EXIT PROBABILITIES OF MARKOV MODULATED CONSTRAINED RANDOM WALKS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF APPLIED MATHEMATICS
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

FATMA BAŞOĞLU KABRAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
FINANCIAL MATHEMATICS

SEPTEMBER 2018

Approval of the thesis:

EXIT PROBABILITIES OF MARKOV MODULATED CONSTRAINED RANDOM WALKS

submitted by **FATMA BAŞOĞLU KABRAN** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Financial Mathematics Department, Middle East Technical University** by,

Prof. Dr. Ömür Uğur
Director, Graduate School of **Applied Mathematics** ______________

Prof. Dr. Sevtap Selçuk Kestel
Head of Department, **Financial Mathematics** ______________

Assoc. Prof. Dr. Ali Devin Sezer
Supervisor, **Financial Mathematics, METU** ______________

Examining Committee Members:

Prof. Dr. Sevtap Selçuk Kestel
Financial Mathematics, METU ______________

Assoc. Prof. Dr. Ali Devin Sezer
Financial Mathematics, METU ______________

Prof. Dr. Sibel Tarı
Computer Engineering, METU ______________

Prof. Dr. Fatih Tank
Insurance and Actuarial Science, Ankara University ______________

Assoc. Prof. Dr. Özge Sezgin Alp
Accounting and Financial Management, Başkent University ______________

Date: ______________

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: FATMA BAŞOĞLU KABRAN
Signature :

ABSTRACT

EXIT PROBABILITIES OF MARKOV MODULATED CONSTRAINED RANDOM WALKS

Başoğlu Kabran, Fatma

Ph.D., Department of Financial Mathematics
Supervisor : Assoc. Prof. Dr. Ali Devin Sezer

September 2018, 77 pages

Let X be the constrained random walk on $\mathbb{Z}_+^2$ with increments $(0,0)$, $(1,0)$, $(-1,1)$, $(0,-1)$ whose jump probabilities are determined by the state of a finite state Markov chain M. X represents the lengths of two queues of customers (or packets, tasks, etc.) waiting for service from two servers working in tandem; the arrival of customers occur with rate $\lambda(M_k)$, service takes place at rates $\mu_1(M_k)$, and $\mu_2(M_k)$ where M_k denotes the current state of the Markov chain M. We assume that the average arrival rate is less than the average service rates, i.e., X is assumed stable. Stability implies that X moves in cycles that restart each time it hits the origin. Let τ_n be the first time X hits the line $\partial A_n = \{x : x(1) + x(2) = n\}$, i.e., when the sum of the queue lengths equals n for the first time ; if the queues share a common buffer, τ_n represents the time of a buffer overflow and $p_n = \mathbb{P}_{(x,m)}(\tau_n < \tau_0)$ is the probability that a given cycle ends with a buffer overflow, i.e., system failure. Let Y be the same random walk as X but only constrained on $\partial_2 = \{y \in \mathbb{Z} \times \mathbb{Z}_+ : y(2) = 0\}$ and its jump probabilities for the first component reversed. Let $B = \{y \in \mathbb{Z}^2 : y(1) = y(2)\}$ and let τ be the first time Y hits B. For $x \in \mathbb{R}_+^2$, with $x(1) + x(2) < 1$ define $x_n = \lfloor nx \rfloor$ and let $m \in \mathcal{M}$ denote the initial point of the Markov chain M. We show that $\mathbb{P}_{((n-x_n(1),x_n(2)),m)}(\tau < \infty)$ approximates $\mathbb{P}_{((x_n(1),x_n(2)),m)}(\tau_n < \tau_0)$ with exponentially vanishing relative error when $x(1) > 0$. We then construct a class of harmonic functions for (Y, M) and use their linear combinations to develop approximate formulas for $\mathbb{P}_{(y,m)}(\tau < \infty)$. The construction is based on points on a characteristic surface associated with Y defined through the eigenvalues of a matrix whose components depend on the transition matrix of the modulating chain and the jump probabilities of Y. We indicate possible applications of our results and approach in finance and insurance.

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

ÖZ

MARKOV MODÜLASYONLU KISITLI RASTGELE YÜRÜYÜŞLERİN ÇIKIŞ OLASILIKLARI

Başoğlu Kabran, Fatma
Doktora, Finansal Matematik Bölümü
Tez Yöneticisi : Doç. Dr. Ali Devin Sezer

Eylül 2018 , 77 sayfa

X'in $\mathbb{Z}_+^2$'de $(0,0)$, $(1,0)$, $(-1,1)$, $(0,-1)$ adımlarını atan kısıtlı bir rastgele yürüyüş olduğunu ve adımların artış olasılıklarının homojen Markov zincirinin durumuyla belirlendiğini söyleyelim. X, ortalama servis oranları $\mu_1(M_k)$ ve $\mu_2(M_k)$ olan arka arkaya çalışan iki hizmet sağlayıcıdan hizmet almak için bekleyen, ortalama varış oranı $\lambda(M_k)$ olan müşterilerin (paket, iş v.b.) her bir servis sağlayıcıyı beklerken oluşturdukları iki kuyruğun uzunluklarını temsil eder. M_k Markov zincirinin şu andaki durumunu gösterir. Ortalama varış oranının ortalama servis oranlarından daha küçük olduğunu varsayalım, yani X'in dengeli bir süreç olduğunu kabul edelim. X dengeli olduğunda, süreç orijine her çarptığında yeniden başlayan döngülerle hareket eder. X sürecinin kuyruk uzunluklarının toplamının n olduğu sınıra, yani $\partial A_n = \{x : x(1) + x(2) = n\}$'e ilk çarpma anını τ_n ile gösterelim. Eğer kuyruklar ortak kapasite kullanıyorsa, $p_n = \mathbb{P}_{(x,m)}(\tau_n < \tau_0)$ sistemin herhangi bir döngüde kapasite aşımı, bir başka deyişle, sistemin başarısız olma olasılığıdır. Y, X ile aynı rastgele yürüyüş olmakla birlikte, sadece $\partial_2 = \{y \in \mathbb{Z} \times \mathbb{Z}_+ : y(2) = 0\}$ üzerinde kısıtlıdır ve birinci bileşeninin artış olasılıkları yer değiştirmiştir. Y'nin bileşenlerinin birbirine eşit olduğu sınırı $B = \{y \in \mathbb{Z}^2 : y(1) = y(2)\}$, B sınırına ilk çarpma anını τ olarak gösterelim. $x \in \mathbb{R}_+^2$ ve $x(1) + x(2) < 1$ için $x_n = \lfloor nx \rfloor$ şeklinde tanımlayalım ve Markov zinciri M'in ilk noktasını $m \in \mathcal{M}$ ile gösterelim. Bu tezde, $x(1) > 0$ için, $\mathbb{P}_{((x_n(1),x_n(2)),m)}(\tau_n < \tau_0)$ olasılığının $\mathbb{P}_{((n-x_n(1),x_n(2)),m)}(\tau < \infty)$ olasılığı ile, üstel olarak 0'a yakınsayan göreli hatayla, yaklaşık olarak hesaplanabileceği gösterilmiştir. (Y, M)-harmonik fonksiyonları oluşturulmuş ve bunların doğrusal kombinasyonlarıyla $\mathbb{P}_{(y,m)}(\tau < \infty)$ için yaklaşık formüller geliştirilmiştir. Harmonik fonksiyonlar, bileşenleri modülasyon zincirinin geçiş olasılıklarına ve Y sürecinin artış olasılıklarına bağlı olan bir matrisin özdeğerleri ile tanımlanan ve Y sürecine ait karakteristik yüzey üzerinde bulunan noktalar ile oluşturulmuştur. Çalışmamızın bulguları ve yaklaşımının finans ve sigortacılık alanlarındaki olası uygulamaları gösterilmiştir.

Anahtar Kelimeler: Markov modülasyonu, rejim değişimi, çok boyutlu kısıtlı rastgele yürüyüş, çıkış olasılıkları, nadir olaylar, sigortacılık sistemleri, piyasa yapıcılığı sistemi, kuyruk sistemleri, karakteristik yüzey, süperharmonik fonksiyonlar, afin dönüşüm

To My Family

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my thesis supervisor Assoc. Prof. Dr. Ali Devin Sezer for his guidance, understanding and encouragement. Without his valuable advices and continuous support, it would not be possible to conduct this thesis.

Special thanks to my progress committee members, Assoc. Prof. Dr. Ali Devin Sezer, Prof. Dr. Sevtap Selçuk Kestel, Prof. Dr. Sibel Tarı and Prof. Dr. Gerhard-Wilhelm Weber for their guidance.

Besides my supervisor Assoc. Prof. Dr. Ali Devin Sezer, I would like to thank my examining committee members Prof. Dr. Sevtap Selçuk Kestel, Prof. Dr. Sibel Tarı, Prof. Dr. Fatih Tank and Assoc. Prof. Dr. Özge Sezgin Alp for their comments and suggestions.

I would like to express my thanks to TÜBİTAK (Scientific and Technologial Research Council of Turkey) for its financial support during my graduate education.

Furthermore, I would like to thank all the members of Institute of Applied Mathematics at METU for their help and support.

I would also like to thank all my friends, especially Kamil Demirberk Ünlü, for their friendship and motivation.

Finally, I would like to thank my family for their emotional support. The final thanks to my husband Alper for his full support, patience and love.

TABLE OF CONTENTS

ABSTRACT vii

ÖZ ix

ACKNOWLEDGMENTS xiii

TABLE OF CONTENTS xv

LIST OF FIGURES xvii

CHAPTERS

1 INTRODUCTION AND PRELIMINARIES 1

2 HARMONIC FUNCTIONS 9

2.1 Harmonic functions of (Y, M) 9

2.1.1 Geometry of the characteristic surface 11

2.1.2 Construction of (Y, M)-harmonic and superharmonic functions 15

3 LIMIT ANALYSIS 29

3.1 Upper bound for $\mathbb{P}_{(y,m)}(\tau < \infty)$ 29

3.2 Upper bound for $\mathbb{P}_{(x,m)}(\sigma_1 < \sigma_{1,2} < \tau_n < \tau_0)$ 31

3.2.1 $\rho_1 > \rho_2$ 31

3.2.2 $\rho_1 < \rho_2$ 37

3.3 Lower bound for $\mathbb{P}_{(x,m)}(\tau_n < \tau_0)$ 39

3.4 Completion of the limit analysis 40

4 COMPUTATION OF $\mathbb{P}(\tau < \infty)$ 45

4.1 Improving the approximation 50

4.2 Numerical example 54

5 APPLICATIONS IN FINANCE AND INSURANCE 59

5.1 Reserve Problem 59

5.2 Market Maker's Inventory Problem 62

6 LITERATURE REVIEW 65

7 CONCLUSION AND OUTLOOK 69

7.1 Comparison with earlier works 69

7.2 Conclusion 71

REFERENCES 73

CURRICULUM VITAE 77

LIST OF FIGURES

FIGURES

Figure 1.1 Markov modulated two tandem queues 2

Figure 1.2 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 3

Figure 1.3 The transformation of (X, M) 6

Figure 2.1 Level curves 15

Figure 2.2 $\rho_1 - \rho_2$ and $\alpha_1^* - \rho_2$ have the same sign (Lemma 2.20); 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) 25

Figure 3.1 A sample path of X_k(left) and $\bar{X}_k$(right) 41

Figure 4.1 Conjugate points of $(\rho_2, 1)$ 47

Figure 4.2 Level curves of $-\log(h^a(n-x(1), x(2), 1))$ and $-\log \mathbb{P}_{(x,m)}(\tau_n < \tau_0)$ 57

Figure 4.3 The relative error $\frac{|\log(h^a(n-x(1),x(2),1)-\log \mathbb{P}_x(\tau_n<\tau_0)|}{|\log \mathbb{P}_x(\tau_n<\tau_0)|}$ 57

Figure 4.4 The relative error (detailed graph around the origin) 58

Figure 5.1 The dynamics of (X, M) 60

Figure 5.2 A sample path of (X, M) 61

Figure 5.3 The dynamics of N 63

Figure 5.4 The dynamics of N when no short selling is allowed 64

Figure 6.1 Boundary layers 66

CHAPTER 1

INTRODUCTION AND PRELIMINARIES

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. Markov modulated stochastic processes are those in which the dynamics of the underlying X process is affected by the changes in a secondary Markov process M modeling the environment within which X operates [23]; if the dynamics of M is independent of that of X the environment is said to be external, otherwise internal. Constrained processes in general, and constrained random walks in particular are used to model processes where there are natural barriers to the dynamics of the underlying processes that keep them within a given set. Most human-made systems are of this type, for example, computer networks, machines, systems serving customers, factories, etc. In finance and insurance applications constraints can represent short selling constraints, dividend payments or the total number of securities available for trading. We give two examples of these in Chapter 5 one in modeling the reserves of an insurance system operating under possible regime switches and another in the inventory management of a market maker, again operating under possible regime switches.

This thesis will concern itself with the approximation of the probability of a failure of a system whose dynamics are modeled by a Markov modulated constrained random walk (the precise probability of interest is given in (1.5)). Analysis, simulation and approximation of this probability, for ordinary (non-modulated) constrained random walks have received great interest at least since [15] and we review some of the related literature below. However there is hardly any study on the same probability for modulated constrained random walks, in fact we are aware of only [27], we comment further on this below. For this reason, this thesis 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.

The tandem walk, which we will denote by X, is the constrained random walk on $\mathbb{Z}_+^2$ with increments $\{I_1, I_2, I_3, ...\}$, constrained to remain in $\mathbb{Z}_+^2$:

$$X_0 = x \in \mathbb{Z}_+^2, \quad X_{k+1} \doteq X_k + \pi(X_k, I_{k-1}), k = 1, 2, 3, ...$$

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

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 $P \in \mathbb{R}_+^{|\mathcal{M}|\times|\mathcal{M}|}$. To ease analysis we will assume P to be irreducible, which implies that it has a unique stationary measure π on $\mathcal{M}$, i.e., $\pi = \pi P$. Let $\mathscr{F}_k \doteq \sigma(\{M_j, j \leq k+1\}, \{X_j, j \leq k\})$, i.e. 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 process (X, M) represents the dynamics of the embedded random walk of a continuous time queueing system (or computer network, or two algorithms running in tandem on a computer) consisting of two tandem queues whose arrival and service rates are determined by a finite state Markov process M (see Figure 1.1).

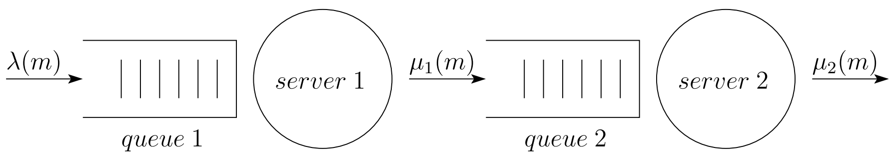

Figure 1.1: Markov modulated two tandem queues

(X, M) has the following dynamics:

- If the current state of M is M_{k-1}, M jumps to state M_k with transition probability $P(M_{k-1}, M_k)$. In this situation, I_k takes value 0, which means there is no increment.
- If the current state of M does not change with transition probability $P(M_{k-1}, M_{k-1})$, i.e. $M_{k-1} = M$, I_k takes value $(1, 0)$ with arrival rate $\lambda(M_k)$, or $(-1, 1)$ with service rate $\mu_1(M_k)$ or $(0, -1)$ with service rate $\mu_2(M_k)$. Hence, I_k takes a nonzero value only when the state of M does not change.

These dynamics are shown in Figure 1.2 where transition between states is represented by the multilayer structure.

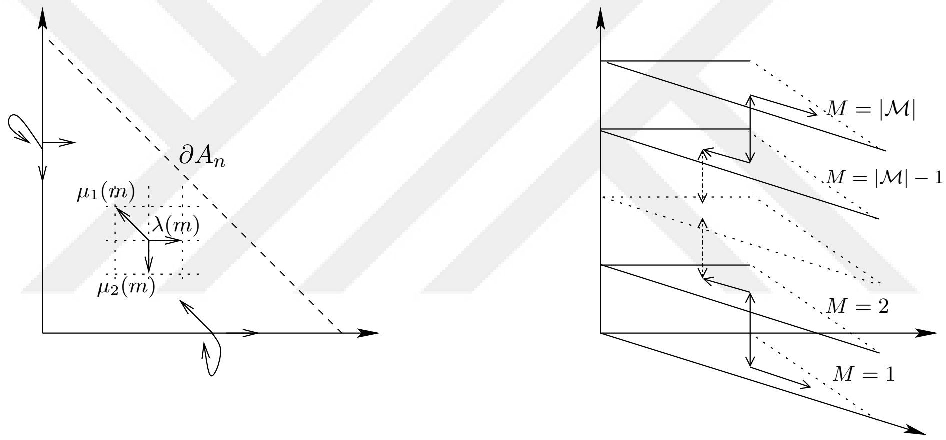

Figure 1.2: Markov modulated Constrained Random Walk (X, M); the left figure shows dynamics in a given layer, the right figure shows jumps between layers representing regime switches

We assume (X, M) to be stable:

$$\sum_{m \in \mathcal{M}} (\lambda(m) - \mu_i(m))\pi(m)P(m, m) < 0, \; i = 1, 2. \tag{1.1}$$

Stability is a natural assumption in any designed system, it means that the designed system can serve on average faster than the service demand that it is expected to meet. If X is used to model the finances of a system of companies stability means that the companies are profitable on average. Throughout this thesis we will work with stable processes.

In addition to (1.1), for our analysis we need two further technical assumptions, see (2.20) and (2.21). Stability means that the queueing system represented by (X, M) serves customers on average faster than the customer arrival rate which 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 get arbitrarily high 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 in tandem) then a natural question is the following: what is the probability that a given cycle ends with a system failure, i.e., a buffer overflow? This problem is represented mathematically as follows: define the region

$$A_n = \left\{x \in \mathbb{Z}_+^2 : x(1) + x(2) \leq n\right\} \tag{1.2}$$

and its boundary

$$\partial A_n = \left\{x \in \mathbb{Z}_+^2 : x(1) + x(2) = n\right\}. \tag{1.3}$$

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

$$\tau_n \doteq \inf\{k \geq 0 : X_k \in \partial A_n\}, n = 0, 1, 2, 3, .. \tag{1.4}$$

Then the buffer overflow probability described above is

$$p_n(x, m) \doteq \mathbb{P}_{(x,m)}(\tau_n < \tau_0). \tag{1.5}$$

The main aim of this thesis is to develop formulas for the approximation of the probability p_n; for this we will follow the approach developed in [30, 31]. Stability implies 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 this probability, in the context of non-modulated random walks, has received great attention over the last thirty years or so using a range of techniques including rare event simulation and large deviations analysis, see [1, 10, 17, 32]. [22] is the first paper to suggest an heuristic importance sampling algorithm to estimate p_n by simulation. They proposed a change of measure, interchanging the arrival rate with the smallest service rate, based on previous large deviation results proposed by [33]. However, the performance of importance sampling measure suggested in [22] was found far from perfection for a range of parameters [15] and even in some cases with infinite variance [9]. A detailed literature review is given in Chapter 6.

The problem, even for the simple two dimensional tandem walk is difficult; there are two sources of difficulty: multidimensionality, and the discontinuous dynamics of the problem on the constraining boundaries. In short, the problem has a nontrivial geometry (if, there were no constraining boundaries, for the exit boundary ∂A_n, the computation of p_n can essentially be reduced to a single dimension (distance from ∂A_n) allowing simple approximative formulas). Asymptotically optimal importance sampling algorithms for the problem were constructed in [11], which proposed a dynamic importance sampling algorithm based on subsolutions of a related Hamilton Jacobi Bellman (HJB) and its boundary conditions. The approach of [11] 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 and taking limits; the limit V of V_n satisfies the HJB equation mentioned above. The works [30, 31] obtained sharp estimates of p_n for non-modulated two dimensional tandem walk using an affine transformation of the process X (see Figure 1.3). One of the goals of the present work is to show that the same transformation approach can be extended to the much more challenging modulated constrained random walk framework. Markov modulation complicates almost every aspect of the problem: the underlying functions, the geometry of the characteristic surfaces, the limit analysis, etc. but remarkably, as this thesis shows, it turns out to be possible to extend the affine transformation approach of [30] to Markov modulated/regime switch dynamics. A detailed comparision with non-modulated case is given in Section 7.1.

To the best our knowledge, there is very limited research on the analysis of this overflow probability p_n for Markov modulated constrained random walks; in fact, we are only aware of the article [27] which develops asymptotically optimal importance sampling algorithms for the approximation of p_n for the two tandem constrained random walk. In doing this, a necessary step is also to compute the large deviation decay rate of p_n; this was also done in [27] and a note [25] accompanying it. The analysis in these works is based on, following [11], the sub and supersolutions of a limit HJB equation.

Now, we summarize our analysis and give the main result.

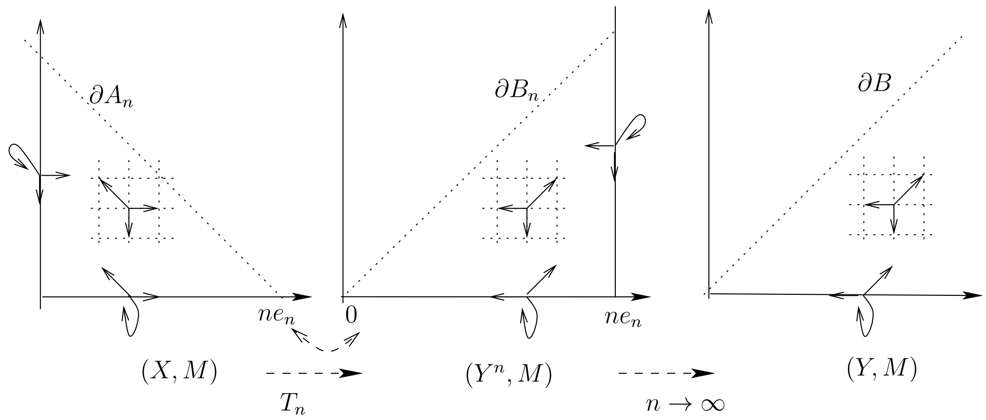

Figure 1.3: The transformation of (X, M)

Denote the constraining boundaries by $\partial_i \doteq \{x \in \mathbb{Z}^2 : x(i) = 0\}$, $i = 1, 2$. Thanks to the Markov property, the following dynamic equations can be written for the process (X, M):

$$\begin{aligned}
&p_n(x, m) = 0 \quad \text{if} \quad X = (0, 0), \\
&p_n(x, m) = 1 \quad \text{if} \quad X \in \partial A_n, \\
&p_n(x, m) = p(x + (1, 0), m)\lambda(m)P(m, m) + p(x + (-1, 1), m)\mu_1(m)P(m, m) \\
&\quad + p(x + (0, -1), m)\mu_2(m)P(m, m) + \sum_{n \in \mathcal{M}, n \neq m} p(x, n)P(m, n) \quad \text{if } X \in A_n^0, \\
&p_n(x, m) = p(x + (1, 0), m)\lambda(m)P(m, m) + p(x, m)\mu_1(m)P(m, m) \\
&\quad + p(x + (0, -1), m)\mu_2(m)P(m, m) + \sum_{n \in \mathcal{M}, n \neq m} p(x, n)P(m, n) \quad \text{if } X \in \partial_1, \\
&p_n(x, m) = p(x + (1, 0), m)\lambda(m)P(m, m) + p(x + (-1, 1), m)\mu_1(m)P(m, m) \\
&\quad + p(x + m)\mu_2(m)P(m, m) + \sum_{n \in \mathcal{M}, n \neq m} p(x, n)P(m, n) \qquad \text{if } X \in \partial_2.
\end{aligned}$$

Unfortunately, it is unfeasible to solve such a system numerically due to the large state space. The number of unknowns increases by n^2. As mentioned above, the work [30, 31] proposed a general approach to the approximation of probabilities of type p_n. Figure 1.3 describes the transformation. The proposed transformation is that move the origin $(0, 0)$ to the corner point $(n, 0)$ on the exit boundary. (Y^n, M) is the same process with (X, M). When n goes to infinity, (Y^n, M) turns into the limit process (Y, M) constrained only on ∂_2.

Define

$$\mathcal{I} \doteq \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix},$$

and the affine transformations

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

where (e_1, e_2) is the standard basis for $\mathbb{R}^2$. Furthermore, define the constraining maps

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

Define Y to be the M-modulated constrained random walk on $\mathbb{Z}_+ \times \mathbb{Z}$ with increments

$$J_k \doteq \mathcal{I} I_k : \tag{1.6}$$

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

Y has the same increments as X, but the probabilities of the increments e_1 and $-e_1$ reversed. Define the region

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

and the exit boundary

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

Let τ is the hitting time

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

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

Our main approximation result is the following:

Theorem (Theorem 3.15). 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

$$\frac{|\mathbb{P}_{(x_n,m)}(\tau_n < \tau_0) - \mathbb{P}_{(T_n(x_n),m)}(\tau < \infty)|}{\mathbb{P}_{(x_n,m)}(\tau_n < \tau_0)} < \rho^{cn} \tag{1.7}$$

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

Theorem 3.15 states that, as n increases, $\mathbb{P}_{(T_n(x_n),m)}(\tau < \infty)$ gives a very good approximation of $\mathbb{P}_{(x_n,m)}(\tau_n < \tau_0)$. The main ideas in the proof of Theorem 3.15 are the same as those used in [30, 31] 1) the difference between the events $\{\tau_n < \tau_0\}$ and $\{\tau < \infty\}$ can be characterized by the occurance of very specific events of the form "X first hitting ∂_1 then ∂_2 and then ∂A_n" 2) the probability of these detailed events are very small compared to the probabilites of the events $\{\tau_n < \tau_0\}$ and $\{\tau < \infty\}$. The challenges in the analysis come from in the implementation of these ideas in the modulated framework. To find our bounds we will use the (Y, M)-harmonic and superharmonic functions constructed in Chapter 2 from single and conjugate points on the corresponding characteristic surface. An upper bound for $\mathbb{P}_{(y,m)}(\tau < \infty)$ using these functions is given in Section 3.1. Section 3.2 constructs an upper bound for a specific event 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 Chapter 2 is given in Section 3.3. These steps are combined in Section 3.4 to prove our main approximation theorem, Theorem 3.15.

With Theorem 3.15 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 an exact formula for $\mathbb{P}_y(\tau < \infty)$ can be constructed by linearly combining only two Y-harmonic functions. This is no longer possible in the modulating case and therefore approximate formulas for $\mathbb{P}_{(y,m)}(\tau < \infty)$ must be constructed. Chapter 4 tackles this problem; to do this, we identify further points on the characteristic surface and derive further (Y, M)-harmonic functions from them to construct increasingly accurate formulas for this probability. A numerical example is also provided showing the effectiveness of the resulting approximations. Chapter 5 gives two examples of applications from insurance and finance. Chapter 6 is a further review of the literature related to the research presented in this thesis. Section 7.1 of the Conclusion (Chapter 7) compares the analysis of the current work with the non-modulated tandem walk treated in [30, 31] and the non-modulated parallel walk treated in [34]. Finally, we summarize our analysis and results with an outlook to future work in Section 7.2 of the Conclusion.

CHAPTER 2

HARMONIC FUNCTIONS

2.1 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

$$\mathbb{E}_{(y,m)}[h(Y_1, M_1)] = h(y, m), (y, m) \in \mathbb{Z}^{1,2} \times \mathcal{M}; \tag{2.1}$$

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

For the case $|\mathcal{M}| = 1$ (i.e., no modulation), [30, 31] looked for Y-harmonic functions which were linear combinations of functions of the form $[(\beta, \alpha), \cdot]$ where

$$y \mapsto [(\beta, \alpha), y] = \beta^{y(1)-y(2)}\alpha^{y(2)}, (\beta, \alpha) \in \mathbb{C}. \tag{2.2}$$

The studies [30, 31] chose (β, α) from the roots of a characteristic polynomial associated with the process Y.

With Markov modulation we have an additional state variable m, this leads to the following generalization of (2.2)

$$(y, m) \mapsto \beta^{y(1)-y(2)}\alpha^{y(2)}d(m), \tag{2.3}$$

where $d : \mathcal{M} \mapsto \mathbb{C}$ is an arbitrary function on $\mathcal{M}$. We will denote the function (2.3) by

$$[(\beta, \alpha, d), \cdot].$$

In the modulated case, there is a local characteristic polynomial for each modulating state $m \in \mathcal{M}$:

$$\boldsymbol{p}(\beta, \alpha, m) \doteq \lambda(m)\frac{1}{\beta} + \mu_1(m)\alpha + \mu_2(m)\frac{\beta}{\alpha}; \tag{2.4}$$

$\boldsymbol{p}$ is rational in β and α, to get a polynomial representation, one must multiply it by $(\beta\alpha)^{|\mathcal{M}|}$. To define the global characteristic polynomial introduce the $|\mathcal{M}| \times |\mathcal{M}|$ matrix A:

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

$(m_1, m_2) \in \mathcal{M}^2$. Let I denote the $|\mathcal{M}| \times |\mathcal{M}|$ identity matrix. Attempting to find functions of the form $[(\beta,\alpha,d),\cdot]$ that satisfy (2.1) leads to the following characteristic equation:

$$A(\beta,\alpha)d = d,$$

i.e,

$$\boldsymbol{p}(\beta,\alpha) \doteq \det(I - A(\beta,\alpha)) = 0, \tag{2.5}$$

and d is an eigenvector of $A(\beta,\alpha)$ for the eigenvalue 1. The $\boldsymbol{p}$ of (2.5) is the global characteristic polynomial for the modulated process (Y,M). Define the characteristic surface for the interior:

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

Points on $\mathcal{H}$ give us (Y,M)-harmonic functions on $\mathbb{Z}\times\mathbb{Z}_+ - \partial_2$.

Proposition 2.1. *If* $(\beta,\alpha,d) \in \mathcal{H}$ *then* $[(\beta,\alpha,d),\cdot]$ *satisfies* (2.1) *for* $y \in \mathbb{Z}\times\mathbb{Z}_+ - \partial_2$.

Proof. We would like to show $[(\beta,\alpha,d),\cdot]$ satisfies (2.1) when $y \in \mathbb{Z}\times\mathbb{Z}_+ - \partial_2$ and $(\beta,\alpha,d) \in \mathcal{H}$. By definition

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

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

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

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

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

the expression in paranthesis equals the m^{th} term of the vector $A(\beta,\alpha)d$, which equals $d(m)$ because $(\beta,\alpha,d)\in\mathcal{H}$ means $A(\beta,\alpha)d=d$. Therefore,

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

This proves the claim of the proposition. □

The previous proposition gives us (Y,M)-harmonic functions on $\mathbb{Z}\times\mathbb{Z}_+-\partial_2$; we will next study the geometry of $\mathcal{H}$, this will be useful in defining fully (Y,M)-harmonic functions.

2.1.1 Geometry of the characteristic surface

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

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

Lemma 2.2. *For each $(\beta,\alpha)\in\mathcal{H}^{\beta\alpha}$ there is at least one paramater family of points $\{(\beta,\alpha,cd),\ c\in\mathbb{C}-\{0\}\}\subset\mathcal{H}$, for some $d\in\mathbb{C}^{|\mathcal{M}|}-\{0\}$. Conversely, for each $(\beta,\alpha,d)\in\mathcal{H}$, we have $(\beta,\alpha)\in\mathcal{H}^{\beta\alpha}$. Furthermore, all points on $\mathcal{H}$ can be obtained from those on $\mathcal{H}^{\beta\alpha}$.*

This follows from basic linear algebra, we provide a proof for completeness.

Proof. $(\beta,\alpha)\in\mathcal{H}^{\beta\alpha}$ implies, by linear algebra, that there exists $d\in\mathbb{C}^{|\mathcal{M}|}$, $d\neq 0$ such that $A(\beta,\alpha)d=d$, i.e., $(\beta,\alpha,d)\in\mathcal{H}$. Therefore, for each point on $\mathcal{H}^{\beta\alpha}$ there is at least one point (β,α,d) on $\mathcal{H}$. Furthermore, all points on $\mathcal{H}$ can be obtained from those on $\mathcal{H}^{\beta\alpha}$ by fixing $(\beta,\alpha)\in\mathcal{H}^{\beta\alpha}$ and solving $A(\beta,\alpha)d=d$ for d. □

$\beta^{|\mathcal{M}|}\alpha^{|\mathcal{M}|}\boldsymbol{p}$ 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 $A(\beta, \alpha)$. In this thesis we will focus on the values of (β, α) for which these eigenvalues are simple:

Proposition 2.3. *Let $D \subset \mathbb{C}^2$ or $D \subset \mathbb{R}^2$ be open and simply connected and suppose $A(\beta, \alpha)$ has simple eigenvalues for all $(\beta, \alpha) \in D$. Then the eigenvalues of 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. For any point $(\beta, \alpha) \in D$ that the eigenvalues Λ_j can be defined smoothly in a neighborhood of (β, α) follows from [24, 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. $\square$

For D and Λ_j as in Proposition 2.3 define

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

Of particular interest to us is the open positive orthant $D = \mathbb{R}^{2,o}_{+} \doteq \mathbb{R}^2_+ - \partial_1 \cup \partial_2$. In the rest of this note we will work under the following assumption:

Assumption 2.1. $A(\beta, \alpha)$ has real distinct eigenvalues for $(\beta, \alpha) \in \mathbb{R}^{2,o}_{+}$.

For $D = \mathbb{R}^{2,o}_{+}$ we will omit the superscript D and write $\mathcal{H}^j$ instead of $\mathcal{H}^{j,D}$.

The last proposition implies

Proposition 2.4. *Let D and Λ_j, $j = 1, 2, 3, ..., |\mathcal{M}|$ be as in Proposition 2.3. Then*

$$D \cap \mathcal{H}^{\beta\alpha} = \sqcup_{j=1}^{|\mathcal{M}|} \mathcal{H}^{j,D}, \tag{2.6}$$

where $\sqcup$ denotes disjoint union.

Proof. The proof follows from the definitions involved. We need to prove the equality in both directions :

$(\beta, \alpha) \in \mathcal{H}^{\beta\alpha} \Rightarrow (\beta, \alpha) \in \mathcal{H}^{j,D}$, $(\beta, \alpha) \in \mathcal{H}^{j,D} \Rightarrow (\beta, \alpha) \in \mathcal{H}^{\beta\alpha}$.

Choose $(\beta,\alpha) \in \mathcal{H}^{\beta\alpha}$, by definition $\boldsymbol{p}(\beta,\alpha) = 0$, i.e. d is an eigenvector of $A(\beta,\alpha)$ for eigenvalue 1. This implies $\Lambda_j(\beta,\alpha) = 1$ for some j which means $(\beta,\alpha) \in \mathcal{H}^{j,D}$. Choose $(\beta,\alpha) \in \mathcal{H}^{j,D}$. Then (β,α) is a root of the polynomial $\boldsymbol{p}$, i.e., $\boldsymbol{p}(\beta,\alpha) = 0$ which means $(\beta,\alpha) \in \mathcal{H}^{\beta\alpha}$. $\square$

The following definitions are from [5, page 57]: A matrix is said to be totally nonnegative (totally positive) if all 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 A is oscillatory, Assumption 2.1 holds:

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

This proposition is a basic fact on oscillating matrices [5, (6.28)]. [5, (6.26)] identifies a particularly simple class of oscillating matrices:

Proposition 2.6. *Suppose $G(1,1)$, $G(1,2)$, $G(|\mathcal{M}|-1,|\mathcal{M}|)$, $G(|\mathcal{M}|,|\mathcal{M}|)$ and $G(j,j-1)$, $G(j,j)$, $G(j,j+1)$, $j = 2,3,4,...,|\mathcal{M}|-1$ are all strictly positive and the rest of the components of G are all zero, i.e., G is tridiagonal with strictly positive entries. Then 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 P is strictly tridiagonal, $A(\beta,\alpha)$ will also be of the same form for all $(\beta,\alpha) \in \mathbb{R}_+^2$; therefore, for such P Assumption 2.1 always holds.

For $(\beta,\alpha) \in \mathbb{R}_+^{2,o}$, $A(\beta,\alpha)$ is a matrix with positive entries. Perron-Frobenius Theorem implies that $A(\beta,\alpha)$ has a simple positive eigenvalue dominating all of the other eigenvalues in absolute value. $\Lambda_1(\beta,\alpha)$ will always denote this largest eigenvalue. Furthermore, if $A(\beta,\alpha)$ 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. If $A(\beta,\alpha)$ has simple real eigenvalues for $(\beta,\alpha) \in \mathbb{R}_+^{2,o}$ we can define

$$\mathcal{R}^j \doteq \{(\beta,\alpha) \in \mathbb{R}_+^{2,o} : \Lambda(\beta,\alpha) \leq 1\}.$$

The continuity of Λ_j implies

$$\mathcal{H}^j = \partial\mathcal{R}^j.$$

Proposition 2.7. *Suppose* $A(\beta,\alpha)$ *has simple real eigenvalues for* $(\beta,\alpha) \in \mathbb{R}_+^{2,o}$. *Then the curve* $\mathcal{H}^j$ *is strictly contained inside the curve* $\mathcal{H}^{j+1}$ *for* $j = 1, 2, 3, ..., |\mathcal{M}|-1$.

Proof. All diagonal entries of $A(\beta,\alpha)$ tend to ∞ when $(\beta,\alpha) \to \partial\mathbb{R}_+^{2,o}$. This and Gershgorin's Theorem [19, 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 $\partial\mathcal{R}^j = \mathcal{H}^j$ lies strictly within $\mathcal{R}^{j+1}$ with strictly positive distance from the boundary $\mathcal{H}^{j+1}$ of $\mathcal{R}^{j+1}$; this proves the claim of the proposition. □

The decomposition of $\mathcal{H}^{\beta\alpha} \cap \mathbb{R}_+^{2,o}$ into $\mathcal{H}^j$ is shown in Figure 2.1 for the transition matrix

$$P = \begin{pmatrix} 0.6 & 0.4 & 0 \\ 0.1 & 0.4 & 0.5 \\ 0 & 0.2 & 0.8 \end{pmatrix}. \tag{2.7}$$

The matrix P of (2.7) is strictly tridiagonal; therefore, Proposition 2.6 applies and $A(\beta,\alpha)$ has distinct real eigenvalues for all $(\beta,\alpha) \in \mathbb{R}_+^{2,o}$ and we have the decomposition (2.6) of $\mathcal{H}^{\beta\alpha} \cap \mathbb{R}_+^{2,o}$ given by Propositions 2.4 and 2.7; Figure 2.1 illustrates $\mathcal{H}^{\beta\alpha}$ and its components $\mathcal{H}^j$; the jump probabilities for this example are

$$\begin{pmatrix} 0.1 & 0.4 & 0.5 \\ 0.12 & 0.41 & 0.47 \\ 0.09 & 0.39 & 0.52 \end{pmatrix} \tag{2.8}$$

where the i^{th} row equals $(\lambda(i), \mu_1(i), \mu_2(i))$.

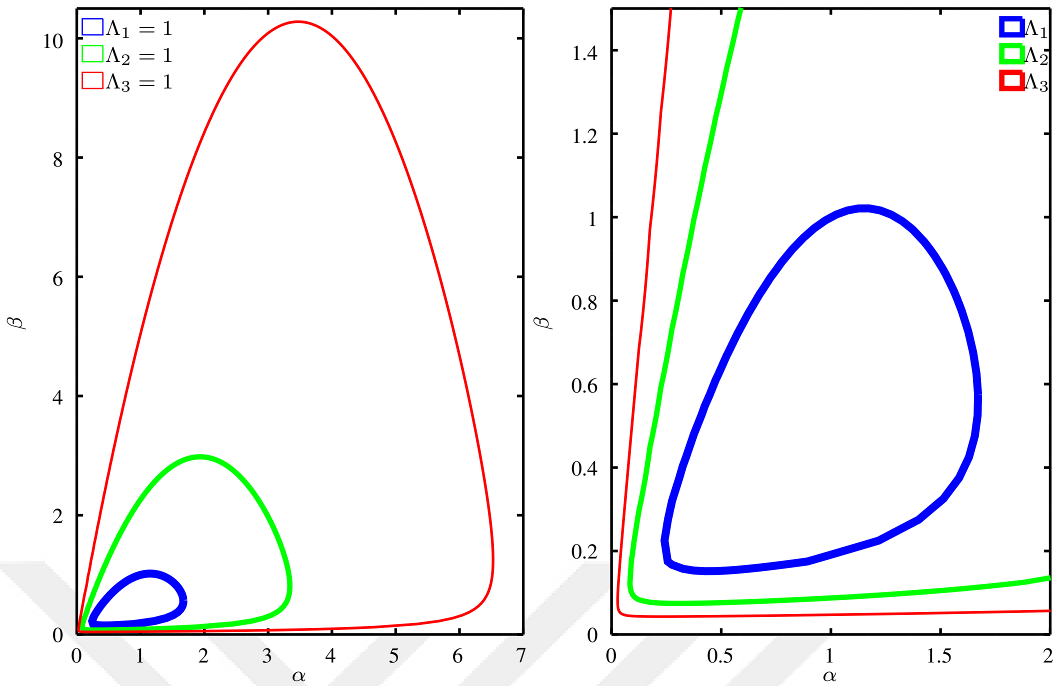

Figure 2.1: Level curves

By Proposition 2.1 and Lemma 2.2 each point on any of the curves depicted in Figure 2.1 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 $\mathcal{H}^1$, 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)-harmonic functions from points on $\mathcal{H}^{\beta\alpha}$.

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

Parallel to [30, 31], we can proceed in two ways to get functions that satisfy (2.1) for $y \in \partial_2$ as well as the interior. The first is by defining the characteristic polynomial

$\boldsymbol{p}_1$, the boundary matrix A_1, and the boundary surface $\mathcal{H}_1$:

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

$$A_1(\beta,\alpha)_{m_1,m_2} \doteq \begin{cases} P(m_1,m_2), & m_1 \neq m_2 \\ P(m_1,m_1)\boldsymbol{p}_1(\beta,\alpha,m), & m_1 = m_2, \end{cases}, (m_1,m_2) \in \mathcal{M}^2,$$

$$\mathcal{H}_1 \doteq \left\{(\beta,\alpha,d) \in \mathbb{C}^{2+|\mathcal{M}|} : A_1(\beta,\alpha)d = d\right\}.$$

Define $\boldsymbol{\Lambda}_1^1(\beta,\alpha)$ to be the largest eigenvalue of $A_1(\beta,\alpha)$. Parallel to the interior case, define

$$\mathcal{H}_1^{\beta\alpha} \doteq \{(\beta,\alpha) \in \mathbb{C}^2 : \boldsymbol{p}_1(\beta,\alpha) = 0\},$$
$$\mathcal{H}_1^1 \doteq \{(\beta,\alpha) \in \mathbb{C}^2 : \boldsymbol{\Lambda}_1^1(\beta,\alpha) = 1\}.$$

Proposition 2.8. $[(\beta,\alpha,d),\cdot]$ *is* (Y,M)*-harmonic if* $(\beta,\alpha,d) \in \mathcal{H} \cap \mathcal{H}_1$.

Proof. Proposition 2.1 says that for $(\beta,\alpha,d) \in \mathcal{H}$, $[(\beta,\alpha,d),\cdot]$ 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 $[(\beta,\alpha,d),\cdot]$ is (Y,M)-harmonic on ∂_2 when $(\beta,\alpha,d) \in \mathcal{H}_1$. By definition

$$\begin{aligned}
&\mathbb{E}_{(y,m)}\left[(\beta,\alpha,d),(Y_1,M_1)\right] \\
&\quad = \sum_{n\in\mathcal{M},n\neq m} P(m,n)[(\beta,\alpha,d),(y,n)] \\
&\quad\quad + P(m,m)(\lambda(m)[(\beta,\alpha,d),(y+(-1,0),m)] + \mu_1(m)[(\beta,\alpha,d),(y+(1,1),m] \\
&\quad\quad + \mu_2(m)[(\beta,\alpha,d),(y,m)]) \\
&\quad = \sum_{n\in\mathcal{M},n\neq m} P(m,n)[(\beta,\alpha,d),(y,n)] \\
&\quad\quad + P(m,m)(\lambda(m)\beta^{y(1)-1}d(m) + \mu_1(m)\beta^{y(1)}\alpha d(m) + \mu_2(m)\beta^{y(1)}d(m)) \\
&\quad = \sum_{n\in\mathcal{M},n\neq m} P(m,n)[(\beta,\alpha,d),(y,n)] + P(m,m)[(\beta,\alpha,d),(y,m)]\boldsymbol{p}_1(\beta,\alpha,m) \\
&\quad = \beta^{y(1)}\left(\sum_{n\in\mathcal{M},n\neq m} P(m,n)d(n) + P(m,m)d(m)\boldsymbol{p}_1(\beta,\alpha,m)\right).
\end{aligned}$$

the expression in paranthesis equals the m^{th} term of the vector $A_1(\beta,\alpha)d$, which equals $d(m)$ because $(\beta,\alpha,d) \in \mathcal{H}_1$ means $A_1(\beta,\alpha)d = d$. Therefore,

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

This argument and Proposition 2.1 proves the claim of the proposition. □

The real sections of $\mathcal{H}^{\beta\alpha}$ and $\mathcal{H}_1^{\beta\alpha}$ 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 turns 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 $|\mathcal{M}|$ nontrivial such points on $\mathcal{H}^{\beta\alpha} \cap \mathcal{H}_1^{\beta\alpha}$; the largest of them is of special significance, and can be identified using the implicit function theorem and the stability assumption (1.1). This point and the (Y, M)-harmonic function it defines are given in Proposition 2.11 and 2.12 below. For the argument we need two auxilary linear algebra results, these are given as Lemmas 2.9 and 2.10.

For a square matrix M, let $M^{i,j}$ denote the same matrix with its i^{th} row and j^{th} column removed.

Lemma 2.9. *Suppose B is an irreducible $|\mathcal{M}| \times |\mathcal{M}|$ matrix. Then* $\det\left((\Lambda_1(B)I - B)^{i,i}\right) > 0$ *for all $i \in \{1, 2, 3, ..., |\mathcal{M}|\}$, where I is the $|\mathcal{M}| \times |\mathcal{M}|$ identity matrix.*

Proof. The argument is the same for all $i \in \{1, 2, 3, ..., |\mathcal{M}|\}$; so it suffices to argue for $i = 1$. Suppose the claim is not true and

$$\det\left((\Lambda_1(B)I - B)^{1,1}\right) \leq 0. \tag{2.9}$$

Consider the function $r \mapsto g(r) = \det\left((rI - B)^{1,1}\right)$, $r \geq 0$. The multilinearity and continuity of $\det$ implies $\lim_{r\nearrow\infty} g(r) = \infty$. This implies that if (2.9) is true there must be $r_0 \geq \Lambda_1(B)$ such that $\det\left((r_0 I - B)^{1,1}\right) = 0$, i.e., $\Lambda_1\left(B^{1,1}\right) = r_0 \geq \Lambda_1(B)$. Let $b^1 \in \mathbb{R}_+^{d-1}$ denote the eigenvector $B^{1,1}$ corresponding to r_0.

For two vectors $x, y \in \mathbb{R}^d$, let $x \geq y$ and $x > y$ denote componentwise comparison. Define $b = [0; b^1] \in \mathbb{R}^d$; it follows from $B^{1,1}b^1 = r_0 b$ and the positivity of the components of B and b^1 that one can choose $\delta > 0$ small enough so that

$$Bb > (r_0 + \delta)b. \tag{2.10}$$

We know that

$$\Lambda_1(B) = \sup\{c : \exists x \in \mathbb{R}_+^d, Bx \geq cx\}, \tag{2.11}$$

(see [19, Proof of Theorem 1, Chapter 16]). This and (2.10) imply $\Lambda_1(B) \geq r_0 + \delta$, which contradicts $r_0 \geq \Lambda_1(B)$. $\square$

The following fact was used in the proof of [27, Lemma 4.4], its proof is elementary and follows from the multilinearity of the determinant function and the previous lemma.

Lemma 2.10. *Let B be an irreducible transition matrix. Then the row vector whose i^{th} component equals $\det\left((I-B)^{i,i}\right)$ is the unique (upto scaling by a positive number) left eigenvector associated with the eigenvalue 1 of B.*

The next proposition identifies the first point on $\mathcal{H}^{\beta\alpha}$ that will play an important role in our analysis. The proof, based on the implicit function theorem, is parallel to that of [27, Lemma 4.4].

Proposition 2.11. *Under assumption (1.1) there exists $0 < \rho_1 < 1$ such that $(\rho_1, \rho_1) \in \mathcal{H}^{\beta\alpha} \cap \mathcal{H}_1^{\beta\alpha}$, i.e., 1 is an eigenvalue of $A_1(\rho_1,\rho_1)$ and $A(\rho_1,\rho_1)$. Furthermore, 1 is the largest eigenvalue of $A(\rho_1,\rho_1)$ and $A_1(\rho_1,\rho_1)$, i.e., $(\rho_1,\rho_1) \in \mathcal{H}^1 \cap \mathcal{H}_1^1$.*

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

$$H(q) \doteq -\log \mathbf{\Lambda}\left(e^{q(1)}, e^{q(2)}\right).$$

By [27, Lemma 4.2, 4.3], H is convex in q. Proceeding parallel to [27, Proof of Lemma 4.4, page 515] define $f(\Lambda, r) \doteq \det(\Lambda I - A(e^r, e^r))$. 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 decreasing at $r=0$. Differentiating f at $(1,0)$ with respect to r gives

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

which equals, by Lemma 2.10, for some constant $c_0 > 0$,

$$\begin{aligned} &= c_0 \sum_{m\in\mathcal{M}} (\lambda(m) - \mu_1(m)) P(m,m) \pi(m) \\ &< 0 \end{aligned}$$

where the last inequality follows from the stability assumption (1.1). 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 $\lambda(s)/\beta$ term appearing in the diagonal terms of A, tending to $+\infty$ with $\beta = e^r$). Then we have $\Lambda_1(e^r, e^r)$ is monotone at $r = 0$ (decreases when r decreases) and tends to infinity as $r \to -\infty$. By the continuity of Λ_1, there must exist at least one point in $(0, \infty)$ 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 $A(\rho_1, \rho_1)$ corresponding to the eigenvalue 1; because 1 is the largest eigenvalue of $A(\rho_1, \rho_1)$ and because $A(\rho_1, \rho_1)$ is irreducible, we can choose $\boldsymbol{d}_1$ so that all of its components are strictly positive; $(\rho_1, \rho_1) \in \mathcal{H}^{\beta\alpha} \cap \mathcal{H}_1^{\beta\alpha}$ implies $(\rho_1, \rho_1, \boldsymbol{d}_1) \in \mathcal{H}^1 \cap \mathcal{H}_1^1$. This and Proposition 2.11 give us our first (Y, M)-harmonic function:

Proposition 2.12.

$$h_{\rho_1} \doteq [(\rho_1, \rho_1, \boldsymbol{d}_1), \cdot] \tag{2.12}$$

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 $\beta \in \mathbb{C}$. We will refer to the points $(\boldsymbol{\alpha}_i(\beta))$, $i = 1, 2, 3, ..., 2|\mathcal{M}|$, as conjugate points. We know by Lemma 2.2 that corresponding to these conjugate points, we have $2|\mathcal{M}|$ conjugate points $(\boldsymbol{\alpha}_i(\beta), d_i)$, $i = 1, 2, 3, ..., 2|\mathcal{M}|$, on $\mathcal{H}$.

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

$$C(\beta, \alpha, d) \in \mathbb{C}^{\mathcal{M}}, \; C(\beta, \alpha, d)(m) = P(m,m)\mu_2(m)d(m)\left(1 - \frac{\beta}{\alpha}\right). \tag{2.13}$$

One can, in general, take linear combinations of conjugate points on $\mathcal{H}$ to define (Y, M)-harmonic functions. This is based on the following lemma

Lemma 2.13. *Suppose* $(\beta, \alpha, d) \in \mathcal{H}$. *Then*

$$\mathbb{E}_{(y,m)}[h(Y_1, M_1)] - h(y, m) = \beta^{y(1)} C(\beta, \alpha, d)(m), \tag{2.14}$$

where C is defined as in (2.13) *for* $y \in \partial_2$.

Proof. The computation in the proof of Proposition 2.1 gives

$$\begin{aligned} &\mathbb{E}_{(y,m)}\left[(\beta, \alpha, d), (Y_1, M_1)\right] \\ &= \beta^{y(1)} \left(\sum_{n \in \mathcal{M}, n \neq m} P(m,n)d(n) + P(m,m)d(m)\boldsymbol{p}_1(\beta, \alpha, m) \right) \end{aligned} \tag{2.15}$$

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

$$\begin{aligned} &[(\beta, \alpha, d), (y, m)] \\ &= \beta^{y(1)} d(m) = \beta^{y(1)} \left(\sum_{n \in \mathcal{M}, n \neq m} P(m,n)d(n) + P(m,m)d(m)\boldsymbol{p}(\beta, \alpha, m) \right). \end{aligned} \tag{2.16}$$

Subtracting the last display from (2.15) gives

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

which proves (2.14). □

We can use the same argument to write the above lemma in the following equivalent form:

Lemma 2.14. *Suppose* $(\beta, \alpha, d) \in \mathcal{H}$. *Then*

$$(A_1(\beta, \alpha)d)(m) = d(m) + C(\beta, \alpha, d)(m), \; m \in \mathcal{M}.$$

The following proposition identifies a family of (Y, M)-harmonic functions constructed from conjugate points on $\mathcal{H}$.

Proposition 2.15. *For $\beta \in \mathbb{C}$ let (β, α_i, d_i) $i = 1, 2, 3, ..., l \leq 2|\mathcal{M}|$ be distinct conjugate points on $\mathcal{H}$. Take any subcollection $\{i_1, i_2, ..., i_k\}$, $k \leq l$ such that $C(\beta, \alpha_{i_j}, d_{i_j})$ are linearly dependent, i.e., there exists $b \in \mathbb{C}^k$ such that*

$$\sum_{j=1}^{k} b(j) C(\beta, \alpha_{i_j}, d_{i_j}) = 0. \tag{2.17}$$

Then

$$h(y, m) = \sum_{j=1}^{|\mathcal{M}|+1} b(j)[(\beta, \alpha_{i_j}, d_{i_j}), \cdot] \tag{2.18}$$

is (Y, M)-harmonic.

For any $\beta \in \mathbb{C}$ such that $\boldsymbol{p}(\beta, \alpha) = 0$ has distinct roots, α_1, α_2,...,$\alpha_{2|\mathcal{M}|}$, that are all different from β, by definition, we have $C(\beta, \alpha_j, d_j) \neq 0$ 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 (2.17).

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 2.13

$$\mathbb{E}_{(y,\cdot)}\left[[(\beta, \alpha_i, d_i), Y_1]\right] - [(\beta, \alpha_i, d_i), (y, \cdot)] = \beta^{y(1)} C(\beta, \alpha_i, d_i). \tag{2.19}$$

Taking linear combinations of these with weight vector b gives:

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

which equals $0 \in \mathbb{R}^{|\mathcal{M}|}$ by (2.17). This proves that h is (Y, M)-harmonic on ∂_2. □

Our primary aim in constructing the (Y, M)-harmonic functions of the present section is to compute/bound the probability $\mathbb{P}_{(y,m)}(\tau < \infty)$, for $(y, m) \in B \times \mathcal{M}$. The function $(y, m) \mapsto \mathbb{P}_{(y,m)}(\tau < \infty)$ is ∂B-determined (Y, M)-harmonic function taking the value 1 on ∂B. For $\alpha = 1$, $[(\beta, 1, d), \cdot]$ takes the value $d(m)$ on ∂B (i.e.

independent of the position on ∂B), which makes $\alpha = 1$ of special significance. The next proposition identifies a point on $\mathcal{H}^1$ of the form $(\rho_2, 1)$ with $0 < \rho_2 < 1$; we will use this point and its conjugates to construct (Y, M)-harmonic functions through Proposition 2.15; we will also construct (Y, M)-superharmonic functions from them to use in our error analysis, which is a new feature of the present setup. Assumption (1.1) gives a point $(\rho_2, 1)$, $0 < \rho_2 < 1$ on $\mathcal{H}^1$:

Proposition 2.16. *Under assumption (1.1) there exists* $0 < \rho_2 < 1$ *such that* $(\rho_2, 1) \in \mathcal{H}^1 \subset \mathcal{H}^{\beta\alpha}$*; i.e.,* 1 *is the largest eigenvalue of* $A(\rho_2, 1)$.

Proof. The proof is parallel to that of Proposition 2.11. We now define $f(\lambda, r) = \det(\lambda I - A(e^r, 1))$ and observe, by assumption (1.1) and Lemma 2.10,

$$\begin{aligned} \left.\frac{\partial f}{\partial r}\right|_{(1,0)} &= \sum_{m \in \mathcal{M}} (\lambda(m) - \mu_2(m)) P(m,m) \det(I - P)^{m,m} \\ &= c_0 \sum_{m \in \mathcal{M}} (\lambda(m) - \mu_2(m)) P(m,m) \pi(m) < 0 \end{aligned}$$

for some constant $c_0 > 0$. The rest of the proof proceeds as in the proof of Proposition 2.11. ☐

Recall that $(\rho_2, 1) \in \mathcal{H}^1$, i.e., 1 is the largest eigenvalue of $A(\rho_2, 1)$; the irreducibility of A implies that the eigenvectors corresponding to 1 has all negative or positive components; let $\boldsymbol{d}_2$ denote a right eigenvector of $A(\rho_2, 1)$ corresponding to the eigenvalue 1 with all 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 ([30, 31, 34]), use a conjugate point of $(\rho_2, 1)$ to construct (Y)-harmonic functions. In the present case, in general, $(\rho_2, 1)$ will have $2|\mathcal{M}| - 1$ conjugate points. Figure 2.1 suggests that only one of these conjugate points lies on $\mathcal{H}^1$; we will use $(\rho_2, 1)$ along with this conjugate to define (Y, M)-superharmonic functions. This will be in two steps. Proposition 2.17 identifies the relevant conjugate point; Proposition 2.21 constructs the superharmonic function. We will use the superharmonic function in Chapter 3, section 3.1 and 3.2 below in our analysis of the relative error (1.7).

The identification of the conjugate point requires the following assumption:

$$\sum_{m \in \mathcal{M}} (\rho_2 \mu_2(m) - \mu_1(m)) P(m,m) \det(I - A(\rho_2, 1))^{m,m} < 0. \tag{2.20}$$

Remark 2.1 comments on this assumption and Proposition 2.18 gives simple conditions under which (2.20) holds.

Proposition 2.17. *Let* $(\rho_2, 1)$*,* $\rho_2 \in (0,1)$ *be the point on* $\mathcal{H}^1$ *identified in Proposition 2.16. Then there exists a unique point* (ρ_2, α_1^*)*,* $\alpha_1^* \in (0,1)$ *if* (2.20) *holds.*

Proof. Set $r_2 = \log(\rho_2)$. Proof is parallel to those of Propositions 2.11 and 2.16 and is based on the analysis of the function H at the point $(r_2, 0)$ via the implicit function theorem. Define $f(\lambda, r) = \det(\lambda I - A(\rho_2, e^r))$ and observe

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

which, by assumption (2.20), is strictly less than 0. The rest of the proof goes as that of Proposition 2.11. □

Remark 2.1. Assumption (2.20) ensures that $(\rho_2, 1)$ has a conjugate point on the principal characteristic surface $\mathcal{H}^1$ 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 [34]) the corresponding assumption is $r^2 < \rho_2$ (see [34, Display (14)]).

The condition $\alpha_1^* < 1$ is needed for the superharmonic function constructed in Proposition 2.21 to be bounded on ∂B, see Proposition 3.2.

Proposition 2.18. *Each of the following conditions are sufficient for* (2.20) *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 $A(\rho_2', 1) = P$. This implies the root ρ_2 identified in Proposition 2.16 must equal ρ_2'. Setting $\rho_2 = \rho_2'$ on the left side of (2.20) gives

$$\begin{aligned}
&\sum_{m\in\mathcal{M}}(\rho_2\mu_2(m)-\mu_1(m))P(m,m)\det(I-A(\rho_2,1))^{m,m}\\
&\quad=\sum_{m\in\mathcal{M}}(\rho_2'\mu_2(m)-\mu_1(m))P(m,m)\det(I-A(\rho_2,1))^{m,m}\\
&\quad=\sum_{m\in\mathcal{M}}(\lambda(m)-\mu_1(m))P(m,m)\det(I-A(\rho_2,1))^{m,m}
\end{aligned}$$

$\det(I-A(\rho_2,1))^{m,m}>0$ by Lemma 2.9, and $\lambda(m)<\mu_1(m)$ by assumption; these and the last line imply (2.20):

$$<0.$$

That the condition $\mu_2(m)<\mu_1(m)$ for all $m\in\mathcal{M}$ implies (2.20) 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 (2.20).

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

$$\rho_1\neq\rho_2, \tag{2.21}$$

where ρ_1 is the first (or the second) component of the point on $\mathcal{H}^1\cap\mathcal{H}_1^1$ identified in Proposition 2.11 and ρ_2 is the β component of the point on $\mathcal{H}^1$ identified in Proposition 2.16. Assumption (2.21) generalizes the assumption $\mu_1\neq\mu_2$ used in [30, 31, 34].The following lemma identifies sufficient conditions for (2.21) to hold.

Lemma 2.19. *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* (2.21) *holds.*

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

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

for some $\epsilon > 0$; here we have used 1) $\boldsymbol{d}_2$ is an eigenvector of $A(\rho_2, 1)$ corresponding to the eigenvalue 1 and 2) AD has strictly positive entries. Then by (2.11), the largest eigenvalue of $A(\rho_2, \rho_2)$ 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 $A(\rho_2, 1)\boldsymbol{d}_2^*$. $\square$

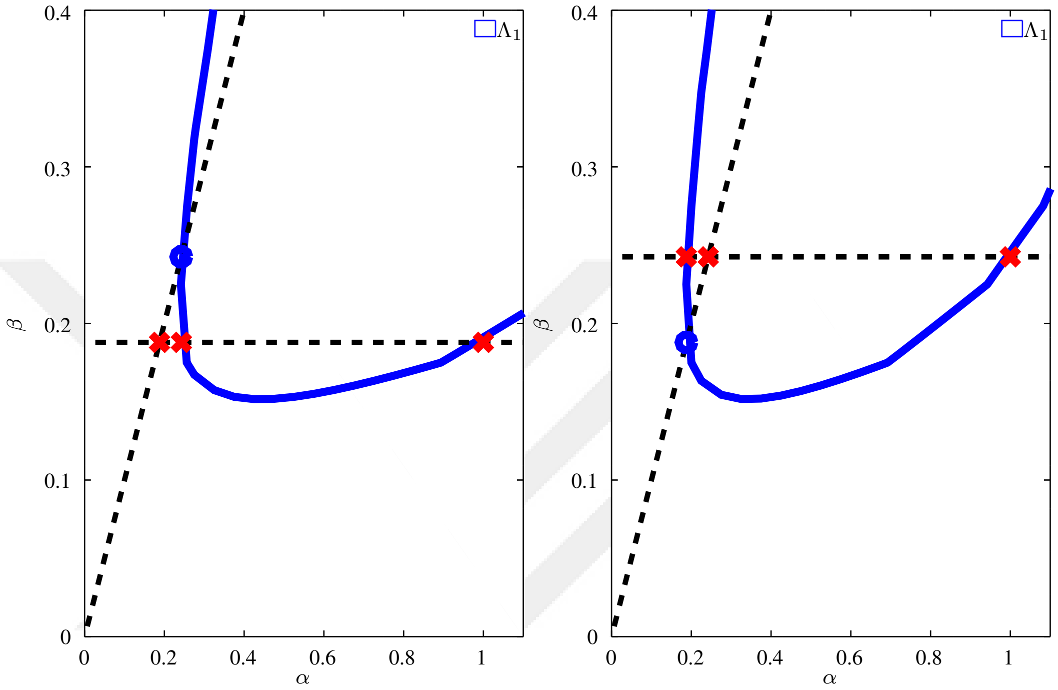

Figure 2.2: $\rho_1 - \rho_2$ and $\alpha_1^* - \rho_2$ have the same sign (Lemma 2.20); the points marked with 'x' are (ρ_2, ρ_2), (ρ_2, α_1^*) and $(1, \rho_2)$; the point marked with 'o' is (ρ_1, ρ_1)

Lemma 2.20. *Let (ρ_2, α_1^*) be the conjugate point of $(\rho_2, 1)$ on $\mathcal{H}^1$ identified in Proposition 2.17. Then $\rho_1 > \rho_2$ implies $\alpha_1^* > \rho_2$ and $\rho_1 < \rho_2$ implies $\alpha_1^* < \rho_2$.*

Figure 2.2 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 $\Lambda_1(\rho_2, \rho) \leq 1$ for $\rho \in (\alpha_1^*, \rho_2]$ It follows that $\rho_2 < \alpha_1^*$.

The argument for the opposite implication is similar. $\square$

Remark 2.3. By the previous lemma the assumption (2.21) is equivalent to

$$\alpha_1^* \neq \rho_2. \tag{2.22}$$

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 $A(\beta,\alpha)$ 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, another way of stating the assumption (2.21) is as follows: "the largest roots of $\boldsymbol{p}(\beta,\beta)$ and $\boldsymbol{p}(\beta,1)$ on $(0,1)$ differ." We will generalize this assumption in Chapter 4 in our computation of $\mathbb{P}_{(y,m)}(\tau < \infty)$.

By definition, 1 is the largest eigenvalue of $A(\rho_2,\alpha_1^*)$; let $\boldsymbol{d}_2^*$ denote a right eigenvector of this matrix with strictly positive entries. Here is one of the key steps of our argument: the construction of a (Y,M)-superharmonic function that will allow us to find upper bounds on approximation errors:

Proposition 2.21. *Under assumption* (2.21) *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*

$$h_{\rho_2} \doteq [(\rho_2, 1, \boldsymbol{d}_2), \cdot] + c_0[(\rho_2, \alpha_1^*, \boldsymbol{d}_2^*), \cdot], \tag{2.23}$$

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

Proof. By their construction, the conjugate points $(\rho_2, 1)$ and (ρ_2, α_1^*) lie on $\mathcal{H}^1$. This and Proposition 2.1 imply that the functions $[(\rho_2, 1, \boldsymbol{d}_2), \cdot]$ and $[(\rho_2, \alpha_1^*, \boldsymbol{d}_2^*), \cdot]$ 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

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

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

By Lemma 2.13,

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

where $C(\cdot,\cdot,\cdot)$ is defined as in (2.13). The last two lines give

$$\mathbb{E}_{(y,m)}[h_{\rho_2}(Y_1,M_1)] - h_{\rho_2}(y,m) = \rho_2^k \left(C(\rho_2,1,\boldsymbol{d}_2)(m) + c_0 C(\rho_2,\alpha_1^*,\boldsymbol{d}_2^*)(m)\right). \tag{2.24}$$

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 the difference

$$C(\rho_2,1,\boldsymbol{d}_2)(m) + c_0 C(\rho_2,\alpha_1^*,\boldsymbol{d}_2^*)(m). \tag{2.25}$$

The definition of 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

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

The sign of the second term in (2.25) depends on whether $\alpha_1^* < \rho_2$ or $\alpha_1^* > \rho_2$. For $\alpha_1^* < \rho_2$, the definition (2.13) of C and $\boldsymbol{d}_2^*(m) > 0$ for all $m \in \mathcal{M}$ imply that the C term in (2.25) is strictly negative for all m. Define

$$d_{\max}^* \doteq \max_{m\in\mathcal{M}} C(\rho_2,\alpha_1^*,\boldsymbol{d}_2^*)(m) < 0. \tag{2.26}$$

If we choose $c_0 > 0$ so that

$$d_{\max} + c_0 d_{\max}^* < 0, \tag{2.27}$$

(2.25) will be strictly less than 0 for all m. This and (2.24) imply that h_{ρ_2} is superharmonic for any c_0 satisfying (2.27).

For $\alpha_1^* > \rho_2$ the argument remains the same except that we replace the $\max$ in (2.26) with $\min$ and $c_0 < 0$. □

In the next chapter we will use h_{ρ_2} to find bounds on the approximation error (1.7).

CHAPTER 3

LIMIT ANALYSIS

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

The works [30, 31, 34] carried out limit analysis using harmonic functions constructed from points on characteristic surfaces. In the present case with a modulating Markov chain, harmonic functions are constructed in general from $|\mathcal{M}| + 1$ points on the characteristic surface $\mathcal{H}^{\beta\alpha}$, which makes analyses based on them more complex. For this reason, we will switch to superharmonic functions whenever we can; this emphasis on superharmonic functions is a new feature of the present work. As we saw in Proposition 2.21 above, (Y, M)-superharmonic functions can be constructed from just two conjugate points on $\mathcal{H}^1 \subset \mathcal{H}^{\beta\alpha}$.

We will need an upper bound on $\mathbb{P}_{(y,m)}(\tau < \infty)$ in our analysis of the relative error (1.7); in the non-modulated tandem walk treated in [30, 31], 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 2.15 and 2.21. 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* (2.12) *and* h_{ρ_2} *be as in* (2.23). *One can choose* $c_1 \geq 0$ *so that*

$$c_2 \doteq \min_{y \in \partial B, m \in \mathcal{M}} h_{\rho_2}(y, m) + c_1 h_{\rho_1}(y, m) > 0; \tag{3.1}$$

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

Proof. By its definition,

$$h_{\rho_2}(y,m) = \boldsymbol{d}_2(m) + c_0(\alpha_1^*)^{y(2)}\boldsymbol{d}_2^*(m), \tag{3.2}$$

for $y \in \partial B$. We know by Proposition 2.21 that $c_0 > 0$ for $\alpha_1^* < \rho_2$. This, $\alpha_1^* > 0$, $\boldsymbol{d}_2^*(m) > 0$ imply

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

which implies (3.1) with $c_1 = 0$.

For $\alpha_1^* > \rho_2$, $c_0 < 0$ and (3.2) can take negative values for small $y(2)$. But $0 < \alpha_1^* < 1$ implies that there exists $k_0 > 0$ such that

$$h_{\rho_2}(y,m) \geq d_{\max}/2 > 0, \quad y \in \partial B, \ y(2) \geq k_0. \tag{3.3}$$

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

$$h_{\rho_1}(y,m) = c_1\boldsymbol{d}_1(m)\rho_1^{y(2)} + \boldsymbol{d}_2(m) + c_0(\alpha_1^*)^{y(2)}\boldsymbol{d}_2^*(m) > d_{\max}/2, \ y \in \partial B, \ y(2) \leq k_0, \tag{3.4}$$

since there are only finitely many inequalities to be satisfied in (3.4). c_1 chosen thus, (3.3) and (3.4) imply (3.1). □

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

$$\mathbb{P}_{(y,m)}(\tau < \infty) \leq \frac{1}{c_2}\left(h_{\rho_2}(y,m) + c_1 h_{\rho_1}(y,m)\right). \tag{3.5}$$

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

$$|f(y,m)| \leq c_2 < \infty, \ \ y \in B, \ m \in \mathcal{M}.$$

Furthermore, by Propositions 2.11 and 2.21 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 ([13, Theorem 5.7.6])

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

this, $Y_\tau \in \partial B$ when $\tau < \infty$ and (3.1) imply

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

which gives (3.5). □

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

Define

$$\rho \doteq \rho_1 \vee \rho_2. \tag{3.6}$$

The goal of the section is to prove

Proposition 3.3. *For any $\epsilon > 0$ there exists $n_0 > 0$ such that*

$$\mathbb{P}_{(x,m)}(\sigma_1 < \sigma_{1,2} < \tau_n < \tau_0) \leq \rho^{n(1-\epsilon)} \tag{3.7}$$

for $n \geq n_0$ 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.

The following fact will be used a number of times; let us record it here:

Lemma 3.4. *The function*

$$(x,m) \mapsto [(\rho_2, 1, \boldsymbol{d}_2), (T_n(x), m)] = \rho_2^{n-(x(1)+x(2))} \boldsymbol{d}_2(m) \tag{3.8}$$

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 (3.8) is (X,M)-harmonic on $\mathbb{Z}_+^{2,o}$; this and $A_1(\rho_2, 1) = A(\rho_2, 1)$ imply the (X,M)-harmonicity of (3.8) on ∂_1. □

3.2.1 $\rho_1 > \rho_2$

To prove (3.7) 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 2.11, $(x, m) \mapsto h_{\rho_1}(T_n(x), m)$ 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

$$
\begin{aligned}
& h_3 : (x, m) \mapsto h_{\rho_2}(T_n(x), m) + c_1 h_{\rho_1}(T_n(x), m) \qquad (3.9)\\
&= h_{\rho_2}((n - x(1), x(2)), m) + c_1 h_{\rho_1}((n - x(1), x(2)), m), \ x \in A_n, \ m \in \mathcal{M},\\
&= \rho_2^{n-(x(1)+x(2))} \left(\boldsymbol{d}_2(m) + c_0 \alpha_1^{*x(2)} \boldsymbol{d}_2^*(m) \right) + c_1 \rho_1^{n-x(1)} \boldsymbol{d}_1(m),
\end{aligned}
$$

where $c_1 \geq 0$ is chosen as in Proposition 3.1 and c_0 is as in Proposition 2.21. The next two propositions imply that h_3 is (X, M)-superharmonic on $\mathbb{Z}_+^2 - \partial_1$.

Proposition 3.5. *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 $\mathbb{Z}_+ \times \mathbb{Z}_+ - \partial_1$ follows from Proposition 2.21 (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:

$$
h_{\rho_2}(T_n(\cdot), \cdot) = [(\rho_2, 1, \boldsymbol{d}_2), (T_n(\cdot), \cdot)] + c_0[(\rho_2, \alpha_1^*, \boldsymbol{d}_2^*), (T_n(\cdot), \cdot)]. \qquad (3.10)
$$

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 3.4. To treat the second term in (3.10) recall the following: $\rho_2 < \rho_1$ implies $\rho_2 < \alpha_1^*$ (Lemma 2.20); then, by Proposition 2.21, $c_0 < 0$. Therefore, if we can show that $[(\rho_2, \alpha_1^*, \boldsymbol{d}_2^*), (T_n(\cdot), \cdot)]$ 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

$$
h(x, m) = [(\rho_2, \alpha_1^*, \boldsymbol{d}_2^*), (T_n(x), m)] = \rho_2^{n-(x(1)+x(2))} \alpha_1^{*x(2)} \boldsymbol{d}_2^*(m).
$$

A calculation parallel to the proof of Proposition 2.1 shows

$$
\mathbb{E}_{(x,m)} [h(X_1, M_1)] - h(x, m) = \boldsymbol{d}_2^*(m) \mu_1(m)(1 - \alpha_1^*) \rho_2^{n-x(2)} > 0, \qquad (3.11)
$$

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

Proposition 3.6. *$h_{\rho_1}(T_n(\cdot),\cdot)$ is harmonic (and therefore superharmonic) on $\mathbb{Z}_+^2 - \partial_1$. It is subharmonic on ∂_1 where it satisfies*

$$\mathbb{E}_{(x,m)}[h_{\rho_1}(T_n(X_1), M_1)] - h_{\rho_1}(T_n(x), m) = \boldsymbol{d}_1(m)\mu_1(m)(1-\rho_1)\rho_1^n > 0. \quad (3.12)$$

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 (3.7):

$$S'_k \doteq \begin{cases} h_1, & k \leq \sigma_1, \\ h_2(X_k, M_k), & \sigma_1 < k \leq \sigma_{1,2}, \\ h_3(X_k, M_k), & k > \sigma_{1,2}, \end{cases}$$

$$S_k \doteq S'_k - c_5 k \rho_1^n,$$

where

$$c_3 \doteq \frac{\max_{m\in\mathcal{M}} \boldsymbol{d}_2(m) + c_1 \max_{m\in\mathcal{M}} \boldsymbol{d}_1(m)}{\min_{m\in\mathcal{M}} \boldsymbol{d}_1(m)}, \quad (3.13)$$

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

$$h_2 \doteq c_3 h_{\rho_1}(T_n(\cdot),\cdot) = c_3[(\rho_1,\rho_1,\boldsymbol{d}_1),(T_n(\cdot),\cdot)] = c_3\rho_1^{n-x(1)}\boldsymbol{d}_1(\cdot), \quad (3.14)$$

$$c_5 \doteq c_3(1-\rho_1)\max_{m\in\mathcal{M}} \boldsymbol{d}_1(m)\mu_1(m). \quad (3.15)$$

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

Proposition 3.7. *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 is $[(\rho_1, \rho_1, \boldsymbol{d}_1, (T_n(X_k), M_k)]$ on ∂_1; the term; $-c_5 k\rho_1^n$ 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 $X_k \in \mathbb{Z}_+^2 - \partial_1 \cup \partial_2$. For $x \in \partial_2$, all of these functions are (X, M)-superharmonic by Propositions 3.5 and 3.6. This implies

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

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

$$\begin{aligned} \mathbb{E}_{(x,m)}[S'_{k+1}|\mathscr{F}_k] &= \mathbb{E}_{(x,m)}[h_3(X_{k+1}, M_{k+1})|\mathscr{F}_k] \\ &\leq h_3(X_k, M_k) \end{aligned} \tag{3.16}$$

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

$$S'_k = h_2(X_k, M_k) \text{ for } k = \sigma_{1,2}. \tag{3.17}$$

The definitions of c_3, h_2 and h_3 in (3.13), (3.14) and (3.9), $\rho_2 < \rho_1$ and $c_0 < 0$ imply

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

for $x \in \partial_1$. This and (3.17) imply

$$h_3(X_k, M_k) \leq h_2(X_k, M_k) = S'_k$$

for $k = \sigma_{1,2}$. The last display and (3.16) imply

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

i.e., S' is an (X, M)-supermartingale for $k = \sigma_{1,2}$ as well.

It remains to prove

$$\mathbb{E}_{(x,m)}[S_{k+1}|\mathscr{F}_k] \leq S_k, \text{ when } X_k \in \partial_1. \tag{3.18}$$

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 $h_2 = h_1$ on ∂_1 imply

$$\begin{aligned} &\mathbb{E}_{(x,m)}[S_{k+1}|\mathscr{F}_k] - S_k \\ &= \mathbb{E}_{(x,m)}[c_3 h_{\rho_1}(T_n(X_{k+1}), M_{k+1})|\mathscr{F}_k] - c_3 \rho_1^n \boldsymbol{d}_1(M_k) - c_5 \rho_1^n, \end{aligned} \tag{3.19}$$

By (3.12) and $\sigma_1 = k$, we compute the first difference as:

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

By the definition of c_5:

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

which proves (3.18) 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:

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

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

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

Proposition 3.6 ((3.12)) 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 3.15, $c_5 > c_1 \boldsymbol{d}_1(m)\mu_1(m)(1-\rho_1)$ for all $m \in \mathcal{M}$, which implies:

$$\le 0.$$

This proves (3.18) for $k > \sigma_{1,2}$ and completes the proof of this proposition. □

We are now ready to give a proof of Proposition 3.3 for $\rho_1 > \rho_2$:

Proof of Proposition 3.3; case $\rho_1 > \rho_2$. By its definition (3.6), ρ of (3.7) equals ρ_1 for $\rho_1 > \rho_2$. We begin by truncating time: [27, Theorem A.2] implies that there exists $c_6 > 0$ and $N_0 > 0$ such that

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

for $n > N_0$. Then:

$$\begin{aligned}
&\mathbb{P}_{(x,m)}(\sigma_1 < \sigma_{1,2} < \tau_n < \tau_0) \qquad (3.20)\\
&\quad = \mathbb{P}_{(x,m)}(\sigma_1 < \sigma_{1,2} < \tau_n < \tau_0, \tau_n \wedge \tau_0 \le c_6 n)\\
&\quad + \mathbb{P}_{(x,m)}(\sigma_1 < \sigma_{1,2} < \tau_n < \tau_0, \tau_n \wedge \tau_0 > c_6 n)\\
&\quad \le \mathbb{P}_{(x,m)}(\sigma_1 < \sigma_{1,2} < \tau_n < \tau_0, \tau_n \wedge \tau_0 \le c_6 n) + \rho_1^{2n}
\end{aligned}$$

for $n > N_0$. Therefore, to prove (3.7) 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 $\tau = \tau_0 \wedge \tau_n \wedge c_6 n$:

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

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

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

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

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

On the set $\{\sigma_1 < \sigma_{1,2} < \tau_n < \tau_0 \le c_6 n\}$, 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 $h_3(x,m) \ge c_2 > 0$ for $x \in \partial A_n$. These and the last display imply

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

Substitute this in (3.20) to get

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

where

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

setting $n_0 \ge N_0$ so that $\epsilon_n < \epsilon$ for $n \ge n_0$ gives (3.7). □

3.2.2 $\rho_1 < \rho_2$

The previous subsection gave a proof of Proposition 3.3 when $\rho_2 < \rho_1$. The only changes needed in this proof when $\rho_1 < \rho_2$ 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.

For $\rho_2 > \rho_1$, $\Lambda_1(\rho_2, \rho_2) < 1$. Let $\boldsymbol{d}_2^+$ be a right eigenvector of $A(\rho_2, \rho_2)$ with strictly positive entries.

Proposition 3.8. *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

$$\mathbb{E}_{(x,m)}[f(X_1, M_1)] - f(x, m) \leq \boldsymbol{d}_2^+(m)\mu_1(m)(1 - \rho_2)\rho_2^n. \tag{3.21}$$

The proof is parallel to that of Proposition 3.6 and follows from $\Lambda_1(\rho_2, \rho_2) < 1$, $A_1(\rho_2, \rho_2) = A(\rho_2, \rho_2)$ and the definitions involved.

Proposition 3.9. *Let h_3 be as in* (3.9)*; h_3 is (X, M)-superharmonic on $\mathbb{Z}_+^2 - \partial_1$; on ∂_1 it satisfies*

$$\mathbb{E}_{(x,m)}\left[h_3(X_1, M_1)\right] - h_3(x, m) = c_0 \boldsymbol{d}_2^*(m)\mu_1(m)(1 - \alpha_1^*)\rho_2^{n-x(2)} > 0. \tag{3.22}$$

Proof. Lemma 2.20 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.,

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

That h_3 is (X,M)-superharmonic on $\mathbb{Z}_+^2 - \partial_1$ follows from the same property of h_{ρ_2} (see Proposition 2.21). On the other hand, again by Proposition 2.21, $\alpha_1^* < \rho_2$ implies that c_0 in the definition of h_{ρ_2} satisfies $c_0 > 0$. By Lemma 3.4 $(x,m) \mapsto [(\rho_2, 1, \boldsymbol{d}_2), (T_n(x), m)]$ is (X,M)-harmonic on ∂_1; (3.22) follows from these and (3.11). □

$\rho_2 > \rho_1$ implies $\rho_2 > \alpha_1^*$ (Lemma 2.20); this and Proposition 3.1 imply $c_1 = 0$; $\rho_2 > \alpha_1^*$ and Proposition 2.21 imply $c_0 > 0$. That $c_0 > 0$ and $c_1 = 0$ lead to the following modifications in the definition of S':

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

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

where

$$c_3 \doteq \frac{\max_{m\in\mathcal{M}} (\boldsymbol{d}_2(m) + c_0 \boldsymbol{d}_2^*(m))}{\min_{m\in\mathcal{M}} \boldsymbol{d}_2^+(m)}, \tag{3.23}$$

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

$$h_4 \doteq c_3 h_{\rho_2}(T_n(\cdot), \cdot) = c_3[(\rho_2, \rho_2, \boldsymbol{d}_2^+), (T_n(\cdot), \cdot)] = c_3 \rho_2^{n-x(1)} \boldsymbol{d}_2^+(\cdot), \tag{3.24}$$

$$c_5 \doteq c_3(1-\rho_2) \max_{m\in\mathcal{M}} \boldsymbol{d}_2^+(m)\mu_1(m) + c_0(1-\alpha_1^*) \max_{m\in\mathcal{M}} \boldsymbol{d}_2^*(m)\mu_1(m). \tag{3.25}$$

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

Proposition 3.10. *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 3.7) and follow from the following facts: $h_1 \geq h_4$ on ∂_1, $h_4 \geq h_3$ 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 3.8 and 3.9), the $-c_5 k \rho_2^n$ term compensating for the potential lack of (X,M)-superharmonicity of h_3 and h_4 on ∂_1 (guaranteed by (3.21) and (3.22) and the choice of the constant c_5). □

Proof of Proposition 3.3; case $\rho_2 > \rho_1$. With S defined as above, the proof given for the case $\rho_1 > \rho_2$ works without change. □

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

To get an upper bound on the relative error (1.7), 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 [34], by constructing a suitable (X, M)-subharmonic function. As opposed to superharmonic functions, subharmonic functions are much simpler to construct.

Proposition 3.11.

$$(x, m) \mapsto [(\rho_2, 1, \boldsymbol{d}_2), (T_n(x), m)] \vee [(\rho_1, \rho_1, \boldsymbol{d}_1), (T_n(x), m)] \\ = \rho_2^{n-(x(1)+x(2))}\boldsymbol{d}_2(m) \vee \rho_1^{n-x(1)}\boldsymbol{d}_1(m) \tag{3.26}$$

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

Proof. We know by Lemma 2.13 that

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

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 3.4. Then, $(x, m) \mapsto [(\rho_2, 1, \boldsymbol{d}_2), (x, m)]$ is (X, M)-subharmonic on all of $\mathbb{Z}_+^2$.

Similarly, Proposition 3.6 and (3.12) 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 (3.26). □

Proposition 3.12.

$$\begin{aligned}
&\mathbb{P}_{(x,m)}(\tau_n < \tau_0) \\
&\geq \left(\max_{m\in\mathcal{M}}(\boldsymbol{d}_2(m) \vee \boldsymbol{d}_1(m))\right)^{-1} \\
&\quad \times \left(\rho_2^{n-(x(1)+x(2))}\boldsymbol{d}_2(m) \vee \rho_1^{n-x(1)}\boldsymbol{d}_1(m) - \rho_2^n \max_{m\in\mathcal{M}} \boldsymbol{d}_2(m) \vee \rho_1^n \max_{m\in\mathcal{M}} \boldsymbol{d}_1(m)\right).
\end{aligned} \tag{3.27}$$

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. 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

$$\mathbb{E}[g(X_{\tau_n\wedge\tau_0}, M_{\tau_n\wedge\tau_0})] = \mathbb{E}[g(X_{\tau_n}, M_{\tau_n})1_{\{\tau_n<\tau_0\}}] + \mathbb{E}[g(X_{\tau_0}, M_{\tau_0})1_{\{\tau_0\leq\tau_n\}}]. \tag{3.28}$$

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

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

This, (3.28) and the optional sampling theorem applied to s at time $\tau_n \wedge \tau_0$ give

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

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

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

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 (3.27). □

3.4 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 (1.7). As in previous works [30, 31, 34], 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 reflected 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

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

Define

$$\sigma_{1,2} \doteq \inf\{k > 0 : X_k \in \partial_2,\ k \geq \sigma_1\} \tag{3.29}$$

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

Lemma 3.13. *Let $\sigma_{1,2}$ be as in* (3.29)*. Then*

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

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

This lemma is the analog of [30, 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. We will give an illustrative example here.

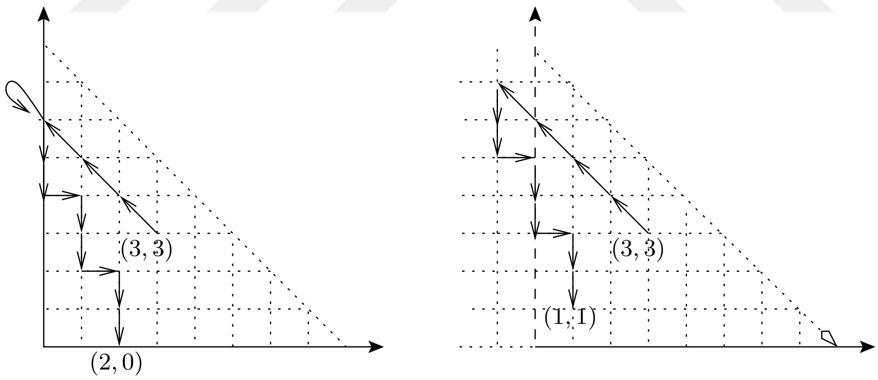

Figure 3.1: A sample path of X_k(left) and $\bar{X}_k$(right)

Example 3.1. For $k \leq \sigma_1$, $X_k = \bar{X}_k$. For $\sigma_1 < k < \sigma_{1,2}$, the process X only differ from $\bar{X}$ if $I_k = (-1, 1)$. As can be seen from Figure 3.1, the difference between increments is equal to

$$\begin{aligned} X_k - \bar{X}_k &= \sum_{j=1}^{k} 1_{\{j \leq k\}} 1_{\{I_k = (-1,1)\}} \cdot (-1, 1) \\ &= (-1, 1) \end{aligned}$$

and the sum of the components at step k is equal to each other as stated in Lemma 3.13.

Define

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

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 (3.7) for the $\bar{X}$ process. This is an adaptation of [30, Proposition 7.5] to the present setup:

Proposition 3.14. *For any $\epsilon > 0$ there exists $n_0 > 0$ such that*

$$\mathbb{P}_{(x,m)}(\sigma_1 < \sigma_{1,2} < \bar{\tau}_n < \infty) \leq \rho^{n(1-\epsilon)} \tag{3.30}$$

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

Proof. As in [30, 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 $\{x \in \mathbb{Z} \times \mathbb{Z}_+ : x(1) + x(2) = 0\}$:

$$\begin{aligned} &\mathbb{P}_{(x,m)}(\sigma_1 < \sigma_{1,2} < \bar{\tau}_n < \infty) \qquad (3.31)\\ &\quad = \mathbb{P}_{(x,m)}(\sigma_1 < \sigma_{1,2} < \bar{\tau}_n < \bar{\sigma}_{1,2} < \infty) + \mathbb{P}_{(x,m)}(\sigma_1 < \sigma_{1,2} < \bar{\sigma}_{1,2} < \bar{\tau}_n < \infty) \end{aligned}$$

Take any $\omega' \in \{\sigma_1 < \sigma_{1,2} < \bar{\tau}_n\}$ such that $\bar{X}$ hits

$$\{x \in \mathbb{Z} \times \mathbb{Z}_+ : x(1) + x(2) = n\}$$

before hitting

$$\{x \in \mathbb{Z} \times \mathbb{Z}_+ : x(1) + x(2) = 0\},$$

Lemma 3.13 implies that

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

i.e., X and $\bar{X}$ will be on the same line $\{x \in \mathbb{Z} \times \mathbb{Z}_+ : x(1) + x(2) = k\}$ for some $k \in \{1, 2, 3, ..., n-1\}$. Then the fully constrained sample path $X(\omega')$ 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 (3.7) on the probability of the last event and (3.31) imply that there exists $n_1 > 0$ such that

$$\mathbb{P}_{(x,m)}(\sigma_1 < \sigma_{1,2} < \bar{\tau}_n < \infty) \leq \rho^{n(1-\epsilon/2)} + \mathbb{P}_{(x,m)}(\sigma_1 < \sigma_{1,2} < \bar{\sigma}_{1,2} < \bar{\tau}_n < \infty) \tag{3.32}$$

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

$$\begin{aligned} &\frac{1}{c_2}(h_{\rho_2}((n - x(1), x(2)), m) + c_1 h_{\rho_1}((n - x(1), x(2)), m)) \\ &\qquad\qquad\qquad \leq \frac{1}{c_2}\left(\rho_2^n \boldsymbol{d}_2(m) + c_1 \rho_1^n \boldsymbol{d}_1(m)\right); \end{aligned}$$

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

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

where c_7 is a constant depending on $\boldsymbol{d}_1$, $\boldsymbol{d}_2$, c_1 and c_2. Substituting this in (3.32) gives

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

for $n > n_1$. This implies the statement of the proposition. □

Finally, we state and prove our main theorem:

Theorem 3.15. *For any $x \in \mathbb{R}_+^2$, $x(1) + x(2) < 1$, and $m \in \mathcal{M}$ (if $\rho_1 > \rho_2$ and $1 - x(2) < \log(\rho_2)/\log(\rho_1)$ we also require $x(1) > 0$) there exists $c_8 > 0$ and $N > 0$ such that*

$$\frac{|\mathbb{P}_{(x_n,m)}(\tau_n < \tau_0) - \mathbb{P}_{(T_n(x_n),m)}(\tau < \infty)|}{\mathbb{P}_{(x_n,m)}(\tau_n < \tau_0)} < \rho^{c_8 n} \tag{3.33}$$

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

Proof. Proposition 3.12, the choice of x (i.e., $x(1) + x(2) < 1$ and $x(1) > 0$ when $\rho 1 > \rho_2$ and $1 - x(2) < \log(\rho_2)/\log(\rho_1)$) imply the lower bound

$$\mathbb{P}_{(x,m)}(\tau_n < \tau_0) \geq \rho^{n(1-2c_8)} \tag{3.34}$$

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

$$\begin{aligned} &\frac{|\mathbb{P}_{(x_n,m)}(\tau_n < \tau_0) - \mathbb{P}_{(T_n(x_n),m)}(\tau < \infty)|}{\mathbb{P}_{(x_n,m)}(\tau_n < \tau_0)} \\ &= \frac{|\mathbb{P}_{(x_n,m)}(\tau_n < \tau_0) - \mathbb{P}_{(x_n,m)}(\bar{\tau}_n < \infty)|}{\mathbb{P}_{(x_n,m)}(\tau_n < \tau_0)} \end{aligned} \tag{3.35}$$

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

$$\begin{aligned} \mathbb{P}_{(x_n,m)}(\tau_n < \tau_0) &= \mathbb{P}_{(x_n,m)}(\tau_n < \sigma_1 < \tau_0) + \mathbb{P}_{(x_n,m)}(\sigma_1 < \tau_n \leq \sigma_{1,2} \wedge \tau_0) \\ &\quad + \mathbb{P}_{(x_n,m)}(\sigma_1 < \sigma_{1,2} < \tau_n < \tau_0) \end{aligned} \tag{3.36}$$

$$\begin{aligned} \mathbb{P}_{(T_n(x_n),m)}(\tau < \infty) &= \mathbb{P}_{(T_n(x_n),m)}(\tau < \sigma_1) + \mathbb{P}_{(T_n(x_n),m)}(\sigma_1 < \tau \leq \sigma_{1,2}) \\ &\quad + \mathbb{P}_{(T_n(x_n),m)}(\sigma_1 < \sigma_{1,2} < \tau < \infty) \end{aligned} \tag{3.37}$$

Lemma 3.13 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 3.13 now gives

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

The last two equalities, Propositions 3.3, 3.14, and partitions (3.36), (3.37) imply that there exists $n_0 > 0$ such that

$$| \mathbb{P}_{(x_n,m)}(\sigma_1 < \sigma_{1,2} < \tau_n < \tau_0) - \mathbb{P}_{(T_n(x_n),m)}(\sigma_1 < \sigma_{1,2} < \tau < \infty) | \leq \rho^{n(1-c_8)} \tag{3.38}$$

for $n > n_0$. Substituting the last bound and (3.34) in (3.35) gives (3.33). □

CHAPTER 4

COMPUTATION OF $\mathbb{P}(\tau < \infty)$

Theorem 3.15 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 chapter we develop approximate formulas for $\mathbb{P}_{(y,m)}(\tau < \infty)$. 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.

$$(y, m) \mapsto \mathbb{P}_{(y,m)}(\tau < \infty) \tag{4.1}$$

is a (Y, M)-harmonic function with the value 1 on ∂B. Furthermore, by definition it is ∂B-determined, (for (4.1), f is the function taking the constant value 1 on ∂B). Our approach to the approximation of $\mathbb{P}_{(y,m)}(\tau < \infty)$ is similar to [30, 31, 34]: take linear combinations of the (Y, M)-harmonic functions identified in Propositions 2.8 and 2.15 approximate the value 1 on ∂B as closely as possible. We need our (Y, M)-harmonic functions to be ∂B-determined; the next identifies a simple condition for functions of the form (4.2) to be ∂B-determined.

Lemma 4.1. *Suppose (β, α_j, d_j) are points on $\mathcal{H}$ and suppose*

$$h(y, m) = \sum_{j=1}^{|\mathcal{M}|+1} b(j)[(\beta, \alpha_j, d_j), \cdot] \tag{4.2}$$

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

This is a version of [30, Proposition 2.2, 4.10] adapted to the Markov modulated setup.

Proof. Define the region $V = \{y \in \mathbb{Z} \times \mathbb{Z}_+ : 0 \le y(1) - y(2) \le n\}$ and the boundaries of V $\partial V_1 = \{y \in \mathbb{Z} \times \mathbb{Z}_+ : y(1) - y(2) = n\}$ and $\partial V_2 = \partial B$. Define $v_n \doteq \inf\{k : Y_k \in \partial V_1\}$. We make the following claim: starting from a point $y \in V$ (Y, M) hits $\partial V_1 \cup \partial V_2$ in finite time, i.e., $v_n \wedge \tau < \infty$ almost surely. Let us first prove this claim.

For each modulating state m, the sample path of the (Y, M) whose increments are only type of $(0, -1)$ hits ∂V_2 in at most n steps and the probability of this event is $(\mu_1(m)P(m, m))^n$. Let us take the minimum of these probabilities for each state as

$$\varepsilon = \min(\mu_1(i)P(i, i))^n, \quad i = 1, 2, ..., |\mathcal{M}|.$$

Then

$$\mathbb{P}_{(y,m)}(\tau \wedge v_n \ge n) \le (1 - \varepsilon).$$

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

$$\mathbb{P}_{(y,m)}(\tau \wedge v_n \ge kn) \le (1 - \varepsilon)^k.$$

Letting $k \to \infty$ gives

$$\mathbb{P}_{(y,m)}(\tau \wedge v_n = \infty) = 0. \tag{4.3}$$

The definition 4.2 and the harmonicity imply that $h(y, m)$ is also bounded on $B - \partial B$. Then $S_k = h(Y_{\tau\wedge v_n\wedge k}, M_{\tau\wedge v_n\wedge k})$ is a bounded martingale. The optional sampling theorem applied to this martingale and (4.3) imply

$$\begin{aligned} h(y, m) &= \mathbb{E}_{(y,m)}[h(Y_{\tau\wedge v_n}, M_{\tau\wedge v_n})] \qquad (4.4) \\ &= \mathbb{E}_{(y,m)}[h(Y_\tau, M_\tau)1_{\{\tau<v_n\}}] + \mathbb{E}_{(y,m)}[h(Y_{v_n}, M_{v_n})1_{\{v_n\le\tau\}}] \end{aligned}$$

For $|\alpha_i| \le 1$ implies $|h(Y_{v_n}, M_{v_n})| \le c_0\beta^n$ for some constant c_0. Then

$$\lim_{n\to\infty} \mathbb{E}_{(y,m)}[h(Y_{v_n}, M_{v_n})1_{\{v_n\le\tau\}}] \le c_0 \lim_{n\to\infty} \beta^n d = 0.$$

The last expression, that $\lim_{n\to\infty} v_n = \infty$ and letting $n \to \infty$ in (4.4) imply

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

i.e, $h(y, m)$ is ∂B-determined. □

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

Lemma 4.2. *h_{ρ_1} is ∂B-determined.*

Recall that we have so far constructed a (Y, M)-superharmonic function corresponding to the root $(\rho_2, 1) \in \mathcal{H}^{\beta\alpha}$. 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 Assumption 2.1 we have sufficient number of conjugate points of $(\rho_2, 1)$ to work with:

Lemma 4.3. *Under Assumption 2.1 and* (2.20)*, there exists $K-1$ additional conjugate points (ρ_2, α_j^*), $j = 2, 3, ..., K$, of $(\rho_2, 1)$ with $0 < \alpha_j^* < \alpha_1^*$.*

Proof. We know that $\Lambda_1(\rho_2, 1) = 1$; then $\Lambda_j(\rho_2, 1) < 1$ for $j = 2, 3, 4..., |\mathcal{M}|$. 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 $\alpha_j^* \in (0, 1)$ such that $\Lambda_j(\rho_2, \alpha_j^*) = 1$. □

Figure 4.1 shows $|\mathcal{M}|$ conjugate points of $(\rho_2, 1)$.

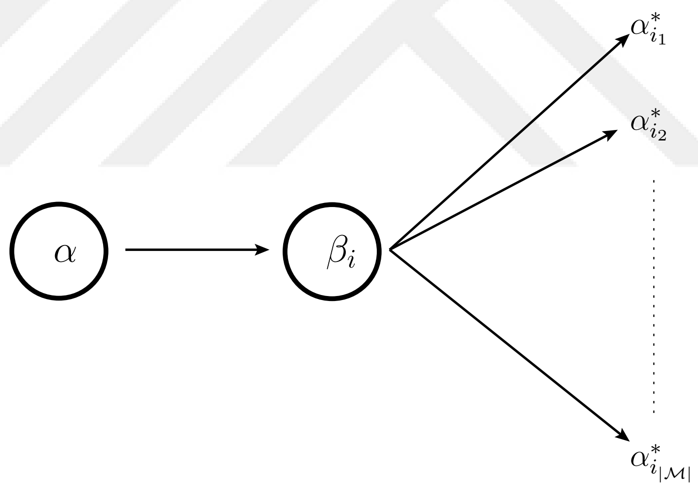

Figure 4.1: Conjugate points of $(\rho_2, 1)$

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

$$\begin{aligned} &\text{There exists } j_1, j_2, ..., j_{n_1}, \;\; n_1 \le |\mathcal{M}|, \text{ such that} \\ &C(\rho_2, 1, \boldsymbol{d}_2) \in \text{Span}\left(C(\rho_2, \alpha^*_{j_k}, \boldsymbol{d}_{2,j_k}), k = 1, 2, ..., n_1\right). \end{aligned} \tag{4.5}$$

Remark 4.1. By definition, $C(\rho_2, \alpha_j, \boldsymbol{d}_{2,j}) = 0$ if $\alpha_j = \rho_2$. Therefore, only those j_k satisfying $\alpha^*_{j_k} \neq \rho_2$ have a role in determining Span $\left(C(\rho_2, \alpha^*_{j_k}, \boldsymbol{d}_{2,j_k}), k = 1, 2, ..., n_1\right)$.

In this sense, assumption (4.5) can be seen as an extension of (2.22) (or, equivalently, of (2.21)).

Remark 4.2. The linear independence of $C(\rho_2, \alpha_j^*, \boldsymbol{d}_{2,j})$, $j = 1, 2, 3, ..., |\mathcal{M}|$, is sufficient for (4.5) to hold. That $C(\beta, \beta, d) = 0$ implies that $\rho_2 \neq \alpha_j$ for all $j = 1, 2, 3, .., |\mathcal{M}|$ is a necessary condition for this independence.

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

Proposition 4.4. *Let* (ρ_2, α_j^*) *be the conjugate points of* $(\rho_2, 1)$ *identified in Proposition 2.17 and Lemma 4.3. Under the additional assumption* (4.5)*, one can find a vector* $b \in \mathbb{R}^{n_1}$ *such that*

$$\boldsymbol{h}_{\rho_2} \doteq [(\rho_2, 1, \boldsymbol{d}_2), \cdot] + \sum_{k=1}^{n_1} b(j)[(\rho_2, \alpha_{j_k}^*, \boldsymbol{d}_{2,j_k}), \cdot] \tag{4.6}$$

is (Y, M)*-harmonic and* ∂B*-determined.*

Proof. Assumption (4.5) implies that the collection of vectors $C(\rho_2, 1, \boldsymbol{d}_2)$, $C(\rho_2, \alpha_{j_k}^*, \boldsymbol{d}_{2,j_k})$, $k = 1, 2, ..., n_1$ are linearly dependent. Therefore, by Proposition 2.15, there exists a vector $b' \in \mathbb{R}^{n_1+1}$ such that

$$b'(0)[(\rho_2, 1, \boldsymbol{d}_2), \cdot] + \sum_{k=1}^{n_1} b'(j)[(\rho_2, \alpha_{j_k}^*, \boldsymbol{d}_{2,j_k}), \cdot]$$

is (Y, M)-harmonic. Assumption (4.5) implies that $b'(0) \neq 0$. Renormalizing the last display by $b'(0)$ gives (4.6). That $\boldsymbol{h}_{\rho_2}$ is ∂B-determined follows from $0 < \alpha_j^* \leq 1$, $\rho_2 < 1$ and Lemma 4.1. □

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

Proposition 4.5. *There exist constants* c_9, c_{10} *and* c_{11} *such that*

$$\mathbb{P}_{(y,m)}(\tau < \infty) < h^{a,0}(y, m) < c_9 \mathbb{P}_{(y,m)}(\tau < \infty) \tag{4.7}$$

where

$$h^{a,0} \doteq c_{11}(\boldsymbol{h}_{\rho_2} + c_{10} h_{\rho_1}). \tag{4.8}$$

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

$$[(\rho_2, \alpha_j^*, \boldsymbol{d}_{2,j}), (k, k, m)] = (\alpha_j^*)^k \boldsymbol{d}_{2,j}(m) \to 0 \tag{4.9}$$

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

$$[(\rho_2, 1, \boldsymbol{d}_2), (k, k, m)] = \boldsymbol{d}_{2,j}(m) > 0, \tag{4.10}$$

for all k. Therefore, there exists $k_0 > 0$ such that

$$\boldsymbol{h}_{\rho_2}(k, k, m) > \min_{m \in \mathcal{M}} \boldsymbol{d}_2(m)/2 \tag{4.11}$$

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

$$h_{\rho_1}(k, k, m) = [(\rho_1, \rho_1, \boldsymbol{d}_1), (k, k, m)] = \boldsymbol{d}_1(m)(\rho_1)^k > 0, \tag{4.12}$$

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

$$\boldsymbol{h}_{\rho_2}(k, k, m) + c_{10} h_{\rho_1}(k, k, m) \geq \min_{m \in \mathcal{M}} \boldsymbol{d}_2(m)/2 \tag{4.13}$$

for all $k \leq k_0$. The last display, (4.11) 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 (4.8). That (4.13) holds for $k \geq 0$ and $m \in \mathcal{M}$ implies

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

By Lemma 4.2 and Proposition 4.4 $h^{a,0}$ is (Y, M)-harmonic and ∂B-determined. This and the last display imply,

$$h^{a,0}(y, m) = \mathbb{E}_{(y,m)}[h^{a,0}(Y_\tau, M_\tau) 1_{\{\tau < \infty\}}] \geq \mathbb{P}_{(y,m)}(\tau < \infty). \tag{4.14}$$

This proves the first inequality in (4.7). To choose c_9 so that the second inequality in (4.7) holds we note the following: (4.9), (4.10) and (4.12) imply

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

Now the same argument giving (4.14) implies the second inequality in (4.7). □

Proposition 4.6. *Fix $m \in \mathcal{M}$ and $x \in \mathbb{R}_+^2$, such that $0 < x(1) + x(2) < 1$ and $x(1) > 0$ if $\rho_1 > \rho_2$ and $1 - x(2) \leq \log(\rho_2)/\log(\rho_1)$, and 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 3.15 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. □

4.1 Improving the approximation

Proposition 4.6 tells us that $h^{a,0}$ of (4.8) approximates $\mathbb{P}_{(y,m)}(\tau < \infty)$ and therefore $\mathbb{P}_{(x,m)}(\tau_n < \tau_0)$ with bounded relative error. The works [30, 31, 34] 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 [30, 31, 34] corresponding to $\boldsymbol{h}_{\rho_2}$, 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 (4.9), (4.10) imply that the value of $\boldsymbol{h}_{\rho_2}$ 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 $\boldsymbol{h}_{\rho_2}$ itself so that we have a function that is almost 1 for most points on ∂B.

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

$$\beta^{|\mathcal{M}|}\boldsymbol{p}(\beta, 1) = 0; \tag{4.15}$$

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 $\boldsymbol{h}_{\rho_2}$. 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 (4.15) in $(0, 1)$; if there are enough of those whose corresponding eigenvectors are linearly independent, we can combine them to obtain the value 1 on ∂B. The next lemma shows that under the stability assumption and the simpleness of all eigenvalues, sufficient number of β roots exist. The proposition after that constructs, under further assumptions, the desired (Y, M)-harmonic function from these β roots.

Lemma 4.7. *Suppose, in addition to the stability assumption (1.1), all eigenvalues of* $A(\beta, \alpha)$ *are real and simple for* $(\beta, \alpha) \in \mathbb{R}_+^{2o}$. *Then there exist* $\rho_{2,k}$, $k = 2, 3, ..., |\mathcal{M}|$, *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*

$$A(\rho_{2,j}, 1)\boldsymbol{e}_j = \boldsymbol{e}_j, j = 2, 3, 4, ..., |\mathcal{M}|,$$

holds.

The proof is parallel to that of Lemma 4.3 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 (4.5) for each $(\rho_{2,j}, 1)$:

Assumption 4.1. For each $j = 2, 3, ..., |\mathcal{M}|$ there exists $m_j \leq |\mathcal{M}|$ conjugate points $(\rho_{2,j}, \alpha_{j,l})$, $l = 1, 2, 3, ..., m_j$ and eigenvectors $0 \neq \boldsymbol{e}_{j,l} \in \mathbb{R}^{\mathcal{M}}$ such that

$$\begin{aligned} |\alpha_{j,l}| &< 1, l = 1, 2, 3, ..., m_j, \\ A(\rho_{2,j}, \alpha_{j,l})\boldsymbol{e}_{j,l} &= \boldsymbol{e}_{j,l} \\ C(\rho_{2,j}, 1, \boldsymbol{e}_j) &\in Span(C(\rho_{2,j}, \alpha_{j,l}, \boldsymbol{e}_{j,l}), l = 1, 2, 3.., m_j). \end{aligned} \tag{4.16}$$

Remark 4.3. Similar to the comments made in Remark 4.2, a set of sufficient conditions for (4.16) is 1) $m_j = |\mathcal{M}|$ and 2) $C(\rho_{2,j}, \alpha_{j,l}, \boldsymbol{e}_{j,l})$, $l = 1, 2, 3, ..., |\mathcal{M}|$ are linearly independent. By $C(\cdot, \cdot, \cdot)$'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 4.4. One can introduce assumptions similar to (2.20) which imply, with an argument similar to the proof of Lemma 4.3, 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 (4.16) to hold and we will allow conjugate points with complex or negative α components. Instead of introducing even more assumptions similar to (2.20), we directly incorporate (4.16) as an assumption.

To get our (Y,M)-harmonic function with the value 1 on ∂B we need one condition:

$$\mathbf{1} \in \mathrm{S}pan(\boldsymbol{d}_2, \boldsymbol{e}_2, ..., \boldsymbol{e}_{|\mathcal{M}|}). \tag{4.17}$$

Obviously, a sufficient condition for (4.17) is that the vectors listed on the right of this display are linearly independent.

Proposition 4.8. *Let $\boldsymbol{e}_j$, $j = 2,3,...,|\mathcal{M}|$ be as in Lemma 4.7. and let $\boldsymbol{d}_2$ be as in Proposition 2.16. Under Assumption 4.1 and* (4.17) *there exist vectors $b_j \in \mathbb{R}^{m_j}$, $j = 2,3,..,|\mathcal{M}|$ and $b^* \in \mathbb{R}^{|\mathcal{M}|}$ such that*

$$\begin{aligned}\boldsymbol{h}_{\rho_{2,j}}(y,m) &\doteq [(\rho_{2,j}, 1, \boldsymbol{e}_j), (y,m)] \\ &\quad + \sum_{l=1}^{m_j} b_j(i)[(\rho_{2,j}, \alpha_{j,l}, \boldsymbol{e}_{j,l}), (y,m)], j = 2,3,...,|\mathcal{M}|,\end{aligned} \tag{4.18}$$

and

$$\boldsymbol{h}^*_{\rho_2} \doteq \sum_{j=1}^{|\mathcal{M}|} b^*(j) \boldsymbol{h}_{\rho_{2,j}} \tag{4.19}$$

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

$$\lim_{k\to\infty} \boldsymbol{h}^*_{\rho_2}(k,k,m) \to 1 \tag{4.20}$$

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

Proof. The existence of the vector b_j so that $\boldsymbol{h}_{\rho_{2,j}}$ defined in (4.18) is (Y,M)-harmonic follows from (4.1) and the argument given in the construction of $\boldsymbol{h}_{\rho_2}$ (see the proof of Proposition 4.4). By (4.17) there is a vector b^* such that

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

for all $m \in \mathcal{M}$. If we choose b^* in this way, $\boldsymbol{h}^*_{\rho_2}$ as defined in (4.19) satisfies

$$\boldsymbol{h}^*_{\rho_2}(k,k,m) = 1 + \sum_{j=1}^{|\mathcal{M}|} b^*(j) \sum_{l=2}^{m_j} b_j(i)\alpha^k_{j,l}\boldsymbol{e}_{j,l}(m)$$

$|\alpha_{j,l}| < 1$ implies the last sum goes to 0 with k. This gives (4.20). □

In Lemma 4.7 we found, in addition to $(\rho_2, 1)$, points on the line $\alpha = 1$ lying on the characteristic surface $\mathcal{H}^{\beta\alpha}$; we used these points above in the construction of $\boldsymbol{h}^*_{\rho_2}$. Similarly, one can go along the line $\beta = \alpha$ to find points, in addition to (ρ_1, ρ_1) lying on $\mathcal{H}^{\beta\alpha}$ giving simple ∂B-determined (Y, M)-harmonic functions:

Lemma 4.9. *Suppose, in addition to the stability assumption (1.1), all eigenvalues of* $A(\beta, \alpha)$ *are real and simple for* $(\beta, \alpha) \in \mathbb{R}_+^{2;o}$ *(same assumption as in Lemma 4.7). Then there exist* $\rho_{1,k}$, $k = 2, 3, 4, .., |\mathcal{M}|$, *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*

$$A(\rho_{1,j}, \rho_{1,j})\boldsymbol{f}_j = \boldsymbol{f}_j, j = 2, 3, 4, ..., |\mathcal{M}|,$$

holds.

The proof is parallel to that of Lemma 4.3 and is based on Gershgorin's Theorem and the fact that $\Lambda_j(\rho_1, \rho_1) < 1$ 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 4.10. *Let* $\rho_{1,j}$, $\boldsymbol{f}_j$, $j = 2, 3, ..., |\mathcal{M}|$ *be as in Lemma 4.9. Then*

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

are ∂B*-determined* (Y, M)*-harmonic.*

Proof. By definition, $(\rho_{1,j}, \rho_{1,j}) \in \mathcal{H}^{\beta\alpha}$ and $A(\rho_{1,j}, \rho_{1,j})\boldsymbol{f}_j = \boldsymbol{f}_j$. Again, $A(\beta, \beta) = A_1(\beta, \beta)$ for all β follows from the definitions of A and A_1. Then $A(\rho_{1,j}, \rho_{1,j})\boldsymbol{f}_j = A_1(\rho_{1,j}, \rho_{1,j})\boldsymbol{f}_j = \boldsymbol{f}_j$, i.e., $(\rho_{1,j}, \rho_{1,j}, \boldsymbol{f}) \in \mathcal{H}^1$. This and Proposition 2.8 imply that $[(\rho_{1,j}, \rho_{1,j}, \boldsymbol{f}), \cdot]$ is (Y, M)-harmonic. That it is ∂B-determined follows from $|\rho_{1,j}| < 1$ and Lemma 4.1. □

The next proposition allows us to compute 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. For any $z \in \mathbb{C}$, let $\Re(z)$ denote its real part.

Proposition 4.11. *Let* $h : \mathbb{Z} \times \mathbb{Z}_+ \mapsto \mathbb{C}$ *be* ∂B*-determined and* (Y, M)*-harmonic. Then*

$$\max_{(y,m)\in B\times\mathcal{M}} \frac{|\Re(h)(y, m) - \mathbb{P}_{(y,m)}(\tau < \infty)|}{\mathbb{P}_{(y,m)}(\tau < \infty)} \leq c^* \tag{4.21}$$

where

$$c^* \doteq \max_{y\in\partial B, m\in\mathcal{M}} |h(y,m)-1|. \tag{4.22}$$

The last proposition allows h to be $\mathbb{C}$ valued because, for any $(\beta,\alpha,d)\in\mathcal{H}$ with complex components the function $[(\beta,\alpha,d),\cdot]$ will be complex valued; we may use such fuctions in improving our approximation, see the next section.

The proof is similar to that of Proposition 4.5:

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 $|\Re(z)-1|\le|z-1|$; these and (4.22) give

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

Then

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

Applying $\mathbb{E}_{(y,m)}[\cdot]$ to all terms above implies (4.21). □

4.2 Numerical example

This section demonstrates the performance of our approximation results on a numerical example. For parameter values P, $\lambda(\cdot)$, $\mu_1(\cdot)$ and $\mu_2(\cdot)$ we take those listed in (2.7) and (2.8), for which $|\mathcal{M}|=3$. We know by Proposition 2.6 that for P as in (2.7), $A(\beta,\alpha)$ has distinct positive eigenvalues for $(\beta,\alpha)\in\mathbb{R}_+^{2,o}$. Furthermore, the rates (2.8) satisfy $\lambda(m)<\mu_1(m),\mu_2(m)$ for all $m\in\mathcal{M}$, therefore, the stability assumption (1.1) is also satisfied. Computing the right side of (2.20) at $(\rho_2,1)$ shows that the parameter values (2.7) and (2.8) satisfy (2.20). Therefore:

1. by Proposition 4.4, the function $\boldsymbol{h}_{\rho_2}$ is well defined and ∂B-determined and (Y,M)-harmonic. Furthermore, Lemma 4.7 implies that we have $|\mathcal{M}|$ points lying on $\mathcal{H}^{\beta\alpha}$ of the form $(\beta^*,1)$ such that $0<\beta^*<1$. Solving the equation

$$\boldsymbol{p}(\beta^*,\alpha)=0$$

for each of these β^* shows that these parameter values satisfy Assumption 4.1, with $m_j = |\mathcal{M}|$ for all j; this and Proposition 4.8 imply that have a (Y, M)-harmonic ∂B-determined function $\boldsymbol{h}^*_{\rho_2}$ of the form (4.19) which satisfies (4.20);

2. Propositions 2.11, 2.12 and Lemma 4.2 apply and give the ∂B-determined (Y, M)-harmonic function $h_{\rho_1} = [(\rho_1, \rho_1, \boldsymbol{d}_1), \cdot]$,

3. Lemmas 4.9 and 4.10 apply and give the ∂B-determined (Y, M)-harmonic functions $h_{\rho_{1,j}} = [(\rho_{1,j}, \rho_{1,j}, \boldsymbol{f}_j), \cdot]$, $j = 2, 3, 4, ...|\mathcal{M}|$.

In addition to these functions, we will use $K \cdot |\mathcal{M}|$ (Y, M)-harmonic functions of the form

$$h^c_{k,j} \doteq \sum_{j=0}^{|\mathcal{M}|} b^c_k(j)[(\beta^c_{k,j}, \alpha^c_{k,l}, d^c_{k,l}), (y, m)], k = 1, 2, 3, ..., K,$$

such that

1. $\alpha^c_{k,0} = R\, e^{ik\frac{2\pi}{K+1}}$;

2. for each k, $\beta^c_{k,j}$, $j = 1, 2, 3, ..., |\mathcal{M}|$, are found by solving

$$\boldsymbol{p}(\alpha^c_{k,0}, \beta) = 0 \tag{4.23}$$

   for β; we choose those solutions which satisfy $|\beta| < 1$.

3. for each k and j, $\alpha^c_{k,j,l}$ are found by solving

$$\boldsymbol{p}(\beta^c_{k,j}, \alpha) = 0; \tag{4.24}$$

   we choose those α satisfying $|\alpha| < 1$;

4. $d^c_{k,j,l}$ is an eigenvector of $A(\beta^c_{k,j}, \alpha^c_{k,j,l})$ i,e., $(\beta^c_{k,j}, \alpha^c_{k,j,l}, d^c_{k,j,l}) \in \mathcal{H}$,

5. the vectors $b^c_{k,j}$ are obtained by solving

$$\sum_{l=0}^{|\mathcal{M}|} b^c_{k,j}(l) C(\beta^c_{k,j}, \alpha^c_{k,j,l}) = 0. \tag{4.25}$$

For $h^c_{k,j}$, $k = 1, 2, ..., K$ and $j = 1, 2, 3, ...|\mathcal{M}|$ to be defined (Y, M)-harmonic and ∂B-determined we need 1) for each k, the equation (4.23) needs to have at least $|\mathcal{M}|$ roots β with absolute value less than 1; 2) for each k and j, the equation (4.24) needs to have at least $|\mathcal{M}|$ solutions with absolute value less than 1; 3) for each k and j the equation (4.25) needs to have a nontrivial solution $b_{k,j}$. 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 (4.23), (4.24) and (4.25) with these parameter values we observe that they have sufficient number of solutions for $h_{j,k}$ 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

$$h^a \doteq \Re(h^{a*}), h^{a*} \doteq \boldsymbol{h}^*_{\rho_2} + \sum_{j=1}^{|\mathcal{M}|} \boldsymbol{c}^d_j[(\rho_{1,j}, \rho_{1,j}, \boldsymbol{d}_{1,j}), \cdot] + \sum_{j=1,k=1}^{K,|\mathcal{M}|} \boldsymbol{c}_{j,k} h^c_{j,k}, \tag{4.26}$$

where $\boldsymbol{c}^d_j$ and $\boldsymbol{c}_{j,k}$ are $\mathbb{C}$ valued coefficients to be chosen so that $h^a|_{\partial B}$ is as close to 1 as possible. As in [30, Section 8.2], one simple way to do this is to choose these $(K + 1)|\mathcal{M}|$ coefficients so that $h^a(y, y, m) = 1$ for $y = 0, 1, 2, 3, .., K$ 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 a 18×18 system, and it does turn out to have a unique solution. Once the $\boldsymbol{c}_j$ and $\boldsymbol{c}_{j,k}$ are determined through this solution, an upper bound on the approximation relative error can be computed via Proposition 4.11; it suffices to compute c^* of (4.22); for h^{a*} of (4.26) this quantity turns out to be

$$c^* = 0.00367;$$

therefore, by Proposition 4.11, h^a approximates $\mathbb{P}_{(y,m)}(\tau < \infty)$ with relative error bounded by this quantity. By Theorem 3.15 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 $h^a(n - x_n(1), x_n(2))$ 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 4.2 gives the level curves of $-\log(h^a(n - x(1), x(2), 1))$ 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.

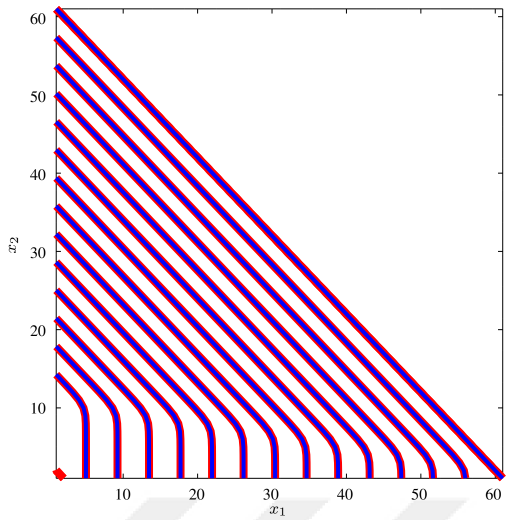

Figure 4.2: Level curves of $-\log(h^a(n - x(1), x(2), 1))$ and $-\log \mathbb{P}_{(x,m)}(\tau_n < \tau_0)$

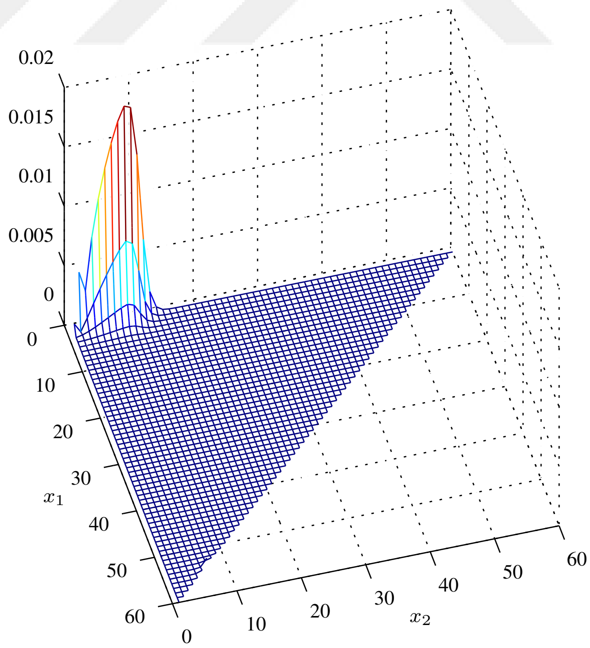

Figure 4.3: The relative error $\frac{|\log(h^a(n-x(1),x(2),1)-\log \mathbb{P}_x(\tau_n<\tau_0)|}{|\log \mathbb{P}_x(\tau_n<\tau_0)|}$

Figure 4.3 shows the relative error

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

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 3.15. Figure 4.4 shows the detailed graph of the relative error around the origin.

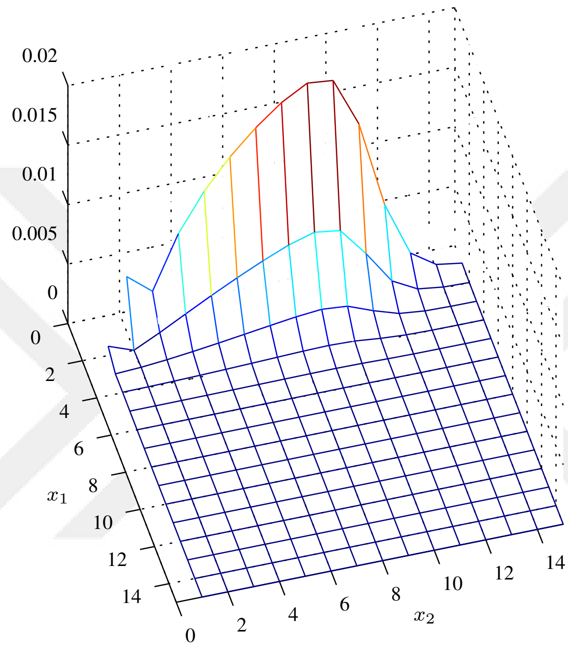

Figure 4.4: The relative error (detailed graph around the origin)

CHAPTER 5

APPLICATIONS IN FINANCE AND INSURANCE

Constrained random walks naturally arise in finance and insurance applications. In Section 5.1 below we give an insurance interpretation of our results; in this context constrained dynamics arise from dividend payments and the hitting time τ_n represents the total reserves of a system consisting of two insurance companies hitting a low threshold n. In Section 5.2 we describe an application of the same ideas to the problem of market making, i.e., to provide buy and sell liquidity for multiple assets. In this context constraints on the dynamics correspond to "no short selling allowed" and hitting time τ_n corresponds to the market makers inventory's getting excessively large, an undesirable state for the market maker. The first application involves multiple regimes modulated by an external Markov chain M, modeling different market conditions under which the companies operate. In the second application modulation is internal.

5.1 Reserve Problem

After the 2008 financial crisis great interest emerged in the modeling of financial and insurance systems consisting of many companies, see [6, 16, 29] and the references in these works. The simplest multidimensional model will consist of two companies; each dimension modeling the financial situation of one of the companies. For the purposes of this example, let us consider two insurance companies. We model their reserves as a Markov modulated random walk in $\mathbb{Z}^2$; the time step corresponds to a unit of time such as a year, quarter or a month, the step increments of the random walk corresponds to some fixed unit of money, e.g., million dollars. The state of the modulating chain M represents, e.g., the business cycle.

So far (X, M) is unconstrained; if these companies pay dividends when their reserves hit certain fixed levels $k(1)$ and $k(2)$, these payments can be represented with a constraining map π:

$$\pi(x, v) = \begin{cases} x + v, \text{ if } x(i) + v(i) \leq k(i), i = 1, 2, \\ 0 \text{ otherwise.} \end{cases}$$

The resulting process has the dynamics

$$X(k + 1) = X(k) + \pi(X(k), I(k)),$$

where $I(k)$ represents the reserve changes from one period to the other whose distribution is determined by the modulating chain M. In general, the possible jumps of I will depend on the income, expenses of the companies as well as the agreements made between them (for example, reinsurance agreements). In the specific case when $I(k)$ can take the values $\{(1, -1), (-1, 0), (0, 1), (0, 0)\}$ this model reduces exactly to a reflection of the tandem walk model given in Chapter 1, see Figure 5.1 which shows the dynamics of X on a single layer.

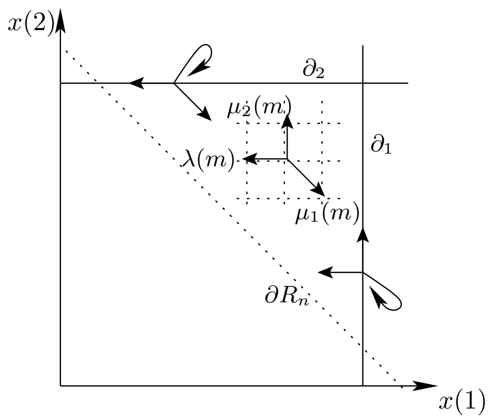

Figure 5.1: The dynamics of (X, M)

The increment $(-1, 0)$ corresponds to the decrease in reserves of the first company, $(1, -1)$ corresponds to the increase in reserves of the first company and a decrease in the reserves of the second company, $(0, 1)$ corresponds to an increase in the reserves of the second company while the reserves of the first company stays constant.

A sample path of the process (X, M) as regime changes occur is depicted in Figure 5.2 where each layer represents a state of the business cycle. For example, during the expansion phase of the business cycle, the reserve level of companies moves on the expansion layer. When the business cycle changes from expansion to recession, the reserve levels remain stable. After the phase change, they continue their move on the recession layer at recession rates.

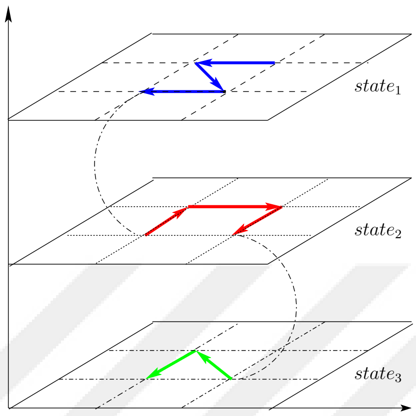

Figure 5.2: A sample path of (X, M)

The stability assumption implies that X moves in cycles which restart every time it hits $(k(1), k(2))$, i.e., the point of maximum reserves for both companies. Let τ_0 denote the first time X hits the points $(k(1), k(2))$. To see the financial interpretation of the probability p_n in this context introduce the absorbing barrier $\partial R_n = \{r \in \mathbb{Z}_+^2,\ x(1) + x(2) = n\}$. The process hits ∂R_n exactly when the total reserves in the system goes down to the level n, see Figure 5.1. Therefore, the probability p_n that we have studied in the last three chapters represents exactly the probability that the total reserves in the system hits level n in a given cycle of the system.

Instead of the hitting time τ_n to the barrier R_n we can also study the probability $\mathbb{P}_{(y,m)}(\sigma_1 \wedge \sigma_2 < \tau_0)$, where σ_i is the first hitting time to the coordinate axis $x(i)$. In our present context σ_i represents the time when company i's reserves hit 0, i.e., the ruin time of company i. Then $\sigma_1 \wedge \sigma_2$ represents the first ruin time for this system and the probability $\mathbb{P}_{(y,m)}(\sigma_1 \wedge \sigma_2 < \tau_0)$ represents the probability that one of the companies is ruined in a given cycle. This problem differs from the problem of earlier chapters only in its rectangular exit boundary. We think that it is possible to extend the approach of [30, 31] and the present work to treat such boundaries by using different points on the exit boundary as origin of the transformed coordinates and characteristic surfaces compatible with the exit boundary. A careful study of these ideas remain for future work. In the next section we present another possible application in finance again making use of rectangular exit boundaries.

5.2 Market Maker's Inventory Problem

Market makers play a significant role in financial markets as liquidity providers. They are always ready to buy and sell an asset at the price they quote. Their profit comes from the difference between the buying price and selling price of the traded asset, i.e. bid-ask spread. We will introduce the market making model described in [16]; this model was initially proposed by [2].

Consider a financial market in which a market maker trades two assets trying to maximize its expected utility. The buying and selling prices are continuously determined by the market maker. The model assumes that one unit of asset is traded at every transaction. The number of each assets held at time t in market maker's inventory is

$$N_t(i) = N_t^{b,i} - N_t^{a,i}, i = 1, 2;$$

where $N_t^{b,i}$ and $N_t^{a,i}$ represents the number of the i^{th} asset bought and the number of the same asset sold, respectively. The inventory process N_t is a continuous time process on $\mathbb{Z}^2$, see Figure 5.3; the first [second] coordinate represents the inventory of the market maker in the first [second] asset.

For each $i = 1, 2$, the processes $N_t^{b,i}$ and $N_t^{a,i}$, are Poisson processes with intensity $\lambda_t^{b,i}$ and $\lambda_t^{a,i}$. The work in [2] assumes that these trading intensities depend on the distance between market maker's buying price (or selling price) and market price of asset:

$$\lambda_t^{b,i} = \Lambda^{b,i}(\delta_t^{b,i}), \quad \delta_t^{b,i} = S_t^i - S_t^{b,i},$$

$$\lambda_t^{a,i} = \Lambda^{a,i}(\delta_t^{a,i}), \quad \delta_t^{a,i} = S_t^{a,i} - S_t^i.$$

The market maker determines the spreads $\delta^{a,i}$ and $\delta^{b,i}$ by solving the optimization problem

$$v(X, S, N, t) = \max_{\delta^{a,i}, \delta^{b,i}, i=1,2} \mathbb{E}[-\exp(-\gamma(X_T + \langle N_T, S_T \rangle))] \tag{5.1}$$

where X is the market maker's wealth and γ is the risk aversion parameter. The optimal bid-ask prices quoted by the market maker depend both on N and X. Note that the expectation on the right is smaller when N_T is large. Larger N_T exposes the market maker to price risk; suggests that under optimal spreads, the process N would be stable and work in cycles restarting each time N hits the origin $(0, 0)$.

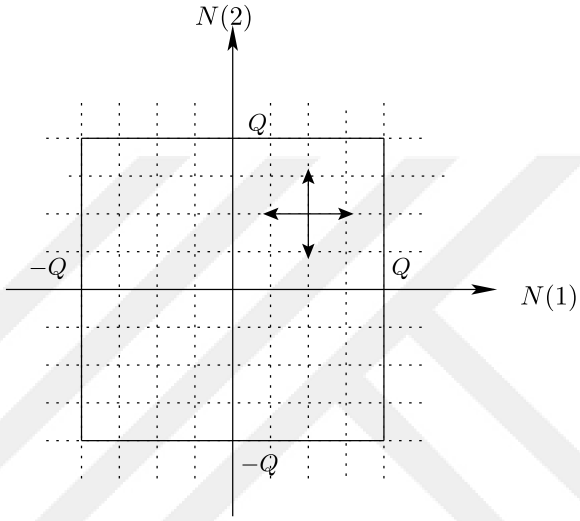

Figure 5.3: The dynamics of N

Observing the process N at the times when trades occur reduces it to a discrete time random walk on $\mathbb{Z}^2$; the possible increments of the random walk are $(1, 0)$ (market maker buys one unit of asset 1), $(-1, 0)$ (sells one unit of asset 1), $(0, 1)$ (market maker buys one unit of asset 2), $(0, -1)$ (sells one unit of asset 2); these jumps occur with probabilities

$$\frac{\lambda^{b,1}}{S}, \frac{\lambda^{a,1}}{S}, \frac{\lambda^{b,2}}{S}, \frac{\lambda^{a,2}}{S},$$

where

$$S = \lambda^{b,1} + \lambda^{a,1} + \lambda^{b,2} + \lambda^{a,2}.$$

Note that these jump probabilities depend on the spreads $\delta^{a,i}$, $\delta^{b,i}$; which can be interpreted as the environment variables within which N operates (i.e., modulation/ regime switch). But these environment processes depend back on N through the optimization problem (5.1), i.e., the regime switch in the present setup is internal.

For inventory level, [16] specifies a threshold $Q \in \mathbb{N}$, i.e., the market maker has an upper bound on its inventories. The process stops once it hits the boundary where the total number of assets 1 or 2 held by the market maker reach Q or $-Q$. The market maker does not buy or sell an asset anymore if the inventory in that asset is Q or $-Q$. Hitting these levels is an undesirable situation because it stops trading in the assets.

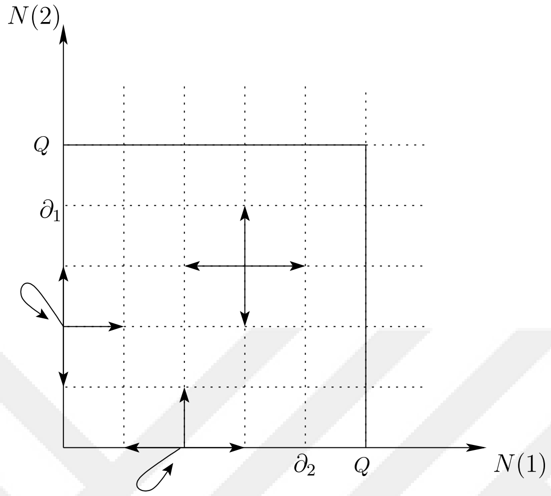

Figure 5.4: The dynamics of N when no short selling is allowed

The model so far has no constraining boundaries; if the market maker is not allowed to short sell, this restriction is modeled by constraining boundaries on the coordinate axes ∂_1 and ∂_2, see Figure 5.4.

Define the hitting time

$$\tau_Q \doteq \inf\{k : |N_k(i)| = Q, i = 1, 2\}, \tau_0 \doteq \inf\{k : N_k = (0, 0)\}.$$

τ_Q is the first time trading in one of the assets halts; the probability

$$\mathbb{P}_{(n_1,n_2)}(\tau_Q < \tau_0) \tag{5.2}$$

is the probability that the market maker is unable to provide liquidity in one of its cycles due to large negative or positive inventory buildup. Note that this is exactly the type of probability studied in the present thesis. In future work we hope to extend the analysis in this thesis to address the approximation of the probability (5.2) of lack of liquidity in a market in a given cycle of the market maker.

CHAPTER 6

LITERATURE REVIEW

As we have already pointed out the difficulty in the computation of p_n is that the state space grows in at least n^2; making the exact and numerical solution of the corresponding linear system difficult. Another difficulty arises from the fact that the stability assumption implies that the components of p_n decay to 0 with n making an asymptotic analysis with classical methods difficult, see [3] for a review discussing these issues. The classisal approach to the asymptotic analysis of quantities such as p_n is large deviations analysis which gives the exponential decay rate of p_n. The large deviations (LD) analysis of p_n for $x = 0$ for non-modulated constrained random walks arising from Jackson networks was done in [17].

The popular technique in the literature for more precise estimates of p_n has been simulation. Because p_n is the probability of a rare event, the ordinary Monte Carlo (MC) method requires too many sample paths to give accurate estimates, this makes MC a poor choice for the approximation of p_n. Therefore, one needs to use variance reduction techniques. One of these that has received great attention in the last several decades is Importance Sampling (IS), which uses a new simulation measure under which the overflow event is no longer rare; the estimator is modified accordingly using the Radon Nikodym derivative of the original measure with respect to the simulation measure (the likelihood ratio of these two measures). The goal in IS is to choose a simulation measure that almost minimizes the second moment of the IS estimator; for p_n the second moment of the optimal estimator decays at twice the rate of p_n. A sequence of IS changes of measures is said to be asymptotically optimal if the decay of their second moment matches this rate. As the relation between the decay rate of the second moment of the estimator and that of p_n itself suggests, there is a close connection between LD analysis of a probability and its optimal importance sampling. Indeed, a heuristic idea that emerged early on is to use an "optimal path" that arises in LD analysis to construct the IS change of measure. To the best of our knowledge, the paper [33] is the first to use this connection to construct almost optimal IS changes of measure for the estimation of a probability very similar to p_n for a single dimensional process.

The work [22] attempted to extend these ideas to multidimensions by applying them to the constrained random walk associated with the two tandem queues. The IS change of measure implied by the LD analysis turns out to be static (exchanging λ with the smaller of μ_i). In [22] this change of measure is modified in boundary layers along the coordinate axes ∂_i leading to state dependent changes of measure, see Figure 6.1.

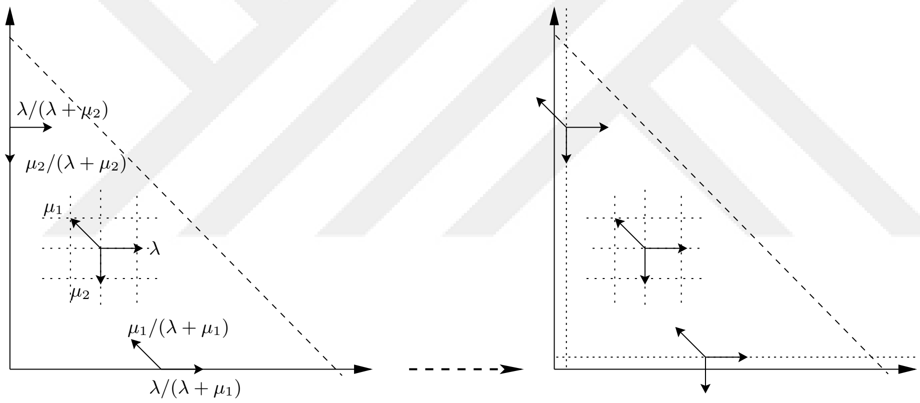

Figure 6.1: Boundary layers

Glasserman and Kou [15] further observed that the static change of measure given by LD optimal path can have very poor performance for the two tandem constrained random walk if when the service rates are nearly equal and the arrival rate is small.

Asymptotically optimal IS changes of measures for the simulation of p_n for ordinary constrained random walks associated with Jackson networks and for the initial condition $x = 0$, were finally constructed in [11, 12, 26, 28]. The construction and analysis in these works rely on a limit HJB equation and its subsolutions. The limit HJB equation in these works arise from the convex transformation $V_n(x) = -\frac{1}{n}\log p_n(\lfloor nx \rfloor)$. The disadvantage of this transformation is that the presence of $\frac{1}{n}\log$ implies that much information about the behavior of p_n is lost upon letting $n \to \infty$ (only the exponential decay rate of p_n remains under this transformation; e.g., if the decay rate is V $n^2 e^{-nV(x)}$ and $e^{-nV(x)}$ are the same, as far as this transformation is concerned). To overcome this difficulty, [30] introduced a simple affine transformation which keeps most of the structure of p_n intact; this was already reviewed in Chapter 1. [30, 31] derived approximation formulas for p_n using this approach and showed that they approximate p_n with exponentially decaying relative error in the case of two dimensional non-modulated constrained tandem walk. For further works [7, 18, 20, 21] we refer the reader to [31].

The works reviewed so far have all considered non-modulated constrained random walks. The study of p_n for modulated constrained random walks is extremely limited. To our knowledge, [27] is the only paper studying the estimation of overflow probability p_n focusing on the Markov modulated two tandem network. The system dynamics are the same as our setup. This work extends the subsolution approach of [11] to the two dimensional Markov modulated tandem walk and constructs asymptotically optimal IS simulation algorithms for the estimation of p_n; to prove asymptotical optimality one also needs the LD decay rate of p_n, this was also computed in the same work. As with earlier works using the subsolution approach [27] focuses on the initial condition $x = (0, 0)$ (or any other sequence of initial condition converging to $(0, 0)$ with n). Just as there is a strong connection between the structures used in [30, 31] and [11] there is a similar connection between the structures appearing in this thesis and [27]. [27] constructs its subsolutions from the roots of a Hamiltonian function H appearing in the limit HJB equation. The H of [27] equals $p \mapsto -\log \Lambda_1(e^{p_1 - p_2}, e^{p_2})$; the subsolutions in [27] are based essentially on two roots of this function, denoted by $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ in [27] (see [27, Figure 2]). The first of these correspond to the point $(\rho_2, 1)$ of Proposition 2.16, the second corresponds to the point (ρ_1, ρ_1) of Proposition 2.11. The proofs of these propositions and that of [27, Lemma 4] use the same argument based on the implicit function theorem; [27] uses these two points to construct subsolutions to a limit HJB equation which yield asymptotically optimal IS simulation algorithms. We use the points $(\rho_2, 1)$, (ρ_1, ρ_1) and many others to construct (Y, M) and (X, M)-(sub,super)harmonic functions 1) to prove that $\mathbb{P}_{(T_n(x_n),m)}(\tau < \infty)$ approximates $\mathbb{P}_{(x_n,m)}(\tau_n < \tau_0)$ with exponentially decaying relative error and 2) to construct approximate formulas for the probability $\mathbb{P}_{(y,m)}(\tau < \infty)$. Our formulas for the last probability uses points on all of the components $\{\Lambda_j = 1\}$, $j = 1, 2, 3, ...|\mathcal{M}|$, of the characteristic surface allowing complex and negative values. As we have already indicated the analysis in [27] is essentially based on the points mentioned above on a curve that curresponds to the primary component $\{\Lambda_1 = 1\}$.

In the previous chapter we have indicated several possible applications of constrained random walk models in insurance and finance. Next we would like to point out further studies in the current literature that uses ideas from queueing theory and constrained random walks in models of financial mathematics. The work [14] proposed a model introducing a random environment where stock price is determined according to the demand of market participants. Following this paper, [4] developed an agent-based approach to stock price fluctuations based on methods and techniques from state dependent Markovian queueing networks. In their model, the agent's orders arrive with an arrival rate depending on the current stock prices and investor characteristics. The paper [8] proposed a credit risk model of a large portfolio by using state dependent queueing networks. The authors made a connection between obligors that moves around rating categories or default with rates change according to the macroeconomic environment and customers visiting service stations where the arrival and service rates depend on the external Markov process.

CHAPTER 7

CONCLUSION AND OUTLOOK

7.1 Comparison with earlier works

In the present work, we extend the limit anaysis approach for the approximation of the probability $\mathbb{P}_x(\tau_n < \tau_0)$ by $\mathbb{P}_y(\tau < \infty)$ developed in [30, 31] to two tandem walk with Markov modulated dynamics. In this section, we compare the Markov modulated two tandem case treated in this thesis with the non-modulated two tandem case treated in [30, 31] and the non-modulated parallel case (the two dimensional simple random walk) treated in [34]. Our main finding is that it is indeed possible to implement the affine transformation approach of [30, 31] in a Markov modulated framework. But introduction Markov modulation leads to many novelties and difficulties in the limit analysis and computation of $\mathbb{P}_{(y,m)}(\tau < \infty)$. We discuss these in the paragraphs below.

Harmonic functions In the non-modulated two tandem case and the parallel case, the harmonic functions are in the form $y \mapsto [(\beta, \alpha), y] = \beta^{y(1)-y(2)}\alpha^{y(2)}$ and (β, α) are chosen from the roots of a characteristic polynomial of second order associated with the process Y. With Markov modulation we have an additional state variable m, leading to the harmonic functions in the form $y \mapsto [(\beta, \alpha, d), (y, m)] = \beta^{(y(1)-y(2)}\alpha^{y(2)}d(m)$. The characteristic surface is now defined in terms of eigenvalue and eigenvector equations. We study the geometry of this characteristic surface through the eigenvalues to identify the points in the analysis.

Geometry of the characteristic surface The characteristic surface in [30, 31, 34] are 1-level curves of second order polynomials; the projection of the characteristic surface to $\mathbb{R}^2_+$ consists of a simple egg shaped curve. Conjugate points on this curve come in pairs and have very elementary formulas. The same curve in modulated case is defined by a $2|\mathcal{M}|$ order polynomial; its projection to $\mathbb{R}^2_+$ consists of $|\mathcal{M}|$ pieces one for each eigenvalue Λ_j of the matrix appearing in the definition of the characteristic polynomial. There are in general no simple formulas for the roots of a polynomial greater than degree 4, the formulas for degree 3 and 4 are fairly complex; therefore, for $|\mathcal{M}| \geq 2$ (i.e., even for the simplest nontrivial Markov modulated constrained random walk with two modulating states) the conjugate points no longer have simple formulas and even to prove that they exist require a lot analysis using properties of matrices, eigenvalues and eigenvectors and the implicit function theorem.

Assumptions We use the point $(\rho_2, 1)$ and its conjugate (ρ_2, α_1^*) lying on the $\mathcal{H}^1$ to define (Y, M)-superharmonic functions to use in our limit analysis. The identification of the conjugate point (ρ_2, α_1^*) requires a technical assumption ensuring $\alpha_1^* < 1$. There is no corresponding assumption in the non-modulated tandem case, because the α component of the conjugate point $\rho_1 = \lambda/\mu_1$ is always less than 1 by the stability assumption. For the parallel case, the corresponding assumption is $r^2 < \rho_2$.

$\rho_1 \neq \rho_2$ This assumption is equivalent to $\alpha_1^* \neq \rho_1$ and generalizes the assumption $\mu_1 \neq \mu_2$ for the non-modulated tandem case and the parallel case. The computation of $\mathbb{P}_{(y,m)}(\tau < \infty)$ needs progressively more general versions of this assumption (see (4.5) and (4.1)).

Analysis The analysis in the non-modulated cases are based on the subsolutions of a limit HJB equation and Y-harmonic functions. These works use the subsolutions to construct supermartingales which are then used to find upper bounds on error probabilities. In this thesis 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; these ideas are new to the present thesis.

In the tandem case there is an explicit formula for $P_y(\tau < \infty)$; this formula is used in the analysis of the error probability. There is obviously no explicit formula for the corresponding probability in the Markov modulated case to compensate for this, we derive an upper bound on this probability in Section 3.1 using again (Y, M)-superharmonic functions.

Computation of the limit probability In the non-modulated tandem case there is an explicit simple formula for $\mathbb{P}_y(\tau < \infty)$ that equal a linear combination of h_{ρ_2} and h_{ρ_1}. In the parallel case, an additional assumption $\rho_1\rho_2 = r^2$ leads to an explicit formula from the linear combinations of h_{ρ_1} and h_r. Without this assumption, they give an approximation for $\mathbb{P}_y(\tau < \infty)$ with exponentially decaying relative error and add more harmonic functions constructed from conjugate points to improve their approximation.

In our case, there is no more explicit formulas. We approximate the probability $\mathbb{P}_{(y,m)}(\tau < \infty)$ with bounded relative error using linear combinations of (Y, M)-harmonic functions; the construction of each of these functions require solution of $2|\mathcal{M}|$ degree polynomial equation, the computation of corresponding eigenvectors and the solution of a linear equation, which, in general, has $|\mathcal{M}|$-unknowns. These tasks either do not exist in the non-modulated case or are trivial, because the polynomial equations are quadratic and the number of unknowns in the linear equation is only 1.

7.2 Conclusion

In this thesis, we develop approximative formulas for the exit probability of two tandem walk with modulated dynamics. The main approximation Theorem 3.15 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 [30], the probability $\mathbb{P}_y(\tau < \infty)$ has an explicit formula for tandem walks in d dimensions. As we have seen in this thesis, even the extension to $d = 2$ entails considerable difficulties. Whether an extension to higher dimensions is possible is a question we would like to tackle in future work.

The work [30] gives also a formula for $\mathbb{P}_y(\tau < \infty)$ for the case $\rho_1 = \rho_2$ based on harmonic functions with polynomial terms for the non-modulated tandem walk. 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 questions for future research.

The computation and error analysis in the present work depend on the dynamics of the process and the geometry of the exit boundary. Another problem for future research is to extend the analyses to walks that have different dynamics than the tandem walk and 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 (which appears in Section 5.2 in the market making problem) appear to be most natural to study in immediate future work.

Another topic for future research is the approximation of large buildup probability for market making problem (Section 5.2) in which the inventory process is constrained in a rectangular domain.

REFERENCES

[1] R. Atar and P. Dupuis, Large deviations and queueing networks: methods for rate function identification, Stochastic Process. Appl., 84(2), pp. 255–296, 1999, ISSN 0304-4149.

[2] M. Avellaneda and S. Stoikov, High-frequency trading in a limit order book, Quantitative Finance, 8(3), pp. 217–224, 2008.

[3] S. Balsamo, V. D. N. Personè, and P. Inverardi, A review on queueing network models with finite capacity queues for software architectures performance prediction, Performance Evaluation, 51(2-4), pp. 269–288, 2003.

[4] E. Bayraktar, U. Horst, and R. Sircar, Queuing theoretic approaches to financial price fluctuations, Handbooks in Operations Research and Management Science, 15, pp. 637–677, 2007.

[5] A. Berman and R. J. Plemmons, *Nonnegative matrices in the mathematical sciences*, volume 9, Siam, 1994.

[6] J. Blanchet, J. Li, and Y. Shi, Stochastic risk networks: Modeling, analysis and efficient Monte Carlo, 2015.

[7] J. Collingwood, R. D. Foley, and D. R. McDonald, Networks with cascading overloads, in *Proceedings of the 6th International Conference on Queueing Theory and Network Applications*, pp. 33–37, ACM, 2011.

[8] M. H. Davis and J. C. Esparragoza-Rodriguez, Large portfolio credit risk modeling, International Journal of Theoretical and Applied Finance, 10(04), pp. 653–678, 2007.

[9] P.-T. de Boer, Analysis of state-independent importance-sampling measures for the two-node tandem queue, ACM Transactions on Modeling and Computer Simulation (TOMACS), 16(3), pp. 225–250, 2006.

[10] P. Dupuis and R. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*, John Wiley & Sons, New York, 1997.

[11] P. Dupuis, A. D. Sezer, and H. Wang, Dynamic Importance Sampling for Queueing Networks, Annals of Applied Probability, 17(4), pp. 1306–1346, 2007.

[12] P. Dupuis and H. Wang, Importance Sampling for Jackson Networks, Queueing Systems, 62, pp. 113–157, 2009.

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

[14] H. Föllmer and M. Schweizer, A microeconomic approach to diffusion models for stock prices, Mathematical Finance, 3(1), pp. 1–23, 1993.

[15] P. Glasserman and S.-G. Kou, Analysis of an importance sampling estimator for tandem queues, ACM Transactions on Modeling and Computer Simulation (TOMACS), 5(1), pp. 22–42, 1995.

[16] O. Guéant, *The Financial Mathematics of Market Liquidity: From optimal execution to market making*, volume 33, CRC Press, 2016.

[17] I. Ignatiouk-Robert, Large deviations of Jackson networks, Annals of Applied Probability, pp. 962–1001, 2000.

[18] I. Ignatiouk-Robert and C. Loree, Martin boundary of a killed random walk on a quadrant, The Annals of Probability, pp. 1106–1142, 2010.

[19] P. D. Lax, Linear Algebra. Pure and Applied Mathematics, 1996.

[20] D. McDonald, Asymptotics of first passage times for random walk in an orthant, Annals of Applied Probability, pp. 110–145, 1999.

[21] M. Miyazawa, Tail decay rates in double QBD processes and related reflected random walks, Mathematics of Operations Research, 34(3), pp. 547–575, 2009.

[22] S. Parekh and J. Walrand, A quick simulation method for excessive backlogs in networks of queues, IEEE Transactions on Automatic Control, 34(1), pp. 54–66, 1989.

[23] N. Prabhu and Y. Zhu, Markov-modulated queueing systems, Queueing systems, 5(1-3), pp. 215–245, 1989.

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

[25] A. D. Sezer, Representation results for largest eigenvalues of matrices with positive entries, Preprint available at: https://sites.google.com/site/devinsezer/.

[26] A. D. Sezer, Asymptotically Optimal Importance Sampling for Jackson Networks with a Tree Topology, arXiv preprint arXiv:0708.3260, 2007.

[27] A. D. Sezer, Importance sampling for a Markov modulated queuing network, Stochastic Processes and their Applications, 119(2), pp. 491–517, 2009.

[28] A. D. Sezer, Asymptotically optimal importance sampling for Jackson networks with a tree topology, Queueing Systems, 64(2), pp. 103–117, 2010, longer (2007) version available at arXiv preprint arXiv:0708.3260.

[29] A. D. Sezer, Modeling of an insurance system and its large deviations analysis, Journal of computational and applied mathematics, 235(3), pp. 535–546, 2010.

[30] A. D. Sezer, Exit Probabilities and Balayage of Constrained Random Walks, arXiv preprint arXiv:1506.08674, 2015.

[31] A. D. Sezer, Approximation of Excessive Backlog Probabilities of Two Tandem Queues, arXiv preprint arXiv:1801.04674, 2018.

[32] A. Shwartz and A. Weiss, *Large deviations for performance analysis*, Stochastic Modeling Series, Chapman & Hall, London, 1995, ISBN 0-412-06311-5, queues, communications, and computing, With an appendix by Robert J. Vanderbei.

[33] D. Siegmund, Importance sampling in the Monte Carlo study of sequential tests, The Annals of Statistics, pp. 673–684, 1976.

[34] K. D. Ünlü and A. D. Sezer, Excessive Backlog Probabilities of Two Parallel Queues, arXiv preprint arXiv:1806.00686, 2018.

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Başoğlu Kabran, Fatma

Nationality: Turkish

Date and Place of Birth: 1985, İstanbul

Phone: +90 (212) 216 09 14

EDUCATION

Degree	Institution	Year of Graduation
Ph.D.	Financial Mathematics, METU	2018
M.S.	Financial Mathematics, METU	2012
B.S.	Business Administration, Marmara University	2008
High School	Cağaloğlu Analotian High School	2003

PROFESSIONAL EXPERIENCE

Year	Place	Enrollment
2018 - Present	İzmir Vocational College of Higher Education	Lecturer
2014 - 2015	İzmir University of Economy	Research Assistant