Approximation of the Exit Probability of a Stable Markov Modulated Constrained Random Walk

Fatma Başoğlu Kabran,[*] Ali Devin Sezer [†]

September 17, 2019

Abstract

Let X be the constrained random walk on $\mathbb{Z}_+^2$ having increments $(1,0)$, $(-1,1)$, $(0,-1)$ with jump probabilities $\lambda(M_k)$, $\mu_1(M_k)$, and $\mu_2(M_k)$ where M is an irreducible aperiodic finite state Markov chain. The process X represents the lengths of two tandem queues with arrival rate $\lambda(M_k)$, and service rates $\mu_1(M_k)$, and $\mu_2(M_k)$. We assume that the average arrival rate with respect to the stationary measure of M is less than the average service rates, i.e., X is assumed stable. Let τ_n be the first time when the sum of the components of X equals n for the first time. Let Y be the random walk on $\mathbb{Z} \times \mathbb{Z}_+$ having increments $(-1,0)$, $(1,1)$, $(0,-1)$ with probabilities $\lambda(M_k)$, $\mu_1(M_k)$, and $\mu_2(M_k)$. Let τ be the first time the components of Y are equal. For $x \in \mathbb{R}_+^2$, $x(1) + x(2) < 1$, $x(1) > 0$, and $x_n = \lfloor nx \rfloor$, we show that $P_{(n-x_n(1),x_n(2)),m)}(\tau < \infty)$ approximates $P_{(x_n,m)}(\tau_n < \tau_0)$ with exponentially vanishing relative error as $n \to \infty$. For the analysis we define a characteristic matrix in terms of the jump probabilities of (X, M). The 0-level set of the characteristic polynomial of this matrix defines the characteristic surface; conjugate points on this surface and the associated eigenvectors of the characteristic matrix are used to define (sub/super) harmonic functions which play a fundamental role both in our analysis and the computation / approximation of $P_{(y,m)}(\tau < \infty)$.

Keywords: Markov modulation, regime switch, multidimensional constrained random walks, exit probabilities, rare events, queueing systems, characteristic surface, superharmonic functions, affine transformation

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

1 Introduction and Definitions

A stochastic processes X is said to be Markov modulated if its dynamics depend on the state of a secondary Markov process M modeling the environment within which X operates [7]. Markov modulation/regime switch is one of the most popular methods of building richer models for a wide range of applications from finance to computer networks to queueing theory. This paper studies the approximation of the probability of a large excursion in the busy cycle of a constrained random walk X whose dynamics are modulated by a Markov process M. We assume M to be external, i.e, the transition probabilities of M do not depend on X. Constrained random walks arise naturally when there are barriers that keep a process

[*]Middle East Technical University, Institute of Applied Mathematics, Ankara, Turkey and İzmir Kavram Vocational School, Department of Finance, Banking and Insurance, İzmir, Turkey

[†]Middle East Technical University, Institute of Applied Mathematics, Ankara, Turkey

1

within a domain, for example: computers/ algorithms sharing resources on a system, financial positions that have shortselling constraints, or queueing systems. If X represents a queueing system, a large excursion corresponds to a buffer overflow event; the analysis, simulation and approximation of probabilities of such events for ordinary (non-modulated) constrained random walks have received considerable attention at least since [6, 4]; for further references and a literature review we refer the reader to [11]. To the best of our knowledge, there is hardly any study on the same probability for modulated constrained random walks: we are aware of only [9] treating the development of asymptotically optimal importance sampling algorithms for the approximation of the buffer overflow event. For this reason, this work will focus on one of the simplest multidimensional constrained random walks, the tandem walk, arising from the modeling of two servers working in tandem. Next we describe the dynamics of this process and give a precise definition of the buffer overflow probability of interest.

Our main process is a random walk X with increments $\{I_1, I_2, I_3, ...\}$, constrained to remain in $\mathbb{Z}_+^2$:

$$X_0 = x \in \mathbb{Z}_+^2, \quad X_{k+1} \doteq X_k + \pi(X_k, I_k), k = 1, 2, 3, ...$$
$$\pi(x, v) \doteq \begin{cases} v, & \text{if } x + v \in \mathbb{Z}_+^2, \\ 0, & \text{otherwise.} \end{cases}$$

The map π ensures that when X is on the constraining boundaries

$$\partial_i \doteq \{x \in \mathbb{Z}^2 : x(i) = 0\}, \quad i = 1, 2,$$

it cannot jump out of $\mathbb{Z}_+^2$. We assume the distribution of the increments I_k to be modulated by a Markov Chain M with state space $\mathcal{M}$ (with finite size $|\mathcal{M}|$) and with transition matrix $\boldsymbol{P} \in \mathbb{R}_+^{|\mathcal{M}| \times |\mathcal{M}|}$. To ease analysis and notation we will assume $\boldsymbol{P}$ to be irreducible and aperiodic, which implies that it has a unique stationary measure $\boldsymbol{\pi}$ on $\mathcal{M}$, i.e., $\boldsymbol{\pi} = \boldsymbol{\pi}\boldsymbol{P}$. Let $\mathscr{F}_k \doteq \sigma(\{M_j, j \leqslant k+1\}, \{X_j, j \leqslant k\})$, i.e., the σ-algebra generated by M and X. The increments I form an independent sequence given M and the increment I_k has the following distribution given $\mathscr{F}_{k-1}$:

$$I_k \in \{(0,0), (1,0), (-1,1), (0,-1)\},$$
$$\mathbb{P}(I_k = (0,0)|\mathscr{F}_{k-1}) = 1_{\{M_k \neq M_{k-1}\}}$$
$$\mathbb{P}(I_k = (1,0)|\mathscr{F}_{k-1}) = \lambda(M_k)1_{\{M_k = M_{k-1}\}}$$
$$\mathbb{P}(I_k = (-1,1)|\mathscr{F}_{k-1}) = \mu_1(M_k)1_{\{M_k = M_{k-1}\}}$$
$$\mathbb{P}(I_k = (0,-1)|\mathscr{F}_{k-1}) = \mu_2(M_k)1_{\{M_k = M_{k-1}\}}.$$

The dynamics of X are shown in Figure 1. The process (X, M) is the embedded random walk of a continuous time queueing system consisting of two tandem queues whose arrival and service rates are determined by a finite state Markov process M.

We assume (X, M) to be stable:

$$\sum_{m \in \mathcal{M}} (\lambda(m) - \mu_i(m))\boldsymbol{\pi}(m)\boldsymbol{P}(m, m) < 0, \ i = 1, 2. \tag{1}$$

In addition to (1), we need two further technical assumptions for our analysis see (27) and (28). Stability means that the queueing system represented by (X, M) serves customers
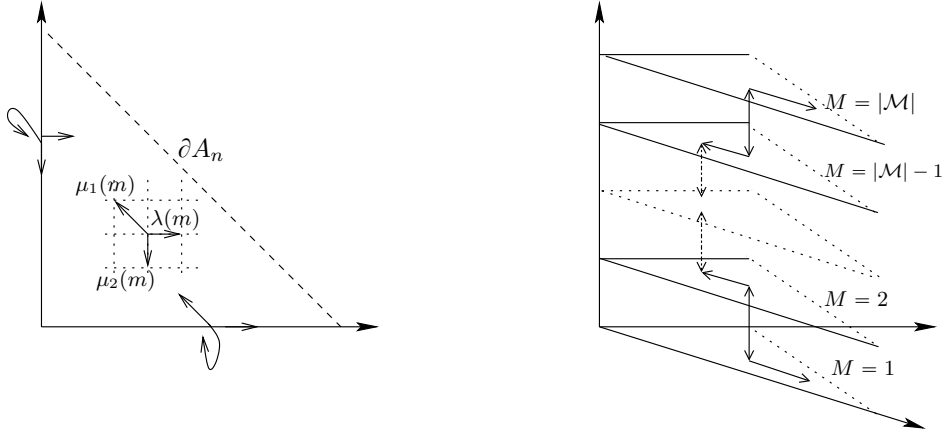
Figure 1: Markov modulated constrained random walk (X, M); the left figure shows dynamics in a given layer, the right figure shows jumps between layers representing regime switches

faster, on average, than the customer arrival rate; this keeps the lengths of both queues close to 0 at all times with high probability; but (X, M) being a random process, components of X can grow arbitrarily large if one waits long enough. For a stable constrained random walk such as (X, M) it is natural to measure time in cycles that restart each time X hits 0. If the system represented by this walk has a shared buffer where all customers wait (or where packets are stored, if, e.g., (X, M) represents a network of two computers / processes) then a natural question is the following: what is the probability that the shared buffer overflows in a given cycle? To express this problem mathematically we introduce the following notation: the region

$$A_n = \left\{ x \in \mathbb{Z}_+^2 : x(1) + x(2) \leqslant n \right\} \tag{2}$$

and the exit boundary

$$\partial A_n = \left\{ x \in \mathbb{Z}_+^2 : x(1) + x(2) = n \right\}. \tag{3}$$

A_n^o denotes the interior $A_n - \partial_1 \cup \partial_2$. Similarly, $\mathbb{Z}_+^{2,o}$ denotes $\mathbb{Z}_+^2 - \partial_1 \cup \partial_2$. Let τ_n be the first time X hits ∂A_n:

$$\tau_n \doteq \inf\{k \geqslant 0 : X_k \in \partial A_n\}, n = 0, 1, 2, 3, .. \tag{4}$$

Then the buffer overflow probability described above is

$$p_n(x, m) \doteq \mathbb{P}_{(x,m)}(\tau_n < \tau_0). \tag{5}$$

The Markov property of (X, M) implies that p_n is (X, M)-harmonic i.e., it satisfies

$$p_n(x, m) = \mathbb{E}_{(x,m)}[p_n(x + \pi(x, I_1), M_1)], x \in A_n - \partial A_n, \quad p_n(x, m) = 1, x \in \partial A_n.$$

This is a system of equations satisfied by p_n, where the number of unknowns is in the order of $|\mathcal{M}|n^2$. More generally, for a d dimensional system the number of unknowns grows like $|\mathcal{M}|n^d$, making the computation of p_n via a direct solution of the linear system resource intensive even for moderate values of n. This justifies the development of approximations of p_n and the main goal of the present work is to find easily computable and accurate approximations of p_n. Stability and the bounded increments of X suggest that when x is away from the exit boundary ∂A_n, p_n decays exponentially in n, making the buffer overflow event rare. The approximation of p_n, even when there is no modulation turns out to be a nontrivial problem.

3

There are two sources of difficulty: multidimensionality, and the discontinuous dynamics of the problem on the constraining boundaries. Asymptotically optimal importance sampling algorithms for the non-modulated setup were constructed in [2], which proposed a dynamic importance sampling algorithm based on subsolutions of a related Hamilton Jacobi Bellman (HJB) and its boundary conditions. The approach of [2] is tightly connected to the large deviations analysis of p_n, which identifies the exponential decay rate of p_n. Large deviations analysis is based on transforming p_n to $V_n = -(1/n) \log p_n$, scaling space by $1/n$ and taking limits; the limit V of V_n satisfies the HJB equation mentioned above. The works [10, 11] obtained sharp estimates of p_n for the non-modulated two dimensional tandem walk using an affine transformation of the process X; see Figure 2 and the summary below. Another goal of the present work is to show that this affine transformation approach can be extended to the analysis of p_n of the Markov modulated constrained random walk. As the present article shows, this extension turns out to be possible but Markov modulation complicates almost every aspect of the problem: the underlying functions, the geometry of the characteristic surfaces, the limit analysis, etc. A detailed comparison with the non-modulated case is given in Section 10.

To the best our knowledge, there is very limited research on the analysis of the overflow probability p_n for Markov modulated constrained random walks; we are only aware of the article [9] which develops asymptotically optimal importance sampling algorithms for the approximation of p_n for the (X, M) process studied in the present work. In doing this, a necessary step is also to compute the large deviation decay rate of p_n; this was also done for $x = 0$ in [9]. The analysis in this work is based on the sub and supersolutions of a limit HJB equation. Next is a summary of our analysis and main results.

1.1 Summary of analysis and main results

The starting point of our analysis is transforming X to another process Y^n by an affine transformation moving the origin to the point $(n, 0)$ on the exit boundary; as n goes to infinity, Y^n converges to the limit process Y constrained only on ∂_2; Figure 2 shows these transformations.
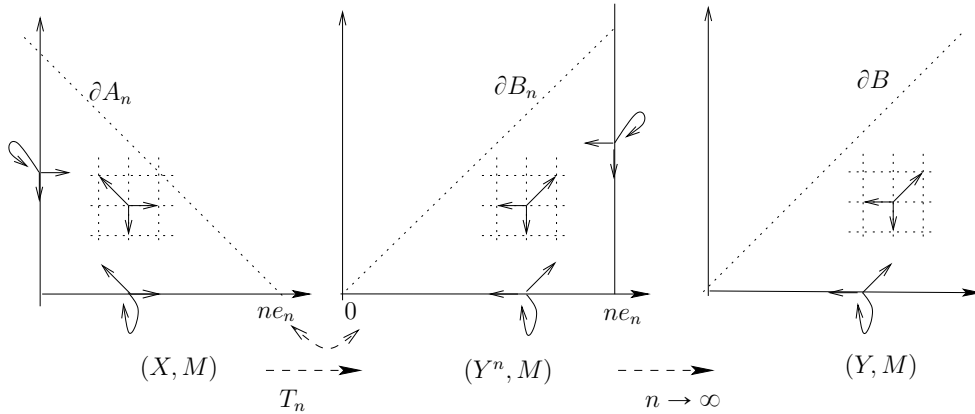

Figure 2: The transformation of (X, M)

The formal definition of the limit process Y is as follows: define

$$\boldsymbol{I}_2 \doteq \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

4

Define the constraining map

$$\pi_1(y,v) = \begin{cases} v, & \text{if } y + v \in \mathbb{Z} \times \mathbb{Z}_+, \\ 0, & \text{otherwise.} \end{cases}$$

Then the limit process Y is the M-modulated constrained random walk on $\mathbb{Z} \times \mathbb{Z}_+$ with increments

$$J_k \doteq \boldsymbol{I}_2 I_k : \tag{6}$$

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

Define the region

$$B \doteq \{y \in \mathbb{Z} \times \mathbb{Z}_+ : y(1) \geqslant y(2)\}$$

and the exit boundary

$$\partial B \doteq \{y \in \mathbb{Z} \times \mathbb{Z}_+ : y(1) = y(2)\}.$$

Let τ be the hitting time

$$\tau \doteq \inf\{k \geqslant 0 : Y_k \in \partial B\}.$$

Y is a process constrained to $\mathbb{Z} \times \mathbb{Z}_+$ with the constraining boundary ∂_2; we will denote the interior $\mathbb{Z} \times \mathbb{Z}_+ - \partial_2$ of this set by $\mathbb{Z} \times \mathbb{Z}_+^o$. Define the affine transformations

$$T_n = ne_1 + \boldsymbol{I}_2$$

where (e_1, e_2) is the standard basis for $\mathbb{R}^2$. Our main approximation result is the following:

Theorem (Theorem 6.1). *For any $x \in \mathbb{R}_+^2$, $x(1) + x(2) < 1$, $x(1) > 0$, and $m \in \mathcal{M}$ there exist constants $c > 0$, $\rho \in (0,1)$ and $N > 0$ such that*

$$\frac{|\mathbb{P}_{(x_n,m)}(\tau_n < \tau_0) - \mathbb{P}_{(T_n(x_n),m)}(\tau < \infty)|}{\mathbb{P}_{(x_n,m)}(\tau_n < \tau_0)} < \rho^{cn}, \tag{7}$$

for $n > N$, where $x_n = \lfloor nx \rfloor$.

Theorem 6.1 states that, as n increases, $\mathbb{P}_{(T_n(x_n),m)}(\tau < \infty)$ gives a very good approximation of $\mathbb{P}_{(x_n,m)}(\tau_n < \tau_0)$. Parallel to the non-modulated case treated in [11], the proof of Theorem 6.1 consists of the following steps 1) the difference between the events $\{\tau_n < \tau_0\}$ and $\{\tau < \infty\}$ can be characterized by the event "X first hits ∂_1 then ∂_2 and then ∂A_n" 2) the probability of this detailed event is very small compared to the probabilities of the events $\{\tau_n < \tau_0\}$ and $\{\tau < \infty\}$. The challenges arise from the implementation of these steps in the Markov modulated framework.

To bound the probabilities appearing in (7) we will use (Y, M)-(super)harmonic functions constructed from single and conjugate points on a characteristic surface $\mathcal{H}$ (see (14)) associated with (Y, M). The characteristic surface is the 0-level set of the characteristic polynomial of the characteristic matrix $\boldsymbol{A}$ (see (12)) defined in terms of the transition matrix $\boldsymbol{P}$ and the jump probabilities $\lambda(\cdot)$, $\mu_1(\cdot)$ and $\mu_2(\cdot)$. The characteristic polynomial is of degree $3|\mathcal{M}|$ and therefore the characteristic curve doesn't have a simple algebraic parametrization; for this reason, in the modulated case, the identification of points on the characteristic surface relies on the decomposition of the the surface into $|\mathcal{M}|$ components, by an eigenvalue analysis of $\boldsymbol{A}$ and the implicit function theorem. The decomposition is given in subsection 2.1 and the points relevant for our analysis are identified in Propositions 2.8, 2.11 and 2.12. These points all lie on the innermost component corresponding to the largest eigenvalue of $\boldsymbol{A}$.

In the presence of a modulating Markov chain, harmonic functions are constructed in general from $|\mathcal{M}| + 1$ points on the characteristic surface, which makes analysis based on them more complex. For this reason, we will switch to superharmonic functions whenever we can, which can be constructed from just two points. An upper bound for $\mathbb{P}_{(y,m)}(\tau < \infty)$ using these functions is given in Section 3. Section 4 constructs an upper bound for the detailed event described above characterizing the difference of the events $\{\tau_n < \tau_0\}$ and $\{\tau < \infty\}$. A lower bound for $\mathbb{P}_{(x,m)}(\tau_n < \tau_0)$ based on subharmonic functions constructed from the functions of Section 2 is given in Section 5. These elements are combined in Section 6 to prove our main approximation theorem, Theorem 6.1.

With Theorem 6.1 we know that $\mathbb{P}_{(x,m)}(\tau_n < \tau_0)$ can be approximated very well with $\mathbb{P}_{(T_n(x),m)}(\tau < \infty)$. In the non-modulated case, a linear combination of two Y-harmonic functions constructed from points on the characteristic surface gives an exact formula for $\mathbb{P}_y(\tau < \infty)$. This is no longer possible when there is modulation; Sections 7 and 8 develops increasingly accurate approximate formulas for $\mathbb{P}_{(y,m)}(\tau < \infty)$ using (Y, M)-harmonic functions constructed from further points on the characteristic surface under further linear independence assumptions (see (77), (89)), see Propositions 7.1 and Propositions 8.1 for the (Y, M)-harmonic functions constructed in these sections. As opposed to the limit analysis which uses points only on the innermost component of the characteristic surface, the construction of harmonic functions uses points on all components of the characteristic surface. Propositions 7.2 and 8.2 find bounds on the relative error of the approximations of $\mathbb{P}_{(y,m)}(\tau < \infty)$ provided by these functions based on the values they take on ∂B. Section 9 gives a numerical example showing the effectiveness of the resulting approximations. Section 10 compares the analysis of the current work with the non-modulated tandem walk treated in [10, 11] and the non-modulated parallel walk treated in [12]. Section 11 comments on future work.

2   (sub/super)Harmonic functions of (Y, M)

A function h on $\mathbb{Z} \times \mathbb{Z}_+ \times \mathcal{M}$ is said to be (Y, M)-harmonic if

$$\mathbb{E}_{(y,m)}[h(Y_1, M_1)] = h(y, m), (y, m) \in \mathbb{Z} \times \mathbb{Z}_+ \times \mathcal{M}; \tag{8}$$

if we replace $=$ with $\geqslant$ [$\leqslant$], h is said to be (Y, M)-subharmonic [superharmonic].

For the case $|\mathcal{M}| = 1$ (i.e., no modulation), [10, 11] use Y-harmonic functions which are linear combinations of exponential functions

$$y \mapsto [(\beta, \alpha), y] = \beta^{y(1)-y(2)} \alpha^{y(2)}, (\beta, \alpha) \in \mathbb{C}, \tag{9}$$

and (β, α) lies on a characteristic surface associated with the process. Markov modulation introduces an additional state variable m, which leads to the following generalization of (9)

$$(y, m) \mapsto \beta^{y(1)-y(2)} \alpha^{y(2)} \boldsymbol{d}(m), \tag{10}$$

where $\boldsymbol{d} : \mathcal{M} \mapsto \mathbb{C}$ is an arbitrary function on $\mathcal{M}$. Let $[(\beta, \alpha, \boldsymbol{d}), \cdot]$ denote the function given (10). We would like to choose $(\beta, \alpha, \boldsymbol{d})$ so that $[(\beta, \alpha, \boldsymbol{d}), \cdot]$ is (Y, M)-harmonic at least over the interior $\mathbb{Z} \times \mathbb{Z}_+^o$. To this end, introduce the local characteristic polynomial for the modulating state $m \in \mathcal{M}$:

$$\boldsymbol{p}(\beta, \alpha, m) \doteq \lambda(m)\frac{1}{\beta} + \mu_1(m)\alpha + \mu_2(m)\frac{\beta}{\alpha}; \tag{11}$$

To define the global characteristic polynomial introduce the $|\mathcal{M}| \times |\mathcal{M}|$ matrix $\boldsymbol{A}$:

$$\boldsymbol{A}(\beta, \alpha)_{m_1, m_2} \doteq \begin{cases} \boldsymbol{P}(m_1, m_2), & m_1 \neq m_2, \\ \boldsymbol{P}(m_1, m_1)\boldsymbol{p}(\beta, \alpha, m), & m_1 = m_2, \end{cases}$$

$(m_1, m_2) \in \mathcal{M}^2$. Let $\boldsymbol{I}$ denote the $|\mathcal{M}| \times |\mathcal{M}|$ identity matrix. Attempting to find functions of the form $[(\beta, \alpha, \boldsymbol{d}), \cdot]$ that satisfy (8) leads to the following characteristic equation:

$$\boldsymbol{A}(\beta, \alpha)\boldsymbol{d} = \boldsymbol{d}, \tag{12}$$

i.e,

$$\boldsymbol{p}(\beta, \alpha) \doteq \det(\boldsymbol{I} - \boldsymbol{A}(\beta, \alpha)) = 0, \tag{13}$$

and $\boldsymbol{d}$ is an eigenvector of $\boldsymbol{A}(\beta, \alpha)$ for the eigenvalue 1. The $\boldsymbol{p}(\cdot, \cdot)$ of (13) is the global characteristic polynomial for the modulated process (Y, M). Define the characteristic surface for the interior:

$$\mathcal{H} \doteq \left\{ (\beta, \alpha, \boldsymbol{d}) \in \mathbb{C}^{2+|\mathcal{M}|} : \boldsymbol{A}(\beta, \alpha)\boldsymbol{d} = \boldsymbol{d}, \ \boldsymbol{d} \neq 0 \right\}. \tag{14}$$

Points on $\mathcal{H}$ give us (Y, M)-harmonic functions on $\mathbb{Z} \times \mathbb{Z}_+^o$.

Proposition 2.1. *If* $(\beta, \alpha, \boldsymbol{d}) \in \mathcal{H}$ *then* $[(\beta, \alpha, \boldsymbol{d}), \cdot]$ *satisfies* (8) *for* $y \in \mathbb{Z} \times \mathbb{Z}_+^o$.

Proof. By definition

$$\mathbb{E}_{(y,m)} \left[(\beta, \alpha, \boldsymbol{d}), (Y_1, M_1) \right]$$
$$= \sum_{n \in \mathcal{M}, n \neq m} \boldsymbol{P}(m, n)[(\beta, \alpha, \boldsymbol{d}), (y, n)]$$
$$+ \boldsymbol{P}(m, m)(\lambda(m)[(\beta, \alpha, \boldsymbol{d}), (y + (-1, 0), m)] + \mu_1(m)[(\beta, \alpha, \boldsymbol{d}), (y + (1, 1), m)]$$
$$+ \mu_2(m)[(\beta, \alpha, \boldsymbol{d}), (y + (0, -1), m)])$$

Expand $[(\beta, \alpha, \boldsymbol{d}), ((y + v), m)]$ terms:

$$= \sum_{n \in \mathcal{M}, n \neq m} \boldsymbol{P}(m, n)[(\beta, \alpha, \boldsymbol{d}), (y, n)]$$
$$+ \boldsymbol{P}(m, m)(\lambda(m)\beta^{y(1)-y(2)-1}\alpha^{y(2)}\boldsymbol{d}(m) + \mu_1(m)\beta^{y(1)-y(2)}\alpha^{y(2)+1}\boldsymbol{d}(m)$$
$$+ \mu_2(m)\beta^{y(1)-y(2)+1}\alpha^{y(2)-1}\boldsymbol{d}(m))$$

Factor out $[(\beta, \alpha, \boldsymbol{d}), (y, m)]$ from the last three terms:

$$= \sum_{n \in \mathcal{M}, n \neq m} \boldsymbol{P}(m, n)[(\beta, \alpha, \boldsymbol{d}), (y, n)] + \boldsymbol{P}(m, m)[(\beta, \alpha, \boldsymbol{d}), (y, m)]\boldsymbol{p}(\beta, \alpha, m)$$
$$= \beta^{y(1)-y(2)}\alpha^{y(2)} \left(\sum_{n \in \mathcal{M}, n \neq m} \boldsymbol{P}(m, n)\boldsymbol{d}(n) + \boldsymbol{P}(m, m)\boldsymbol{d}(m)\boldsymbol{p}(\beta, \alpha, m) \right).$$

The expression in parenthesis equals the mth term of the vector $\boldsymbol{A}(\beta, \alpha)\boldsymbol{d}$, which equals $\boldsymbol{d}(m)$ because $(\beta, \alpha, \boldsymbol{d}) \in \mathcal{H}$ means $\boldsymbol{A}(\beta, \alpha)\boldsymbol{d} = \boldsymbol{d}$. Therefore,

$$= \beta^{y(1)-y(2)}\alpha^{y(2)}\boldsymbol{d}(m) = [(\beta, \alpha, \boldsymbol{d}), (y, m)].$$

This proves the claim of the proposition. $\square$

The previous proposition gives us (Y, M)-harmonic functions on $\mathbb{Z} \times \mathbb{Z}_+^o$. we next study the geometry of $\mathcal{H}$, this will be useful in defining (Y, M)-(super/sub) harmonic functions over all of $\mathbb{Z} \times \mathbb{Z}_+$.

2.1 Geometry of the characteristic surface

Define $\mathcal{H}^{\beta\alpha}$, the projection of $\mathcal{H}$ onto its first two dimensions:

$$\mathcal{H}^{\beta\alpha} \doteq \{(\beta, \alpha) \in \mathbb{C}^2 : \boldsymbol{p}(\beta, \alpha) = 0\};$$

we will to refer to $\mathcal{H}^{\beta\alpha}$ as the characteristic surface for the interior as well, which is justified by the next lemma; its proof follows from basic linear algebra.

Lemma 2.1. *For each $(\beta, \alpha) \in \mathcal{H}^{\beta\alpha}$ there is at least one parameter family of points $\{(\beta, \alpha, c\boldsymbol{d}), \ c \in \mathbb{C} - \{0\}\} \subset \mathcal{H}$, for some $\boldsymbol{d} \in \mathbb{C}^{|\mathcal{M}|} - \{0\}$. Conversely, for each $(\beta, \alpha, \boldsymbol{d}) \in \mathcal{H}$, we have $(\beta, \alpha) \in \mathcal{H}^{\beta\alpha}$. Furthermore, all points on $\mathcal{H}$ can be obtained from those on $\mathcal{H}^{\beta\alpha}$.*

$\beta^{|\mathcal{M}|}\alpha^{|\mathcal{M}|}\boldsymbol{p}$ is a polynomial of degree $3|\mathcal{M}|$ in (β, α), which makes, in general, the analysis of the geometry of $\mathcal{H}^{\beta\alpha}$ nontrivial. A natural approach to the study of the geometry of this curve is through the eigenvalues of $\boldsymbol{A}(\beta, \alpha)$. The next two propositions show that the curve $\mathcal{H}^{\beta\alpha}$ decomposes into $|\mathcal{M}|$ distinct pieces over any region D where $\boldsymbol{A}(\beta, \alpha)$ has simple eigenvalues.

Proposition 2.2. *Let $D \subset \mathbb{C}^2$ or $D \subset \mathbb{R}^2$ be open and simply connected and suppose $\boldsymbol{A}(\beta, \alpha)$ has simple eigenvalues for all $(\beta, \alpha) \in D$. Then the eigenvalues of $\boldsymbol{A}$ can be written as $|\mathcal{M}|$ distinct smooth functions $\Lambda_j(\beta, \alpha)$ on D.*

Proof. The argument is the same for both real and complex variables. That the eigenvalues Λ_j can be defined smoothly in a neighborhood of any $(\beta, \alpha) \in D$ follows from [8, Theorem 5.3] and the assumption that they are distinct. Once defined locally, one extends them to all of D through continuous extension, which is possible because D is simply connected. $\square$

Most of our analysis will be based on $(\beta, \alpha) \in D = \mathbb{R}_+^{2,o} \doteq \mathbb{R}_+^2 - \partial_1 \cup \partial_2$. For $(\beta, \alpha) \in \mathbb{R}_+^{2,o}$, $\boldsymbol{A}(\beta, \alpha)$ is an irreducible matrix with positive entries. Perron-Frobenius Theorem implies that $\boldsymbol{A}(\beta, \alpha)$ has a simple positive eigenvalue dominating all of the other eigenvalues in absolute value with an eigenvector with strictly positive entries. $\Lambda_1(\beta, \alpha)$ will always denote this largest eigenvalue. Furthermore, if $\boldsymbol{A}(\beta, \alpha)$ has distinct real eigenvalues for $(\beta, \alpha) \in \mathbb{R}_+^{2,o}$, we will label them so that

$$\Lambda_j(\beta, \alpha) > \Lambda_i(\beta, \alpha), \text{ for } j < i,$$

i.e., the eigenvalues are assumed to be sorted in descending order.

For D and Λ_j as in Proposition 2.2 define

$$\mathcal{L}_j^D \doteq \{(\beta, \alpha) \in D : \Lambda_j(\beta, \alpha) = 1\}.$$

The last proposition implies

Proposition 2.3. *Let D and Λ_j, $j = 1, 2, 3, ..., |\mathcal{M}|$, be as in Proposition 2.2. Then*

$$D \cap \mathcal{H}^{\beta\alpha} = \sqcup_{j=1}^{|\mathcal{M}|} \mathcal{L}_j^D \tag{15}$$

where $\sqcup$ denotes disjoint union.

The proof follows from the definitions involved. For $D = \mathbb{R}_+^{2,o}$ we will omit the superscript D and write $\mathcal{L}_j$ instead of $\mathcal{L}_j^{\mathbb{R}_+^{2,o}}$.

If $\boldsymbol{A}(\beta, \alpha)$ has simple real eigenvalues for $(\beta, \alpha) \in \mathbb{R}_+^{2,o}$ we can define

$$\mathcal{R}_j \doteq \{(\beta, \alpha) \in \mathbb{R}_+^{2,o} : \Lambda_j(\beta, \alpha) \leqslant 1\}.$$

The continuity of Λ_j implies $\mathcal{L}_j = \partial\mathcal{R}_j$.

8

Proposition 2.4. *Suppose $\boldsymbol{A}(\beta, \alpha)$ has simple real eigenvalues for $(\beta, \alpha) \in \mathbb{R}_+^{2,o}$. Then the curve $\mathcal{L}_j$ is strictly contained inside the curve $\mathcal{L}_{j+1}$ for $j = 1, 2, ..., |\mathcal{M}| - 1$.*

Proof. All diagonal entries of $\boldsymbol{A}(\beta, \alpha)$ tend to ∞ when $(\beta, \alpha) \to \partial \mathbb{R}_+^{2,o}$. This and Gershgorin's Theorem [5, Appendix 7], imply $\Lambda_j(\beta, \alpha) \to \infty$ for $(\beta, \alpha) \to \partial \mathbb{R}_+^{2,o}$. This implies in particular that $\mathcal{R}_j$ is a compact subset of $\mathbb{R}_+^{2,o}$. Secondly, $\Lambda_{j+1} < \Lambda_j$ implies $\mathcal{R}_j \subset \mathcal{R}_{j+1}$; the compactness of these sets, the strictness of the inequality $\Lambda_{j+1}(\beta, \alpha) < \Lambda_j(\beta, \alpha)$ imply that $\partial \mathcal{R}_j = \mathcal{L}_j$ lies strictly within $\mathcal{R}_{j+1}$ with strictly positive distance from the boundary $\mathcal{L}_{j+1}$ of $\mathcal{R}_{j+1}$; this proves the claim of the proposition. $\square$

In Sections 7 and 8 we will employ the following assumption and the above decomposition of $\mathcal{H}^{\beta\alpha}$ to identify points on $\mathcal{H}^{\beta\alpha}$ to be used in the construction of (Y, M)-harmonic functions:

Assumption 1. $\boldsymbol{A}(\beta, \alpha)$ *has real distinct eigenvalues for* $(\beta, \alpha) \in \mathbb{R}_+^{2,o}$.

To show that Assumption 1 is not vacuous, we now give a class of matrices $\boldsymbol{A}$ that satisfies it. The following definitions are from [1, page 57]: a matrix is said to be totally nonnegative (totally positive) if all of its minors of any degree are nonnegative (positive). A totally nonnegative matrix is said to be *oscillatory* if some positive integer power of the matrix is totally positive. If $\boldsymbol{A}$ is oscillatory, Assumption 1 holds:

Proposition 2.5. *Suppose $\boldsymbol{A}(\beta, \alpha)$ is an oscillatory matrix for all $(\beta, \alpha) \in \mathbb{R}_+^{2,o}$, then $\boldsymbol{A}(\beta, \alpha)$ has $|\mathcal{M}|$ distinct eigenvalues over $\mathbb{R}_+^{2,o}$.*

This proposition is a basic fact on oscillatory matrices [1, (6.28)]. [1, (6.26)] identifies a particularly simple class of oscillatory matrices:

Proposition 2.6. *Suppose $\boldsymbol{G}(1, 1)$, $\boldsymbol{G}(1, 2)$, $\boldsymbol{G}(|\mathcal{M}| - 1, |\mathcal{M}|)$, $\boldsymbol{G}(|\mathcal{M}|, |\mathcal{M}|)$ and $\boldsymbol{G}(j, j - 1)$, $\boldsymbol{G}(j, j)$, $\boldsymbol{G}(j, j + 1)$, $j = 2, 3, ..., |\mathcal{M}| - 1$ are all strictly positive and the rest of the components of $\boldsymbol{G}$ are all zero, i.e., $\boldsymbol{G}$ is tridiagonal with strictly positive entries. Then $\boldsymbol{G}$ is an oscillatory matrix.*

We will call any tridiagonal matrix with strictly positive entries on the three diagonals "strictly tridiagonal." By the above proposition any strictly tridiagonal matrix is oscillatory. In particular, if the transition matrix $\boldsymbol{P}$ is strictly tridiagonal, $\boldsymbol{A}(\beta, \alpha)$ will also be of the same form for all $(\beta, \alpha) \in \mathbb{R}_+^2$; therefore, for such $\boldsymbol{P}$ Assumption 1 holds.

The decomposition of $\mathcal{H}^{\beta\alpha} \cap \mathbb{R}_+^{2,o}$ into $\mathcal{L}_j$ is shown in Figure 3 for the transition matrix

$$\boldsymbol{P} = \begin{pmatrix} 0.6 & 0.4 & 0 \\ 0.1 & 0.4 & 0.5 \\ 0 & 0.2 & 0.8 \end{pmatrix}. \tag{16}$$

The matrix $\boldsymbol{P}$ of (16) is strictly tridiagonal; therefore, Proposition 2.6 applies and $\boldsymbol{A}(\beta, \alpha)$ has distinct real eigenvalues for all $(\beta, \alpha) \in \mathbb{R}_+^{2,o}$ and we have the decomposition (15) of $\mathcal{H}^{\beta\alpha} \cap \mathbb{R}_+^{2,o}$ given by Propositions 2.3 and 2.4; Figure 3 shows $\mathcal{H}^{\beta\alpha}$ and its components $\mathcal{L}_j$; the jump probabilities for this example are

$$\begin{pmatrix} 0.1 & 0.4 & 0.5 \\ 0.12 & 0.41 & 0.47 \\ 0.09 & 0.39 & 0.52 \end{pmatrix} \tag{17}$$

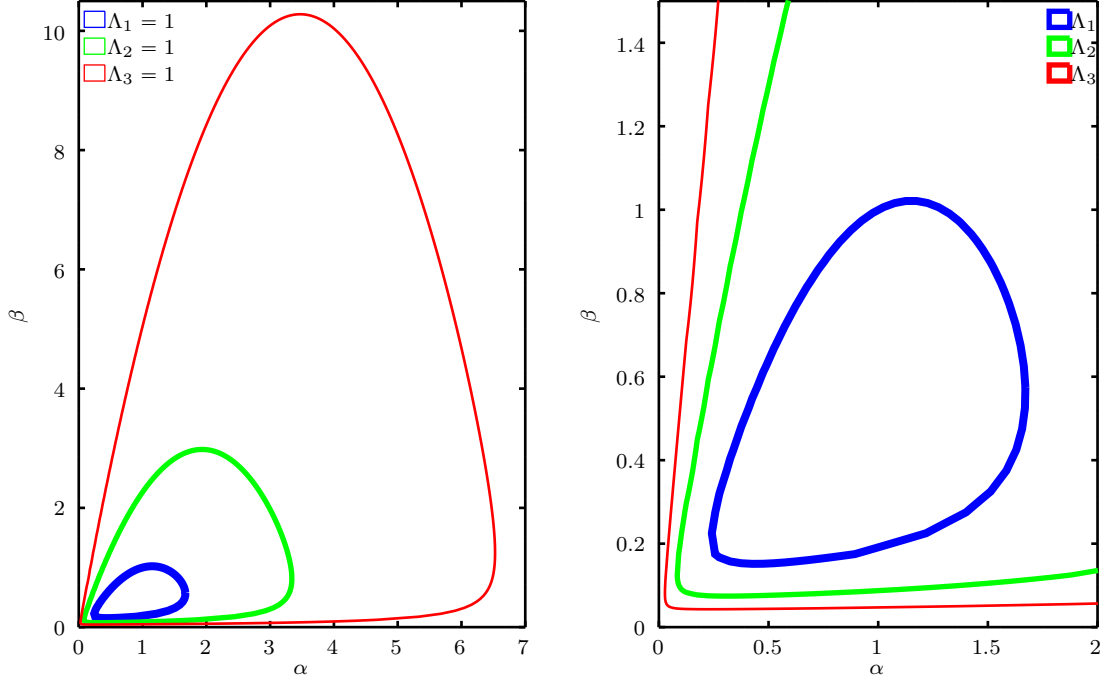where the ith row equals $(\lambda(i), \mu_1(i), \mu_2(i))$.

9

Figure 3: Real section of the characteristic surface $\mathcal{H}^{\beta\alpha} = \cup_{j=1}^3 \mathcal{L}_j$ for the parameter values given in (16) and (17). On the right: detailed graph around the origin

By Proposition 2.1 and Lemma 2.1 each point on any of the curves depicted in Figure 3 gives a Y-harmonic function on $\mathbb{Z} \times \mathbb{Z}_+^o$. Most of our analysis will be based on points on the innermost curve $\mathcal{L}_1$, the 1-level curve of the largest eigenvalue Λ_1; before identifying the relevant points, let us look at two different methods of constructing (Y, M)-(super)harmonic functions from points on $\mathcal{H}^{\beta\alpha}$.

2.2 Construction of (Y, M)-harmonic and superharmonic functions

We can proceed in two ways to get functions that satisfy
$\mathbb{E}_{(y,m)}[h(Y_1, M_1)] = h(y, m)$ or $\mathbb{E}_{(y,m)}[h(Y_1, M_1)] \leqslant h(y, m)$ for $y \in \partial_2$ as well as the interior. The first is by defining the characteristic polynomial $\boldsymbol{p}_2$, the boundary matrix $\boldsymbol{A}_2$, and the boundary surface $\mathcal{H}_2$ associated with ∂_2 and using points on $\mathcal{H} \cap \mathcal{H}_2$:

$$\boldsymbol{p}_2(\beta, \alpha, m) \doteq \lambda(m)\frac{1}{\beta} + \mu_1(m)\alpha + \mu_2(m), \quad m \in \mathcal{M},$$

$$\boldsymbol{A}_2(\beta, \alpha)_{m_1, m_2} \doteq \begin{cases} \boldsymbol{P}(m_1, m_2), & m_1 \neq m_2 \\ \boldsymbol{P}(m_1, m_1)\boldsymbol{p}_2(\beta, \alpha, m), & m_1 = m_2, \end{cases}, (m_1, m_2) \in \mathcal{M}^2,$$

$$\mathcal{H}_2 \doteq \left\{ (\beta, \alpha, \boldsymbol{d}) \in \mathbb{C}^{2+|\mathcal{M}|} : \boldsymbol{A}_2(\beta, \alpha))\boldsymbol{d} = \boldsymbol{d}, \boldsymbol{d} \neq 0 \right\}. \tag{18}$$

Define $\Lambda_{2,1}(\beta, \alpha)$ to be the largest eigenvalue of $\boldsymbol{A}_2(\beta, \alpha)$. Parallel to the interior case, define

$$\mathcal{H}_2^{\beta\alpha} \doteq \{(\beta, \alpha) \in \mathbb{C}^2 : \boldsymbol{p}_2(\beta, \alpha) = 0\},$$
$$\mathcal{L}_{2,1} \doteq \{(\beta, \alpha) \in \mathbb{R}_+^2 : \Lambda_{2,1}(\beta, \alpha) = 1\}.$$

Proposition 2.7. $[(\beta, \alpha, \boldsymbol{d}), \cdot]$ *is* (Y, M)*-harmonic if* $(\beta, \alpha, \boldsymbol{d}) \in \mathcal{H} \cap \mathcal{H}_2$.

Proof. Proposition 2.1 says that for $(\beta, \alpha, \boldsymbol{d}) \in \mathcal{H}$, $[(\beta, \alpha, \boldsymbol{d}), \cdot]$ satisfies the harmonicity condition when $y \in \mathbb{Z} \times \mathbb{Z}_+ - \partial_2$. Similar to the proof of Proposition 2.1, we would like to show that $[(\beta, \alpha, \boldsymbol{d}), \cdot]$ is (Y, M)-harmonic on ∂_2 when $(\beta, \alpha, \boldsymbol{d}) \in \mathcal{H}_2$. By definition

$$\begin{aligned}
&\mathbb{E}_{(y,m)}\left[(\beta, \alpha, \boldsymbol{d}), (Y_1, M_1)\right] \\
&= \sum_{n \in \mathcal{M}, n \neq m} \boldsymbol{P}(m, n)[(\beta, \alpha, \boldsymbol{d}), (y, n)] \\
&\quad + \boldsymbol{P}(m, m)(\lambda(m)[(\beta, \alpha, \boldsymbol{d}), (y + (-1, 0), m)] + \mu_1(m)[(\beta, \alpha, \boldsymbol{d}), (y + (1, 1), m] \\
&\quad + \mu_2(m)[(\beta, \alpha, \boldsymbol{d}), (y, m)]) \\
&= \sum_{n \in \mathcal{M}, n \neq m} \boldsymbol{P}(m, n)[(\beta, \alpha, \boldsymbol{d}), (y, n)] \\
&\quad + \boldsymbol{P}(m, m)(\lambda(m)\beta^{y(1)-1}\boldsymbol{d}(m) + \mu_1(m)\beta^{y(1)}\alpha\boldsymbol{d}(m) + \mu_2(m)\beta^{y(1)}\boldsymbol{d}(m)) \\
&= \sum_{n \in \mathcal{M}, n \neq m} \boldsymbol{P}(m, n)[(\beta, \alpha, \boldsymbol{d}), (y, n)] + \boldsymbol{P}(m, m)[(\beta, \alpha, \boldsymbol{d}), (y, m)]\boldsymbol{p}_2(\beta, \alpha, m) \\
&= \beta^{y(1)}\left(\sum_{n \in \mathcal{M}, n \neq m} \boldsymbol{P}(m, n)\boldsymbol{d}(n) + \boldsymbol{P}(m, m)\boldsymbol{d}(m)\boldsymbol{p}_2(\beta, \alpha, m)\right).
\end{aligned}$$

The expression in parenthesis equals the mth term of the vector $\boldsymbol{A}_2(\beta, \alpha))\boldsymbol{d}$, which equals $\boldsymbol{d}(m)$ because $(\beta, \alpha, \boldsymbol{d}) \in \mathcal{H}_2$ means $\boldsymbol{A}_2(\beta, \alpha))\boldsymbol{d} = \boldsymbol{d}$. Therefore,

$$= \beta^{y(1)}\boldsymbol{d}(m) = [(\beta, \alpha, \boldsymbol{d}), (y, m)].$$

This argument and Proposition 2.1 prove the claim of the proposition. $\square$

The real sections of $\mathcal{H}^{\beta\alpha}$ and $\mathcal{H}_2^{\beta\alpha}$ are 1 dimensional curves and their intersection will in general consist of finitely many points. In the analysis of the tandem walk with no modulation, these points can easily be identified explicitly. There turn out to be three of them, of which only one is nontrivial (i.e., different from 0 and 1). In the present case, there will in general be $3|\mathcal{M}| - 2$ nontrivial points on $\mathcal{H}^{\beta\alpha} \cap \mathcal{H}_2^{\beta\alpha}$; one of these which lies on $\mathcal{L}_1 \cap \mathcal{L}_{2,1}$ can be identified using the implicit function theorem and the stability assumption (1); this point and the (Y, M)-harmonic function it defines are given in Proposition 2.8 and 2.9 below. For the argument we need two auxiliary linear algebra results, Lemmas A.1 and A.2 given in the appendix.

Proposition 2.8. *Under the stability assumption (1) there exists unique* $0 < \rho_1 < 1$ *such that* $(\rho_1, \rho_1) \in \mathcal{L}_1 \cap \mathcal{L}_{2,1} \subset \mathcal{H}^{\beta\alpha} \cap \mathcal{H}_2^{\beta\alpha}$, *i.e.,* 1 *is the largest eigenvalue of* $\boldsymbol{A}(\rho_1, \rho_1)$ *and* $\boldsymbol{A}_2(\rho_1, \rho_1)$.

Proof. For $q \in \mathbb{R}^2$ define

$$H(q) \doteq -\log \Lambda_1 \left(e^{q(1)}, e^{q(2)} \right). \tag{19}$$

By [9, Lemma 4.2, 4.3], H is convex in q. Proceeding parallel to [9, Proof of Lemma 4.4, page 515] define $f(\Lambda, r) \doteq \det(\Lambda \boldsymbol{I} - \boldsymbol{A}(e^r, e^r))$. We know that $f(\Lambda_1(e^r, e^r), r) = 0$ for $r \in \mathbb{R}$. To prove our proposition, we will apply the implicit function theorem to f at $(1,0)$ to prove that $r \mapsto \Lambda_1(e^r, e^r)$ is strictly increasing at $r = 0$. Differentiating f at $(1,0)$ with respect to r gives

$$\left. \frac{\partial f}{\partial r} \right|_{(1,0)} = \sum_{m \in \mathcal{M}} (\lambda(m) - \mu_1(m)) \boldsymbol{P}(m,m) \det(\boldsymbol{I} - \boldsymbol{P})^{m,m},$$

which equals, by Lemma A.2, for some constant $c > 0$,

$$= c \sum_{m \in \mathcal{M}} (\lambda(m) - \mu_1(m)) \boldsymbol{P}(m,m) \boldsymbol{\pi}(m)$$

$$< 0$$

where the last inequality follows from the stability assumption (1). Similarly, differentiation of f at $(1,0)$ with respect to Λ gives:

$$\left. \frac{\partial f}{\partial \Lambda} \right|_{(1,0)} = 1.$$

This implies that the implicit function theorem is applicable to f; the last two display give:

$$\frac{d}{dr} \Lambda_1(e^r, e^r)|_{(0,0)} > 0.$$

On the other hand, Gershgorin's Theorem implies $\Lambda_1(e^r, e^r) \to \infty$ as $r \to -\infty$ (because of the $\lambda(m)/\beta$ term appearing in the diagonal terms of $\boldsymbol{A}$, tending to $+\infty$ with $\beta = e^r$). To sum up: we have that $\Lambda_1(e^r, e^r)$ is strictly monotone at $r = 0$ (decreases when r decreases) and it tends to infinity as $r \to -\infty$. Then, by the continuity of Λ_1, there must exist at least one point in $(-\infty, 0)$ where $\Lambda_1(e^r, e^r)$ takes the value 1; the convexity of H implies that such a point is unique, i.e., there is a unique point $r^* < 0$ such that $\Lambda_1(e^{r^*}, e^{r^*}) = 1$. Setting $\rho_1 = e^{r^*}$ proves the proposition. □

Let $\boldsymbol{d}_1$ be an eigenvector of $\boldsymbol{A}(\rho_1, \rho_1)$ corresponding to the eigenvalue 1; because 1 is the largest eigenvalue of $\boldsymbol{A}(\rho_1, \rho_1)$ and because $\boldsymbol{A}(\rho_1, \rho_1)$ is irreducible and aperiodic, we can choose $\boldsymbol{d}_1$ so that all of its components are strictly positive. The point $(\rho_1, \rho_1, \boldsymbol{d}_1) \in \mathcal{H} \cap \mathcal{H}_2$ and Proposition 2.7 give us our first (Y, M)-harmonic function:

Proposition 2.9.

$$h_{\rho_1} \doteq [(\rho_1, \rho_1, \boldsymbol{d}_1), \cdot] \tag{20}$$

is (Y, M)-harmonic.

The second way of obtaining (Y, M)-harmonic functions is through conjugate points on $\mathcal{H}^{\beta\alpha}$. The function $\alpha^{|\mathcal{M}|} \boldsymbol{p}$ is a polynomial of degree $2|\mathcal{M}|$ in α. By the fundamental of theorem of algebra, $\alpha^{|\mathcal{M}|} \boldsymbol{p}$ has $2|\mathcal{M}|$ roots, $\boldsymbol{\alpha}_1(\beta), ..., \boldsymbol{\alpha}_2(\beta), ..., \boldsymbol{\alpha}_{2|\mathcal{M}|}(\beta)$, in $\mathbb{C}$ for each fixed $\beta \in \mathbb{C}$; points $(\beta, \boldsymbol{\alpha}_i) \in \mathcal{H}^{\beta\alpha}$, $i = 1, 2, ..., 2|\mathcal{M}|$ are said to be conjugate points. In the non-modulated case, i.e., when $|\mathcal{M}| = 1$, $\alpha \boldsymbol{p}$ is only of second order, therefore, the conjugate points come

in pairs, and given one of the points in the pair, the other can be computed easily; in the modulated case, there are obviously no simple formulas to obtain all of the conjugate points given one among them, because computation of conjugate points involves finding the roots of a polynomial of degree $2|\mathcal{M}|$.

For $(\beta, \alpha, \boldsymbol{d}) \in \mathcal{H}$ define

$$\boldsymbol{c}(\beta, \alpha, \boldsymbol{d}) \in \mathbb{C}^{\mathcal{M}}, \ \boldsymbol{c}(\beta, \alpha, \boldsymbol{d})(m) \doteq \boldsymbol{P}(m, m)\mu_2(m)\boldsymbol{d}(m)\left(1 - \frac{\beta}{\alpha}\right). \tag{21}$$

One can take linear combinations of functions defined by conjugate points to define (Y, M)-harmonic functions. This is based on the following lemma

Lemma 2.2. *Suppose* $(\beta, \alpha, \boldsymbol{d}) \in \mathcal{H}$. *Then, for* $(y, m) \in \partial_2 \times \mathcal{M}$,

$$\mathbb{E}_{(y,m)}\left[(\beta, \alpha, \boldsymbol{d}), (Y_1, M_1)\right] - \left[(\beta, \alpha, \boldsymbol{d}), (y, m)\right] = \beta^{y(1)}\boldsymbol{c}(\beta, \alpha, \boldsymbol{d})(m), \tag{22}$$

where $\boldsymbol{c}$ *is defined as in* (21).

Proof. The computation in the proof of Proposition 2.1 gives

$$\mathbb{E}_{(y,m)}\left[(\beta, \alpha, \boldsymbol{d}), (Y_1, M_1)\right] \tag{23}$$
$$= \beta^{y(1)}\left(\sum_{n \in \mathcal{M}, n \neq m} \boldsymbol{P}(m, n)\boldsymbol{d}(n) + \boldsymbol{P}(m, m)\boldsymbol{d}(m)\boldsymbol{p}_2(\beta, \alpha, m)\right).$$

On the other hand, $(\beta, \alpha, \boldsymbol{d}) \in \mathcal{H}$ means

$$\left[(\beta, \alpha, \boldsymbol{d}), (y, m)\right] \tag{24}$$
$$= \beta^{y(1)}\boldsymbol{d}(m) = \beta^{y(1)}\left(\sum_{n \in \mathcal{M}, n \neq m} \boldsymbol{P}(m, n)\boldsymbol{d}(n) + \boldsymbol{P}(m, m)\boldsymbol{d}(m)\boldsymbol{p}(\beta, \alpha, m)\right).$$

Subtracting the last display from (23) gives

$$\mathbb{E}_{(y,m)}[h(Y_1, M_1)] - h(y, m)$$
$$= \beta^{y(1)}\boldsymbol{P}(m, m)\mu_2(m)\boldsymbol{d}(m)\left(1 - \frac{\beta}{\alpha}\right) = \beta^{y(1)}\boldsymbol{c}(\beta, \alpha, \boldsymbol{d})(m),$$

which proves (22). $\qquad\square$

We now identify a family of (Y, M)-harmonic functions constructed from conjugate points on $\mathcal{H}$:

Proposition 2.10. *For* $\beta \in \mathbb{C}$ *let* (β, α_i, d_i) $i = 1, 2, ..., l \leqslant 2|\mathcal{M}|$ *be distinct conjugate points on* $\mathcal{H}$. *Take any subcollection* $\{i_1, i_2, ..., i_k\}$, $k \leqslant l$ *such that* $\boldsymbol{c}(\beta, \alpha_{i_j}, d_{i_j})$ *are linearly dependent, i.e., there exists* $b \in \mathbb{C}^k$, $b \neq 0$, *such that*

$$\sum_{j=1}^{k} b(j)\boldsymbol{c}(\beta, \alpha_{i_j}, d_{i_j}) = 0. \tag{25}$$

Then

$$h(y, m) = \sum_{j=1}^{k} b(j)[(\beta, \alpha_{i_j}, d_{i_j}), \cdot] \tag{26}$$

is (Y, M)-*harmonic.*

Proof. We already know from Proposition 2.1, harmonic functions of the form $[(\beta, \alpha_i, d_i), \cdot]$ are (Y, M)-harmonic in the interior $\mathbb{Z} \times \mathbb{Z}_+ - \partial_2$. So, their linear combinations are also (Y, M)-harmonic in the interior and we need to check the harmonicity for $y \in \partial_2$. By Lemma 2.2

$$\mathbb{E}_{(y,\cdot)}\left[(\beta, \alpha_i, d_i), (Y_1, M_1)\right] - \left[(\beta, \alpha_i, d_i), (y, \cdot)\right] = \beta^{y(1)} \boldsymbol{c}(\beta, \alpha_i, d_i).$$

Taking linear combinations of these with weight vector b gives:

$$\mathbb{E}_{(y,\cdot)}\left[h(Y_1, M_1)\right] - h(y, \cdot) = \beta^{y(1)} \left(\sum_{j=1}^{k} b(j) \boldsymbol{c}(\beta, \alpha_{i_j}, d_{i_j})\right)$$

which equals $0 \in \mathbb{R}^{|\mathcal{M}|}$ by (25). This proves that h is (Y, M)-harmonic on ∂_2. $\qquad\square$

For any $\beta \in \mathbb{C}$ such that $\boldsymbol{p}(\beta, \alpha) = 0$ has distinct roots, $\alpha_1, \alpha_2, ..., \alpha_{2|\mathcal{M}|}$, all different from β, we have, by definition, $\boldsymbol{c}(\beta, \alpha_j, d_j) \neq 0$ for all $j = 1, 2, ..., 2|\mathcal{M}|$. Therefore, for such β, and for any subcollection $\alpha_{j_1}, \alpha_{j_2}, ..., \alpha_{j_k}$, with $k \geq |\mathcal{M}| + 1$, we can find a nonzero vector b satisfying (25).

We will call a (Y, M)-harmonic function ∂B-determined if it of the form,

$$(y, m) \mapsto \mathbb{E}_{(y,m)}[f(Y_\tau, M_\tau) 1_{\{\tau < \infty\}}]$$

for some function f. The function $(y, m) \mapsto \mathbb{P}_{(y,m)}(\tau < \infty)$ is the unique ∂B-determined (Y, M)-harmonic function taking the value 1 on ∂B. Among the functions of the form $[(\beta, \alpha, \boldsymbol{d}), \cdot]$, the closest we get to this type of behavior is when $\alpha = 1$: for $\alpha = 1$, $[(\beta, 1, \boldsymbol{d}), (y, m)]$ depends only on m for $y \in \partial B$. Therefore, $\alpha = 1$ play a key role in computing/approximating $\mathbb{P}_{(y,m)}(\tau < \infty)$. The next proposition identifies a point on $\mathcal{L}_1$ of the form $(\rho_2, \alpha = 1)$ with $0 < \rho_2 < 1$.

Proposition 2.11. *Under assumption (1) there exists $0 < \rho_2 < 1$ such that $(\rho_2, 1) \in \mathcal{L}_1 \subset \mathcal{H}^{\beta\alpha}$; i.e., 1 is the largest eigenvalue of $\boldsymbol{A}(\rho_2, 1)$.*

Proof. The proof is parallel to that of Proposition 2.8. We now define $f(\Lambda, r) = \det(\Lambda \boldsymbol{I} - \boldsymbol{A}(e^r, 1))$ and observe, by assumption (1) and Lemma A.2,

$$\left.\frac{\partial f}{\partial r}\right|_{(1,0)} = \sum_{m \in \mathcal{M}} (\lambda(m) - \mu_2(m)) \boldsymbol{P}(m, m) \det(\boldsymbol{I} - \boldsymbol{P})^{m,m}$$

$$= c \sum_{m \in \mathcal{M}} (\lambda(m) - \mu_2(m)) \boldsymbol{P}(m, m) \boldsymbol{\pi}(m) < 0$$

for some constant $c > 0$. The rest of the proof proceeds as in the proof of Proposition 2.8. $\qquad\square$

Recall that $(\rho_2, 1) \in \mathcal{L}_1$, i.e., 1 is the largest eigenvalue of $\boldsymbol{A}(\rho_2, 1)$; the irreducibility of $\boldsymbol{A}$ implies that the eigenvectors corresponding to 1 have strictly negative or positive components; let $\boldsymbol{d}_2$ denote a right eigenvector of $\boldsymbol{A}(\rho_2, 1)$ corresponding to the eigenvalue 1 with strictly positive components. Proposition 2.1 and the previous proposition imply that $[(\rho_2, 1, \boldsymbol{d}_2), \cdot]$ is (Y, M)-harmonic on $\mathbb{Z} \times \mathbb{Z}_+ - \partial_2$. All of the prior works ([10, 11, 12]), use a conjugate point of $(\rho_2, 1)$ to construct a Y-harmonic function. In the present case, in general, $(\rho_2, 1)$ will have $2|\mathcal{M}| - 1$ conjugate points. Figure 3 suggests that only one of these conjugate points lies on $\mathcal{L}_1$; we will use $(\rho_2, 1)$ along with this conjugate to define a (Y, M)-*super*harmonic

function. This will be in two steps. Proposition 2.12 identifies the relevant conjugate point; Proposition 2.14 constructs the superharmonic function. We will use the superharmonic function in Sections 3 and 4 below in our analysis of the relative error (7).

The identification of the conjugate point requires the following assumption:

$$\sum_{m \in \mathcal{M}} (\rho_2 \mu_2(m) - \mu_1(m)) \boldsymbol{P}(m,m) \det(\boldsymbol{I} - \boldsymbol{A}(\rho_2, 1))^{m,m} < 0. \tag{27}$$

Remark 2.1 comments on this assumption and Proposition 2.13 gives simple conditions under which (27) holds.

Proposition 2.12. *Let $(\rho_2, 1)$, $\rho_2 \in (0,1)$, be the point on $\mathcal{L}_1$ identified in Proposition 2.11. Then there exists a unique point $(\rho_2, \alpha_1^*) \in \mathcal{L}_1$, $\alpha_1^* \in (0,1)$ if (27) holds.*

Proof. Set $r_2 = \log(\rho_2)$. Proof is parallel to those of Propositions 2.8 and 2.11 and is based on the analysis of the function H of (19) at the point $(r_2, 0)$ via the implicit function theorem. Define $f(\Lambda, r) = \det(\Lambda \boldsymbol{I} - \boldsymbol{A}(\rho_2, e^r))$ and observe

$$\left. \frac{\partial f}{\partial r} \right|_{(r_2, 0)} = \sum_{m \in \mathcal{M}} (\rho_2 \mu_2(m) - \mu_1(m)) \boldsymbol{P}(m,m) \det(\boldsymbol{I} - \boldsymbol{A}(\rho_2, 1))^{m,m},$$

which, by assumption (27), is strictly less than 0. The rest of the proof goes as that of Proposition 2.8. □

Remark 2.1. *Assumption (27) ensures that $(\rho_2, 1)$ has a conjugate point on the principal characteristic surface $\mathcal{L}_1$ with α component less than 1. There is no corresponding assumption in the non-modulated tandem case, because, in that setup, the conjugate of $(\rho_2, 1)$ is (ρ_2, ρ_1) whose α component ρ_1 is always less than 1 by the stability assumption. In the simple constrained random walk case (treated in [12]) the corresponding assumption is $r^2 < \rho_1 \rho_2$ (see [12, Display (14)]).*

The condition $\alpha_1^* < 1$ is needed for the superharmonic function constructed in Proposition 2.14 to be bounded on ∂B, see Proposition 3.2.

Proposition 2.13. *Each of the following conditions is sufficient for (27) to hold:*

1. *$\lambda(m)/\mu_2(m) < 1$, $\lambda(m) < \mu_1(m)$ for all $m \in \mathcal{M}$ and the ratio $\lambda(m)/\mu_2(m)$ does not depend on m,*

2. *$\mu_2(m) < \mu_1(m)$ for all $m \in \mathcal{M}$.*

Proof. If $\lambda(m)/\mu_2(m) < 1$ does not depend on m we can denote the common ratio by $\rho_2' < 1$. Substituting $(\beta, \alpha) = (\rho_2', 1)$ we see that $\boldsymbol{A}(\rho_2', 1) = \boldsymbol{P}$. This implies that the root ρ_2 identified in Proposition 2.11 must equal ρ_2'. Setting $\rho_2 = \rho_2'$ on the left side of (27) gives

$$\sum_{m \in \mathcal{M}} (\rho_2 \mu_2(m) - \mu_1(m)) \boldsymbol{P}(m,m) \det(\boldsymbol{I} - \boldsymbol{A}(\rho_2, 1))^{m,m}$$

$$= \sum_{m \in \mathcal{M}} (\rho_2' \mu_2(m) - \mu_1(m)) \boldsymbol{P}(m,m) \det(\boldsymbol{I} - \boldsymbol{A}(\rho_2, 1))^{m,m}$$

$$= \sum_{m \in \mathcal{M}} (\lambda(m) - \mu_1(m)) \boldsymbol{P}(m,m) \det(\boldsymbol{I} - \boldsymbol{A}(\rho_2, 1))^{m,m}$$

$\det(\boldsymbol{I} - \boldsymbol{A}(\rho_2, 1))^{m,m} > 0$ by Lemma A.1, and $\lambda(m) < \mu_1(m)$ by assumption; these and the last line imply (27):

$$< 0.$$

That the condition $\mu_2(m) < \mu_1(m)$ for all $m \in \mathcal{M}$ implies (27) follows from a similar argument. □

Remark 2.2. *The argument used in the proof above can be used to prove that the conjugate point (ρ_2, α_1^*) satisfies $\alpha_1^* > 1$ if one replaces $<$ with $>$ in (27).*

For the rest of our analysis we will need a further assumption:

$$\rho_1 \neq \rho_2, \tag{28}$$

where ρ_1 is the first (or the second) component of the point on $\mathcal{L}_1 \cap \mathcal{L}_{2,1}$ identified in Proposition 2.8 and ρ_2 is the β component of the point on $\mathcal{L}_1$ identified in Proposition 2.11. Assumption (28) generalizes the assumption $\mu_1 \neq \mu_2$ in [10, 11, 12]. The following lemma identifies sufficient conditions for (28) to hold.

Lemma 2.3. *If $\mu_1(m) > \mu_2(m)$ for all $m \in \mathcal{M}$, or $\mu_1(m) < \mu_2(m)$ for all $m \in \mathcal{M}$, then (28) holds.*

Proof. The matrix $\boldsymbol{D} = \boldsymbol{A}(\rho_2, \rho_2) - \boldsymbol{A}(\rho_2, 1)$ is a diagonal matrix whose m^{th} entry equals $(1 - \rho_2)(\mu_2(m) - \mu_1(m))$. Suppose $\mu_2(m) > \mu_1(m)$ for all $m \in \mathcal{M}$; then $\rho_2 \in (0, 1)$ implies that $\boldsymbol{D}$ has strictly positive entries.

We have then:

$$\begin{aligned}
\boldsymbol{A}(\rho_2, \rho_2)\boldsymbol{d}_2 &= \boldsymbol{A}(\rho_2, 1)\boldsymbol{d}_2 + \boldsymbol{D}\boldsymbol{d}_2 \\
&= \boldsymbol{d}_2 + \boldsymbol{D}\boldsymbol{d}_2 \\
&> (1 + \epsilon)\boldsymbol{d}_2
\end{aligned} \tag{29}$$

for some $\epsilon > 0$; here we have used 1) $\boldsymbol{d}_2$ is an eigenvector of $\boldsymbol{A}(\rho_2, 1)$ corresponding to the eigenvalue 1 and 2) $\boldsymbol{D}$ has strictly positive entries. We know by [5, Proof of Theorem 1, Chapter 16] that

$$\Lambda_1(\boldsymbol{A}(\rho_2, \rho_2)) = \sup\{c : \exists x \in \mathbb{R}_+^{|\mathcal{M}|}, \boldsymbol{A}(\rho_2, \rho_2)x \geqslant cx\}. \tag{30}$$

This and (29) imply that the largest eigenvalue of $\boldsymbol{A}(\rho_2, \rho_2)$ is strictly greater than 1. This implies $\rho_2 < \rho_1$. That $\mu_1(m) > \mu_2(m)$ for all $m \in \mathcal{M}$ implies $\rho_2 > \rho_1$ follows from the same argument applied to $\boldsymbol{A}(\rho_2, 1)\boldsymbol{d}_{2,1}$. □
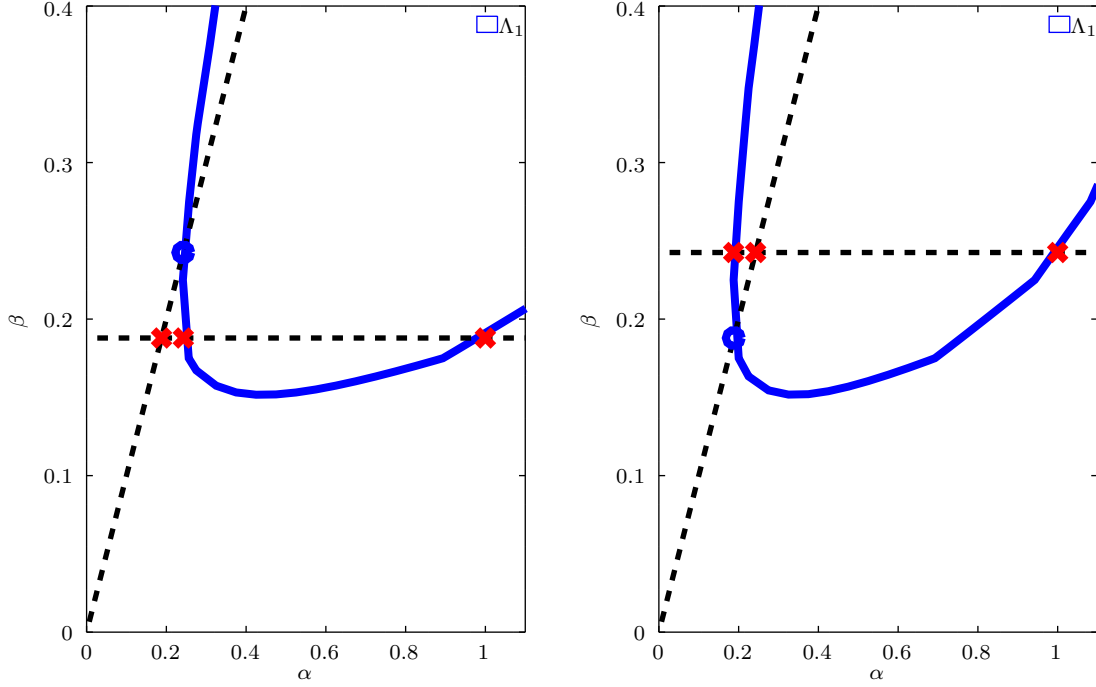
Figure 4: $\rho_1 - \rho_2$ and $\alpha_1^* - \rho_2$ have the same sign (Lemma 2.4); the points marked with 'x' are (ρ_2, ρ_2), (ρ_2, α_1^*) and $(1, \rho_2)$; the point marked with 'o' is (ρ_1, ρ_1)

Lemma 2.4. *Let (ρ_2, α_1^*) be the conjugate point of $(\rho_2, 1)$ on $\mathcal{L}_1$ identified in Proposition 2.12. Then $\rho_1 > \rho_2$ implies $\alpha_1^* > \rho_2$ and $\rho_1 < \rho_2$ implies $\alpha_1^* < \rho_2$.*

Figure 4 illustrates this lemma.

Proof. By definition ρ_1 is the unique positive number strictly less than 1 satisfying $\Lambda_1(\rho_1, \rho_1) = 1$; $\rho_2 < \rho_1$ implies $\Lambda_1(\rho_2, \rho_2) > 1$. But α_1^* satisfies $\Lambda_1(\rho_2, \alpha_1^*) = 1$ and $\Lambda_1(\rho_2, \rho) \leqslant 1$ for $\rho \in (\alpha_1^*, \rho_2]$. It follows that $\rho_2 < \alpha_1^*$. The argument for the opposite implication is similar. □

Remark 2.3. *By the previous lemma the assumption* (28) *is equivalent to*

$$\alpha_1^* \neq \rho_2. \tag{31}$$

Remark 2.4. *ρ_1 is the unique solution of $\Lambda_1(\beta, \beta) = 1$ on $(0, 1)$; similarly ρ_2 is the unique solution of $\Lambda_1(\beta, 1) = 1$ on $(0, 1)$. That Λ_1 is the largest eigenvalue of $\boldsymbol{A}(\beta, \alpha)$ and the above facts imply that ρ_1 [ρ_2] is the largest root of $\boldsymbol{p}(\beta, \beta)$ [$\boldsymbol{p}(\beta, 1)$] on $(0, 1)$. Therefore, one can state the assumption* (28) *also as follows: "the largest roots of $\boldsymbol{p}(\beta, \beta)$ and $\boldsymbol{p}(\beta, 1)$ on $(0, 1)$ differ."*

By definition, 1 is the largest eigenvalue of $\boldsymbol{A}(\rho_2, \alpha_1^*)$; let $\boldsymbol{d}_{2,1}$ denote a right eigenvector of this matrix with strictly positive entries. Next proposition constructs a (Y, M)-superharmonic function that we will use to find upper bounds on approximation errors; this is one of the key steps of our argument.

Proposition 2.14. *Under assumption* (28) *one can choose a constant* $c_0 \in \mathbb{R}$
($c_0 > 0$ *for* $\alpha_1^* < \rho_2$ *and* $c_0 < 0$ *for* $\alpha_1^* > \rho_2$*) so that*

$$h_{\rho_2} \doteq [(\rho_2, 1, \boldsymbol{d}_2), \cdot] + c_0[(\rho_2, \alpha_1^*, \boldsymbol{d}_{2,1}, \cdot], \tag{32}$$

is a (Y, M)*-superharmonic function.*

Proof. By their construction, the conjugate points $(\rho_2, 1)$ and (ρ_2, α_1^*) lie on $\mathcal{L}_1$. This and Proposition 2.1 imply that the functions $[(\rho_2, 1, \boldsymbol{d}_2), \cdot]$ and $[(\rho_2, \alpha_1^*, \boldsymbol{d}_{2,1}), \cdot]$ are (Y, M)-harmonic on $\mathbb{Z} \times \mathbb{Z}_+ - \partial_2$. This implies the same for their linear combination h_{ρ_2}. Therefore, to prove that h_{ρ_2} is (Y, M)-superharmonic, it suffices to check this on ∂_2.

By definition h_{ρ_2} is superharmonic on ∂_2 if

$$\mathbb{E}_{(y,m)}[h_{\rho_2}(Y_1, M_1)] \leqslant h_{\rho_2}(y, m)$$

for $y = (k, 0)$ and $m \in \mathcal{M}$. By Lemma 2.2,

$$\mathbb{E}_{(y,m)}[(\rho_2, 1, \boldsymbol{d}_2), (Y_1, M_1)] - [(\rho_2, 1, \boldsymbol{d}_2), (y, m)] = \rho_2^k \boldsymbol{c}(\rho_2, 1, \boldsymbol{d}_2)(m),$$
$$\mathbb{E}_{(y,m)}[(\rho_2, \alpha_1^*, \boldsymbol{d}_{2,1}), (Y_1, M_1)] - [(\rho_2, \alpha_1^*, \boldsymbol{d}_{2,1}), (y, m)] = \rho_2^k \boldsymbol{c}(\rho_2, \alpha_1^*, \boldsymbol{d}_{2,1})(m),$$

where $\boldsymbol{c}(\cdot, \cdot, \cdot)$ is defined as in (21). The last two lines give

$$\mathbb{E}_{(y,m)}[h_{\rho_2}(Y_1, M_1)] - h_{\rho_2}(y, m) = \rho_2^k \left(\boldsymbol{c}(\rho_2, 1, \boldsymbol{d}_2)(m) + c_0 \boldsymbol{c}(\rho_2, \alpha_1^*, \boldsymbol{d}_{2,1})(m) \right). \tag{33}$$

For h_{ρ_2} to be superharmonic, the right side of the last display must be negative. The sign of this expression is determined by

$$\boldsymbol{c}(\rho_2, 1, \boldsymbol{d}_2)(m) + c_0 \boldsymbol{c}(\rho_2, \alpha_1^*, \boldsymbol{d}_{2,1})(m). \tag{34}$$

The definition (21) of $\boldsymbol{c}$ and $\rho_2 < 1$ and $\boldsymbol{d}_2(m) > 0$ for all $m \in \mathcal{M}$ imply that the first term is strictly positive for all $m \in \mathcal{M}$. Define

$$d_{\max} \doteq \max_{m \in \mathcal{M}} \boldsymbol{c}(\rho_2, 1, \boldsymbol{d}_2)(m) > 0.$$

The sign of the second term in (34) depends on whether $\alpha_1^* < \rho_2$ or $\alpha_1^* > \rho_2$. For $\alpha_1^* < \rho_2$, the definition (21) of $\boldsymbol{c}$ and $\boldsymbol{d}_{2,1}(m) > 0$ for all $m \in \mathcal{M}$ imply that the $\boldsymbol{c}$ term in (34) is strictly negative for all m. Define

$$d_{\max}^* \doteq \max_{m \in \mathcal{M}} \boldsymbol{c}(\rho_2, \alpha_1^*, \boldsymbol{d}_{2,1})(m) < 0. \tag{35}$$

If we choose $c_0 > 0$ so that
$$d_{\max} + c_0 d_{\max}^* < 0, \tag{36}$$

(34) will be strictly less than 0 for all m. This and (33) imply that h_{ρ_2} is superharmonic for any c_0 satisfying (36).

For $\alpha_1^* > \rho_2$ the argument remains the same except that we replace the max in (35) with min and $c_0 < 0$. $\qquad\square$

In the next section we will use h_{ρ_2} to find bounds on the approximation error (7).

3 Upper bound for $\mathbb{P}_{(y,m)}(\tau < \infty)$

As we saw in Proposition 2.14 above, (Y, M)-superharmonic functions can be constructed from just two conjugate points on $\mathcal{L}_1 \subset \mathcal{H}^{\beta\alpha}$.

We will need an upper bound on $\mathbb{P}_{(y,m)}(\tau < \infty)$ in our analysis of the relative error (7); in the non-modulated tandem walk treated in [10, 11], this probability can be represented exactly using the harmonic functions constructed from points on the characteristic surface, which also obviously serves as an upper bound. In the present case, we will construct an upper bound for $\mathbb{P}_{(y,m)}(\tau < \infty)$ from (Y, M)-harmonic and superharmonic functions constructed in Propositions 2.10 and 2.14. The next proposition constructs the necessary function the one following it derives the upper bound.

Proposition 3.1. *Let* $h_{\rho_1} = [(\rho_1, \rho_1, \boldsymbol{d}_1), \cdot]$ *be as in* (20) *and* h_{ρ_2} *be as in* (32). *One can choose* $c_1 \geqslant 0$ *so that*

$$c_2 \doteq \min_{y \in \partial B, m \in \mathcal{M}} h_{\rho_2}(y, m) + c_1 h_{\rho_1}(y, m) > 0; \tag{37}$$

for $\alpha_1^* < \rho_2$ *one can choose* $c_1 = 0$.

Proof. By its definition,

$$h_{\rho_2}(y, m) = \boldsymbol{d}_2(m) + c_0(\alpha_1^*)^{y(2)}\boldsymbol{d}_{2,1}(m), \tag{38}$$

for $y \in \partial B$. We know by Proposition 2.14 that $c_0 > 0$ for $\alpha_1^* < \rho_2$. This, $\alpha_1^* > 0$, $\boldsymbol{d}_{2,1}(m) > 0$ imply

$$\min_{y \in \partial B} h_{\rho_2}(y, m) \geqslant \min_{m \in \mathcal{M}} \boldsymbol{d}_2(m) > 0,$$

which implies (37) with $c_1 = 0$.

For $\alpha_1^* > \rho_2$, $c_0 < 0$ and (38) can take negative values for small $y(2)$. But $0 < \alpha_1^* < 1$ implies that there exists $k_0 > 0$ such that

$$h_{\rho_2}(y, m) \geqslant \min_{m \in \mathcal{M}} \boldsymbol{d}_2(m)/2 > 0, \quad y \in \partial B, \ y(2) \geqslant k_0. \tag{39}$$

On the other hand, $\boldsymbol{d}_1(m) > 0$ for all $m \in \mathcal{M}$ and $\rho_1 > 0$ imply that $h_{\rho_1}(y, m) > 0$ for all $y \in \partial B$, $m \in \mathcal{M}$. Then one can choose $c_1 > 0$ so that

$$c_1 \boldsymbol{d}_1(m)\rho_1^{y(2)} + \boldsymbol{d}_2(m) + c_0(\alpha_1^*)^{y(2)}\boldsymbol{d}_{2,1}(m) > \min_{m \in \mathcal{M}} \boldsymbol{d}_2(m)/2, \quad y \in \partial B, \ y(2) \leqslant k_0, \tag{40}$$

since this inequality concerns only finitely many $y \in \partial B$. c_1 chosen thus, (39) and (40) imply (37). $\square$

Proposition 3.2. *Let* $c_1 \geqslant 0$, $c_2 > 0$ *be as in Proposition 3.1*

$$\mathbb{P}_{(y,m)}(\tau < \infty) \leqslant \frac{1}{c_2}\left(h_{\rho_2}(y, m) + c_1 h_{\rho_1}(y, m)\right). \tag{41}$$

Proof. For ease of notation set

$$f = h_{\rho_2} + c_1 h_{\rho_1};$$

$\rho_1, \rho_2, \alpha_1^* \in (0, 1)$ implies

$$\sup_{y \in B, m \in \mathcal{M}} |f(y, m)| < \infty.$$

19

Furthermore, by Propositions 2.8 and 2.14 f is (Y, M)-superharmonic. These imply that $k \mapsto f(Y_{k \wedge \tau}, M_{k \wedge \tau})$ is a bounded supermartingale. Then by the optional sampling theorem ([3, Theorem 5.7.6])

$$\mathbb{E}_{(y,m)}[f(Y_\tau, M_\tau) 1_{\{\tau < \infty\}}] \leqslant f(y, m);$$

this, $Y_\tau \in \partial B$ when $\tau < \infty$ and (37) imply

$$c_2 \mathbb{P}_{(y,m)}(\tau < \infty) \leqslant f(y, m),$$

which gives (41). $\qquad \square$

4 Upper bound for $\mathbb{P}_{(x,m)}(\sigma_1 < \sigma_{1,2} < \tau_n < \tau_0)$

Define

$$\sigma_i \doteq \inf\{k \geqslant 0 : X_k \in \partial_i\}; i = 1, 2, \tag{42}$$

and

$$\sigma_{1,2} \doteq \inf\{k \geqslant 0 : X_k \in \partial_2, \ k \geqslant \sigma_1\}; \tag{43}$$

σ_i is the first time X hits ∂_i and $\sigma_{1,2}$ is the first time X hits ∂_2 after hitting ∂_1. In the next section we find an upper bound on the probability $\mathbb{P}_{(x,m)}(\sigma_1 < \sigma_{1,2} < \tau_n < \tau_0)$, we will use this bound in the analysis of the approximation error in the proof of Theorem 6.1. Define

$$\rho \doteq \rho_1 \vee \rho_2. \tag{44}$$

The goal of the section is to prove

Proposition 4.1. *For any $\epsilon > 0$ there exists $n_0 > 0$ such that*

$$\mathbb{P}_{(x,m)}(\sigma_1 < \sigma_{1,2} < \tau_n < \tau_0) \leqslant \rho^{n(1-\epsilon)} \tag{45}$$

for $n \geqslant n_0$ and $(x, m) \in A_n$.

We split the proof into cases $\rho_1 > \rho_2$ and $\rho_2 > \rho_1$. The first subsection below treats the first case $\rho_1 > \rho_2$, the next gives the changes needed for the latter.

Let $\boldsymbol{A}_1$ denote the characteristic matrix for ∂_1:

$$\boldsymbol{p}_1(\beta, \alpha, m) \doteq \lambda(m) \frac{1}{\beta} + \mu_1(m) + \mu_2(m) \frac{\beta}{\alpha}, \quad m \in \mathcal{M},$$

$$\boldsymbol{A}_1(\beta, \alpha)_{m_1, m_2} \doteq \begin{cases} \boldsymbol{P}(m_1, m_2), & m_1 \neq m_2 \\ \boldsymbol{P}(m_1, m_1) \boldsymbol{p}_1(\beta, \alpha, m), & m_1 = m_2, \end{cases}, (m_1, m_2) \in \mathcal{M}^2.$$

We will use the following fact several times in our analysis.

Lemma 4.1. *The function*

$$(x, m) \mapsto [(\rho_2, 1, \boldsymbol{d}_2), (T_n(x), m)] = \rho_2^{n-(x(1)+x(2))} \boldsymbol{d}_2(m) \tag{46}$$

is (X, M)-harmonic on $\mathbb{Z}_+^2 - \partial_2$.

Proof. We know by Proposition 2.1 and $(\rho_2, 1, \boldsymbol{d}_2) \in \mathcal{H}$ that $[(\rho_2, 1, \boldsymbol{d}_2), \cdot]$ is (Y, M)-harmonic on $\mathbb{Z} \times \mathbb{Z}_+^o$, which implies that (46) is (X, M)-harmonic on $\mathbb{Z}_+^{2,o}$; this and $\boldsymbol{A}_1(\rho_2, 1) = \boldsymbol{A}(\rho_2, 1)$ imply the (X, M)-harmonicity of (46) on ∂_1. $\qquad \square$

20

4.1 $\rho_1 > \rho_2$

To prove (45) we will construct a corresponding supermartingale; applying the optional sampling theorem to the supermartingale will give our desired bound. The event $\{\sigma_1 < \sigma_{1,2} < \tau_n < \tau_0\}$ consists of three stages: X first hits ∂_1 then ∂_2 and finally ∂A_n without ever hitting 0. If h is an (X, M)-superharmonic function, it follows from the definitions that $h(X, M)$ is a supermartingale. We will construct our supermartingale by applying three functions (one for each of the above stages) to (X, M): the function for the first stage is the constant ρ_1^n, which is trivially superharmonic. The function for the second stage will be a constant multiple of $(x, m) \mapsto h_{\rho_1}(T_n(x), m)$. By Proposition 2.8, $(x, m) \mapsto h_{\rho_1}(T_n(x), m)$ is (X, M)-harmonic on $\mathbb{Z}_+^2 - \partial_1$. One can check directly that it is in fact subharmonic on ∂_1. The definition of the supermartingale S will involve terms to compensate for this. The function for the third stage is

$$
\begin{aligned}
h_3 : (x, m) &\mapsto h_{\rho_2}(T_n(x), m) + c_1 h_{\rho_1}(T_n(x), m), \ x \in A_n, \ m \in \mathcal{M}, \\
&= h_{\rho_2}((n - x(1), x(2)), m) + c_1 h_{\rho_1}((n - x(1), x(2)), m), \\
&= \rho_2^{n-(x(1)+x(2))} \left(\boldsymbol{d}_2(m) + c_0 \alpha_1^{*x(2)} \boldsymbol{d}_{2,1}(m) \right) + c_1 \rho_1^{n-x(1)} \boldsymbol{d}_1(m),
\end{aligned}
\tag{47}
$$

where $c_1 \geqslant 0$ is chosen as in Proposition 3.1 and c_0 is as in Proposition 2.14. The next two propositions imply that h_3 is (X, M)-superharmonic on $\mathbb{Z}_+^2 - \partial_1$.

Proposition 4.2. For $\rho_1 > \rho_2$, $h_{\rho_2}(T_n(\cdot), \cdot)$ is superharmonic on all of $\mathbb{Z}_+^2$.

Proof. That $h_{\rho_2}(T_n(\cdot), \cdot)$ is (X, M)-superharmonic on $\mathbb{Z}_+^2 - \partial_1$ follows from Proposition 2.14 (i.e., from the fact that $h_{\rho_2}(\cdot, \cdot)$ is (Y, M)-harmonic). Therefore, it suffices to prove that $h_{\rho_2}(T_n(\cdot), \cdot)$ is superharmonic on ∂_1. $h_{\rho_2}(T_n(\cdot), \cdot)$ is a sum of two functions:

$$
h_{\rho_2}(T_n(\cdot), \cdot) = [(\rho_2, 1, \boldsymbol{d}_2), (T_n(\cdot), \cdot)] + c_0[(\rho_2, \alpha_1^*, \boldsymbol{d}_{2,1}), (T_n(\cdot), \cdot)].
\tag{48}
$$

Let us show that each of these summands is (X, M)- superharmonic on ∂_1. The first summand is (X, M)-harmonic (and therefore, superharmonic) on ∂_1 by Lemma 4.1. To treat the second term in (48) recall the following: $\rho_2 < \rho_1$ implies $\rho_2 < \alpha_1^*$ (Lemma 2.4); then, by Proposition 2.14, $c_0 < 0$. Therefore, if we can show that $[(\rho_2, \alpha_1^*, \boldsymbol{d}_{2,1}), (T_n(\cdot), \cdot)]$ is (X, M)-subharmonic on ∂_1 we will be done. Let us now see that this is indeed the case.

For ease of notation set

$$
h(x, m) = [(\rho_2, \alpha_1^*, \boldsymbol{d}_{2,1}), (T_n(x), m)] = \rho_2^{n-(x(1)+x(2))} \alpha_1^{*x(2)} \boldsymbol{d}_{2,1}(m).
$$

A calculation parallel to the proof of Proposition 2.1 shows

$$
\mathbb{E}_{(x,m)} \left[h(X_1, M_1) \right] - h(x, m) = \boldsymbol{d}_{2,1}(m) \mu_1(m)(1 - \alpha_1^*) \rho_2^{n-x(2)} > 0,
\tag{49}
$$

for $x \in \partial_1$, i.e., h is (X, M)-subharmonic on ∂_1. This completes the proof of this proposition. □

Proposition 4.3. $h_{\rho_1}(T_n(\cdot), \cdot)$ is harmonic (and therefore superharmonic) on $\mathbb{Z}_+^2 - \partial_1$. It is subharmonic on ∂_1 where it satisfies

$$
\mathbb{E}_{(x,m)}[h_{\rho_1}(T_n(X_1), M_1)] - h_{\rho_1}(T_n(x), m) = \boldsymbol{d}_1(m) \mu_1(m)(1 - \rho_1) \rho_1^n > 0.
\tag{50}
$$

21

The proof is parallel to the computation given in the proof of Proposition 2.1 and is omitted. We can now define the supermartingale that we will use to prove (45):

$$S_k' \doteq \begin{cases} h_1, & k \leqslant \sigma_1, \\ h_2(X_k, M_k), & \sigma_1 < k \leqslant \sigma_{1,2}, \\ h_3(X_k, M_k), & k > \sigma_{1,2}, \end{cases}$$
$$S_k \doteq S_k' - c_5 k \rho_1^n,$$

where

$$c_3 \doteq \frac{\max_{m \in \mathcal{M}} \boldsymbol{d}_2(m) + c_1 \max_{m \in \mathcal{M}} \boldsymbol{d}_1(m)}{\min_{m \in \mathcal{M}} \boldsymbol{d}_1(m)} > 0, \tag{51}$$

$$h_1 \doteq c_4 \rho_1^n, \qquad c_4 \doteq c_3 \max_{m \in \mathcal{M}} \boldsymbol{d}_1(m) > 0,$$

$$h_2 \doteq c_3 h_{\rho_1}(T_n(\cdot), \cdot) = c_3[(\rho_1, \rho_1, \boldsymbol{d}_1), (T_n(\cdot), \cdot)] = c_3 \rho_1^{n-x(1)} \boldsymbol{d}_1(\cdot), \tag{52}$$

$$c_5 \doteq c_3 (1 - \rho_1) \max_{m \in \mathcal{M}} \boldsymbol{d}_1(m) \mu_1(m). \tag{53}$$

Two comments: h_1 is a constant function, independent of x and m, and $h_1 \geqslant h_2$ on ∂_1.

Proposition 4.4. *S is a supermartingale.*

Proof. The claim follows mostly from the fact that the functions involved in the definition of S' are (X, M)-superharmonic away from ∂_1. The term that breaks superharmonicity on ∂_1 is $[(\rho_1, \rho_1, \boldsymbol{d}_1), (T_n(X_k), M_k)]$; the $-c_5 k \rho_1^n$ term in the definition of S is introduced to compensate for this. The details are as follows.

The (X, M)-harmonicity of h_1, h_2 and h_3 implies

$$\mathbb{E}_{(x,m)}[S_{k+1}' | \mathscr{F}_k] = S_k'$$

for $X_k \in \mathbb{Z}_+^2 - \partial_1 \cup \partial_2$; i.e., S_k' satisfies the martingale equality condition for $X_k \in \mathbb{Z}_+^2 - \partial_1 \cup \partial_2$; this implies that S_k satisfies the supermartingale inequality condition over the same event.

h_2 and h_3 are (X, M)-superharmonic on ∂_2 by Propositions 4.2 and 4.3 (h_1 is trivially so because it is constant); this implies

$$\mathbb{E}_{(x,m)}[S_{k+1}' | \mathscr{F}_k] \leqslant S_k'$$

for $X_k \in \partial_2$ and $k \neq \sigma_{1,2}$. For $k = \sigma_{1,2}$ we have $S_{k+1}' = h_3(X_{k+1}, M_{k+1})$. This, the (X, M)-superharmonicity of h_3 on ∂_2 implies

$$\mathbb{E}_{(x,m)}[S_{k+1}' | \mathscr{F}_k] = \mathbb{E}_{(x,m)}[h_3(X_{k+1}, M_{k+1}) | \mathscr{F}_k]$$
$$\leqslant h_3(X_k, M_k) \tag{54}$$

for $k = \sigma_{1,2}$. On the other hand,

$$S_k' = h_2(X_k, M_k) \text{ for } k = \sigma_{1,2}. \tag{55}$$

The definitions of c_3, h_2 and h_3 in (51), (52) and (47), $\rho_2 < \rho_1$ and $c_0 < 0$ imply

$$h_3(x, m) \leqslant h_2(x, m)$$

for $x \in \partial_2$. This and (55) imply

$$h_3(X_k, M_k) \leqslant h_2(X_k, M_k) = S_k'$$

for $k = \sigma_{1,2}$. The last display and (54) imply

$$\mathbb{E}_{(x,m)}[S_{k+1}'|\mathscr{F}_k] \leqslant S_k',$$

i.e., S' and S are (X, M)-supermartingales for $k = \sigma_{1,2}$ as well.

It remains to prove

$$\mathbb{E}_{(x,m)}[S_{k+1}|\mathscr{F}_k] \leqslant S_k, \text{ when } X_k \in \partial_1. \tag{56}$$

The cases to be treated here are: $k = \sigma_1$, $\sigma_1 < k < \sigma_{1,2}$ and $k > \sigma_{1,2}$.

For $k = \sigma_1$, we have $S_k' = h_1(X_k, M_k) = c_3 \rho_1^n \boldsymbol{d}_1(M_k)$ and $S_{k+1}' = h_2(X_{k+1}, M_{k+1})$; these and $h_1 \geqslant h_2$ on ∂_1 imply

$$\mathbb{E}_{(x,m)}[S_{k+1}|\mathscr{F}_k] - S_k \tag{57}$$
$$= \mathbb{E}_{(x,m)}[c_3 h_{\rho_1}(T_n(X_{k+1}), M_{k+1})|\mathscr{F}_k] - c_3 \rho_1^n \boldsymbol{d}_1(M_k) - c_5 \rho_1^n,$$

By (50) and $\sigma_1 = k$, this equals

$$\leqslant c_3 \boldsymbol{d}_1(M_k)\mu_1(M_k)(1 - \rho_1)\rho_1^n - c_5 \rho_1^n.$$

By the definition of c_5:

$$= \rho_1^n c_3(1 - \rho_1)(\boldsymbol{d}_1(M_k)\mu_1(M_k) - \max_{m \in \mathcal{M}} \boldsymbol{d}_1(m)\mu_1(m)) \leqslant 0,$$

which proves (56) for $k = \sigma_1$.

For $\sigma_1 < k < \sigma_{1,2}$, $S_k' = h_2(X_k, M_k) = c_3 h_{\rho_1}(T_n(X_k), M_k)$; therefore the above argument applies to this case as well (except for the last step which is not needed here because S_k' and S_{k+1}' are defined by applying the same function h_2 to (X_{k+1}, M_{k+1}) and (X_k, M_k)).

Finally, to treat the case $X_k \in \partial_1$ and $k > \sigma_{1,2}$ we start with

$$\mathbb{E}_{(x,m)}[S_{k+1}|\mathscr{F}_k] - S_k = \mathbb{E}_{(x,m)}[S_{k+1}'|\mathscr{F}_k] - S_k' - c_5 \rho_1^n,$$

$S_k' = h_3(X_k, M_k)$ for $k > \sigma_{1,2}$. Then by the definition of h_3:

$$= \mathbb{E}_{(x,m)}[h_{\rho_2}(T_n(X_{k+1}), M_{k+1}) + c_1 h_{\rho_1}(T_n(X_{k+1}), M_{k+1})|\mathscr{F}_k]$$
$$- h_{\rho_2}(T_n(X_k), M_k) - c_1 h_{\rho_1}(T_n(X_k), M_k) - c_5 \rho_1^n,$$
$$= (\mathbb{E}_{(x,m)}[h_{\rho_2}(T_n(X_{k+1}), M_{k+1})|\mathscr{F}_k] - h_{\rho_2}(T_n(X_k), M_k))$$
$$+ \mathbb{E}[c_1 h_{\rho_1}(T_n(X_{k+1}), M_{k+1})|\mathscr{F}_k] - c_1 h_{\rho_1}(T_n(X_k), M_k) - c_5 \rho_1^n.$$

The (X, M)-superharmonicity of $h_{\rho_2}(T_n(\cdot), \cdot)$ implies that the difference inside the parenthesis is negative, therefore:

$$\leqslant \mathbb{E}[c_1 h_{\rho_1}(T_n(X_{k+1}), M_{k+1})|\mathscr{F}_k] - c_1 h_{\rho_1}(T_n(X_k), M_k) - c_5 \rho_1^n.$$

Proposition 4.3 ((50)) now gives

$$= c_1 \boldsymbol{d}_1(M_k)\mu_1(M_k)(1-\rho_1)\rho_1^n - c_5\rho_1^n.$$

By its definition (53), $c_5 > c_1\boldsymbol{d}_1(m)\mu_1(m)(1-\rho_1)$ for all $m \in \mathcal{M}$, which implies:

$$\leqslant 0.$$

This proves (56) for $k > \sigma_{1,2}$ and completes the proof of this proposition. $\qquad\square$

We are now ready to give a proof of Proposition 4.1 for $\rho_1 > \rho_2$:

Proof of Proposition 4.1; case $\rho_1 > \rho_2$. By its definition (44), ρ of (45) equals ρ_1 for $\rho_1 > \rho_2$. We begin by truncating time: [9, Theorem A.2] implies that there exists $c_6 > 0$ and $N_0 > 0$ such that

$$\mathbb{P}_{(x,m)}(\tau_n \wedge \tau_0 > c_6 n) \leqslant \rho_1^{2n},$$

for $n > N_0$. Then:

$$\mathbb{P}_{(x,m)}(\sigma_1 < \sigma_{1,2} < \tau_n < \tau_0) \qquad\qquad (58)$$
$$= \mathbb{P}_{(x,m)}(\sigma_1 < \sigma_{1,2} < \tau_n < \tau_0, \tau_n \wedge \tau_0 \leqslant c_6 n)$$
$$+ \mathbb{P}_{(x,m)}(\sigma_1 < \sigma_{1,2} < \tau_n < \tau_0, \tau_n \wedge \tau_0 > c_6 n)$$
$$\leqslant \mathbb{P}_{(x,m)}(\sigma_1 < \sigma_{1,2} < \tau_n < \tau_0, \tau_n \wedge \tau_0 \leqslant c_6 n) + \rho_1^{2n}$$

for $n > N_0$. Therefore, to prove (45) it suffices to bound the first term on the right side of the last inequality. Now apply the optional sampling theorem to the supermartingale S at the bounded stopping time $\tau = \tau_0 \wedge \tau_n \wedge c_6 n$:

$$\mathbb{E}_{(x,m)}\left[S_{\tau_0 \wedge \tau_n \wedge c_6 n}\right] \leqslant S_0 = c_4\rho_1^n.$$

By definition, $S_k = S'_k - c_5 k \rho_1^n$; substituting this in the last display gives:

$$-c_5 c_6 n \rho_1^n + \mathbb{E}_{(x,m)}[S'_\tau] \leqslant c_4\rho_1^n$$
$$\mathbb{E}_{(x,m)}[S'_\tau] \leqslant (c_4 + n c_5 c_6)\rho_1^n.$$

By its definition, $S'_k > 0$, therefore restricting it to an event makes the last expectation smaller:

$$\mathbb{E}_{(x,m)}[S'_\tau 1_{\{\sigma_1 < \sigma_{1,2} < \tau_n < \tau_0 \leqslant c_6 n\}}] \leqslant (c_4 + n c_5 c_6)\rho_1^n.$$

On the set $\{\sigma_1 < \sigma_{1,2} < \tau_n < \tau_0 \leqslant c_6 n\}$, we have $\tau = \tau_n$ and $S'_{\tau_n} = h_3(X_{\tau_n}, M_{\tau_n})$; by definition $X_{\tau_n} \in \partial A_n$. By definition of h_3 and by Proposition 3.1 $h_3(x,m) \geqslant c_2 > 0$ for $x \in \partial A_n$. These and the last display imply

$$c_2 \mathbb{P}_{(x,m)}(\sigma_1 < \sigma_{1,2} < \tau_n < \tau_0 \leqslant c_6 n) \leqslant (c_4 + n c_5 c_6)\rho_1^n.$$

Substitute this in (58) to get

$$\mathbb{P}_{(x,m)}(\sigma_1 < \sigma_{1,2} < \tau_n < \tau_0) \leqslant \rho_1^{n(1-\epsilon_n)}$$

where

$$\epsilon_n = \frac{1}{n}\log_{1/\rho_1}\left(\frac{c_4 + n c_5 c_6}{c_2}\right);$$

setting $n_0 \geqslant N_0$ so that $\epsilon_n < \epsilon$ for $n \geqslant n_0$ gives (45). $\qquad\square$

24

4.2 $\rho_1 < \rho_2$

The previous subsection gave a proof of Proposition 4.1 for $\rho_2 < \rho_1$. The only changes needed in this proof for $\rho_1 < \rho_2$ concern the functions used in the definition of the supermartingale S; the needed changes are:

1. Modify the function h_2 for the second stage,

2. The function h_3 is no longer superharmonic on ∂_1; quantify how much it deviates from superharmonicity on ∂_1,

3. Modify the constants used in the definition of S in accordance with these changes.

The next two propositions deal with the first two items above; the definition of the supermartingale (taking also care of the third item) is given after them.

The convexity of $q \mapsto -\log(\Lambda_1(e^q, e^q))$ and $\Lambda_1(\rho_1, \rho_1) = 1$ imply $\Lambda_1(\rho_2, \rho_2) < 1$ for $\rho_2 > \rho_1$. Let $\boldsymbol{d}_2^+$ be a right eigenvector of $\boldsymbol{A}(\rho_2, \rho_2)$ with strictly positive entries.

Proposition 4.5. *The function*

$$f : (x, m) \mapsto [(\rho_2, \rho_2, \boldsymbol{d}_2^+), (T_n(x), m)]$$

is superharmonic on $\mathbb{Z}_+^2 - \partial_1$. On ∂_1 it satisfies

$$\mathbb{E}_{(x,m)}[f(X_1, M_1)] - f(x, m) \leqslant \boldsymbol{d}_2^+(m)\mu_1(m)(1 - \rho_2)\rho_2^n. \tag{59}$$

The proof is parallel to that of Proposition 4.3 and follows from $\Lambda_1(\rho_2, \rho_2) < 1$, $\boldsymbol{A}_2(\rho_2, \rho_2) = \boldsymbol{A}(\rho_2, \rho_2)$ and the definitions involved.

Proposition 4.6. *Let h_3 be as in (47); h_3 is (X, M)-superharmonic on $\mathbb{Z}_+^2 - \partial_1$; on ∂_1 it satisfies*

$$\mathbb{E}_{(x,m)}[h_3(X_1, M_1)] - h_3(x, m) = c_0\boldsymbol{d}_{2,1}(m)\mu_1(m)(1 - \alpha_1^*)\rho_2^{n-x(2)} > 0. \tag{60}$$

Proof. Lemma 2.4 and $\rho_2 > \rho_1$ imply $\alpha_1^* < \rho_2$; this and Proposition 3.1 imply that c_1 in the definition of h_3 is 0; i.e.,

$$h_3(x, m) = h_{\rho_2}(T_n(x), m) = \rho_2^{n-(x(1)+x(2))}\left(\boldsymbol{d}_2(m) + c_0(\alpha_1^*)^{x(2)}\boldsymbol{d}_{2,1}(m)\right);$$

That h_3 is (X, M)-superharmonic on $\mathbb{Z}_+^2 - \partial_1$ follows from the same property of h_{ρ_2} (see Proposition 2.14). On the other hand, again by Proposition 2.14, $\alpha_1^* < \rho_2$ implies that c_0 in the definition of h_{ρ_2} satisfies $c_0 > 0$. By Lemma 4.1 $(x, m) \mapsto [(\rho_2, 1, \boldsymbol{d}_2), (T_n(x), m)]$ is (X, M)-harmonic on ∂_1; (60) follows from these and (49). $\qquad\square$

$\rho_2 > \rho_1$ implies $\rho_2 > \alpha_1^*$ (Lemma 2.4); this and Proposition 3.1 imply $c_1 = 0$; $\rho_2 > \alpha_1^*$ and Proposition 2.14 imply $c_0 > 0$. That $c_0 > 0$ and $c_1 = 0$ lead to the following modifications in the definition of S':

$$S_k' \doteq \begin{cases} h_1, & k \leqslant \sigma_1, \\ h_4(X_k, M_k), & \sigma_1 < k \leqslant \sigma_{1,2}, \\ h_3(X_k, M_k), & k > \sigma_{1,2}, \end{cases}$$

$$S_k \doteq S_k' - c_5 k \rho_2^n,$$

25

where

$$c_3 \doteq \frac{\max_{m \in \mathcal{M}} \left(\boldsymbol{d}_2(m) + c_0 \boldsymbol{d}_{2,1}(m) \right)}{\min_{m \in \mathcal{M}} \boldsymbol{d}_2^+(m)},$$

$$h_1 \doteq c_4 \rho_2^n, \qquad c_4 \doteq c_3 \max_{m \in \mathcal{M}} \boldsymbol{d}_2^+(m),$$

$$h_4 \doteq c_3[(\rho_2, \rho_2, \boldsymbol{d}_2^+), (T_n(\cdot), \cdot)] = c_3 \rho_2^{n-x(1)} \boldsymbol{d}_2^+(\cdot),$$

$$c_5 \doteq c_3(1 - \rho_2) \max_{m \in \mathcal{M}} \boldsymbol{d}_2^+(m)\mu_1(m) + c_0(1 - \alpha_1^*) \max_{m \in \mathcal{M}} \boldsymbol{d}_{2,1}(m)\mu_1(m).$$

The modification in c_3 ensures $h_4 \geqslant h_3$ on ∂_2; $c_0 > 0$ implies that h_3 is no longer superharmonic on ∂_1; the second term in c_5 compensates for this.

Proposition 4.7. *S as defined above is a supermartingale for $\rho_2 > \rho_1$.*

Proof. With the modifications made as above, the proof proceeds exactly as in the case $\rho_1 > \rho_2$ (Proposition 4.4) and follow from the following facts: $h_1 \geqslant h_4$ on ∂_1, $h_4 \geqslant h_3$ on ∂_2 (these are guaranteed by the choices of the constants c_4, c_3); (X, M)-superharmonicity of h_4 and h_3 on $\mathbb{Z}_+^2 - \partial_1$ (guaranteed by Propositions 4.5 and 4.6), the $-c_5 k \rho_2^n$ term compensating for the lack of (X, M)-superharmonicity of h_3 and h_4 on ∂_1 (guaranteed by (59) and (60) and the choice of the constant c_5). $\qquad \square$

Proof of Proposition 4.1; case $\rho_2 > \rho_1$. With S defined as above, the proof given for the case $\rho_1 > \rho_2$ works without change. $\qquad \square$

5 Lower bound for $\mathbb{P}_{(x,m)}(\tau_n < \tau_0)$

To get an upper bound on the relative error (7), we need a lower bound on the probability $\mathbb{P}_{(x,m)}(\tau_n < \tau_0)$. We will get the desired bound by applying the optional sampling theorem, this time to an (X, M)-submartingale. This we will do, following [12], by constructing a suitable (X, M)-subharmonic function. As opposed to superharmonic functions, subharmonic functions are simpler to construct.

Proposition 5.1.

$$(x, m) \mapsto [(\rho_2, 1, \boldsymbol{d}_2), (T_n(x), m)] \vee [(\rho_1, \rho_1, \boldsymbol{d}_1), (T_n(x), m)] \tag{61}$$
$$= \rho_2^{n-(x(1)+x(2))} \boldsymbol{d}_2(m) \vee \rho_1^{n-x(1)} \boldsymbol{d}_1(m)$$

is (X, M)-subharmonic on $\mathbb{Z}_+^2$.

Proof. We know by Lemma 2.2 that

$$\mathbb{E}_{(x,m)}[(\rho_2, 1, \boldsymbol{d}_2), (T_n(X_1), M_1)] - [(\rho_2, 1, \boldsymbol{d}_2), (x, m)]$$
$$= \rho_2^{n-x(1)} \boldsymbol{P}(m, m)\mu_2(m)\boldsymbol{d}_2(m)(1 - \rho_2) > 0,$$

i.e, $(x, m) \mapsto [(\rho_2, 1, \boldsymbol{d}_2), (x, m)]$ is (X, M)-subharmonic on ∂_2.
That $(x, m) \mapsto [(\rho_2, 1, \boldsymbol{d}_2), (T_n(x), m)]$ is (X, M)-subharmonic on $\mathbb{Z}_+^2 - \partial_2$ follows from Lemma 4.1. Then, $(x, m) \mapsto [(\rho_2, 1, \boldsymbol{d}_2), (x, m)]$ is (X, M)-subharmonic on all of $\mathbb{Z}_+^2$.

Similarly, Proposition 4.3 and (50) imply that $(x, m) \mapsto [(\rho_1, \rho_1, \boldsymbol{d}_1), (x, m)]$ is (X, M)-subharmonic on all of $\mathbb{Z}_+^2$.

The maximum of two subharmonic functions is again subharmonic. This and the above facts imply the (X, M)-subharmonicity of (61). $\qquad \square$

Proposition 5.2.

$$\mathbb{P}_{(x,m)}(\tau_n < \tau_0)$$
$$\geqslant \left(\max_{m \in \mathcal{M}} (\boldsymbol{d}_2(m) \vee \boldsymbol{d}_1(m)) \right)^{-1} \tag{62}$$
$$\times \left(\rho_2^{n-(x(1)+x(2))} \boldsymbol{d}_2(m) \vee \rho_1^{n-x(1)} \boldsymbol{d}_1(m) - \rho_2^n \max_{m \in \mathcal{M}} \boldsymbol{d}_2(m) \vee \rho_1^n \max_{m \in \mathcal{M}} \boldsymbol{d}_1(m) \right).$$

Proof. Set

$$g(x,m) = \rho_2^{n-(x(1)+x(2))} \boldsymbol{d}_2(m) \vee \rho_1^{n-x(1)} \boldsymbol{d}_1(m);$$

by the previous proposition g is (X, M)-subharmonic. By its definition, g is positive and bounded from above for $x \in \mathbb{Z}_+^2$. It follows that

$$s_k = g(X_{\tau_n \wedge \tau_0 \wedge k}, M_{\tau_n \wedge \tau_0 \wedge k})$$

is a bounded positive submartingale. By definition

$$\mathbb{E}[g(X_{\tau_n \wedge \tau_0}, M_{\tau_n \wedge \tau_0})] = \mathbb{E}[g(X_{\tau_n}, M_{\tau_n}) 1_{\{\tau_n < \tau_0\}}] + \mathbb{E}[g(X_{\tau_0}, M_{\tau_0}) 1_{\{\tau_0 \leqslant \tau_n\}}]. \tag{63}$$

That $X_{\tau_n} \in \partial A_n$ implies $g(X_{\tau_n}, M_{\tau_n}) = g(k, n-k)$ for some $k < n$; then

$$g(X_{\tau_n}, M_{\tau_n}) \leqslant \max_{m \in \mathcal{M}} (\boldsymbol{d}_2(m) \vee \boldsymbol{d}_1(m)).$$

This, (63) and the optional sampling theorem applied to s at time $\tau_n \wedge \tau_0$ give

$$\mathbb{P}_{(x,m)}(\tau_n < \tau_0) \left(\max_{m \in \mathcal{M}} (\boldsymbol{d}_2(m) \vee \boldsymbol{d}_1(m)) \right) + g(0, m) \mathbb{P}_{(x,m)}(\tau_0 \leqslant \tau_n) \geqslant g(x, m).$$

$\mathbb{P}_{(x,m)}(\tau_0 \leqslant \tau_n) \leqslant 1$ implies

$$\left(\max_{m \in \mathcal{M}} (\boldsymbol{d}_2(m) \vee \boldsymbol{d}_1(m)) \right) \mathbb{P}_{(x,m)}(\tau_n < \tau_0) \geqslant g(x, m) - \max_{m \in \mathcal{M}} [g(0, m)];$$

this and $\max_{m \in \mathcal{M}} [g(0, m)] = \rho_2^n \max_{m \in \mathcal{M}} \boldsymbol{d}_2(m) \vee \rho_1^n \max_{m \in \mathcal{M}} \boldsymbol{d}_1(m)$ give (62). $\qquad \square$

6 Completion of the limit analysis

This section puts together the results of the last two sections to derive an exponentially decaying upper bound on the relative error (7). As in previous works [10, 11, 12], this task is simplified if we express the Y process in the x coordinates thus:

$$\bar{X}_k \doteq T_n(Y_k);$$

$\bar{X}$ has the same dynamics as X, except that it is not constrained on ∂_1. In this section we will set the initial condition using the scaled coordinate $x \in \mathbb{R}_+^2$, $x(1) + x(2) < 1$, the initial condition for the X and $\bar{X}$ will be

$$X_0 = \bar{X}_0 = \lfloor nx \rfloor.$$

As in the non-modulated case, the following relation between $\bar{X}$ and X will be very useful:

Lemma 6.1. *Let $\sigma_{1,2}$ be as in (43). Then*

$$X_k(1) + X_k(2) = \bar{X}_k(1) + \bar{X}_k(2)$$

for $k \leqslant \sigma_{1,2}$.

This lemma is the analog of [10, Proposition 7.2], which expresses the same fact for the non-modulated two dimensional tandem walk; the proof is unchanged because it does not depend on the modulating process. Example sample paths of X and $\bar{X}$ up to time $\sigma_{1,2}$ demonstrating Lemma 6.1 are shown in Figure 5.
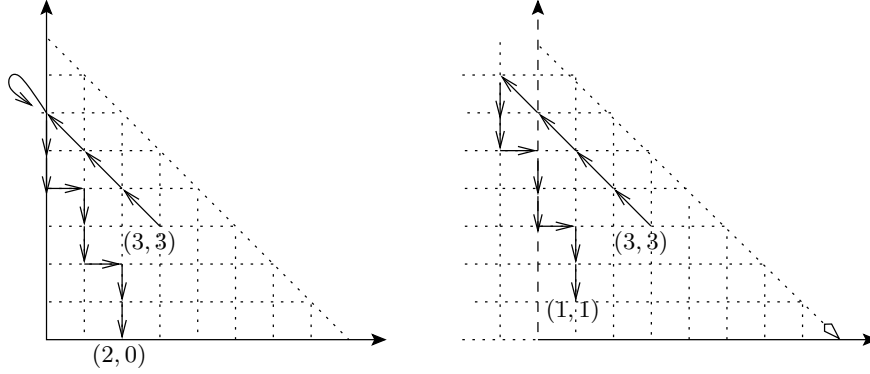

Figure 5:   A sample path of X_k(left) and $\bar{X}_k$(right)

Define

$$\bar{\tau}_n \doteq \inf\{k > 0 : \bar{X}_k \in \partial A_n\},$$
$$\bar{\sigma}_{1,2} \doteq \inf\{k > 0 : \bar{X}_k(1) + \bar{X}_k(2) = 0, \ k \geqslant \sigma_1\}.$$

X and $\bar{X}$ have identical dynamics upto time σ_1; $\bar{\sigma}_{1,2}$ is the first time after (σ_1, i.e., the first time X and $\bar{X}$ hit ∂_1) that the sum of the components of $\bar{X}$ equals 0. By the definitions of $\bar{X}$ and Y, $\bar{\tau}_n = \tau$.

What follows is an upper bound similar to (45) for the $\bar{X}$ process. This is a generalization of [10, Proposition 7.5] to the present setup:

Proposition 6.1. *For any $\epsilon > 0$ there exists $n_0 > 0$ such that*

$$\mathbb{P}_{(x,m)}(\sigma_1 < \sigma_{1,2} < \bar{\tau}_n < \infty) \leqslant \rho^{n(1-\epsilon)} \tag{64}$$

for $n > n_0$ and $(x, m) \in A_n$.

Proof. As in [10, Proposition 7.5] we partition the event $\{\sigma_1 < \sigma_{1,2} < \bar{\tau}_n < \infty\}$ into whether $\bar{X}$ hits ∂A_n before or after it hits $\{x \in \mathbb{Z} \times \mathbb{Z}_+ : x(1) + x(2) = 0\}$:

$$\mathbb{P}_{(x,m)}(\sigma_1 < \sigma_{1,2} < \bar{\tau}_n < \infty) \tag{65}$$
$$= \mathbb{P}_{(x,m)}(\sigma_1 < \sigma_{1,2} < \bar{\tau}_n < \bar{\sigma}_{1,2} < \infty) + \mathbb{P}_{(x,m)}(\sigma_1 < \sigma_{1,2} < \bar{\sigma}_{1,2} < \bar{\tau}_n < \infty)$$

Lemma 6.1 implies

$$\bar{X}_{\sigma_{1,2}}(1) + \bar{X}_{\sigma_{1,2}}(2) = X_{\sigma_{1,2}}(1) + X_{\sigma_{1,2}}(2)$$

i.e., at time $\sigma_{1,2}$, X and $\bar{X}$ will be on the same line $\{x \in \mathbb{Z} \times \mathbb{Z}_+ : x(1) + x(2) = k\}$ for some $k \in \{1, 2, ..., n-1\}$. Then for $\omega \in \{\sigma_1 < \sigma_{1,2} < \bar{\tau}_n\}$ the fully constrained sample path $X(\omega)$

28

cannot hit 0 before the path $\bar{X}(\omega)$ hits $\{x \in \mathbb{Z} \times \mathbb{Z}_+ : x(1) + x(2) = 0\}$ and it cannot hit ∂A_n after $\bar{X}$ hits $\{x \in \mathbb{Z} \times \mathbb{Z}_+ : x(1) + x(2) = n\}$ (intuitively: more constraints on X push it faster to ∂A_n and slower to 0 than less constraints do the process $\bar{X}$): these give

$$\{\sigma_1 < \sigma_{1,2} < \bar{\tau}_n < \bar{\sigma}_{1,2} < \infty\} \subset \{\sigma_1 < \sigma_{1,2} < \tau_n < \tau_0\};$$

the bound (45) on the probability of the last event and (65) imply that there exists $n_1 > 0$ such that

$$\mathbb{P}_{(x,m)}(\sigma_1 < \sigma_{1,2} < \bar{\tau}_n < \infty) \leqslant \rho^{n(1-\epsilon/2)} + \mathbb{P}_{(x,m)}(\sigma_1 < \sigma_{1,2} < \bar{\sigma}_{1,2} < \bar{\tau}_n < \infty) \qquad (66)$$

for $n > n_1$.

To bound the last probability we observe that $\bar{X}_{\bar{\sigma}_{1,2}}$ lies on $\{x \in \mathbb{Z} \times \mathbb{Z}_+ : x(1) + x(2) = 0\}$; by Proposition 3.2, starting from this line, the probability of $\bar{X}$ ever hitting $\{x \in \mathbb{Z} \times \mathbb{Z}_+ : x(1) + x(2) = n\}$ is bounded from above by

$$\frac{1}{c_2}(h_{\rho_2}((n - x(1), x(2)), m) + c_1 h_{\rho_1}((n - x(1), x(2)), m))$$

$$\leqslant \frac{1}{c_2}\left(\rho_2^n \boldsymbol{d}_2(m) + c_1 \rho_1^n \boldsymbol{d}_1(m)\right);$$

this and the strong Markov property of $\bar{X}$ give:

$$\mathbb{P}_{(x,m)}(\sigma_1 < \sigma_{1,2} < \bar{\sigma}_{1,2} < \bar{\tau}_n < \infty) \leqslant c_7 \rho^n$$

where c_7 is a constant depending on $\boldsymbol{d}_1$, $\boldsymbol{d}_2$, c_1 and c_2. Substituting this in (66) gives

$$\mathbb{P}_{(x,m)}(\sigma_1 < \sigma_{1,2} < \bar{\tau}_n < \infty) \leqslant \rho^{n(1-\epsilon/2)} + c_7 \rho^n,$$

for $n > n_1$. This implies the statement of the proposition. $\qquad \square$

Finally, we state and prove our main theorem:

Theorem 6.1. *For any $x \in \mathbb{R}_+^2$, $x(1) + x(2) < 1$, and $m \in \mathcal{M}$ (if $\rho_1 > \rho_2$ and $x(2) < 1 - \log(\rho_1)/\log(\rho_2)$ we also require $x(1) > 0$) there exists $c_8 > 0$ and $N > 0$ such that*

$$\frac{|\mathbb{P}_{(x_n,m)}(\tau_n < \tau_0) - \mathbb{P}_{(T_n(x_n),m)}(\tau < \infty)|}{\mathbb{P}_{(x_n,m)}(\tau_n < \tau_0)} < \rho^{c_8 n} \qquad (67)$$

for $n > N$, where $x_n = \lfloor nx \rfloor$.

Proof. Proposition 5.2, the choice of x (i.e., $x(1) + x(2) < 1$ and $x(1) > 0$ and furthermore $x(2) < 1 - \log(\rho_1)/\log(\rho_2)$) when $\rho_1 > \rho_2$) imply the lower bound

$$\mathbb{P}_{(x,m)}(\tau_n < \tau_0) \geqslant \rho^{n(1-2c_8)} \qquad (68)$$

for some constant $1/2 > c_8 > 0$ depending on x.

By definition $\bar{X}$ hits ∂A_n exactly when Y hits ∂B, i.e., $\bar{\tau}_n = \tau$; therefore, $\mathbb{P}_{(x_n,m)}(\bar{\tau}_n < \infty) = \mathbb{P}_{(T_n(x_n),m)}(\tau < \infty)$ and

$$\frac{|\mathbb{P}_{(x_n,m)}(\tau_n < \tau_0) - \mathbb{P}_{(T_n(x_n),m)}(\tau < \infty)|}{\mathbb{P}_{(x_n,m)}(\tau_n < \tau_0)} \qquad (69)$$

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

29

We partition the probabilities of events $\{\tau_n < \tau_0\}$ and $\{\tau < \infty\}$ as follows

$$
\begin{aligned}
\mathbb{P}_{(x_n,m)}(\tau_n < \tau_0) = {} & \mathbb{P}_{(x_n,m)}(\tau_n < \sigma_1 < \tau_0) + \mathbb{P}_{(x_n,m)}(\sigma_1 < \tau_n \leqslant \sigma_{1,2} \wedge \tau_0) \\
& + \mathbb{P}_{(x_n,m)}(\sigma_1 < \sigma_{1,2} < \tau_n < \tau_0)
\end{aligned} \tag{70}
$$

$$
\begin{aligned}
\mathbb{P}_{(T_n(x_n),m)}(\tau < \infty) = {} & \mathbb{P}_{(T_n(x_n),m)}(\tau < \sigma_1) + \mathbb{P}_{(T_n(x_n),m)}(\sigma_1 < \tau \leqslant \sigma_{1,2}) \\
& + \mathbb{P}_{(T_n(x_n),m)}(\sigma_1 < \sigma_{1,2} < \tau < \infty).
\end{aligned} \tag{71}
$$

Lemma 6.1 says the processes X and $\bar{X}$ move together until they hit ∂_1, so

$$
\mathbb{P}_{(x_n,m)}(\tau_n < \sigma_1 < \tau_0) = \mathbb{P}_{(T_n(x_n),m)}(\tau < \sigma_1).
$$

After hitting ∂_1, the sum of the components of X and $\bar{X}$ are still equal until one of the processes hits ∂_2. Lemma 6.1 now gives

$$
\mathbb{P}_{(x_n,m)}(\sigma_1 < \tau_n \leqslant \sigma_{1,2} \wedge \tau_0) = \mathbb{P}_{(T_n(x_n),m)}(\sigma_1 < \tau \leqslant \sigma_{1,2}).
$$

The last two equalities, Propositions 4.1, 6.1, and partitions (70), (71) imply that there exists $n_0 > 0$ such that

$$
\mid \mathbb{P}_{(x_n,m)}(\sigma_1 < \sigma_{1,2} < \tau_n < \tau_0) - \mathbb{P}_{(T_n(x_n),m)}(\sigma_1 < \sigma_{1,2} < \tau < \infty) \mid \leqslant \rho^{n(1-c_8)} \tag{72}
$$

for $n > n_0$. Substituting the last bound and (68) in (69) gives (67). $\qquad \square$

7　Computation of $\mathbb{P}(\tau < \infty)$

Theorem 6.1 tells us that $\mathbb{P}_{(T_n(x_n),m)}(\tau < \infty)$ approximates $\mathbb{P}_{(x_n,m)}(\tau_n < \tau_0)$ very well. In this section we develop approximate formulas for $\mathbb{P}_{(y,m)}(\tau < \infty)$. Recall that a (Y,M)-harmonic function is said to be ∂B-determined if it of the form

$$
(y,m) \mapsto \mathbb{E}_{(y,m)}[f(Y_\tau, M_\tau) 1_{\{\tau < \infty\}}]
$$

for some function f. The function

$$
(y,m) \mapsto \mathbb{P}_{(y,m)}(\tau < \infty) \tag{73}
$$

is (Y,M)-harmonic with $f = 1$ on ∂B. Furthermore, by definition it is ∂B-determined, (for (73), f is the function taking the constant value 1 on ∂B). Our approach to the approximation of $\mathbb{P}_{(y,m)}(\tau < \infty)$ is based on the classical superposition principle: take linear combinations of the (Y,M)-harmonic functions identified in Propositions 2.7 and 2.10 to approximate the value 1 on ∂B as closely as possible. We need our (Y,M)-harmonic functions to be ∂B-determined; the next lemma identifies a simple condition for functions of the form (74) to be ∂B-determined.

Lemma 7.1. *Suppose* (β, α_j, d_j) *are points on* $\mathcal{H}$ *and suppose*

$$
h(y,m) = \sum_{j=1}^{k} b(j)[(\beta, \alpha_j, d_j), \cdot], \tag{74}
$$

$k \geqslant 1$, *is* (Y,M)*-harmonic. If* $|\beta| < 1$ *and* $|\alpha_j| \leqslant 1$ *then* h *is* ∂B*-determined.*

This generalizes [10, Proposition 2.2, 4.10] to the Markov modulated setup.

Proof. Define the region $U = \{y \in \mathbb{Z} \times \mathbb{Z}_+ : 0 \leqslant y(1) - y(2) \leqslant n\}$ and the boundaries of U $\partial U_1 = \{y \in \mathbb{Z} \times \mathbb{Z}_+ : y(1) - y(2) = n\}$ and $\partial U_2 = \partial B$. Define $\upsilon_n \doteq \inf\{k : Y_k \in \partial U_1\}$. We make the following claim: starting from a point $y \in U$, (Y, M) hits $\partial U_1 \cup \partial U_2$ in finite time, i.e., $\upsilon_n \wedge \tau < \infty$ almost surely. Let us first prove this claim. For each modulating state m, the sample path of (Y, M) consisting only of increments $(0, -1)$ hits ∂U_2 in at most n steps and the probability of this path is $(\lambda(m) \boldsymbol{P}(m, m))^n$. Then if we set

$$\varepsilon = \min_{m \in \mathcal{M}} (\lambda(m) \boldsymbol{P}(m, m))^n$$

we have

$$\mathbb{P}_{(y,m)}(\tau \wedge \upsilon_n \geqslant n) \leqslant (1 - \varepsilon).$$

An iteration of this inequality and the Markov property of (Y, M) give

$$\mathbb{P}_{(y,m)}(\tau \wedge \upsilon_n \geqslant kn) \leqslant (1 - \varepsilon)^k.$$

Letting $k \to \infty$ gives

$$\mathbb{P}_{(y,m)}(\tau \wedge \upsilon_n = \infty) = 0. \tag{75}$$

Definition (74) and $|\alpha_j| \leqslant 1$, $|\beta| < 1$ imply that h is bounded on B. This and that h is (Y, M)-harmonic imply that

$$S_k = h(Y_{\tau \wedge \upsilon_n \wedge k}, M_{\tau \wedge \upsilon_n \wedge k})$$

is a bounded martingale. The optional sampling theorem applied to this martingale and (75) imply

$$\begin{aligned} h(y, m) &= \mathbb{E}_{(y,m)}[h(Y_{\tau \wedge \upsilon_n}, M_{\tau \wedge \upsilon_n})] \\ &= \mathbb{E}_{(y,m)}[h(Y_\tau, M_\tau) 1_{\{\tau < \upsilon_n\}}] + \mathbb{E}_{(y,m)}[h(Y_{\upsilon_n}, M_{\upsilon_n}) 1_{\{\upsilon_n \leqslant \tau\}}]. \end{aligned} \tag{76}$$

That $|\alpha_j| \leqslant 1$ implies $|h(Y_{\upsilon_n}, M_{\upsilon_n})| \leqslant c\beta^n$ for some constant $c > 0$. Therefore,

$$\lim_{n \to \infty} \mathbb{E}_{(y,m)}[h(Y_{\upsilon_n}, M_{\upsilon_n}) 1_{\{\upsilon_n \leqslant \tau\}}] \leqslant c \lim_{n \to \infty} \beta^n = 0.$$

The last expression, that $\lim_{n \to \infty} \upsilon_n = \infty$ and letting $n \to \infty$ in (76) imply

$$h(y, m) = \mathbb{E}_{(y,m)}[h(Y_\tau, M_\tau) 1_{\{\tau < \infty\}}],$$

i.e, $h(y, m)$ is ∂B-determined. $\qquad\square$

The last lemma and $0 < \rho_1 < 1$ imply

Lemma 7.2. h_{ρ_1} *is ∂B-determined.*

Recall that we have constructed a (Y, M)-superharmonic function, h_{ρ_2} from the roots $(\rho_2, 1), (\rho_2, \alpha_1^*) \in \mathcal{H}^{\beta\alpha}$. We would like to strengthen this to a (Y, M)-harmonic function. This requires the use of further conjugate points of $(\rho_2, 1)$ (in addition to (ρ_2, α_1^*)). The next lemma shows that under Assumptions 1 and (27) we have sufficient number of conjugate points of $(\rho_2, 1)$ to work with:

Lemma 7.3. *Let (ρ_2, α_1^*) be the point conjugate to $(\rho_2, 1)$ identified in Proposition 2.12. Under Assumptions 1 and (27), there exists $|\mathcal{M}| - 1$ additional conjugate points (ρ_2, α_j^*), $j = 2, 3, ..., |\mathcal{M}|$, of $(\rho_2, 1)$ with $0 < \alpha_j^* < \alpha_1^*$.*

Proof. We know that $\Lambda_1(\rho_2, \alpha_1^*) = 1$; then $\Lambda_j(\rho_2, \alpha_1^*) < 1$ for $j = 2, 3, ..., |\mathcal{M}|$. On the other hand, Gershgorin's Theorem implies $\lim_{\alpha \to 0} \Lambda_j(\rho_2, \alpha) = \infty$. These and the continuity of Λ_j imply the existence of $\alpha_j^* \in (0, \alpha_1^*)$ such that $\Lambda_j(\rho_2, \alpha_j^*) = 1$. $\square$

To construct our (Y, M)-harmonic functions from the points identified in the previous lemma we need the following assumption:

$$c(\rho_2, 1, d_2) \in \text{Span}\left(c(\rho_2, \alpha_j^*, d_{2,j}), j = 1, 2, ..., |\mathcal{M}|\right). \tag{77}$$

Remark 7.1. *By definition, $c(\rho_2, \alpha_j, d_{2,j}) = 0$ if $\alpha_j = \rho_2$. Therefore, only those j satisfying $\alpha_j^* \neq \rho_2$ have a role in determining $\text{Span}\left(c(\rho_2, \alpha_j^*, d_{2,j}), j = 1, 2, ..., |\mathcal{M}|\right)$. In this sense, assumption (77) can be seen as an extension of (31) (or, equivalently, of (28)).*

Remark 7.2. *The linear independence of $c(\rho_2, \alpha_j^*, d_{2,j})$, $j = 1, 2, ..., |\mathcal{M}|$, is sufficient for (77) to hold. That $c(\beta, \beta, d) = 0$ implies that $\rho_2 \neq \alpha_j^*$ for all $j = 1, 2, .., |\mathcal{M}|$ is a necessary condition for this independence.*

Now on to the (Y, M)-harmonic function:

Proposition 7.1. *Let (ρ_2, α_j^*) be the conjugate points of $(\rho_2, 1)$ identified in Proposition 2.12 and Lemma 7.3. Under the additional assumption (77), one can find a vector $b_{2,1} \in \mathbb{R}^{m_1}$ such that*

$$\mathfrak{h}_{\rho_2} \doteq [(\rho_2, 1, d_2), \cdot] + \sum_{j=1}^{|\mathcal{M}|} b_{2,1}(j)[(\rho_2, \alpha_j^*, d_{2,j}), \cdot] \tag{78}$$

is (Y, M)-harmonic and ∂B-determined.

Proof. Assumption (77) implies that the collection of vectors $c(\rho_2, 1, d_2)$, $c(\rho_2, \alpha_j^*, d_{2,j})$, $j = 1, 2, ..., |\mathcal{M}|$ are linearly dependent. Therefore, by Proposition 2.10, there exists a vector $b' \in \mathbb{R}^{|\mathcal{M}|+1}$ such that

$$b'(0)[(\rho_2, 1, d_2), \cdot] + \sum_{k=1}^{n_1} b'(j)[(\rho_2, \alpha_{j_k}^*, d_{2,j_k}), \cdot]$$

is (Y, M)-harmonic. Assumption (77) implies that one can choose b' so that $b'(0) \neq 0$. Renormalizing the last display by $b'(0)$ gives (78). That $\mathfrak{h}_{\rho_2}$ is ∂B-determined follows from $0 < \alpha_j^* \leqslant 1$, $\rho_2 < 1$ and Lemma 7.1. $\square$

Next proposition constructs an approximation of $\mathbb{P}_{(y,m)}(\tau < \infty)$ with bounded relative error from functions $\mathfrak{h}_{\rho_2}$ and h_{ρ_1}.

Proposition 7.2. *There exist constants c_9, c_{10} and c_{11} such that*

$$\mathbb{P}_{(y,m)}(\tau < \infty) < h^{a,0}(y, m) < c_9 \mathbb{P}_{(y,m)}(\tau < \infty) \tag{79}$$

where

$$h^{a,0} \doteq c_{11}(\mathfrak{h}_{\rho_2} + c_{10} h_{\rho_1}). \tag{80}$$

Proof. The proof is similar to that of Proposition 3.1. That $0 < \alpha_j^* < 1$, $j = 2, 3, ..., |\mathcal{M}|$ imply that

$$[(\rho_2, \alpha_j^*, \boldsymbol{d}_{2,j}), (k, k, m)] = (\alpha_j^*)^k \boldsymbol{d}_{2,j}(m) \to 0 \tag{81}$$

as $k \to \infty$. We further have

$$[(\rho_2, 1, \boldsymbol{d}_2), (k, k, m)] = \boldsymbol{d}_2(m) > 0, \tag{82}$$

for all $k \geqslant 0$. This and (81) imply that there exists $k_0 > 0$ such that

$$\mathfrak{h}_{\rho_2}(k, k, m) > \min_{m \in \mathcal{M}} \boldsymbol{d}_2(m)/2 \tag{83}$$

for all $k > k_0$. On the other hand,

$$h_{\rho_1}(k, k, m) = [(\rho_1, \rho_1, \boldsymbol{d}_1), (k, k, m)] = \boldsymbol{d}_1(m)\rho_1^k > 0, \tag{84}$$

for all k. Then we can choose $c_{10} > 0$ large enough so that

$$\mathfrak{h}_{\rho_2}(k, k, m) + c_{10} h_{\rho_1}(k, k, m) \geqslant \min_{m \in \mathcal{M}} \boldsymbol{d}_2(m)/2 \tag{85}$$

for all $k \leqslant k_0$. The last display, (83) and the positivity of $c_{10} h_{\rho_1}$ imply that the last display holds for all k and $m \in \mathcal{M}$. Set

$$c_{11} \doteq \left(\min_{m \in \mathcal{M}} \boldsymbol{d}_2(m)/2 \right)^{-1},$$

and $h^{a,0}$ be as in (80). That (85) holds for $k \geqslant 0$ and $m \in \mathcal{M}$ implies

$$h^{a,0}|_{\partial B} \geqslant 1.$$

By Lemma 7.2 and Proposition 7.1 $h^{a,0}$ is (Y, M)-harmonic and ∂B-determined. This and the last display imply,

$$h^{a,0}(y, m) = \mathbb{E}_{(y,m)}[h^{a,0}(Y_\tau, M_\tau)1_{\{\tau < \infty\}}] \geqslant \mathbb{P}_{(y,m)}(\tau < \infty). \tag{86}$$

This proves the first inequality in (79). To choose c_9 so that the second inequality in (79) holds we note the following: (81), (82) and (84) imply

$$c_9 \doteq \max_{k \geqslant 0, m \in \mathcal{M}} h^{a,0}(k, k, m) < \infty.$$

Now the same argument giving (86) implies the second inequality in (79). $\qquad\square$

Proposition 7.3. *Fix* $m \in \mathcal{M}$ *and* $x \in \mathbb{R}_+^2$, *such that* $0 < x(1) + x(2) < 1$; *furthermore assume* $x(1) > 0$ *if* $\rho_1 > \rho_2$ *and* $x(2) \leqslant 1 - \log(\rho_1)/\log(\rho_2)$; *set* $x_n = \lfloor nx \rfloor$. *Then* $h^{a,0}(T_n(x_n))$ *approximates* $\mathbb{P}_{(x_n,m)}(\tau_n < \tau_0)$ *with relative error whose* $\limsup$ *in* n *is bounded by* $|c_9 - 1|$.

Proof. We know by the previous proposition that $h^{a,0}$ approximates $\mathbb{P}_{(y,m)}(\tau < \infty)$ with relative error bounded by $|c_9 - 1|$; we also know by Theorem 6.1 that $\mathbb{P}_{(T_n(x_n),m)}(\tau < \infty)$ approximates $\mathbb{P}_{(x_n,m)}(\tau_n < \tau_0)$ with vanishing relative error. These two imply the statement of the proposition. $\qquad\square$

8 Improving the approximation

Proposition 7.3 tells us that $h^{a,0}$ of (80) approximates $\mathbb{P}_{(y,m)}(\tau < \infty)$ and therefore $\mathbb{P}_{(x,m)}(\tau_n < \tau_0)$ with bounded relative error. The works [10, 11, 12] covering the non-modulated case are able to construct progressively better approximations (i.e., reduction of the relative error) by using more harmonic functions constructed from conjugate points (in the tandem case with no modulation, one is able to construct an exact representation of $\mathbb{P}_y(\tau < \infty)$ so no reduction in relative error is necessary). This is possible because the function in [10, 11, 12] corresponding to $\mathfrak{h}_{\rho_2}$, takes the value 1 on ∂B away from the origin. Thus, by and large, that single function provides an excellent approximation of $\mathbb{P}_y(\tau < \infty)$ for points away from ∂_2. Rest of the harmonic functions are added to the approximation to improve the approximation along ∂_2.

When a modulating chain is present, the situation is different. Note that (81), (82) imply that the value of $\mathfrak{h}_{\rho_2}$ on ∂B, away from the origin, is determined by the eigenvector $\boldsymbol{d}_2$ and in general, the components of $\boldsymbol{d}_2$ will change with m. We need to improve $\mathfrak{h}_{\rho_2}$ itself so that we have a (Y, M)-harmonic function that is close to 1 on ∂B away from the origin.

How is this to be done? Remember that the construction of $\mathfrak{h}_{\rho_2}$ began with fixing $\alpha = 1$ and solving

$$\beta^{|\mathcal{M}|}\boldsymbol{p}(\beta, 1) = 0; \tag{87}$$

ρ_2 is the largest root of this equation in the interval $(0, 1)$. Then we fixed $\beta = \rho_2$ in $\alpha^{|\mathcal{M}|}\boldsymbol{p}(\rho_2, \alpha) = 0$ and solved for α to find the conjugate points (ρ_2, α_j^*) of $(\rho_2, 1)$; from these points we constructed $\mathfrak{h}_{\rho_2}$. Now to get our (Y, M)-harmonic function that almost takes the value 1 on ∂B away from the origin we will use the rest of the roots of (87) in $(0, 1)$. The next lemma shows that under the stability assumption and the simpleness of all eigenvalues, $|\mathcal{M}| - 1$ real β roots exist that lies in the interval $(0, \rho_2)$. The proposition after that constructs the desired (Y, M)-harmonic function from these roots.

Lemma 8.1. *Under the stability assumption (1), and Assumption 1 (all eigenvalues of $\boldsymbol{A}(\beta, \alpha)$ are real and simple for $(\beta, \alpha) \in \mathbb{R}_+^{2o}$) there exist $\rho_{2,j}$, $j = 2, 3, ..., |\mathcal{M}|$, such that $\rho_2 > \rho_{2,2} > \rho_{2,3} > \cdots > \rho_{2,|\mathcal{M}|} > 0$ and $\{\boldsymbol{e}_2 \neq 0,\ \boldsymbol{e}_3 \neq 0,...,\boldsymbol{e}_{|\mathcal{M}|} \neq 0\} \subset \mathbb{R}^{\mathcal{M}}$ such that*

$$\boldsymbol{A}(\rho_{2,j}, 1)\boldsymbol{e}_j = \boldsymbol{e}_j,\ j = 2, 3, ..., |\mathcal{M}|,$$

holds.

The proof is parallel to that of Lemma 7.3 and is based on Gershgorin's Theorem and the fact that $\Lambda_j(\rho_2, 1) < 1$ for $j = 2, 3, ..., |\mathcal{M}|$.

Each of the points $(\rho_{2,j}, 1)$ will in general have $2|\mathcal{M}| - 1$ conjugate points. To get ∂B-determined (Y, M)-harmonic functions from these we need the analog of (77) for each $(\rho_{2,j}, 1)$:

Assumption 2. *For each $j = 2, 3, ..., |\mathcal{M}|$ there exists $m_j \leqslant |\mathcal{M}|$ conjugate points $(\rho_{2,j}, \alpha_{j,l}^*)$, $l = 1, 2, ..., m_j$, of $(\rho_{2,j}, 1)$ and eigenvectors $0 \neq \boldsymbol{e}_{j,l} \in \mathbb{R}^{\mathcal{M}}$ such that*

$$|\alpha_{j,l}^*| < 1,\ l = 1, 2, ..., m_j,$$
$$\boldsymbol{A}(\rho_{2,j}, \alpha_{j,l}^*)\boldsymbol{e}_{j,l} = \boldsymbol{e}_{j,l}$$
$$\boldsymbol{c}(\rho_{2,j}, 1, \boldsymbol{e}_j) \in Span(\boldsymbol{c}(\rho_{2,j}, \alpha_{j,l}^*, \boldsymbol{e}_{j,l}), l = 1, 2, .., m_j). \tag{88}$$

Remark 8.1. *Similar to the comments made in Remark 7.2, a set of sufficient conditions for (88) is 1) $m_j = |\mathcal{M}|$ and 2) $\boldsymbol{c}(\rho_{2,j}, \alpha^*_{j,l}, \boldsymbol{e}_{j,l})$, $l = 1, 2, ..., |\mathcal{M}|$ are linearly independent. By $\boldsymbol{c}(\cdot, \cdot, \cdot)$'s definition, linear independence of these vectors require $\alpha^*_{j,l} \neq \rho_{2,j}$, which is, yet another generalization of the assumption $\rho_1 \neq \rho_2$.*

Remark 8.2. *One can introduce assumptions similar to (27) which imply, with an argument similar to the proof of Lemma 7.3, that $(\rho_{2,j}, 1)$ has $|\mathcal{M}| - j$ conjugate points in the interval $(0, 1)$. But in general, this number of conjugate points will not suffice for (88) to hold and when constructing (Y, M)-harmonic functions with $\beta = \rho_{2,j}$, $j = 2, 3, ..., |\mathcal{M}|$, we will use conjugate points with complex or negative α components. Instead of introducing even more assumptions similar to (27), we directly incorporate (88) as an assumption.*

To get our (Y, M)-harmonic function converging to 1 on the tail of ∂B (see (92) below for the precise statement) we need one more condition:

$$\mathbf{1} \in \text{Span}(\boldsymbol{d}_2, \boldsymbol{e}_2, ..., \boldsymbol{e}_{|\mathcal{M}|}). \tag{89}$$

A sufficient condition for (89) is that the vectors listed on the right of this display are linearly independent.

Proposition 8.1. *Let $\boldsymbol{e}_j$, $j = 2, 3, ..., |\mathcal{M}|$ be as in Lemma 8.1. and let $\boldsymbol{d}_2$ be as in Proposition 2.11. Under Assumptions 2 and (89) there exist vectors $\boldsymbol{b}_{2,j} \in \mathbb{R}^{m_j}$, $j = 2, 3, .., |\mathcal{M}|$ and $\boldsymbol{b}_2 \in \mathbb{R}^{|\mathcal{M}|}$ such that*

$$\mathfrak{h}_{\rho_{2,j}}(y, m) \doteq [(\rho_{2,j}, 1, \boldsymbol{e}_j), (y, m)] \tag{90}$$

$$+ \sum_{l=1}^{m_j} \boldsymbol{b}_{2,j}(l)[(\rho_{2,j}, \alpha^*_{j,l}, \boldsymbol{e}_{j,l}), (y, m)], \ j = 2, 3, ..., |\mathcal{M}|,$$

and

$$\mathfrak{h} \doteq \boldsymbol{b}_2(1)\mathfrak{h}_{\rho_2} + \sum_{j=2}^{|\mathcal{M}|} \boldsymbol{b}_2(j)\mathfrak{h}_{\rho_{2,j}} \tag{91}$$

are all (Y, M)-harmonic and ∂B-determined; furthermore

$$\lim_{k \to \infty} \mathfrak{h}(k, k, m) \to 1 \tag{92}$$

for all $m \in \mathcal{M}$.

Proof. The existence of the vector $\boldsymbol{b}_{2,j}$, $j = 2, 3, ..., |\mathcal{M}|$, so that $\mathfrak{h}_{\rho_{2,j}}$ defined in (90) is (Y, M)-harmonic follows from (2) and the argument given in the construction of $\mathfrak{h}_{\rho_2}$ (see the proof of Proposition 7.1). By (89) there is a vector $\boldsymbol{b}_2$ such that

$$\boldsymbol{b}_2(1)\boldsymbol{d}_2(m) + \sum_{j=2}^{|\mathcal{M}|} \boldsymbol{b}_2(j)\boldsymbol{e}_j(m) = 1$$

for all $m \in \mathcal{M}$. Then $\mathfrak{h}$ as defined in (91) satisfies

$$\mathfrak{h}(k, k, m) = 1 + \boldsymbol{b}_2(1) \sum_{j=1}^{|\mathcal{M}|} \boldsymbol{b}_{2,1}(j)(\alpha^*_j)^k \boldsymbol{d}_{2,j}(m) + \sum_{j=2}^{|\mathcal{M}|} \boldsymbol{b}_2(j) \sum_{l=2}^{m_j} \boldsymbol{b}_{2,j}(i)(\alpha^*_{j,l})^k \boldsymbol{e}_{j,l}(m);$$

$|\alpha^*_j| < 1$, $|\alpha^*_{j,l}| < 1$ imply that the last two sums go to 0 with k. This gives (92). $\qquad \square$

In Lemma 8.1 we found points on $\{\alpha = 1\} \cap \mathcal{H}^{\beta\alpha}$ in addition to $(\rho_2, 1)$ identified in, we used these points above in the construction of $\mathfrak{h}$. Similarly, one can go along the line $\beta = \alpha$ to find points on $\mathcal{H}^{\beta\alpha}$ other than (ρ_1, ρ_1) defining further simple ∂B-determined (Y, M)-harmonic functions:

Lemma 8.2. *Under the stability assumption (1), and Assumption 1 ($\boldsymbol{A}(\beta, \alpha)$ has real distinct eigenvalues for $(\beta, \alpha) \in \mathbb{R}_+^{2,o}$) there exist $\rho_{1,k}$, $k = 2, 3, .., |\mathcal{M}|$, such that $\rho_1 > \rho_{1,2} > \rho_{1,3} > \cdots > \rho_{1,|\mathcal{M}|} > 0$ and $\{\boldsymbol{f}_2 \neq 0, \boldsymbol{f}_3 \neq 0, ..., \boldsymbol{f}_{|\mathcal{M}|} \neq 0\} \subset \mathbb{R}^{|\mathcal{M}|}$ such that*

$$\boldsymbol{A}(\rho_{1,j}, \rho_{1,j})\boldsymbol{f}_j = \boldsymbol{f}_j, \ j = 2, 3, ..., |\mathcal{M}|,$$

holds.

The proof is parallel to that of Lemma 7.3 and is based on Gershgorin's Theorem and the fact that $\Lambda_j(\rho_1, \rho_1) < \Lambda_1(\rho_1, \rho_1) = 1$ for $j = 2, 3, ..., |\mathcal{M}|$.

One can use the points identified in the previous lemma to construct further ∂B-determined (Y, M)-harmonic functions.

Lemma 8.3. *Let $\rho_{1,j}$, $\boldsymbol{f}_j$, $j = 2, 3, ..., |\mathcal{M}|$, be as in Lemma 8.2. Then*

$$[(\rho_{1,j}, \rho_{1,j}, \boldsymbol{f}_j), \cdot], \ j = 2, 3, ..., |\mathcal{M}|,$$

are ∂B-determined (Y, M)-harmonic.

Proof. By definition, $(\rho_{1,j}, \rho_{1,j}) \in \mathcal{H}^{\beta\alpha}$ and $\boldsymbol{A}(\rho_{1,j}, \rho_{1,j})\boldsymbol{f}_j = \boldsymbol{f}_j$. Again, $\boldsymbol{A}(\beta, \beta) = \boldsymbol{A}_2(\beta, \beta)$ for all β follows from and $\boldsymbol{p}_2(\beta, \beta, m) = \boldsymbol{p}(\beta, \beta)$ and the definitions of $\boldsymbol{A}$ and $\boldsymbol{A}_2$. Then $\boldsymbol{A}(\rho_{1,j}, \rho_{1,j})\boldsymbol{f}_j = \boldsymbol{A}_2(\rho_{1,j}, \rho_{1,j})\boldsymbol{f}_j = \boldsymbol{f}_j$, i.e., $(\rho_{1,j}, \rho_{1,j}, \boldsymbol{f}) \in \mathcal{H}_2$ (i.e., the characteristic surface of ∂_2, see (18)). This and Proposition 2.7 imply that $[(\rho_{1,j}, \rho_{1,j}, \boldsymbol{f}_j), \cdot]$ is (Y, M)-harmonic. That it is ∂B-determined follows from $|\rho_{1,j}| < 1$ and Lemma 7.1. $\qquad\square$

The function $[(\beta, \alpha, d), \cdot]$ is complex valued for any $(\beta, \alpha, d) \in \mathcal{H}$ with complex components and such points and the functions they define can also be used to improve the approximation; see the next section for an example. The next proposition gives an upper bound on the relative error of an approximation of $\mathbb{P}_{(y,m)}(\tau < \infty)$ in terms of the values the approximation takes on the boundary ∂B; it covers cases when complex valued $(\beta, \alpha, d) \in \mathcal{H}$ is used in the construction of the approximation. For any $z \in \mathbb{C}$, let $\Re(z)$ denote its real part.

Proposition 8.2. *Let $h : \mathbb{Z} \times \mathbb{Z}_+ \mapsto \mathbb{C}$ be ∂B-determined and (Y, M)-harmonic. Then*

$$\max_{(y,m) \in B \times \mathcal{M}} \frac{|\Re(h)(y, m) - \mathbb{P}_{(y,m)}(\tau < \infty)|}{\mathbb{P}_{(y,m)}(\tau < \infty)} \leqslant c^* \tag{93}$$

where

$$c^* \doteq \max_{y \in \partial B, m \in \mathcal{M}} |h(y, m) - 1|. \tag{94}$$

The proof is similar to that of Proposition 7.2:

Proof. That h is ∂B-determined (Y, M)-harmonic implies the same for its real and imaginary parts. For any complex number z we have $|\Re(z) - 1| \leqslant |z - 1|$; these and (94) give

$$\max_{y' \in \partial B, m \in \mathcal{M}} |\Re(h)(y', m) - 1| \leqslant c^*.$$

Then

$$(1 - c^*)1_{\{\tau < \infty\}} \leqslant \Re(h)(Y_\tau, M_\tau)1_{\{\tau < \infty\}} \leqslant (1 + c^*)1_{\{\tau < \infty\}}.$$

Applying $\mathbb{E}_{(y,m)}[\cdot]$ to all terms above implies (93). $\qquad\square$

9 Numerical example

This section demonstrates the performance of our approximation results on a numerical example. For parameter values $\boldsymbol{P}$, $\lambda(\cdot)$, $\mu_1(\cdot)$ and $\mu_2(\cdot)$ we take those listed in (16) and (17), for which $|\mathcal{M}| = 3$. We know by Proposition 2.6 that for $\boldsymbol{P}$ as in (16), $\boldsymbol{A}(\beta, \alpha)$ has distinct positive eigenvalues for $(\beta, \alpha) \in \mathbb{R}_+^{2,o}$. Furthermore, the rates (17) satisfy $\lambda(m) < \mu_1(m), \mu_2(m)$ for all $m \in \mathcal{M}$, therefore, the stability assumption (1) is also satisfied. Computing the right side of (27) at $(\rho_2, 1)$ shows that the parameter values (16) and (17) satisfy (27). Therefore:

1. By Proposition 7.1, the function $\mathfrak{h}_{2,1}$ is well defined and ∂B-determined and (Y, M)-harmonic. Furthermore, we know by Lemma 8.1 that there are $\rho_{2,j}$, $j = 2, 3, ..., |\mathcal{M}|$, such that $0 < \rho_{2,j} < \rho_2$ and $(\rho_{2,j}, 1) \in \mathcal{H}^{\beta\alpha}$ for all j. We solve

$$\boldsymbol{p}(\rho_{2,j}, \alpha) = 0$$

   for α for the parameter values assumed in the section and verify that Assumption 2 holds with $m_j = |\mathcal{M}|$ for all j; this and Proposition 8.1 imply that the (Y, M)-harmonic ∂B-determined function $\mathfrak{h}$ defined in (91) and satisfying (92) is well defined.

2. Propositions 2.8, 2.9 and Lemma 7.2 apply and give the ∂B-determined (Y, M)-harmonic function $h_{\rho_1} = [(\rho_1, \rho_1, \boldsymbol{d}_1), \cdot]$,

3. Lemmas 8.2 and 8.3 apply and give the ∂B-determined (Y, M)-harmonic functions $h_{\rho_{1,j}} = [(\rho_{1,j}, \rho_{1,j}, \boldsymbol{f}_j), \cdot]$, $j = 2, 3, ...|\mathcal{M}|$.

In addition to these functions, we can fix an integer $K > 0$, and construct $K \cdot |\mathcal{M}|$ further (Y, M)-harmonic functions of the form

$$\mathfrak{h}_{k,j} \doteq \sum_{l=0}^{|\mathcal{M}|} \boldsymbol{b}_{k,j}(l)[(\beta_{k,j}, \alpha_{k,j,l}, \boldsymbol{d}_{k,j,l}), \cdot], \tag{95}$$

for $k = 1, 2, ..., K$, $\beta_{k,j}$ and $j = 1, 2, ..., |\mathcal{M}|$, as follows:

1. Set $\alpha_{k,j,0} = R \, e^{ik\frac{2\pi}{K+1}}$, $R \in (0, 1)$ to be determined below; note that $\alpha_{k,j,0}$ depends only on k; including j as an index simplifies notation in (95) and below.

2. For each k, $\beta_{k,j}$, $j = 1, 2, ..., |\mathcal{M}|$, are the β-roots of

$$\boldsymbol{p}(\alpha_{k,j,0}, \beta) = 0 \tag{96}$$

   satisfying $|\beta| < 1$;

3. $\alpha_{k,j,l}$, $l = 1, 2, ..., |\mathcal{M}|$, are the α-roots of

$$\boldsymbol{p}(\beta_{k,j}, \alpha) = 0; \tag{97}$$

   with $|\alpha| < 1$ which are distinct from $\alpha_{k,j,0}$.

4. $\boldsymbol{d}_{k,j,l}$ is an eigenvector of $\boldsymbol{A}(\beta_{k,j}, \alpha_{k,j,l})$ i.e., $(\beta_{k,j}, \alpha_{k,j,l}, \boldsymbol{d}_{k,j,l}) \in \mathcal{H}$,

5. for each (k, j) the vector $\boldsymbol{b}_{k,j}$ solves

$$\sum_{l=0}^{|\mathcal{M}|} \boldsymbol{b}_{k,j}(l)\boldsymbol{c}(\beta_{k,j}, \alpha_{k,j,l}) = 0, \tag{98}$$

where $\boldsymbol{b}_{k,j}(l)$ is the lth component of the vector $\boldsymbol{b}_{k,j}$. For $\mathfrak{h}_{k,j}$, $k = 1, 2, ..., K$ and $j = 1, 2, ...|\mathcal{M}|$ to be well defined, (Y, M)-harmonic and ∂B-determined we need 1) for each k, the equation (96) needs to have at least $|\mathcal{M}|$ β-roots with absolute value less than 1; 2) for each k and j, the equation (97) needs to have at least $|\mathcal{M}|$ solutions different from $\alpha_{k,j,0}$ with absolute value less than 1; 3) for each k and j the equation (98) needs to have a nontrivial solution $\boldsymbol{b}_{k,j}$. Here we have two parameters to set: K and R; for the purposes of this numerical example we set $R = 0.7$, and $K = 5$. Upon solving (96), (97) and (98) with these parameter values we observe that they have sufficient number of solutions for $\mathfrak{h}_{k,j}$ to be well defined and (Y, M)-harmonic and ∂B-determined.

We have now $1 + 6|\mathcal{M}|$, ∂B-determined (Y, M)-harmonic functions to construct our approximation of $\mathbb{P}_{(y,m)}(\tau < \infty)$; the approximation will be of the form

$$h^{a,K} \doteq \Re(h^{a*,K}), \ h^{a*,K} \doteq \mathfrak{h} + \phi_1 h_{\rho_1} + \sum_{j=2}^{|\mathcal{M}|} \phi_j[(\rho_{1,j}, \rho_{1,j}, \boldsymbol{d}_{1,j}), \cdot] + \sum_{k=1,j=1}^{K,|\mathcal{M}|} \phi_{j,k}\mathfrak{h}_{j,k}, \quad (99)$$

where ϕ_j and $\phi_{j,k}$ are $\mathbb{C}$ valued coefficients to be chosen so that $h^{a,K}|_{\partial B}$ is as close to 1 as possible. As in [10, Section 8.2], one simple way to do this is to choose these $(K + 1)|\mathcal{M}|$ coefficients so that $h^{a,K}(y, y, m) = 1$ for $y = 0, 1, 2, .., K$ and $m \in \mathcal{M}$. This defines a $(K + 1)|\mathcal{M}| \times (K+1)|\mathcal{M}|$ system; for our parameter values ($K = 5$ and $|\mathcal{M}| = 3$) this is an 18×18 system, and it does turn out to have a unique solution. Once the ϕ_j and $\phi_{j,k}$ are determined through this solution, an upper bound on the approximation relative error can be computed via Proposition 8.2; it suffices to compute c^* of (94); for $h^{a*,K}$ of (99) it turns out to be

$$c^* = 0.00367;$$

therefore, by Proposition 8.2, $h^{a,K}$ approximates $\mathbb{P}_{(y,m)}(\tau < \infty)$ with relative error bounded by this quantity. By Theorem 6.1 we know that $\mathbb{P}_{(T_n(x_n),m)}(\tau < \infty)$ approximates $\mathbb{P}_{(x_n,m)}(\tau_n < \tau_0)$ with vanishing relative error for $x_n = \lfloor nx \rfloor$, $x(1) > 0$; it follows from these that $h^{a,K}(n - x_n(1), x_n(2))$ will approximate $\mathbb{P}_{(x_n,m)}(\tau_n < \tau_0)$ with relative error bounded by c^* for n large. Let us see how well this approximation works in practice. Figure 6 gives the level curves of $-\log(h^{a,K}(n - x(1), x(2), 1))$ and $-\log \mathbb{P}_{(x,m)}(\tau_n < \tau_0)$; $\mathbb{P}_{(x,m)}(\tau_n < \tau_0)$ is computed by iterating the harmonic equation satisfied by this probability; for $n = 60$, this iteration converges in less than 1000 steps. As can be seen, and agreeing with the analysis above, these lines completely overlap except for a narrow region around the origin.
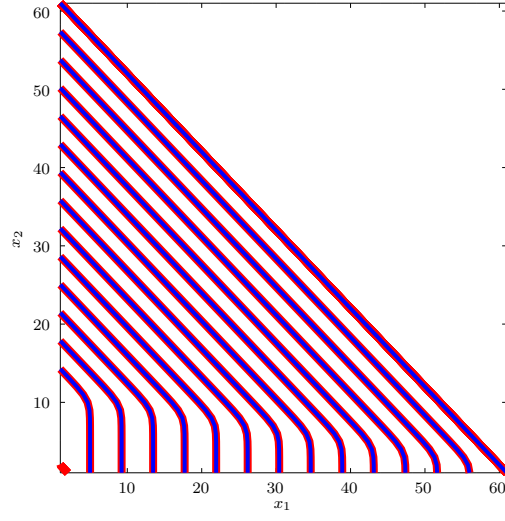
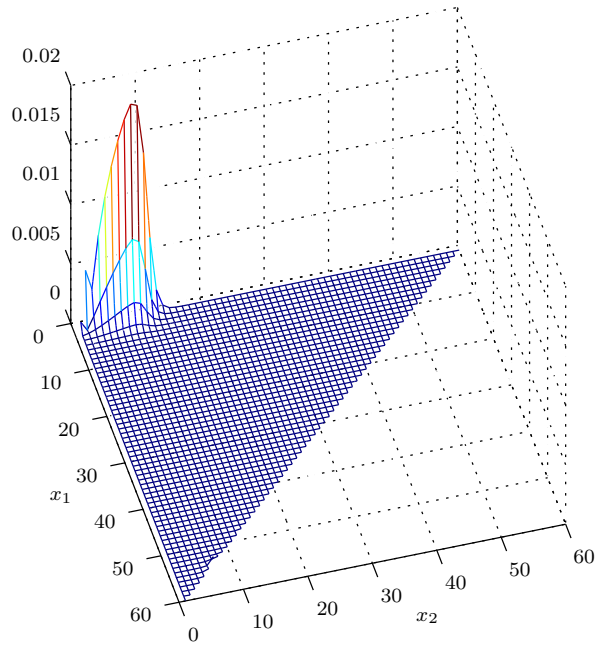Figure 6: Level curves of $-\log(h^{a,K}(n-x(1),x(2),1))$ and $-\log \mathbb{P}_{(x,m)}(\tau_n < \tau_0)$

Figure 7: The relative error $\frac{|\log(h^{a,K}(n-x(1),x(2),1)-\log \mathbb{P}_x(\tau_n < \tau_0)|}{|\log \mathbb{P}_x(\tau_n < \tau_0)|}$

Figure 7 shows the relative error

$$\frac{|\log(h^{a,K}(n - x(1), x(2), 1)) - \log \mathbb{P}_{(x,m)}(\tau_n < \tau_0)|}{|\log \mathbb{P}_{(x,m)}(\tau_n < \tau_0)|},$$

we see that it is virtually 0 except for the same region around 0 where it is bounded by 0.02. This narrow layer of where the relative error spikes corresponds to the region $1 - x(2) < \log(\rho_2)/\log(\rho_1)$ identified in Theorem 6.1.

10 Comparison with earlier works

The present work shows how one can approximate the probability $\mathbb{P}_{(x,m)}(\tau_n < \tau_0)$ by $\mathbb{P}_{(y,m)}(\tau < \infty)$ with exponentially vanishing relative error and constructs analytical approximation formulas for the latter. This is done by extending the approach of [10, 11] to Markov modulated dynamics. In this section, we compare the analysis of the modulated case treated in this work with the non-modulated two tandem case treated in [10, 11] and the non-modulated two dimensional simple random walk treated in [12].

Harmonic functions The nonmodulated analysis uses functions of the form $y \mapsto [(\beta, \alpha), y] = \beta^{y(1)-y(2)}\alpha^{y(2)}$ where (β, α) are chosen from the roots of a characteristic polynomial of second order associated with the process Y. Markov modulation brings an additional state variable m, leading to functions of the form $(y, m) \mapsto [(\beta, \alpha, \boldsymbol{d}), (y, m)] = \beta^{(y(1)-y(2)}\alpha^{y(2)}\boldsymbol{d}(m)$. The characteristic surface is now defined in terms of eigenvalue and eigenvector equations of a characteristic matrix depending on $(\beta, \alpha) \in \mathbb{C}^2$.

Geometry of the characteristic surface The characteristic surface in [10, 11, 12] is the 1-level curve of a rational function which can be represented as a second degree polynomial in each of the β, α variables; the projection of the characteristic surface to $\mathbb{R}_+^2$ is a smooth closed curve bounding a convex region. Conjugate points on this curve come in pairs and have elementary formulas. The characteristic curve in the modulated case is the 0-level curve of the characteristic polynomial of a characteristic matrix and can be represented as a $2|\mathcal{M}|$ degree polynomial in each of the variables; its projection to $\mathbb{R}_+^2$ consists of $|\mathcal{M}|$ components, one for each eigenvalue Λ_j of the characteristic matrix. The error analysis is based on the level curve of the largest eigenvalue while the computation of $\mathbb{P}_{(y,m)}(\tau < \infty)$ uses points on all components. There are in general no simple formulas for the roots of a polynomial greater than degree 4 and the formulas for degree 4 are fairly complex; therefore, for $|\mathcal{M}| \geqslant 2$ (i.e., even for the simplest nontrivial Markov modulated constrained random walk with two modulating states) the points on these curves no longer have simple formulas and identification of the relevant points (Propositions 2.8, 2.11 and 2.12, Lemmas 7.3, 8.1 and 8.2) requires matrix / eigenvalue analysis and the implicit function theorem.

Assumptions We use the point $(\rho_2, 1)$ and its conjugate (ρ_2, α_1^*) lying on $\mathcal{L}_1$ to define (Y, M)-superharmonic functions to use in our limit analysis. The existence of $(\rho_2, 1) \in \mathcal{L}_1$, follows from the stability assumption (1). The identification of the conjugate point (ρ_2, α_1^*) requires the additional assumption (27) ensuring $\alpha_1^* < 1$. A similar assumption is not needed in the non-modulated tandem case, because when there is no modulation, i.e., when $|\mathcal{M}| = 1$, the conjugate of $(\rho_2, 1)$ is (ρ_2, ρ_1) and $\rho_1 < 1$ by the stability assumption. For the constrained

random walk representing two parallel queues treated in [12], the assumption corresponding to (27) is $r^2 < \rho_2 \rho_1$, where r is utilization rate of the whole system.

The assumption $\rho_1 \neq \rho_2$ (see (28)) generalizes the assumption $\mu_1 \neq \mu_2$ from the non-modulated tandem case and the parallel case treated in [11, 12]. The computation of $\mathbb{P}_{(y,m)}(\tau < \infty)$ needs progressively more general versions of this assumption (see (77), Remark 7.2 and Assumption 2).

Analysis The approximation error analysis in the non-modulated case is based on the subsolutions of a limit HJB equation and Y-harmonic functions. These works use these subsolutions to construct supermartingales which are then used to find upper bounds on error probabilities. In this work we construct the supermartingales directly using (Y, M)-*super*harmonic functions constructed from points on the characteristic surface. Because Y has one less constraint compared to X, these functions can be subharmonic on the boundary where Y is not constrained. To overcome this, we introduce a decreasing term to the definition of the supermartingale.

In the tandem case there is an explicit formula for $P_y(\tau < \infty)$; this formula is used directly in the analysis of the error probability. There is obviously no explicit formula for the corresponding probability in the Markov modulated case. Instead, we derive an upper bound on it in Section 3 using again (Y, M)-superharmonic functions; this upperbound is used in the error analysis of Section 6.

Computation of the limit probability In the non-modulated tandem case treated in [11], $\mathbb{P}_y(\tau < \infty)$ can be represented exactly as a linear combination of h_{ρ_2} and h_{ρ_1}; so the computation of $\mathbb{P}_y(\tau < \infty)$ is trivial for the nonmodulated two dimensional tandem walk. In the parallel case treated in [12], $P_y(\tau < \infty)$ can be represented exactly as a linear combination of h_{ρ_1} and h_r when $r^2 = \rho_1 \rho_2$; when this doesn't hold [12] develops approximations of $P_y(\tau < \infty)$ from harmonic functions constructed from conjugate points on the characteristic surface, which is an application of the principle of superposition. For the modulated case we use the same principle but Markov modulation complicates the construction of the functions used in the approximation. The identification of the points on the characteristic surface requires the solution of $2|\mathcal{M}|$ degree polynomial equations (first the α component is fixed to identify possible β components; then for each of the identified β's, the polynomial is solved in α to find the relevant conjugate points). Eigenvectors corresponding to these points are then computed and finally we solve a linear equation to find the coefficients of the exponential functions (see, for example, the $\boldsymbol{b}_{k,j}$ vector in (95) and (98)). The corresponding process is trivial when there is no modulation. In [11] and [12] the function h_{ρ_2} plays a central role in the approximation of $\mathbb{P}_y(\tau < \infty)$ because it equals approximately 1 away from the origin; due to Markov modulation there can be in general no function constructed from a single point and its conjugates that takes a fixed value on ∂B. To deal with this, we use an appropriate linear combination of functions constructed from multiple points and their conjugates on the characteristic surface so that the linear combination takes the value 1 away from the origin (Proposition 8.1).

11   Conclusion

The current work develops approximate formulas for the exit probability of the two dimensional tandem walk with modulated dynamics. Our main approximation Theorem 6.1 says

that $\mathbb{P}_{(T_n(x_n),m)}(\tau < \infty)$ approximates $\mathbb{P}_{(x_n,m)}(\tau_n < \tau_0)$ with relative error vanishing exponentially fast with n. To compute the exit probability, we first construct ∂B-determined (Y, M)-harmonic functions from single and conjugate points on the corresponding characteristic surface and then with their linear combinations, approximate the boundary value 1 of the harmonic function $\mathbb{P}_{(y,m)}(\tau < \infty)$. In the non-modulated tandem case treated in [10], the probability $\mathbb{P}_y(\tau < \infty)$ can be represented in any dimension exactly using harmonic functions constructed from points on the characteristic surface. As is seen in the present work, even dimension two entails considerable difficulties. Whether an extension to higher dimensions is possible is a question we would like to tackle in future work.

The work [10] gives a formula for $\mathbb{P}_y(\tau < \infty)$ for the non-modulated tandem walk when $\rho_1 = \rho_2$ based on harmonic functions with polynomial terms. Whether similar computations can be carried out for $\mathbb{P}_{(y,m)}(\tau < \infty)$ in the modulated case when $\rho_1 = \rho_2$ is another question for future research.

The assumption (27) plays a key role in our analysis; it ensures that various functions such as h_{ρ_2} whose construction involves the point (ρ_2, α_1^*) remain bounded on ∂B. We think that new ideas will be needed to treat the case when (27) doesn't hold; this remains for future work.

The computations and the error analysis in the present work depend on the dynamics of the process and the geometry of the exit boundary. A significant problem for future research is to extend these to other dynamics in two or higher dimensions and to other exit boundaries. The simple random walk dynamics (i.e., increments $(1, 0)$, $(-1, 0)$, $(0, 1)$ and $(0, -1)$) and the rectangular exit boundary appear to be the most natural to study in immediate future work.

A Two lemmas

For a square matrix $\boldsymbol{G}$, let $\boldsymbol{G}^{i,j}$ denote the matrix obtained by removing the ith row and j^{th} column of $\boldsymbol{G}$.

Lemma A.1. *For $n_0 \in \{2, 3, 4, ...\}$, suppose $\boldsymbol{G}$ is an $n_0 \times n_0$ irreducible and aperiodic matrix with nonnegative entries. Then $\det\left((\Lambda_1(\boldsymbol{G})\boldsymbol{I} - \boldsymbol{G})^{i,i}\right) > 0$ for all $i \in \{1, 2, ..., n_0\}$, where $\boldsymbol{I}$ is the $n_0 \times n_0$ identity matrix.*

Proof. The argument is the same for all $i \in \{1, 2, ..., n_0\}$; so it suffices to argue for $i = 1$. Suppose the claim is not true and

$$\det\left((\Lambda_1(\boldsymbol{G})\boldsymbol{I} - \boldsymbol{G})^{1,1}\right) \leqslant 0. \tag{100}$$

Consider the function $u \mapsto g(u) = \det\left((u\boldsymbol{I} - \boldsymbol{G})^{1,1}\right)$, $u \geqslant 0$. The multilinearity and continuity of $\det$ implies $\lim_{u \nearrow \infty} g(u) = \infty$. This implies that if (100) is true there must be $u_0 \geqslant \Lambda_1(\boldsymbol{G})$ such that

$$\det\left((u_0\boldsymbol{I} - \boldsymbol{G})^{1,1}\right) = 0. \tag{101}$$

The matrix $\boldsymbol{G}^{1,1}$ is nonnegative, therefore, it has a largest eigenvalue $\Lambda_1(\boldsymbol{G}^{1,1})$ with an eigenvector $\boldsymbol{v}_1 \geqslant 0$. The equality (101) implies

$$\Lambda_1(\boldsymbol{G}^{1,1}) \geqslant u_0 \geqslant \Lambda_1(\boldsymbol{G}). \tag{102}$$

That $\boldsymbol{G}$ is irreducible and aperiodic implies that $\boldsymbol{G}^{n_0}$ is strictly positive; its largest eigenvalue is

$$\Lambda_1(\boldsymbol{G}^{n_0}) = \Lambda_1(\boldsymbol{G})^{n_0}.$$

The matrix $(\boldsymbol{G}^{n_0})^{1,1}$ has strictly positive entries and therefore its largest eigenvalue $\Lambda_1((\boldsymbol{G}^{n_0})^{1,1})$ has an eigenvalue $\boldsymbol{v}_2$ with strictly positive entries. For two vectors $x, y \in \mathbb{R}^d$, let $x \geqslant y$ and $x > y$ denote componentwise comparison. The inequality

$$(\boldsymbol{G}^{n_0})^{1,1} \geqslant (\boldsymbol{G}^{1,1})^{n_0}$$

implies

$$(\boldsymbol{G}^{n_0})^{1,1} \boldsymbol{v}_1 \geqslant \Lambda_1(\boldsymbol{G}^{1,1})^{n_0} \boldsymbol{v}_1. \tag{103}$$

On the other hand

$$\Lambda_1((\boldsymbol{G}^{n_0})^{1,1}) = \sup\{c : \exists x \in \mathbb{R}_+^{n_0-1}, (\boldsymbol{G}^{n_0})^{1,1} x \geqslant cx\}, \tag{104}$$

(see [5, Proof of Theorem 1, Chapter 16]). This and (103) imply

$$\Lambda_1((\boldsymbol{G}^{n_0})^{1,1}) \geqslant \Lambda_1(\boldsymbol{G}^{1,1})^{n_0}. \tag{105}$$

Define $\boldsymbol{v}_3 = [1; \boldsymbol{v}_2] \in \mathbb{R}^{n_0}$; it follows from $(\boldsymbol{G}^{n_0})^{1,1} \boldsymbol{v}_2 = \Lambda_1((\boldsymbol{G}^{n_0})1, 1)\boldsymbol{v}_2$, the strict positivity of the components of $\boldsymbol{G}^{n_0}$ and $\boldsymbol{v}_2$ that one can choose $\delta > 0$ small enough so that

$$\boldsymbol{G}^{n_0} \boldsymbol{v}_3 > \left(\Lambda_1((\boldsymbol{G}^{n_0})^{1,1} + \delta\right) \boldsymbol{v}_3;$$

This and

$$\Lambda_1(\boldsymbol{G}^{n_0}) = \sup\{c : \exists x \in \mathbb{R}_+^{n_0}, \boldsymbol{G}^{n_0} x \geqslant cx\}$$

imply

$$\Lambda_1(\boldsymbol{G}^{n_0}) > \Lambda_1\left((\boldsymbol{G}^{n_0})^{1,1}\right).$$

The last inequality, (105) and (102) imply

$$\Lambda_1(\boldsymbol{G})^{n_0} = \Lambda_1(\boldsymbol{G}^{n_0}) > \Lambda_1((\boldsymbol{G}^{n_0})^{1,1}) \geqslant \Lambda_1(\boldsymbol{G}^{1,1})^{n_0} \geqslant \Lambda_1(\boldsymbol{G})^{n_0},$$

which is a contradiction. $\qquad\square$

In our analysis we need the following fact from [9]; its proof is elementary and follows from the multilinearity of the determinant function and the previous lemma.

Lemma A.2. *Let $\boldsymbol{G}$ be an aperiodic and irreducible transition matrix. Then the row vector whose i^{th} component equals $\det\left((\boldsymbol{I} - \boldsymbol{G})^{i,i}\right)$ is the unique (upto scaling by a positive number) left eigenvector associated with the eigenvalue 1 of $\boldsymbol{G}$.*

References

[1] Abraham Berman and Robert J Plemmons, *Nonnegative matrices in the mathematical sciences*, vol. 9, Siam, 1994.

[2] Paul Dupuis, Ali Devin Sezer, and Hui Wang, *Dynamic importance sampling for queueing networks*, Annals of Applied Probability **17** (2007), no. 4, 1306–1346.

[3] Rick Durrett, *Probability: theory and examples, 4th edition*, Cambridge University press, 2010.

[4] Paul Glasserman and Shing-Gang Kou, *Analysis of an importance sampling estimator for tandem queues*, ACM Transactions on Modeling and Computer Simulation (TOMACS) **5** (1995), no. 1, 22–42.

[5] Peter D Lax, *Linear Algebra. Pure and Applied Mathematics*, (1996).

[6] Shyam Parekh and Jean Walrand, *A quick simulation method for excessive backlogs in networks of queues*, IEEE Transactions on Automatic Control **34** (1989), 54–66.

[7] NU Prabhu and Yixin Zhu, *Markov-modulated queueing systems*, Queueing systems **5** (1989), no. 1-3, 215–245.

[8] Denis Serre, *Matrices: theory and applications, volume 216 of Graduate Texts in Mathematics, second edition*, 2010.

[9] Ali Devin Sezer, *Importance sampling for a Markov modulated queuing network*, Stochastic Processes and their Applications **119** (2009), no. 2, 491–517.

[10] _____, *Exit probabilities and balayage of constrained random walks*, arXiv preprint arXiv:1506.08674 (2015).

[11] _____, *Approximation of excessive backlog probabilities of two tandem queues*, Journal of Applied Probability **55** (2018), no. 3, 968–997.

[12] Kamil Demirberk Ünlü and Ali Devin Sezer, *Excessive backlog probabilities of two parallel queues*, Annals of Operations Research (2019).