-V\left(S_k,\frac{X_k}{n},\varepsilon_n\right)\right)\right]\le 1.$$

https://doi.org/10.1017/jpr.2018.60 Published online by Cambridge University Press

Over Γ_n, X first hits ∂_1 and then ∂_2 and finally ∂A_n. Furthermore, the sum inside the expectation is telescoping across this whole trajectory; these imply that the last inequality reduces to

$$\mathbb{E}_{x_n}\left[\mathbf{1}_{\Gamma_n}\exp\left(-\frac{c_3}{n\varepsilon_n}\tau_{0,n}\right)\exp(-n(V(2, X_{\tau_{0,n}}, \varepsilon_n) - V(0, X_0, \varepsilon_n))\right] \le 1.$$

Then $\tau_{0,n} = \tau_n$ on Γ_n and, therefore, on the same set $X_{\tau_{0,n}} \in \partial_n$. This, $V(0, \cdot, \epsilon_n) = \gamma - 3\epsilon_n$, (43), and the previous inequality yield

$$\mathbb{E}_{x_n}\left[\mathbf{1}_{\Gamma_n}\exp\left(-\frac{c_3}{n\varepsilon_n}\tau_{0,n}\right)\right] \le \mathrm{e}^{-n(\gamma-3\varepsilon_n)}. \tag{46}$$

Now suppose that the statement of Theorem 10 does not hold, i.e. there exist $\epsilon > 0$ and a sequence n_k such that

$$\mathbb{P}_{x_{n_k}}(\Gamma_{n_k}) > \mathrm{e}^{-n_k(\gamma-\epsilon)} \quad \text{for all } k. \tag{47}$$

Let us pass to this subsequence and drop the subscript k. Using [51, Theorem A.1.1] we see that we can choose $c_4 > 0$ so that $\mathbb{P}(\tau_{0,n} > nc_4) \le \mathrm{e}^{-n(\gamma+1)}$ for large n. Then

$$\begin{aligned}\mathbb{E}_{x_n}\left[\mathbf{1}_{\Gamma_n}\exp\left(-\frac{c_3}{n\varepsilon_n}\tau_{0,n}\right)\right] &\ge \mathbb{E}_{x_n}\left[\mathbf{1}_{\Gamma_n}\exp\left(-\frac{c_3}{n\varepsilon_n}\tau_{0,n}\right)\mathbf{1}_{\{\tau_{0,n}\le nc_4\}}\right]\\ &\ge \exp\left(-\frac{c_4c_3}{n\varepsilon_n}n\right)\mathbb{E}_{x_n}[\mathbf{1}_{\Gamma_n} j_{\{\tau_{0,n}\le nc_4\}}],\end{aligned}$$

$\mathbb{P}(E_1 \cap E_2) \ge \mathbb{P}(E_1) - \mathbb{P}(E_2^c)$ for any two events E_1 and E_2; this and the previous line imply that this is greater than or equal to

$$\exp\left(\frac{-c_3c_4}{n\varepsilon_n}n\right)(\mathbb{P}_{x_n}(\Gamma_n) - \mathbb{P}_{x_n}(\tau_{0,n} > nc_4)) \ge \exp\left(-\frac{c_3c_4}{n\varepsilon_n}n\right)(\mathrm{e}^{-n(\gamma-\varepsilon)} - \mathrm{e}^{-(\gamma+1)n}).$$

By assumption, $n\varepsilon_n \to \infty$, which implies that $c_3c_4/n\varepsilon_n \to 0$; this and the last inequality mean that $\mathbb{E}_{x_n}[\mathbf{1}_{\Gamma_n}\exp(-(c_3/(n\varepsilon_n))\tau_{0,n})]$ cannot decay at an exponential rate faster than $\gamma - \epsilon$, but this contradicts (46) since $\varepsilon_n \to 0$. Then, there cannot be $\epsilon > 0$ and a sequence $\{n_k\}$ for which (47) holds and this implies the statement of Proposition 10. □

Define $\boldsymbol{r}_3 \doteq \log(\rho_2)(1, 1)$ and $V(x) \doteq (-\log(\rho_1) + \langle \boldsymbol{r}_1, x\rangle) \wedge (-\log(\rho_2) + \langle \boldsymbol{r}_3, x\rangle)$ for $x \in \mathbb{R}^2$.

Proposition 11. *It holds that*

$$\lim_{n\to\infty} -\frac{1}{n}\log\mathbb{P}_{x_n}(\tau_n < \tau_0) = V(x) \quad \textit{for } x \in \mathbb{R}_+^2,\ 0 < x(1) + x(2) < 1,\ x_n = \lfloor nx \rfloor.$$

The omitted proof is a one-step version of the argument used in the proof of Proposition 10 and uses a mollification of V as the subsolution.

Proposition 12. *For any $\epsilon > 0$, there is $N > 0$ such that if $n > N$,*

$$\mathbb{P}_{x_n}(\sigma_1 < \sigma_{1,2} < \bar{\tau}_n < \infty) \le \mathrm{e}^{-n(\gamma-\epsilon)}, \tag{48}$$

where $x_n = \lfloor nx \rfloor$ and $x \in \mathbb{R}_+^2$, $x(1) + x(2) < 1$.

https://doi.org/10.1017/jpr.2018.60 Published online by Cambridge University Press

Proof. Recall that $\bar{\tau}_n = \tau$ (see (33)). Write

$$\mathbb{P}_{x_n}(\sigma_1 < \sigma_{1,2} < \bar{\tau}_n < \infty)$$
$$= \mathbb{P}_{x_n}(\sigma_1 < \sigma_{1,2} < \bar{\sigma}_{1,2} < \bar{\tau}_n < \infty) + \mathbb{P}_{x_n}(\sigma_1 < \sigma_{1,2} < \bar{\tau}_n < \bar{\sigma}_{1,2}).$$

The definitions of X and $\bar{X}$ imply that $\tau_0 \geq \bar{\sigma}_{1,2}$. Then, if a sample path ω satisfies $\sigma_1(\omega) < \sigma_{1,2}(\omega) < \bar{\tau}_n(\omega) < \bar{\sigma}_{1,2}$, it must also satisfy $\sigma_1(\omega) < \sigma_{1,2}(\omega) < \tau_n(\omega) < \tau_0(\omega)$. This and Proposition 10 imply that there is an N such that

$$\mathbb{P}_{x_n}(\sigma_1 < \sigma_{1,2} < \bar{\tau}_n < \bar{\sigma}_{1,2}) \leq \mathrm{e}^{-n(\gamma-\epsilon)} \quad \text{for } n > N.$$

On the other hand, Proposition 7 and the Markov property of $\bar{X}$ imply that

$$\mathbb{P}_{x_n}(\sigma_1 < \sigma_{1,2} < \bar{\sigma}_{1,2} < \bar{\tau}_n < \infty) \leq c_5\mathrm{e}^{-n(\gamma-\epsilon)} \quad \text{for some constant } c_5 > 0.$$

These imply (48). □

Proof of Proposition 8. Decompose $\mathbb{P}_{x_n}(\tau_n < \tau_0)$ and $\mathbb{P}_{x_n}(\bar{\tau}_n < \infty)$ as

$$\begin{aligned}\mathbb{P}_{x_n}(\tau_n < \tau_0) &= \mathbb{P}_{x_n}(\tau_n < \sigma_1 < \tau_0) + \mathbb{P}_{x_n}(\sigma_1 < \tau_n \leq \sigma_{1,2} \wedge \tau_0)\\ &\quad + \mathbb{P}_{x_n}(\sigma_1 < \sigma_{1,2} < \tau_n < \tau_0), \qquad (49)\\ \mathbb{P}_{x_n}(\bar{\tau}_n < \infty) &= \mathbb{P}_{x_n}(\bar{\tau}_n < \sigma_1) + \mathbb{P}_{x_n}(\sigma_1 < \bar{\tau}_n < \sigma_{1,2}) + \mathbb{P}_{x_n}(\sigma_1 < \sigma_{1,2} < \bar{\tau}_n < \infty). \qquad (50)\end{aligned}$$

By definition, X and $\bar{X}$ are identical until they hit ∂_1; therefore, $\{\tau_n < \sigma_1\} = \{\bar{\tau}_n < \sigma_1\}$ and

$$\mathbb{P}_{x_n}(\tau_n < \sigma_1) = \mathbb{P}_{x_n}(\bar{\tau}_n < \sigma_1). \qquad (51)$$

The processes X and $\bar{X}$ begin to differ after they hit ∂_1; but from Proposition 9 we see that the sums of their components remain equal before time $\sigma_{1,2}$. This implies $\bar{\tau}_n = \tau_n$ on $\tau_n \leq \sigma_{1,2}$ and, therefore,

$$\mathbb{P}_{x_n}(\sigma_1 < \bar{\tau}_n \leq \sigma_{1,2}) = \mathbb{P}_{x_n}(\sigma_1 < \tau_n \leq \sigma_{1,2} \wedge \tau_0).$$

This, (51), and the decompositions (49) and (50) imply that

$$|\mathbb{P}_{x_n}(\tau_n < \tau_0) - \mathbb{P}_{x_n}(\bar{\tau}_n < \infty)| = |\mathbb{P}_{x_n}(\sigma_1 < \sigma_{1,2} < \tau_n < \tau_0) - \mathbb{P}_{x_n}(\sigma_1 < \sigma_{1,2} < \bar{\tau}_n < \infty)|.$$

By Propositions 10 and 12 for $\epsilon > 0$ arbitrarily small, the right-hand side of this equality is bounded above by $\mathrm{e}^{-n(\gamma-\epsilon)}$ when n is large. On the other hand, from Proposition 11 we see that for $\epsilon_0 > 0$ arbitrarily small, $\mathbb{P}_{x_n}(\tau_n < \tau_0) \geq \mathrm{e}^{-n(\gamma_1+\epsilon_0)}$ for large n, where $\gamma_1 \doteq V(x) < \gamma$. Choose ϵ and ϵ_0 to satisfy $\gamma - \gamma_1 > \epsilon + \epsilon_0$. These imply that, for $c_6 = (\epsilon + \epsilon_0) + \gamma_1 - \gamma < 0$,

$$\frac{|\mathbb{P}_{x_n}(\tau_n < \tau_0) - \mathbb{P}_{x_n}(\bar{\tau}_n < \infty)|}{|\mathbb{P}_{x_n}(\tau_n < \tau_0)|} < \mathrm{e}^{c_6 n}$$

when n is large; our proof is complete. □

It is possible to generalize Proposition 8 in many directions. In particular, one expects it to hold for any tandem walk of finite dimension with the same exit boundary; the proof is almost identical but requires a generalization of Proposition 11, which, we believe, will involve the same ideas given in its proof. We leave this task to a future work.

https://doi.org/10.1017/jpr.2018.60 Published online by Cambridge University Press

5. Numerical example

From Proposition 8 we see that for $x \in \mathbb{R}_+^2$ and $x_n = \lfloor nx \rfloor$, the relative error

$$\frac{|W(T_n(x_n)) - \mathbb{P}_{x_n}(\tau_n < \tau_0)|}{\mathbb{P}_{x_n}(\tau_n < \tau_0)}$$

decays exponentially in n. We now examine numerically to see how well this approximation works. Define

$$p_n^k(x) = \mathbb{P}_x(\tau_n < \tau_0 \le k).$$

By the Markov property of X, p^k satisfies the following recursion:

$$p_n^{k+1}(x) = \begin{cases} 1 & \text{if } x \in \partial A_n, \\ 0 & \text{if } x = 0, \\ \lambda p_n^k(x + (1,0)) + \mu_1 p_n^k(x + (-1,1)) + \mu_2 p_n^k(x) & \text{if } x \in \partial_2, \\ \lambda p_n^k(x + (1,0)) + \mu_1 p_n^k(x) + \mu_2 p_n^k(x + (0,-1)) & \text{if } x \in \partial_1, \\ \lambda p_n^k(x + (1,0)) + \mu_1 p_n^k(x + (-1,1)) + \mu_2 p_n^k(x + (0,-1)) & \text{otherwise.} \end{cases}$$

By the monotone convergence theorem, $\lim_{k\to\infty} p_n^k(x) = p_n(x)$. Furthermore, by the Perron–Frobenius theorem, this convergence occurs exponentially fast. Therefore, p_k^n provides an excellent approximation of p^n for large k. All the prior works [16], [18], [50], [51], [53], [54], and [56] use this type of approximation to compute the quantity of interest 'exactly' to illustrate the numerical performance of the algorithms under consideration and we will do the same. For our numerical study, we set $\mu_1 = 0.4$, $\mu_2 = 0.5$, $\lambda = 0.1$, and $n = 60$. For $n = 60$, the above iteration can be run quickly on a computer and $k = 600$ is more than enough for its convergence (convergence can be observed easily by checking that p_n^k no longer changes as k increases). Recall that $W(T_n(x_n))$ is a sum of functions decaying exponentially in $n - x(1)$ and $x(2)$. This suggests that it is visually simpler to compare the functions $W(T_n(x_n))$ and $\mathbb{P}_x(\tau_n < \tau_0)$ by representing them in the log scale:

$$V_n \doteq -\frac{1}{n}\log \mathbb{P}_x(\tau_n < \tau_0) \quad \text{and} \quad W_n \doteq -\frac{1}{n}\log W(T_n(x_n)).$$

In the left panel of Figure 4 we depict the level curves of W_n of V_n; they all completely overlap except for the first one along the $x(2)$ axis. In the right panel of Figure 4, we present the relative error $(W_n - V_n)/V_n$; we see that it appears to be 0 except for a narrow layer around 0 where it is bounded by 0.02.

For $x = (1,0)$, the exact value for the probability $\mathbb{P}_x(\tau_{60} < \tau_0)$ is 1.1285×10^{-35} and the approximate value given by $W(T_n(x))$ is equal to 1.2037×10^{-35}. Moving away from the origin, these quantities quickly converge to each other. For example,

$$\mathbb{P}_x(\tau_{60} < \tau_0) = 4.8364 \times 10^{-35}, \qquad W(T_n(x)) = 4.8148 \times 10^{-35} \quad \text{for } x = (2,0),$$

and

$$\mathbb{P}_x(\tau_{60} < \tau_0) = 7.8886 \times 10^{-31}, \qquad W(T_n(x)) = 7.8885 \times 10^{-31} \quad \text{for } x = (9,0).$$

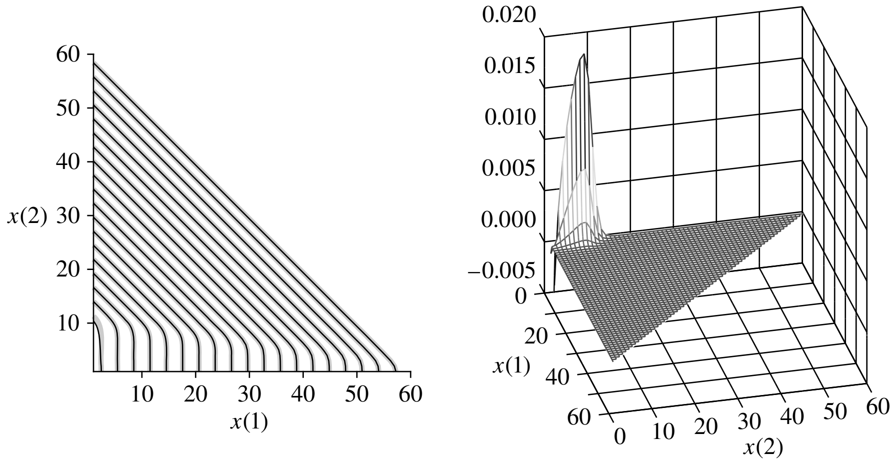

FIGURE 4: (*Left*): level curves of V_n (*thin shaded dark*) and W_n (*thick shaded light*). (*Right*): the graph of $(W_n - V_n)/W_n$.

6. Literature review

There is a vast literature related to the analysis presented in this paper. Below we review a number of related works and point out the connections between them and our paper.

There is a clear correspondence between the structures which appear in the LD analysis and the subsolution approach to IS estimation of p_n of [15], [16] [18], [51], and [54], and those involved in the methods developed in this paper. This connection is best expressed in the following equation (in the context of two tandem walk just studied). For $q = (q_1, q_2) \in \mathbb{R}^2$, set $\beta = \mathrm{e}^{q_1}$ and $\alpha = \mathrm{e}^{q_1 - q_2}$; then

$$H(q) = -\log(\boldsymbol{p}(\beta, \alpha)),$$

where $\boldsymbol{p}$ is the characteristic polynomial defined in (16). A similar relation exists between H_2 and $\boldsymbol{p}_2$. In the LD analysis, H and H_2 appear as two of the Hamiltonians of the limit deterministic continuous time control problem; the gradient of the limit value function lies on their 0-level sets. Parallel to our construction in Subsection 3.2.1, the articles using the subsolution approach construct subsolutions to a limit HJB equation using points on or inside the 0-level curve of the hamiltonians H and H_2 or their intersection; e.g. the gradient $\boldsymbol{r}_1$ defined following (39) lies exactly on this intersection and corresponds to the point (ρ_1, ρ_1) lying on $\mathcal{H} \cap \mathcal{H}_2$; see [18, Figure 9], the point r_1 lying on the intersection of the 0-level sets of the Hamiltonians H and H_2 corresponds again to the point (ρ_1, ρ_1) lying on $\mathcal{H} \cap \mathcal{H}_2$ identified in Subsection 3.2.1). These works use subsolutions to estimate variances of IS estimators (again based on the same subsolution) for buffer overflow probabilities of the form $\mathbb{P}_x(\tau_n < \tau_0)$ and concentrate on the initial point $x = 0$. Concentrating on the initial point $x = 0$ allows great flexibility on the choice of the exit boundary ∂A_n.

McDonald [39] considered the buffer overflow of a chosen node in a given stable network. The process W considered in [39] is $(r + m)$-dimensional: the first dimension represents the node whose overflow event is to be studied and the dimensions $2, 3, \ldots, r$ represent nodes that become unstable when the first node overflows. For $n > 0$, let τ_n be the first time the first component of W hits n, i.e. $\tau_n = \inf\{k \geq 0\colon W(k) \in F_n\}$, $F_n = \{x \in \mathbb{Z}_+^{r+m}\colon x_1 \geq n\}$, and let τ_0 denote the first time W hits the origin $\mathbf{0}$. Finally, let $\tau_\triangle$ denote the first time after

https://doi.org/10.1017/jpr.2018.60 Published online by Cambridge University Press

time 0 that one of the nodes from 1 to r hits 0, i.e. $\tau_\triangle = \inf\{k\colon k > 0,\ W \in \triangle\}$, where $\triangle = \{x : x_j = 0 \text{ for some } j \in \{1, 2, 3, \dots, r\}\}$ is the constraining boundary of the state space for the components 1 to r. McDonald [39] derived the following approximation result: let $\pi_\triangle$ denote the stationary measure conditioned on $\triangle$ and $\mathbb{E}_{\pi_\triangle}$ denote expectation conditioned on $W(0)$ having initial distribution $\pi_\triangle$. Let $\boldsymbol{\tau}_0$ be the first return time to 0, i.e. $\boldsymbol{\tau}_0 = \inf\{k > 0\colon W_k = 0\}$. From [39, Lemma 1.8], under the assumptions made in the paper, we see that

$$\lim_{n\to\infty} \frac{|\pi(\mathbf{0})\mathbb{P}_{\mathbf{0}}(\tau_n < \boldsymbol{\tau}_{\mathbf{0}}) - \pi(\Delta)\mathbb{P}_{\pi_\triangle}(\tau_n < \tau_\triangle)|}{\pi(\mathbf{0})\mathbb{P}_{\mathbf{0}}(\tau_n < \boldsymbol{\tau}_{\mathbf{0}})} = 0.$$

The analysis that leads to this result is based on the h-transform of W, where h is a harmonic function of W away from the set Δ taking the form $h(x) = \mathrm{e}^{\alpha x_1} a(x)$; [39] gives conditions under which such an h function exists based on results from [43]. McDonald [39] developed the following representation for $\pi(\Delta)\mathbb{P}_{\pi_\triangle}(\tau_n < \tau_\triangle)$:

$$\pi(\Delta)\mathbb{P}_{\pi_\triangle}(\tau_n < \tau_\triangle) = \mathrm{e}^{-\alpha n}\mathbb{E}_{\pi_\triangle}[h(W(1))\Psi(W(1))],$$
$$\Psi(x) = \mathbb{E}_x[a^{-1}(\mathfrak{W}(\tau_n))\mathrm{e}^{-\alpha(\mathfrak{W}(\tau_n)-n)}\mathbf{1}_{\{\tau_n<\tau_\triangle\}}], \tag{52}$$

where $\mathfrak{W}$ is the h-transform of the process W (if W is not a nearest-neighbor random walk on $\mathbb{Z}_+^{r+m}$, the formula for Ψ needs to be slightly modified; see [39] for details). For the computation of the expectation appearing in (52), McDonald [39] suggested simulation. In [39, Section 3], the two-dimensional constrained random walk on $\mathbb{Z}_+^2$ was treated with increments $(-1, 0)$, $(1, 0)$, $(0, -1)$, $(0, 1)$, $(1, 1)$; for this process, the author constructed explicitly an h function of the form $h(x) = a_1^{x_1} a_2^{x_2}$, where $(a_1, a_2) \in \mathbb{R}^2$ is a point on a curve whose definition is analogous to the definition of the characteristic surface $\mathcal{H}$.

Miyazawa [41] employed the idea of removing constraints on one of the boundaries and using points on curves associated with the resulting process to study the tail asymptotics of the stationary distribution of a two-dimensional nearest-neighbor random walk $\boldsymbol{L}$ constrained to remain in $\mathbb{Z}_+^2$. To study the asymptotic decay rate of $\boldsymbol{\nu}(n, k)$ in n for a fixed k, the author considered the random walk $\boldsymbol{L}^{(1)}$, which has the same dynamics as $\boldsymbol{L}$ except that it is not constrained on the vertical axis. Associated with this process, the author defined two curves, whose definitions are parallel to the definition of $\mathcal{H}$ and $\mathcal{H}_1$ (see the definition of $\mathfrak{D}_1$ in [41, p. 554]) and used points on and inside these curves to define solutions to an eigenvalue/eigenvector problem associated with the problem; see [41, Theorem 3.1]. For the study of tail asymptotics along the vertical axis, the author used the same analysis but this time removing the constraint on the horizontal axis. For further works along this line of research, we refer the reader to [9], [32], and [42].

Ignatiouk-Robert [27] developed an explicit formula for the large deviation local rate function $L(x, v)$ of a general Jackson network, starting from representations of these rates as limits derived in [3] and [13]. For this, the author employed 'free processes', these are versions of the original process obtained by removing those constraints from the original process that are not involved in a given direction v at a given point $x \in \mathbb{R}_+^d$. The proofs in [27] use fluid limits for the free process under a change of measure (i.e. a twisted/h-transformed version of the free process); the changes of measures used here correspond to using h-functions of the form $\mathrm{e}^{\langle\theta,x\rangle}$, where θ is a point on a characteristic surface (analogous to $\mathcal{H}$ in this work or H in [18]) associated with the process being transformed; see [27, Section 6]. As an application of its results, the author computed the limit $\lim_{n\to\infty}(1/n)\log\mathbb{E}_0[\tau_n]$ by noting from [45] that this limit is equal to

$$\lim_{n\to\infty} -\frac{1}{n}\log\mathbb{P}_0(\tau_n < \tau_0),$$

https://doi.org/10.1017/jpr.2018.60 Published online by Cambridge University Press

which is the LD decay rate of the probability studied in this paper for general stable Jackson networks; Ignatiouk-Robert [27] derived the explicit formula $\min_{1\le i\le d} -\log(\rho_i)$ for the above LD rate using the explicit local rate functions developed in the same work and the explicit formulas available for the stationary distribution of the underlying process.

The Martin boundary of an unstable process is a characterization of the directions through which the process may diverge to ∞. The idea of using points on characteristic surfaces, and the idea of removing constraints from the process to simplify analysis, appear also in works devoted to identifying Martin boundaries of constrained or stopped processes. An example is [28], in which the authors identified the Martin boundary of two-dimensional random walks in $\mathbb{Z}_+^2$ and which are stopped as soon as they hit the boundary of $\mathbb{Z}_+^2$. The authors considered

1. the directions $q \in \mathbb{R}_+^2$, where both components of q are nonzero,
2. the directions q such that $q(1) = 0$, and
3. directions such that $q(2) = 0$.

For each of these cases, the authors worked with what they called *local processes*; the local process for the first case is a completely unconstrained random walk, the local process for the second case is a process keeping the horizontal axis (i.e. the vertical boundary is removed), and the third case is the reverse of the second case. The authors used LD analysis of the local processes, harmonic functions of the form

$$h_a(x) = \begin{cases} x_1 \mathrm{e}^{\langle a,x\rangle} - \mathbb{E}_x[S_1(\tau)\mathrm{e}^{\langle a,x\rangle}\mathbf{1}_{\{\tau<\infty\}}] & \text{if } q(a) = (0,1), \\ x_2 \mathrm{e}^{\langle a,x\rangle} - \mathbb{E}_x[S_2(\tau)\mathrm{e}^{\langle a,x\rangle}\mathbf{1}_{\{\tau<\infty\}}] & \text{if } q(a) = (1,0), \\ \mathrm{e}^{\langle a,x\rangle} - \mathbb{E}_x[\mathrm{e}^{\langle a,x\rangle}\mathbf{1}_{\{\tau<\infty\}}] & \text{otherwise,} \end{cases}$$

where S is the underlying process, τ is the first hitting time to the boundary of $\mathbb{Z}_+^2$, a is a given point on a surface associated with S (defined analogous to $\mathcal{H}$), and $q(a)$ is the mean direction of S under an exponential change of measure defined by a; see [28, p. 1108]. In this connection, we also cite [33], in which the authors used geometry and complex analysis to identify the Martin boundary of random walks on $\mathbb{Z}^2$, $\mathbb{Z}\times\mathbb{Z}_+$, and $\mathbb{Z}_+^2$.

Let X be the constrained random walk in $\mathbb{Z}_+^2$ with increments $(1,0)$, $(-1,0)$, $(0,1)$, and $(0,-1)$ and let τ_n be as in (1). A classical problem in computer science going back to [31, Section 2.2.2, Exercise 13] is the analysis of the expectation

$$\mathbb{E}[\max(X_1(\tau_n), X_2(\tau_n))], \tag{53}$$

i.e. the expected size of the longest queue at the time of buffer overflow. This expectation was computed in [31] for the cases of $\mathbb{P}(I_k = (1,0)) = \mathbb{P}(I_k = (0,1)) = \frac{1}{2}$ and $\mathbb{P}(I_k = (-1,0)) = \mathbb{P}(I_k = (0,-1)) = 0$. Various versions of this problem have since been treated in [20], [25], [36], [37], and [57]. Maier [37] treated a generalization of this problem where the dynamics of the random walk depend on its position; the approach of [37] used LD techniques from [22]. Yao [57] treated the approximation of (53) for the case when the increments have a symmetric distribution as follows: $\mathbb{P}(I_k = (1,0)) = \mathbb{P}(I_k(0,1)) = \frac{1}{2}(1-p)$ and $\mathbb{P}(I_k = (-1,0)) = \mathbb{P}(I_k(0,-1)) = \frac{1}{2}p$; furthermore, $p < \frac{1}{2}$ was assumed, i.e. the process was assumed to be unstable. Under these assumptions, the author developed an approximation for the expectation in (53) as $n\to\infty$. The main idea of Yao [57] is the following: under the assumptions of the paper one can ignore both the constraining boundaries of the process. To prove this, the author used LD bounds on i.i.d. Bernoulli sequences; see [57, Lemma 3].

https://doi.org/10.1017/jpr.2018.60 Published online by Cambridge University Press

Then an explicit computation for the unconstrained process using elementary techniques yields the desired approximation.

7. Conclusion

In this section we point out several implications of our results, work in progress, and possible extensions.

7.1. The $\mu_1 = \mu_2$ case

Equation (2) for $\mathbb{P}_y(\tau < \infty)$ (derived in Proposition 7) requires that $\mu_1 \neq \mu_2$. The $\mu_1 = \mu_2$ case can be handled by letting $\mu_2 \to \mu_1$ in (2). Then

$$\mathbb{P}_y(\tau < \infty) = \rho^{y(1)-y(2)} + \frac{\mu - \lambda}{\mu} \rho^{y(1)}(y(1) - y(2)),$$

where $\rho = \lambda/\mu$ and $\mu_1 = \mu_2 = \mu$. The $\mu_1 = \mu_2$ case leads to the linear term $y(1) - y(2)$.

7.2. Constrained diffusions with drift and elliptic equations with Neumann boundary conditions

Diffusion processes are weak limits of random walks. Thus, the results of the previous sections can be used to compute/approximate Balayage and exit probabilities of constrained unstable diffusions. We present an example demonstrating this possibility.

For $a, b > 0$, let X be the constrained diffusion on $\mathbb{R} \times \mathbb{R}_+$ with infinitesimal generator L defined as

$$f \to Lf, \qquad Lf = \langle \nabla f, ((2a + b), (a - b)) \rangle + \frac{1}{6} \nabla^2 f \cdot \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix},$$

where ∇^2 denotes the Hessian operator, mapping f to its matrix of second-order partial derivatives. On $\{x : x(2) = 0\}$, X is pushed up to remain in $\mathbb{R} \times \mathbb{R}_+$ (the precise definition involves the Skorokhod map; see, e.g. [34]). Then $a, b > 0$ implies that, starting from $B = \{x : x(1) > x(2)\}$, X has a positive probability of never hitting $\partial B = \{x : x(1) = x(2)\}$. Let τ be the first time that X hits $\{x : x(1) = x(2)\}$. Proposition 7 for $d = 2$ suggests

$$\begin{aligned}\mathbb{P}_x(\tau < \infty) = {} & \mathrm{e}^{-(a+2b)3(x(1)-x(2))} + \frac{a + 2b}{a - b} \mathrm{e}^{-(a+2b)3(x(1)-x(2))} \mathrm{e}^{-(2a+b)3x(2)} \\ & - \frac{a + 2b}{a - b} \mathrm{e}^{-3(2a+b)x(1)}, \qquad x \in B. \end{aligned} \tag{54}$$

One can check directly that the right-hand side of this display satisfies

$$LV = 0, \qquad \langle \nabla V, (0, 1) \rangle = 0, \quad x \in \partial_2.$$

This and a verification argument similar to the proof of Proposition 5 imply (54).

7.3. General Jackson networks

7.3.1. *Multiple approximations.* We have seen with Proposition 8 that $\mathbb{P}_{y_n}(\tau < \infty)$ approximates $\mathbb{P}_{x_n}(\tau_n < \tau_0)$, $x_n = \lfloor nx \rfloor$, very well (i.e. with exponentially decaying relative error) for all $x \in A \doteq \{x \in \mathbb{R}^2_+, 0 < x(1) + x(2) < 1\}$ when n is large. When X is the constrained random walk associated with a general two-dimensional Jackson network, this does not hold in general and to obtain a good approximation across all A we will have to use the transformation $T_n^2(x) = (x(1), (n - x(2)))$ as well as T_n. Then T_n^2 moves the origin of the coordinate system to the corner $(0, n)$ of ∂A_n. Thus, for general two-dimensional X, we will have to construct two limit processes Y^1 and Y^2; Y^1 is as above and Y^2 is the limit of $Y^{2,n} \doteq T_n^2(X)$. The limit probability is, as before, $\mathbb{P}_y(\tau^2 < \infty)$, where τ^2 is the first time Y^2 hits ∂B. In d-dimensions, we will have d possible limit processes, one for each corner of ∂A_n providing precise approximations for initial points which lie away from the boundaries missing in the limit problem. For a numerical example, see the preprint of this paper [55, Section 8.2]. One work in progress, based on the approach of Section 4, gives details of these ideas in the context of Jackson networks consisting of parallel queues. The same work also considers the approximation of the expectation (53) using the techniques of this paper.

7.3.2. *Approximation of* $\mathbb{P}_y(\tau < \infty)$ *in general.* The second issue is the generalization of the computation of the limit probability $\mathbb{P}_y(\tau < \infty)$. As we have seen in Proposition 7, in the case of two tandem queues, it is possible to compute this probability exactly as the superposition of two Y-harmonic functions: $[(\rho_1, \rho_1), \cdot]$ and h_{ρ_2}. For general two-dimensional Jackson networks, superposition of these two functions will yield only an approximation of $\mathbb{P}_y(\tau < \infty)$; to construct better approximations one should proceed as indicated in Remark 3 and use a linear combination of a finite number of functions in the class of Y-harmonic functions constructed in Subsections 3.2.1 and 3.2.2 to approximate the constant function j on the boundary ∂B; the error made in this approximation on ∂B will provide an upper bound for the error made in the approximation of $\mathbb{P}_y(\tau < \infty)$ for any $y \in B$. The numerical example in [55, Section 8.2] demonstrates this point.

7.3.3. ∂B*-determined* Y*-harmonic functions.* In the above section we have noted that, in general, to construct improved approximations of $\mathbb{P}_y(\tau < \infty)$, we will need to use additional Y-harmonic functions of the form

$$h_\beta = \beta^{y(1)-y(2)}\big(C(\beta, \alpha_2)\alpha_1^{y(2)} - C(\beta, \alpha_1)\alpha_2^{y(2)}\big),$$

where (β, α_1) and (β, α_2) are conjugate and $\Delta(\beta) \neq 0$. We know from Proposition 5 that h_β is ∂B-determined if $|\alpha_1|, |\alpha_2| \leq 1$, and $|\beta| < 1$. Suppose we fix $\alpha \in \{z \in \mathbb{C},\ |z| = 1\}$ and compute β and α^* so that (β, α) and (β, α^*) are conjugate (β and α^* are computed by solving the characteristic equation $\boldsymbol{p} = 1$). In view of Proposition 5, and in view of the fact that h_β is used in the approximation of a ∂B-determined Y-harmonic function, a natural question is under what conditions on the parameters of the model do $|\alpha^*| \leq 1$ and $|\beta| < 1$ hold. This problem was studied for the general two-dimensional Jackson network in [55, Section 4] (in particular, see Propositions 4.12 and 4.13). These propositions require simplifying conditions on the system parameters; see, e.g. [55, Condition (56), p. 18]. The derivation of more precise conditions remains an open problem.

7.3.4. *Harmonic systems.* In Subsection 3.3 we pointed out that the classes of Y-harmonic functions constructed in Subsections 3.2.1 and 3.2.2 have graph representations, as presented in Figure 3; we refer to these graphs and the system of equations they represent as ‘harmonic systems.’ It is possible to generalize these graphs to walks in d-dimensions, and corresponding to each solution, to the system of equations represented by the graph one can define a Y-harmonic function; this was addressed in the preprint of this paper [55, Section 5]; see Definitions 5.1 and 5.2, Proposition 5.2 generalizing our Proposition 4, and Proposition 5.3 generalizing our Proposition 5.

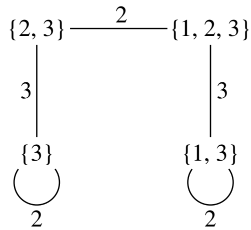

FIGURE 5: A harmonic system for $d = 3$.

7.3.5. *d tandem queues.* Remarkably, it turns out to be possible to define a class of harmonic systems and explicitly solve them to generalize (2) for $\mathbb{P}_y(\tau < \infty)$ to d tandem queues. This was addressed in [55, Section 6]. As an example, consider $d = 3$. To compute $\mathbb{P}(\tau < \infty)$, we use, in addition to the graphs in Figure 3, the graph presented in Figure 5. Using [55, Proposition 6.3] implies that, for

$$\mu_i \neq \mu_j, \qquad i, j \in \{1, 2, 3\}, \tag{55}$$

the following function solves the harmonic system presented in Figure 5:

$$h_{\rho_3}(y) = \rho_3^{y(1)-(y(2)+y(3))}\big(1 - c_3\rho_2^{y(3)} - c_3c_1\rho_1^{y(2)}\rho_1^{y(3)} + c_3c_2\rho_1^{y(2)}\rho_2^{y(3)}\big), \tag{56}$$

where

$$c_2 = \frac{\mu_2 - \lambda}{\mu_2 - \mu_1}, \qquad c_3 = \frac{\mu_3 - \lambda}{\mu_3 - \mu_2}, \qquad c_1 = \frac{\mu_3 - \lambda}{\mu_3 - \mu_1}.$$

The Y process for the three tandem queues is a random walk on $\mathbb{Z} \times \mathbb{Z}_+^2$ with increments $(-1, 0, 0), (1, 1, 0), (0, -1, 1)$, and $(0, 0, -1)$. The function h of (56) is a Y-harmonic function. There are four terms in the sum (56) defining h_{ρ_3}, each of these terms corresponds to a node of the graph in Figure 5. None of them is Y-harmonic individually. But the particular linear combination in (56) is indeed Y-harmonic. Two additional Y-harmonic functions used in the calculation of $\mathbb{P}_y(\tau < \infty)$ are

$$h_{\rho_2} = \rho_2^{y(1)-(y(2)+y(3))}\big(\rho_2^{y(3)} - c_2\rho_1^{y(2)}\rho_2^{y(3)}\big), \qquad h_{\rho_1} = \rho_1^{y(1)-(y(2)+y(3))}\rho_1^{y(2)}\rho_1^{y(3)};$$

the harmonic systems for these functions are 'edge-completions' of those presented in Figure 3; see [55, Definition 5.4]. The exact formula for $\mathbb{P}_y(\tau < \infty)$ for $y \in \mathbb{Z}\times\mathbb{Z}_+^2,\ y(1) \geq y(2)+y(3)$ is stated in [55, Proposition 6.5] as

$$\mathbb{P}_y(\tau < \infty) = h_{\rho_3} + c_3h_{\rho_2} + c_1c_3h_{\rho_1}.$$

To treat the case when (55) does not hold, it suffices to take limits in the last formula, which leads to polynomial terms in y.

7.4. Extension to other processes and domains

In the foregoing sections, we approximated $\mathbb{P}_x(\tau_n < \tau_0)$ in three stages:

1. used an affine change of coordinates to move the origin to a point on the exit boundary and took limits; as a result, some of the constraints in the prelimit process disappeared and we obtained as a limit process an unstable constrained random walk, and as a limit problem the probability of the return $\mathbb{P}_y(\tau < \infty)$ of the unstable process;

2. proved that, for a set of initial conditions, the resulting approximation had exponentially decaying relative error,

https://doi.org/10.1017/jpr.2018.60 Published online by Cambridge University Press

3. found a class of basis functions on the exit boundary on which the Balayage operator of the limit process had a simple action; then tried to approximate the constant function j (i.e. the value of $\mathbb{P}_y(\tau < \infty)$ on the exit boundary) on the exit boundary with linear combinations of the functions in the basis class.

The last two stages obviously depend on the particular dynamics of the original process and the geometry of the exit boundary. In ongoing research we consider two tandem queues with Markov modulated dynamics; optimal IS simulation for this process was developed in [53]. For Markov modulated dynamics, one needs a more general class of Y-harmonic functions than those constructed in Section 3 and the resulting equations are of higher degree and more challenging to analyze but the main ideas of Section 3 do generalize. In the present work we have focused on the exit boundary ∂A_n; another natural exit boundary is $\{y\colon y(i) \leq \lfloor a_i n \rfloor\}$ for $a_i > 0$, $i = 1, 2$. We expect our approach in this paper to generalize to this exit boundary with the following important modification: for this boundary, there are three points on the exit boundary from which one must conduct a limit analysis: the corners $n(0, a_2)$, $n(0, a_1)$, and $n(a_1, a_2)$. For the last point, the limit process is the completely unconstrained version of the random walk. Providing the details of this and further extensions to other processes and exit boundaries remain problems for future research.

7.5. Loss of stability in random perturbations of stable dynamical systems

The type of problem we have studied is of the following form: there is a process X with a certain law of large numbers limit which takes X away from a boundary ∂A_n towards a stable point or a region; τ_0 is the first time the process enters this stable region. The probability of interest is $\mathbb{P}(\tau_n < \tau_0)$. This setup is closely related to the study of random perturbations of stable dynamical systems whose LD analysis was treated in [22, Chapter 4]. In this framework, one starts with a dynamical system $\dot{x} = b(x)$ around a stable equilibrium point (taken to be $0 \in \mathbb{R}^d$). Stability implies that smooth trajectories of x move toward 0. Random perturbations x^ϵ of this system can be used as models for systems subject to noise in real-life situations. Trajectories of x^ϵ will no longer converge to 0 deterministically but may go arbitrarily away from it. Leaving a certain open set D containing 0 is considered to be a loss of stability for the perturbed system; then $\tau_D = \inf\{t > 0\colon x^\epsilon(t) \in D^c\}$ is the time when the perturbed system becomes unstable. Let τ_δ be the first time x^ϵ hits a δ neighborhood of 0. Probabilities of the form $\mathbb{P}(\tau_D < \tau_\delta)$ naturally arise in the analysis of τ_D. The parallels between this framework and the question treated in the present work suggest that our approach may be useful in the analysis of the loss of stability in random perturbations of stable dynamical systems. A study of this possible connection can also be the subject of future research.

Acknowledgements

A large part of this work was supported by the Rbuce-up COFUND programme while the author was at L'Université d'Evry, Department of Mathematics, Probability and Analysis Laboratory. The author thanks the Probability and Analysis Laboratory of L'Université dEvry. The author thanks wife Tuçe Değirmenci and his parents Erol and Hafize Sezer for their infinite support; this work is dedicated to them.

References

[1] Asmussen, S. (2003). *Applied Probability and Queues*, 2nd edn. Springer, New York.

[2] Asmussen, S. and Glynn, P. W. (2007). *Stochastic Simulation: Algorithms and Analysis*. Springer, New York.

https://doi.org/10.1017/jpr.2018.60 Published online by Cambridge University Press

[3] Atar, R. and Dupuis, P. (1999). Large deviations and queueing networks: methods for rate function identification. *Stoch. Process. Appl.* **84,** 255–296.
[4] Blanchet, J. (2013). Optimal sampling of overflow paths in Jackson networks. *Math. Operat. Res.* **38,** 698–719.
[5] Blanchet, J. H., Leder, K. and Glynn, P. W. (2008). Efficient simulation of light-tailed sums: an old-folk song sung to a faster new tune. In *Monte Carlo and Quasi-Monte Carlo Methods 2008*, Springer, Berlin, pp. 227–258.
[6] Borovkov, A. A. and Mogul'skiĭ, A. A. (2001). Large deviations for Markov chains in the positive quadrant. *Russian Math. Surveys* **56,** 803–916.
[7] Chen, H. and Yao, D. D. (2001). *Fundamentals of Queueing Networks: Performance, Asymptotics, and Optimization*. Springer, New York.
[8] Comets, F., Delarue, F. and Schott, R. (2007). Distributed algorithms in an ergodic Markovian environment. *Random Structures Algorithms* **30,** 131–167.
[9] Dai, J. G. and Miyazawa, M. (2011). Reflecting Brownian motion in two dimensions: exact asymptotics for the stationary distribution. *Stoch. Systems* **1,** 146–208.
[10] De Boer, P.-T. (2006). Analysis of state-independent importance-sampling measures for the two-node tandem queue. *ACM Trans. Model. Comput. Simul.* **16,** 225–250.
[11] Dean, T. and Dupuis, P. (2009). Splitting for rare event simulation: a large deviation approach to design and analysis. *Stoch. Process. Appl.* **119,** 562–587.
[12] Dupuis, P. and Ellis, R. S. (1997). *A Weak Convergence Approach to the Theory of Large Deviations*. John Wiley, New York.
[13] Dupuis, P. and Ellis, R. S. (1995). The large deviation principle for a general class of queueing systems. I. *Trans. Amer. Math. Soc.* **347,** 2689–2751.
[14] Dupuis, P. and Wang, H. (2004). Importance sampling, large deviations, and differential games. *Stoch. Stoch. Reports* **76,** 481–508.
[15] Dupuis, P. and Wang, H. (2007). Subsolutions of an Isaacs equation and efficient schemes for importance sampling. *Math. Operat. Res.* **32,** 723–757.
[16] Dupuis, P. and Wang, H. (2009). Importance sampling for Jackson networks. *Queueing Systems* **62,** 113–157.
[17] Dupuis, P., Leder, K. and Wang, H. (2007). Importance sampling for sums of random variables with regularly varying tails. *ACM Trans. Model. Comput. Simul.* **17,** 14.
[18] Dupuis, P., Sezer, A. D. and Wang, H. (2007). Dynamic importance sampling for queueing networks. *Ann. Appl. Prob.* **17,** 1306–1346.
[19] Durrett, R. (1996). *Probability: Theory and Examples*, 2nd edn. Duxbury, Belmont, CA.
[20] Flajolet, P. (1986). The evolution of two stacks in bounded space and random walks in a triangle. In *Mathematical Foundations of Computer Science, 1986*, Springer, Berlin, pp. 325–340.
[21] Foley, R. D. and McDonald, D. R. (2012). Constructing a harmonic function for an irreducible nonnegative matrix with convergence parameter $R > 1$. *Bull. London Math. Soc.* **44,** 533–544.
[22] Freidlin, M. I. and Wentzell, A. D. (2012). *Random Perturbations of Dynamical Systems*, 2nd edn. Springer, Heidelberg.
[23] Glasserman, P. and Kou, S.-G. (1995). Analysis of an importance sampling estimator for tandem queues. *ACM Trans. Model. Comput. Simul.* **5,** 22–42.
[24] Griffiths, P. A. (1989). *Introduction to Algebraic Curves*. American Mathematical Society, Providence, RI.
[25] Guillotin-Plantard, N. and Schott, R. (2006). *Dynamic Random Walks: Theory and Applications*. Elsevier, Amsterdam.
[26] Henderson, S. G. and Nelson, B. L. (eds) (2006). *Handbooks in Operations Research and Management Science:* Vol. 13, *Simulation*. North-Holland, Amsterdam.
[27] Ignatiouk-Robert, I. (2000). Large deviations of Jackson networks. *Ann. Appl. Prob.* **10,** 962–1001.
[28] Ignatiouk-Robert, I. and Loree, C. (2010). Martin boundary of a killed random walk on a quadrant. *Ann. Prob.* **38,** 1106–1142.
[29] Ignatyuk, I. A., Malyshev, V. A. and Scherbakov, V. V. (1994). Boundary effects in large deviation problems. *Russian. Math. Surveys* **49,** 41–99.
[30] Juneja, S. and Nicola, V. (2005). Efficient simulation of buffer overflow probabilities in Jackson networks with feedback. *ACM Trans. Model. Comput. Simul.* **15,** 281–315.
[31] Knuth, D. E. (1969). *The Art of Computer Programming*, Vol. 1, *Fundamental Algorithms*. Addison-Wesley, Reading, MA.
[32] Kobayashi, M. and Miyazawa, M. (2013). Revisiting the tail asymptotics of the double QBD process: refinement and complete solutions for the coordinate and diagonal directions. In *Matrix-Analytic Methods in Stochastic Models*, Springer, New York, pp. 145–185.
[33] Kurkova, I. A. and Malyshev, V. A. (1998). Martin boundary and elliptic curves. *Markov Process. Relat. Fields* **4,** 203–272.
[34] Kushner, H. J. and Dupuis, P. (2001). *Numerical Methods for Stochastic Control Problems in Continuous Time*, 2nd edn. Springer, New York.

[35] Louchard, G. and Schott, R. (1991). Probabilistic analysis of some distributed algorithms. *Random Structures Algorithms* **2,** 151–186.
[36] Louchard, G., Schott, R., Tolley, M. and Zimmermann, P. (1994). Random walks, heat equation and distributed algorithms. *J. Comput. Appl. Math.* **53,** 243–274.
[37] Maier, R. S. (1991). Colliding stacks: a large deviations analysis. *Random Structures Algorithms* **2,** 379–420.
[38] Maier, R. S. (1993). Large fluctuations in stochastically perturbed nonlinear systems: applications in computing. In *1992 Lectures in Complex Systems*, Addison-Wesley, Reading, MA, pp. 501–517.
[39] McDonald, D. R. (1999). Asymptotics of first passage times for random walk in an orthant. *Ann. Appl. Prob.* **9,** 110–145.
[40] Miretskiy, D., Scheinhardt, W. and Mandjes, M. (2010). State-dependent importance sampling for a Jackson tandem network. *ACM Trans. Model. Comput. Simul.* **20,** 15.
[41] Miyazawa, M. (2009). Tail decay rates in double QBD processes and related reflected random walks. *Math. Operat. Res.* **34,** 547–575.
[42] Miyazawa, M. (2011). Light tail asymptotics in multidimensional reflecting processes for queueing networks. *TOP* **19,** 233–299.
[43] Ney, P. and Nummelin, E. (1987). Markov additive processes. I. Eigenvalue properties and limit theorems. *Ann. Prob.* **15,** 561–592.
[44] Nicola, V. F. and Zaburnenko, T. S. (2007). Efficient importance sampling heuristics for the simulation of population overflow in Jackson networks. *ACM Trans. Model. Comput. Simul.* **17,** 10.
[45] Parekh, S. and Walrand, J. (1989). A quick simulation method for excessive backlogs in networks of queues. *IEEE Trans. Automatic Control* **34,** 54–66.
[46] Revuz, D. (1984). *Markov Chains*, 2nd edn. North-Holland, Amsterdam.
[47] Ridder, A. (2009). Importance sampling algorithms for first passage time probabilities in the infinite server queue. *Europ. J. Operat. Res.* **199,** 176–186.
[48] Robert, P. (2003). *Stochastic Networks and Queues*. Springer, Berlin.
[49] Rubino, G. and Tuffin, B. (2009). *Rare Event Simulation using Monte Carlo Methods*. John Wiley, New York.
[50] Setayeshgar, L. and Wang, H. (2013). Efficient importance sampling schemes for a feed-forward network. *ACM Trans. Model. Comput. Simul.* **23,** 21.
[51] Sezer, A. D. (2006). *Dynamic Importance Sampling for Queueing Networks*. Doctoral thesis, Division of Applied Mathematics, Brown University.
[52] Sezer, A. D. (2007). Asymptotically optimal importance sampling for Jackson networks with a tree topology. Preprint. Available at https://arxiv.org/abs/0708.3260.
[53] Sezer, A. D. (2009). Importance sampling for a Markov modulated queuing network. *Stoch. Process. Appl.* **119,** 491–517.
[54] Sezer, A. D. (2010). Asymptotically optimal importance sampling for Jackson networks with a tree topology. *Queueing Systems* **64,** 103–117.
[55] Sezer, A. D. (2015). Exit probabilities and balayage of constrained random walks. Preprint. Available at https://arxiv.org/abs/1506.08674.
[56] Sezer, A. D. and Özbudak, F. (2011). Approximation of bounds on mixed-level orthogonal arrays. *Adv. Appl. Prob.* **43,** 399–421.
[57] Yao, A. C. (1981). An analysis of a memory allocation scheme for implementing stacks. *SIAM J. Comput.* **10,** 398–403.