INTERTWININGS, INTERLACING EIGENVALUES, AND STRONG STATIONARY DUALITY FOR DIFFUSIONS

by

Vince Lyzinski

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

January, 2013

Abstract

In part one of this dissertation, we develop a systematic matrix-analytic approach, based on intertwinings of Markov semigroups, for proving theorems about hitting-time distributions for finite-state Markov chains—an approach that (sometimes) deepens understanding of the theorems by providing corresponding sample-path-by-sample-path stochastic constructions. We employ our approach to give new proofs and constructions for two theorems due to Mark Brown, theorems giving two quite different representations of hitting-time distributions for finite-state Markov chains started in stationarity. The proof, and corresponding construction, for one of the two theorems elucidates an intriguing connection between hitting-time distributions and the interlacing eigenvalues theorem for bordered symmetric matrices.

In part two, we develop the theory of strong stationary duality for diffusion processes on finite intervals. We analytically derive the generator and boundary behavior of the dual process and recover a central tenet of the classical theory by proving that the separation mixing time in the primal diffusion is equal in law to the absorption time in the dual diffusion. We also exhibit our strong stationary dual as the natural limiting process of the strong station-

ABSTRACT

ary dual sequence of a well chosen sequence of approximating birth-and-death Markov chains, allowing for simultaneous numerical simulations of our primal and dual diffusion processes. Lastly, we show how our new definition of diffusion duality allows the spectral theory of cutoff phenomena to extend naturally from birth-and-death Markov chains to the present diffusion context.

Primary Reader: Professor James Allen Fill

Secondary Reader: Professor Avanti Athreya

Acknowledgements

It has been wonderful working with Professor James Allen Fill. His brilliant insights have consistently helped me look at mathematics from new and ever challenging perspectives, and have immeasurably aided in my mathematical growth. I thank Professor Avanti Athreya for her helpful feedback on this dissertation. I express gratitude for the support of the faculty and staff in the department of Applied Mathematics and Statistics at Johns Hopkins; of note, Professors Fred Torcaso, Daniel Naiman, Carey Priebe, and Donniell Fishkind have mentored me throughout my graduate studies. I thank Professor Mark Brown of City College of New York for his helpful discussions which helped inspire the first part of this dissertation, and also express gratitude to an anonymous referee whose comments helped refine the first part of this dissertation greatly. I acknowledge the generous financial support I received while doing my research, both from the Acheson J. Duncan Fund for the Advancement of Research in Statistics and from the U.S. Department of Education GAANN grant P200A090128. I very much thank my family for their support and encouragement throughout this process, and I especially thank my wife Anela who kept me grounded, sane, and well-fed throughout. Finally, it is with greatest humility that I sub-

mit this dissertation and acknowledge God's grace and blessing throughout, without which none of this would be possible.

Contents

Chapter 1

Introduction and Summary

In Chapter 2, we develop a systematic matrix-analytic approach, based on intertwinings of Markov semigroups, for proving theorems about hitting-time distributions for finite-state Markov chains—an approach that (sometimes) deepens understanding of the theorems by providing corresponding sample-path-by-sample-path stochastic constructions. In Sections 2.1.1–2.1.3 we describe a systematic approach, using intertwinings of Markov semigroups, for obtaining simple stochastic decompositions of the distributions of hitting times for Markov chains and also providing sample-path-by-sample-path constructions for the individual components in these decompositions.

Our approach is essentially matrix-analytic, but if certain conditions elaborated in Sections 2.1.1–2.1.2 are met, then our method also yields a decomposition for each sample path. For the applications discussed in this dissertation, our approach provides new matrix-analytic proofs for hitting-time results which were previously only known via analytic

methods (such as computation of Laplace transforms), and these new proofs provide new insights into the evolution of the Markov chain. A simple example of our approach, with an application to the Moran model in population genetics, is presented in Section 2.2.

We then employ our intertwinings approach to provide new proofs for two theorems due to Mark Brown, providing two quite different representations of hitting-time distributions for Markov chains started in stationarity. The proof, and subsequent construction, for the first theorem (Section 2.3) will elucidate an interesting connection between hitting-time distributions and the interlacing eigenvalues theorem for bordered symmetric matrices. Application of our approach obtains a construction for the second theorem (Section 2.4) that results in a bonus: We are able to extend Brown's theorem from reversible chains to more general ones. The material in Chapter 2 has been published in [24], and is reproduced here with Springer's permission.

In Chapter 3, we systematically develop the theory of strong stationary duality for diffusion processes on finite intervals. The theory of strong stationary duality was first developed in the setting of discrete-state Markov chains in [12] and [19]. In the Markov chain setting, strong stationary duality gives that the separation mixing time in the primal chain is equal in law to a suitable absorption time in the dual chain. By studying and bounding the absorption time, which is sometimes more tractable than direct consideration of the mixing time, we can tightly bound the separation mixing time in our primal chain. See [12] for further detail and background. This mixing-time/hitting-time duality played a leading role in the development of such diverse techniques as perfect sampling of Markov chains

(see [20], [23]) to characterizing separation cut-offs in birth and death chains (see [14]), just to name a few.

In Section 3.2.1, we define the strong stationary dual in the diffusion setting and in Section 3.2.2 we analytically derive the form of the dual diffusion's generator; in the process we explicitly derive the boundary behavior of the dual diffusion. In Section 3.3, we show that a suitably defined sequence of Markov chains and their strong stationary duals converge respectively to our primal diffusion and its strong stationary dual. By establishing the newly defined strong stationary dual diffusion as a limit of a sequence of Markov chain strong stationary duals, we ground our definition and our present work in the classical theory and allow for simultaneous discrete approximations of our primal and dual diffusions using the appropriately defined Markov chains. In Section 3.4, we recover a central tenet of the classical theory by proving that the separation mixing time in the primal diffusion is equal in law to the absorption time in the dual diffusion. We exploit this connection in Section 3.5 to derive the analogue to the birth-and-death cut-off phenomenon theory of [14] in the present diffusion setting.

Chapter 2

Hitting Times and Interlacing Eigenvalues: A Stochastic Approach Using Intertwinings

2.1 Introduction and Outline of our General Technique

Recently, stochastic proofs and constructions have been provided for some theorems that give explicit descriptions of Markov chain hitting-time distributions; previously known proofs of the theorems had been analytic in nature. Specifically, Fill [22] and Diaconis and Miclo [13] both give stochastic constructions for a famous birth-and-death hitting-time re-

sult first proven analytically by Karlin and McGregor [28] in 1959. Fill [21] (see also Miclo [39]) extends to upward-skip-free and more general chains, in particular giving a (sometimes) stochastic proof for a hitting-time theorem for upward-skip-free chains established analytically by Brown and Shao [10].

In Sections 2.1.1–2.1.3 we describe a systematic approach, using intertwinings of Markov semigroups, for obtaining simple stochastic decompositions of the distributions of hitting times for Markov chains and also providing sample-path-by-sample-path constructions for the individual components in these decompositions. For example, if one can prove a theorem that the law of a certain Markov chain hitting time T is a convolution of Geometric distributions with certain parameters, our additional goal is to decompose T explicitly—sample path by sample path—as a sum of independent Geometric random variables with the specified parameters; this deepens understanding as to "why" the theorem is true. See Fill [21] for a class of examples using this approach. Our approach is essentially matrix-analytic, but if certain conditions elaborated in Sections 2.1.1–2.1.2 are met, then our method also yields a decomposition for each sample path. For the applications discussed in this chapter, our approach provides new matrix-analytic proofs for hitting-time results which were previously only known via analytic methods (such as computation of Laplace transforms), and these new proofs provide new insights into the evolution of the Markov chain. A simple example of our approach, with an application to the Moran model in population genetics, is presented in Section 2.2.

We then employ our intertwinings approach to provide new proofs for two theorems

due to Mark Brown, providing two quite different representations of hitting-time distributions for Markov chains started in stationarity. The proof, and subsequent construction, for the first theorem (Section 2.3) will elucidate an interesting connection between hitting-time distributions and the interlacing eigenvalues theorem for bordered symmetric matrices. Application of our approach obtains a construction for the second theorem (Section 2.4) that results in a bonus: We are able to extend Brown's theorem from reversible chains to more general ones.

Notation: Throughout this chapter, all vectors used are by default row vectors. We write δ_j for the vector of 0's except for a 1 in the jth position, and $\vec{1}$ for the vector of 1's. The transpose of a matrix A is denoted by A^T. The notation $A(:, j) := A\delta_j^T$ is used to denote the jth column of A, and $A(i, :) := \delta_i A$ to denote the ith row of A. For any matrix A, we let A_0 denote the principal submatrix of A obtained by deleting the topmost row and leftmost column.

2.1.1 Intertwinings and sample-path linking

The main conceptual tool in our approach is the notion of an intertwining of Markov semigroups, for which we now provide the needed background in the context (sufficient for our purposes) of finite-state Markov chains. For further background on intertwinings, see [6], [11], [41]. Suppose that we have two state spaces, the first ("primary") of size n and the second ("dual") of size $\hat{n}$. Let P be the transition matrix of a Markov chain X, begun in distribution π_0, on the primary state space. [We write $X \sim (\pi_0, P)$ as shorthand.]

Similarly, let $\widehat{P}$ be the transition matrix of a Markov chain $\widehat{X}$, begun in $\hat{\pi}_0$, on the dual state space. Let Λ be an $\hat{n}$-by-n stochastic matrix.

Definition 2.1.1. We say that the Markov semigroups $(P^t)_{t=0,1,2,\ldots}$ and $(\widehat{P}^t)_{t=0,1,2,\ldots}$ are *intertwined by the link* Λ (or, for short, that P and $\widehat{P}$ are intertwined by the link Λ) if

$$\Lambda P = \widehat{P}\Lambda;$$

and we say that (π_0, P) and $(\hat{\pi}_0, \widehat{P})$ are *intertwined by* Λ if additionally

$$\pi_0 = \hat{\pi}_0\Lambda.$$

Here are three consequences when (π_0, P) and $(\hat{\pi}_0, \widehat{P})$ are intertwined by Λ (with the first two immediate—for example, $\Lambda P^2 = \widehat{P}\Lambda P = \widehat{P}^2\Lambda$—and the third *crucial* for our purposes):

- For $t = 0, 1, 2, \ldots$, we have $\Lambda P^t = \widehat{P}^t\Lambda$.
- For $t = 0, 1, 2, \ldots$, the distributions π_t and $\hat{\pi}_t$ at time t satisfy $\pi_t = \hat{\pi}_t\Lambda$.
- Given $X \sim (\pi_0, P)$, one can build $\widehat{X}_t$ from $X_0, \ldots, X_t$ and randomness independent of X so that $\widehat{X} \sim (\hat{\pi}_0, \widehat{P})$ and the conditional law of X_t given $(\widehat{X}_0, \ldots, \widehat{X}_t)$ has probability mass function given by the $\widehat{X}_t$-row of Λ:

$$\mathcal{L}(X_t \,|\, \widehat{X}_0, \ldots, \widehat{X}_t) = \Lambda(\widehat{X}_t, \cdot), \quad t = 0, 1, 2, \ldots. \tag{2.1.1}$$

We call this last consequence *sample-path linking*, and will explain next, once and for all, (a) how it is done and (b) why it is useful for hitting-time (or mixing-time) constructions.

We will then have no need to repeat this discussion when we turn to applications, each of which will therefore culminate with the explicit construction of an intertwining (or at least of a quasi-intertwining, as discussed in Section 2.1.3).

Whenever we have an intertwining of (π_0, P) and $(\hat{\pi}_0, \widehat{P})$, Section 2.4 of the strong stationary duality paper [12] by Diaconis and Fill gives a family of ways to create sample-path linking. Here is one [12, eq. (2.36)], with $\Delta := \widehat{P}\Lambda = \Lambda P$:

- Set $\widehat{X}_0 \leftarrow \hat{x}_0$ with probability $\hat{\pi}_0(\hat{x}_0)\Lambda(\hat{x}_0, x_0)/\pi_0(x_0)$.
- Inductively, for $t \geq 1$, set $\widehat{X}_t \leftarrow \hat{x}_t$ with probability

$$\widehat{P}(\hat{x}_{t-1}, \hat{x}_t)\Lambda(\hat{x}_t, x_t)/\Delta(\hat{x}_{t-1}, x_t).$$

Suppose (π_0, P) and $(\hat{\pi}_0, \widehat{P})$ are intertwined and that, given $X \sim (\pi_0, P)$, we have created linked sample paths for $\widehat{X} \sim (\hat{\pi}_0, \widehat{P})$, as at (2.1.1). Suppose further that there are states, call them 0 and $\hat{0}$, such that 0 (respectively, $\hat{0}$) is the unique absorbing state for P (resp., $\widehat{P}$) and that

$$\Lambda\delta_0^T = \delta_{\hat{0}}^T, \tag{2.1.2}$$

i.e., that $\Lambda(\hat{0}, 0) = 1$ and $\Lambda(\hat{x}, 0) = 0$ for $\hat{x} \neq \hat{0}$. Then, for the bivariate process $(\widehat{X}, X)$, we see that absorption times agree: $T_0(X) = T_{\hat{0}}(\widehat{X})$. For a parallel explanation of how sample-path linking can be used to connect the *mixing* time for an ergodic primary chain with a hitting time for a dual chain, consult [12]; very closely related is the FMMR perfect sampling algorithm [20, 23].

2.1.2 Strategy for absorption-time decompositions

The two hitting-time theorems discussed in Sections 2.3–2.4 both concern ergodic Markov chains. However, since for these theorems we have no interest in the chain after the specified target state 0 has been hit, the hitting-time distribution for such a chain is the same as the absorption-time distribution for the corresponding chain for which the target state is converted to absorbing by replacing the row of P corresponding to state 0 by the row vector δ_0.

It should also be noted that hitting-time theorems and stochastic constructions are easily extended to hitting times of general subsets A, by the standard trick of collapsing A to a single state.

Here is then a general strategy for obtaining a decomposition of the time to absorption in state 0 of a Markov chain $X \sim (\pi_0, P)$ from a decomposition of its distribution:

1. Discover another chain $\widehat{X} \sim (\hat{\pi}_0, \widehat{P})$ for which the sample-point-wise decomposition of the time to absorption in state $\hat{0}$ is *clearly* of the form specified for X. (For example, for a pure-death chain started at d with absorbing state $\hat{0} = 0$, the time to absorption is clearly the sum of independent Geometric random variables.)

2. Find a link Λ that intertwines (π_0, P) and $(\hat{\pi}_0, \widehat{P})$.

3. Prove the condition (2.1.2).

4. Conclude from the preceding discussion that (after sample-path linking) $T_0(X) = T_{\hat{0}}(\widehat{X})$ and use the sample-point-wise decomposition for $T_{\hat{0}}(\widehat{X})$ as the decomposition

for $T_0(X)$.

An early use of our strategy (adapted for mixing times, rather than absorption times) was in connection with the theory of strong stationary duality [12], for which the fullest development has resulted in the case of set-valued strong stationary duality (see especially [12, Secs. 3–4] and [20]; very closely related is the technique of *evolving sets* [40]). For a very recent application to hitting times and fastest strong stationary times for birth and death chains, see [22] and [21].

2.1.3 Quasi-intertwinings

Suppose that the (algebraic) intertwining conditions $\Lambda P = \widehat{P}\Lambda$ and $\pi_0 = \hat{\pi}_0\Lambda$ hold for some *not necessarily stochastic* matrix Λ with rows summing to unity. We call this a *quasi-intertwining* of (π_0, P) and $(\hat{\pi}_0, \widehat{P})$ by the *quasi-link* Λ. Then we again have the identities $\Lambda P^t = \widehat{P}^t\Lambda$ and $\pi_t = \hat{\pi}_t\Lambda$. As before, suppose further that (2.1.2) holds. Then, although (if Λ is not stochastic) we cannot do sample-path linking and so cannot achieve $T_0(X) = T_{\hat{0}}(\widehat{X})$, we can still conclude that $T_0(X)$ and $T_{\hat{0}}(\widehat{X})$ have the same distribution, because

$$\mathbf{P}(T_0(X) \le t) = \pi_t(0) = \textstyle\sum_{\hat{x}} \hat{\pi}_t(\hat{x})\Lambda(\hat{x}, 0) = \hat{\pi}_t(\hat{0}) = \mathbf{P}(T_{\hat{0}}(\widehat{X}) \le t).$$

Remark 2.1.2. The following easily-verified observations will be used in our application in Section 2.3.

(a) If Λ_1 is a quasi-link providing a quasi-intertwining of (π_0, P) and (π_0^*, P^*) and Λ_2 is similarly a quasi-link from (π_0^*, P^*) to $(\hat{\pi}_0, \widehat{P})$, then $\Lambda := \Lambda_2\Lambda_1$ is a quasi-link from (π_0, P) to $(\hat{\pi}_0, \widehat{P})$.

(b) If, additionally, the chains have respective unique absorbing states $0, 0^*, \hat{0}$ and (2.1.2) holds for Λ_1 and for Λ_2 (i.e., $\Lambda_1\delta_0^T = \delta_{0^*}^T$ and $\Lambda_2\delta_{0^*}^T = \delta_{\hat{0}}^T$), then (2.1.2) holds also for Λ (i.e., $\Lambda\delta_0^T = \delta_{\hat{0}}^T$).

(c) If Λ_1 and Λ_2 in (a) are both links, then so is Λ.

2.2 An illustrative example: Block chains and the Moran model

2.2.1 Block chains

In this section we warm up to the main applications of Sections 2.3–2.4 by providing a simple application of the technique outlined in Section 2.1. Let P be a Markov kernel on finite state space $\mathcal{X}$ with the following block structure:

$$P = \begin{pmatrix} P_{00} & P_{01} & P_{02} & \dots & P_{0k} \\ P_{10} & P_{11} & P_{12} & \dots & P_{1k} \\ P_{20} & P_{21} & P_{22} & \dots & P_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_{k0} & P_{k1} & P_{k2} & \dots & P_{kk} \end{pmatrix}. \tag{2.2.1}$$

For $i = 0, \ldots, k$, let μ_i be a Perron left eigenvector of P_{ii} [that is, a nonzero row vector with nonnegative entries such that

$$\mu_i P_{ii} = \rho(P_{ii})\mu_i,$$

where $\rho(A)$ denotes the spectral radius of a matrix A], normalized to sum to 1. It is well known (e.g., [26, Theorem 8.3.1]) that such an eigenvector exists; when, additionally, P_{ii} is irreducible, the vector μ_i is unique (e.g., [26, Theorem 8.4.4]) and is often called the quasi-stationary distribution for P_{ii}. We make the following special assumption concerning P: For every i and j, the vector $\mu_i P_{ij}$ is proportional to μ_j, say $\mu_i P_{ij} = \widehat{P}(i,j)\mu_j$. In words, the chain with transition matrix P, started in distribution μ_i over block i, moves in one step to block j with probability $\widehat{P}(i,j)$; and, conditionally given that it moves to block j, it "lands" in block j with distribution μ_j. We note in passing that $\widehat{P}$ is a $(k+1)$-by-$(k+1)$ matrix, and that $\widehat{P}(i,i) = \rho(P_{ii})$ for every i. Define a $(k+1)$-by-$|\mathcal{X}|$ stochastic matrix Λ by setting

$$\Lambda := \begin{pmatrix} \mu_0 & 0 & 0 & \ldots & 0 \\ 0 & \mu_1 & 0 & \ldots & 0 \\ 0 & 0 & \mu_2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \mu_k \end{pmatrix}. \tag{2.2.2}$$

Now consider a chain X with transition matrix P and initial distribution π_0; suppose that π_0 is a mixture, say $\sum_{i=0}^{k} \hat{\pi}_0(i)\mu_i$, of the distributions μ_i (each of which can be regarded naturally as a distribution on the entire state space).

Proposition 2.2.1. *In the block-chain setting described above,* (π_0, P) *and* $(\hat{\pi}_0, \widehat{P})$ *are intertwined by the link* Λ*.*

Proof. The proof is a simple matter of checking Definition 2.1.1 by checking that the identity $\mu_i P_{ij} \equiv \widehat{P}(i,j)\mu_j$ gives $\Lambda P = \widehat{P}\Lambda$ and that the assumption $\pi_0 = \sum_{i=0}^{k} \hat{\pi}_0(i)\mu_i$ gives $\pi_0 = \hat{\pi}_0\Lambda$. □

The sample-path linking developed in Section 2.1.1 is very simple to describe in our present block-chain setting: $\widehat{X}_t$ is simply the block ($\in \{0, \ldots, k\}$) to which X_t belongs. This simple description is due to the very simple nature of the link (2.2.2); the sample-path linking is more complicated for the applications in Sections 2.3–2.4.

2.2.2 The Moran model

We now apply the block-chain development in the preceding subsection to a Markov chain on partitions of the positive integer n introduced in [38] as a somewhat light-hearted model for collaboration among mathematicians. Their model is precisely the Moran model from population genetics according to the following definition [17, Definition 2.26] modified (a) to switch in natural fashion from continuous time to discrete time and (b) to limit the description of the state at each unit of time by distinguishing between genes with different labels but otherwise ignoring the values of the labels:

> A population of N genes evolves according to the Moran model if at exponential rate $\binom{N}{2}$ a pair of genes is sampled uniformly at random from the population, one dies and the other splits in two.

The chain we will consider here is a simple example of a coalescent chain, a class popularized in the seminal works of Kingman (see for example [31], [32], [33]). For a more complete modern picture of the application and study of coalescing chains, see [15].

Let S be a set of n indistinguishable objects. (The objects are gene labels in the Moran model and are mathematicians in [38].) The Markov chain of interest in [38] is more easily described if we make use of the natural bijection between partitions of the integer n and set partitions of S obtained by identifying a partition $(n_1, n_2, \ldots, n_r)$ (with $1 \leq r < \infty$ and $n_1 \geq n_2 \geq \cdots \geq n_r \geq 1$) of the integer n with a partition of S into r indistinguishable subsets where the subsets are of sizes $n_1, n_2, \ldots, n_r$. Accordingly, if the present state of the Markov chain is the partition $(n_1, n_2, \ldots, n_r)$, then, viewing this as a partition of S, uniformly select an ordered pair of unequal objects from S, and suppose that the first and second objects are currently in subsets of size n_i and n_j, respectively. The transition is realized by moving the second object from the second subset to the first, resulting in two new subsets of sizes $n_i + 1$ and $n_j - 1$. For example, if $n = 6$ and the Markov chain is currently in the partition $(4, 1, 1)$, then with probability $8/30$ the chain transitions to $(5, 1)$; with probability $2/30$, to $(4, 2)$; with probability $8/30$, to $(3, 2, 1)$; and with probability $12/30$ the chain stays in $(4, 1, 1)$. The authors of [38] are concerned with the distribution of the hitting time of state (n), the (absorbing) single-part partition, when the chain is begun in the n-parts partition $(1, \ldots, 1)$.

Collecting partitions into blocks, where block i contains all partitions with i parts ($1 \leq i \leq n$), it is clear that the transition matrix P for this chain is block upper bidiagonal, since

a one-step transition can only change the number of parts by 0 or -1. For example, in the simple case $n = 4$, one possible ordering of the partitions by decreasing number of parts is $(1,1,1,1)$, $(2,1,1)$, $(2,2)$, $(3,1)$, (4) and the corresponding P is given by

$$P = \begin{pmatrix} P_{44} & P_{43} & 0 & 0 \\ 0 & P_{33} & P_{32} & 0 \\ 0 & 0 & P_{22} & P_{21} \\ 0 & 0 & 0 & P_{11} \end{pmatrix} = \left(\begin{array}{c|c|cc|c} 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 6/12 & 2/12 & 4/12 & 0 \\ \hline 0 & 0 & 4/12 & 8/12 & 0 \\ 0 & 0 & 3/12 & 6/12 & 3/12 \\ \hline 0 & 0 & 0 & 0 & 1 \end{array}\right).$$

We will make use of results in [38] to see that P satisfies the assumptions of Section 2.2.1. To describe the results, let $1 \leq t \leq n$ and consider a partition $\mathbf{r}$ of n with t parts. For $i = 1, \ldots, n$, let r_i be the number of parts of $\mathbf{r}$ equal to i, so that $\sum_i i r_i = n$. Let $m_{\mathbf{r}} := \binom{t}{r_1, r_2, \ldots, r_n}$. Define μ_t to be the row vector, supported on partitions of size t, whose entry corresponding to partition $\mathbf{r}$ is $\binom{n-1}{t-1}^{-1} m_{\mathbf{r}}$. For $1 \leq t \leq n$, define $\lambda_t := 1 - \frac{t(t-1)}{n(n-1)}$. For example, if $n = 4$ and $t = 2$ and partitions with 2 parts are listed (as above) in the order $(2,2)$, $(3,1)$, then $\mu_2 = (1/3, 2/3)$ and $\lambda_2 = 5/6$. Let the dual state space be ordered $n, n-1, \ldots, 1$ (corresponding naturally to the ordering we have used for the primary state space). Define Λ by (2.2.2), but with the nonzero blocks correspondingly in *decreasing*

order $\mu_n, \mu_{n-1}, \ldots, \mu_1$ of subscript. Let

$$\widehat{P} := \begin{pmatrix} \lambda_n & 1-\lambda_n & 0 & \cdots & 0 \\ 0 & \lambda_{n-1} & 1-\lambda_{n-1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_1 \end{pmatrix}.$$

From Theorems 2 and 4 of [38] we can use our Proposition 2.2.1 to derive easily the following intertwining result.

Proposition 2.2.2. *Let π_0 be unit mass at the partition $(1, \ldots, 1)$. Then (π_0, P) and $(\delta_n, \widehat{P})$ are intertwined by the link Λ.*

As a direct consequence of Proposition 2.2.2, we get the following hitting-time result.

Corollary 2.2.3. *For fixed n, the law of the time to absorption in state (n) for the partitions-chain started in $(1, \ldots, 1)$ is that of $\sum_{t=2}^{n} Y_{n,t}$ where $Y_{n,t} \sim \mathrm{Geo}(1-\lambda_{n,t})$, with $\lambda_{n,t} = 1 - \frac{t(t-1)}{n(n-1)}$, are independent.*

In [38], the authors were able to identify a simple expression for the expected hitting time of state (n) when the chain is started in $\pi_0 = \delta_{(1,\ldots,1)}$, and challenged the reader to discover a pattern for the associated variance. The authors found that $\mathbf{E}_{\pi_0} T_{(n)} = (n-1)^2$. This is confirmed by our Corollary 2.2.3, as

$$\mathbf{E}_{\pi_0} T_{(n)} = \mathbf{E} \sum_{k=2}^{n} Y_{n,k} = \sum_{k=2}^{n} \frac{n(n-1)}{k(k-1)} = (n-1)^2.$$

Similarly, letting $H_n^{(2)} := \sum_{j=1}^n j^{-2}$ denote the nth second-order harmonic number, we find

$$\begin{aligned}\mathbf{Var}_{\pi_0} T_{(n)} = \mathbf{Var} \sum_{k=2}^n Y_{n,k} &= \sum_{k=2}^n \left(\left[\frac{n(n-1)}{k(k-1)} \right]^2 - \frac{n(n-1)}{k(k-1)} \right) \\ &= 2[n(n-1)]^2 H_n^{(2)} - (n-1)^2(3n^2 - 2n + 2) \\ &\sim (\tfrac{\pi^2}{3} - 3)n^4 \text{ as } n \to \infty.\end{aligned}$$

Proceeding further, it is not difficult to show that, when the partition chain is started in π_0, we have

$$\frac{T_{(n)}}{n^2} \xrightarrow{\mathcal{L}} S_\infty := \sum_{j=2}^\infty X_j$$

for independent random variables

$$X_j \sim \text{Exp}(j(j-1)), \quad j = 2, 3, \ldots,$$

with convergence of moments of all orders and (pointwise) of moment generating functions. We omit the details.

2.3 Hitting times and interlacing eigenvalues

2.3.1 Brown's theorem

Our next construction will provide insight into a hitting-time result of Mark Brown [9] that elegantly connects the hitting time of a state for a reversible Markov chain started in stationarity to the celebrated interlacing eigenvalues theorem of linear algebra (see, e.g., Theorem 4.3.8 in [26]). We now proceed to set up Brown's result.

Let $(X_t)_{t=0,1,2,\ldots}$ be a time-reversible ergodic discrete-time Markov chain with transition matrix P on finite state space $\mathcal{X} = \{0, 1, \ldots, n\}$ with stationary distribution π. If we let $D := \text{diag}(\pi(0), \ldots, \pi(n))$, then reversibility of P implies that $S := D^{1/2}PD^{-1/2}$ is a symmetric matrix and thus P has a real spectrum and a basis of real eigenvectors. Denote the eigenvalues of P by $1 = \theta_0 > \theta_1 \geq \cdots \geq \theta_n > -1$.

Recall that, for any matrix A, the principal submatrix of A obtained by deleting row 0 and column 0 is denoted A_0. Denote the eigenvalues of P_0 by $\eta_1 \geq \cdots \geq \eta_n$. Note that $S_0 = D_0^{1/2}P_0D_0^{-1/2}$ is symmetric; by the interlacing eigenvalues theorem for bordered symmetric matrices (e.g., [26, Theorem 4.3.8]), the eigenvalues of P and P_0 interlace: $\theta_0 > \eta_1 \geq \theta_1 \geq \cdots \geq \eta_n \geq \theta_n$. Cancel out common pairs of eigenvalues from the spectra $\sigma(P)$ and $\sigma(P_0)$ as follows. Consider $\sigma(P)$ and $\sigma(P_0)$ as multisets and remove the multiset $\sigma(P) \cap \sigma(P_0)$ from each of $\sigma(P)$ and $\sigma(P_0)$. Relabel the reduced set of eigenvalues of P as $\{\lambda_i\}_{i=0}^r$ with $\lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_r$ and of P_0 as $\{\gamma_i\}_{i=1}^r$ with $\gamma_1 \geq \cdots \geq \gamma_r$. After this cancellation, it is clear that the remaining eigenvalues strictly interlace: $1 = \lambda_0 > \gamma_1 > \lambda_1 > \cdots > \gamma_r > \lambda_r > -1$.

In what follows we need to assume that $\lambda_r \geq 0$. This is a rather harmless assumption, since we can if necessary shift attention from P to $\frac{1}{1+c}(P + cI)$ for suitably large c.

Brown found it convenient to work in continuous time, but he could just as easily have proven the analogous result in our present discrete-time setting. To state Brown's original continuous-time result, we make use of a very standard technique to produce a continuous-time chain from a discrete-time chain, by using independent and identically distributed (iid)

Exp(1) holding times (in place of unit times) between transitions. This continuous-time chain is sometimes called the *continuization* of the Markov chain with one-step transition matrix P, and it has generator matrix $Q = P - I$.

Brown's original result can be stated as follows.

Theorem 2.3.1. *Let $Q = P - I$ be the generator of the continuization of a Markov chain with one-step transition matrix P. In the continuized chain, the distribution (or law) $\mathcal{L}_\pi T_0$ of the hitting time of state 0 when the chain is started in stationarity, is that of $\sum_{i=1}^{r} Y_i$, where $Y_1, Y_2, \ldots, Y_r$ are independent and the distribution of Y_i is the "modified Exponential" mixture*

$$Y_i \sim \frac{1-\gamma_i}{1-\lambda_i}\delta_0 + \left(1 - \frac{1-\gamma_i}{1-\lambda_i}\right) \mathrm{Exp}(1-\gamma_i)$$

of unit mass at 0 and the Exponential distribution with parameter $1-\gamma_i$; the λ's and γ's are defined as above.

We find it more convenient to work in discrete time, where the corresponding theorem (involving Geometric, rather than Exponential, distributions) is as follows.

Theorem 2.3.2. *In the discrete-time setting outlined above, $\mathcal{L}_\pi T_0$ is the distribution of $\sum_{i=1}^{r} Y_i$, where $Y_1, Y_2, \ldots, Y_r$ are independent with the following "modified Geometric" distributions:*

$$Y_i \sim \frac{1-\gamma_i}{1-\lambda_i}\delta_0 + \left(1 - \frac{1-\gamma_i}{1-\lambda_i}\right) \mathrm{Geo}(1-\gamma_i). \tag{2.3.1}$$

We have our choice of working in discrete or continuous time because, fortunately, for any finite-state Markov chain and any target state 0 there is a simple relationship between

hitting-time distributions in the two cases. Let T_0^{d} be the time to hit state 0 in the discrete-time chain $(X_t)_{t=0,1,2,\ldots}$ with transition matrix P, and let T_0^{c} be the corresponding hitting time in the continuized chain. Then the Laplace transform $\psi_{T_0^{\mathrm{c}}}(s) := \mathbf{E}\exp(-s\,T_0^{\mathrm{c}})$ and the probability generating function $G_{T_0^{\mathrm{d}}}(z) := \mathbf{E}\,z^{T_0^{\mathrm{d}}}$ of the hitting times satisfy a simple relationship:

Lemma 2.3.3. *For any finite-state discrete-time Markov chain and any target state* 0*, we have the following identity relating the distributions of the hitting time of state* 0 *for the continued chain and the discrete-time chain:*

$$\psi_{T_0^{\mathrm{c}}}(s) = G_{T_0^{\mathrm{d}}}\left(\frac{1}{1+s}\right),\ s \geq 0.$$

Proof. Let $X_i \sim \mathrm{Exp}(1)$ be iid and independent of T_0^{d}. By definition of the continuized chain, we have $T_0^{\mathrm{c}} \stackrel{\mathcal{L}}{=} \sum_{i=1}^{T_0^{\mathrm{d}}} X_i$. Then

$$\psi_{T_0^{\mathrm{c}}}(s) = \mathbf{E}\exp(-sT_0^{\mathrm{c}}) = \mathbf{E}\exp\left(-s\sum_{i=1}^{T_0^{\mathrm{d}}} X_i\right) = \mathbf{E}\left(\frac{1}{1+s}\right)^{T_0^{\mathrm{d}}} = G_{T_0^{\mathrm{d}}}\left(\frac{1}{1+s}\right).$$

□

This lemma allows us to easily derive Theorem 2.3.1 from Theorem 2.3.2 (and vice versa), since for $s \geq 0$ we have

$$\begin{aligned}\psi_{T_0^{\mathrm{c}}}(s) = G_{T_0^{\mathrm{d}}}\left(\frac{1}{1+s}\right) &= \prod_i \left[\frac{1-\gamma_i}{1-\lambda_i} + \left(1 - \frac{1-\gamma_i}{1-\lambda_i}\right)\frac{\frac{1-\gamma_i}{1+s}}{1-\frac{\gamma_i}{1+s}}\right] \\ &= \prod_i \left[\frac{1-\gamma_i}{1-\lambda_i} + \left(1 - \frac{1-\gamma_i}{1-\lambda_i}\right)\frac{1-\gamma_i}{1-\gamma_i+s}\right].\end{aligned}$$

Our main result of Section 2.3 is another proof for Theorem 2.3.2, culminating in our Theorem 2.3.16 (see also the last paragraph of Section 2.3.4). Our proof provides—at least

when the quasi-link Λ we construct is a *bona fide* link—an explicit stochastic construction of the hitting time of state 0 from a stationary start as a sum of independent modified Geometric random variables. We tackle our proof of Theorem 2.3.2 in two stages: in Section 2.3.2 we build a certain "star chain" (random walk on a weighted star graph) from the given chain and prove Theorem 2.3.2 when this star chain is substituted for the given chain, and in Section 2.3.3 we attempt to "link" the given chain with the star chain of Section 2.3.2. In Section 2.3.4 we combine the results of Sections 2.3.2–2.3.3 and provide our complete proof of Theorem 2.3.2. We could equally well prove the continuous-time analogues of all of our theorems and then apply the analogous intertwining results outlined in Section 2.3 of [19] to provide (again when Λ is a link) an explicit continuous-time stochastic construction for Theorem 2.3.1. We choose to work in discrete time for convenience and because, we believe, the ideas behind our constructions are easier to grasp in discrete time.

2.3.2 A stochastic construction for the star chain

Carrying out step 1 of the four-step strategy outlined in Section 2.1.2 (finding a chain $\widehat{X}$ for which the hitting time of state $\hat{0}$ can be decomposed as a sum of independent modified Geometric random variables) turns out not to be too difficult; this step is carried out later, in Lemma 2.3.10. However, step 2 (finding a link Λ between the given X and $\widehat{X}$) proved challenging to us, so we break it down into two substeps, as described at the end of the preceding subsection. In this subsection we build an ergodic star chain X^* from the given

chain X and show that the Markov semigroups for X^* (with the target state 0^* converted to absorbing) and $\widehat{X}$ are intertwined by a link Λ_2. The state spaces for X^* and $\widehat{X}$ will both be $\{0, \ldots, r\}$, and the roles of $\hat{0}$ and 0^* will both be played by state 0. For the star chain, we make full use of the notation in Section 2.3.1. The "star" has "hub" at 0 and "spokes" terminating at vertices $1, \ldots, r$. The r-spoke star chain we build has previously been constructed in [2].

For the sake of brevity it is convenient to establish some additional notation. Define

$$\rho_i := \frac{1-\gamma_i}{1-\lambda_i} \quad \text{for } i = 1, \ldots, r,$$

and for $0 \le k \le r$ define

$$\pi_k^*(i) := \begin{cases} (1-\rho_i) \prod_{1 \le j \le k,\ j \ne i} \frac{1-\gamma_j - \rho_j(1-\gamma_i)}{\gamma_i - \gamma_j} & \text{for } i = 1, \ldots, k \\ \prod_{j=1}^k \rho_j & \text{for } i = 0. \end{cases} \tag{2.3.2}$$

Set $\pi_k^* := (\pi_k^*(0), \ldots, \pi_k^*(k), 0, \ldots, 0) \in \mathbb{R}^{r+1}$ and note that $\pi_0^* = \delta_0$. The following lemma lays out the ergodic star chain of interest corresponding to the given chain.

Lemma 2.3.4.

(a) *For all* $0 \le k \le r$ *we have* $\pi_k^*(i) > 0$ *for* $i = 0, \ldots, k$ *and* $\sum_{i=0}^k \pi_k^*(i) = 1$.

(b) *The row vector* $\pi^* := \pi_r^*$ *is the stationary distribution of the ergodic* r*-spoke star chain with transition matrix* P^* *satisfying, for* $i = 1, \ldots, r$,

$$P^*(i,0) = 1-\gamma_i \quad \textit{and} \quad P^*(i,i) = \gamma_i.$$
$$P^*(0,i) = \frac{(1-\gamma_i)\pi^*(i)}{\pi^*(0)} \quad \textit{and} \quad P^*(0,0) = 1 - \frac{1}{\pi^*(0)} \sum_{i=1}^r (1-\gamma_i)\pi^*(i).$$

Proof.

(a) Fix $k \in \{0, \ldots, r\}$. Clearly $\pi_k^*(0) > 0$, so we begin by showing that $\pi_k^*(i) > 0$ for $i = 1, \ldots, k$. Since $1 - \rho_i > 0$, we'll do this by showing that each factor in the product $\prod_{j \neq i}$ in (2.3.2) is strictly positive. Indeed, if $j > i$ this is clear because $0 < \rho_j < 1$. If $j < i$, then we use

$$\frac{1 - \gamma_j - \rho_j(1 - \gamma_i)}{\gamma_i - \gamma_j} = \frac{\rho_j(1 - \gamma_i) - (1 - \gamma_j)}{\gamma_j - \gamma_i} > \frac{\left(\frac{1-\gamma_j}{1-\gamma_i}\right)(1 - \gamma_i) - (1 - \gamma_j)}{\gamma_j - \gamma_i} = 0,$$

where the inequality holds because $\lambda_j > \gamma_i$ by the interlacing condition. To show $\sum_{i=1}^k \pi_k^*(i) = 1$, we repeat the argument in the proof of Lemma 2.1 in [9] and include it for completeness. Define

$$\psi(s) := \prod_{i=1}^{k} \frac{1 - \gamma_i + \rho_i s}{1 - \gamma_i + s}. \tag{2.3.3}$$

Then $\psi(0) = 1$, and we will show

$$\begin{aligned} \psi(s) &= \pi_k^*(0) + \sum_{i=1}^{k} \pi_k^*(i) \frac{1 - \gamma_i}{1 - \gamma_i + s} \text{ for general s} \\ &= \sum_{i=0}^{k} \pi_k^*(i) \text{ at } s = 0, \end{aligned} \tag{2.3.4}$$

which will complete the argument. To show (2.3.4), first set

$$f(s) := \prod_{j=1}^{k}(1 - \gamma_j + \rho_j s), \quad g(s) := \prod_{j=1}^{k}(1 - \gamma_j + s), \quad \tilde{f}(s) := f(s) - \left(\prod_{j=1}^{k} \rho_j\right) g(s).$$

Note that $\tilde{f}(s)$ is a polynomial of degree $\leq k - 1$ and that

$$\tilde{f}(-1 + \gamma_i) = f(-1 + \gamma_i), \quad i = 1, \ldots, k.$$

Define

$$h(s) := \sum_{i=1}^{k} \left(\pi_k^*(i)(1-\gamma_i) \prod_{j \neq i : 1 \leq j \leq k} (1 - \gamma_j + s) \right).$$

A brief calculation yields

$$\pi_k^*(i)(1-\gamma_i) = \frac{f(\gamma_i - 1)}{g'(\gamma_i - 1)},$$

and we see that

$$h(\gamma_i - 1) = f(\gamma_i - 1) = \tilde{f}(\gamma_i - 1), \quad i = 1, \ldots, k.$$

But $h(s)$, like $\tilde{f}(s)$, is a polynomial of degree $\leq k-1$, and so $h(s) = \tilde{f}(s)$ for all s. Finally, we see

$$\begin{aligned} \psi(s) = \frac{f(s)}{g(s)} &= \frac{1}{g(s)} \left[\left(\prod_{i=1}^{k} \rho_i \right) g(s) + \tilde{f}(s) \right] = \prod_{i=1}^{k} \rho_i + \frac{h(s)}{g(s)} \\ &= \pi_k^*(0) + \sum_{i=1}^{k} \pi_k^*(i) \frac{1-\gamma_i}{1-\gamma_i+s}, \end{aligned}$$

establishing (2.3.4) and completing the proof of part (a).

(b) Clearly, $P^* \vec{1}^T = \vec{1}^T$. To show that P^* is stochastic, we need only show that $P^* \geq 0$ entrywise. This is clear except perhaps for the entry $P^*(0,0)$. To see $P^*(0,0) > 0$, we first note that $P^*(0,0) = \operatorname{tr} P^* - \operatorname{tr} P_0^*$; Lemma 2.6 in [9] then gives $\operatorname{tr} P^* - \operatorname{tr} P_0^* = \sum_{i=0}^{r} \lambda_i - \sum_{i=1}^{r} \gamma_i = \sum_{i=0}^{r-1} (\lambda_i - \gamma_{i+1}) + \lambda_r > 0$. Part (a) establishes that $\pi^* = \pi_r^*$ is a distribution, and one sees immediately that π^* satisfies the detailed balance equations for the transition matrix P^*. □

Remark 2.3.5. It would seem natural to define a k-spokes star chain with transition matrix $P^{*(k)}$ and stationary distribution π_k^* for general k just as is done for $k = r$ in Lemma 2.3.4.

However, it is then not clear whether $P^{*(k)}(0,0) \geq 0$. Moreover, in our construction we use only the P^* of Lemma 2.3.4(b) (with $k = r$).

Define P^*_{abs} to be the chain $(X^*_t)_{t=0,1,\ldots}$ modified so that 0 is an absorbing state and note that

$$\sigma(P^*_{\text{abs}}) = \{1, \gamma_1, \ldots, \gamma_r\}.$$

We now begin to head towards Theorem 2.3.11, which will show that $\mathcal{L}_{\pi^*}(T^*_0) = \mathcal{L}(\sum_{i=1}^r Y_i)$ for the Y_i's described in Theorem 2.3.2. To do this, we will construct a link Λ_2 between the absorbing star chain and a dual chain $(\widehat{X}_t)_{t=0,1,\ldots}$ for which the hitting time for state 0 is explicitly given as an independent sum of the modified Geometric random variables Y_i.

Remark 2.3.6. If the given chain *is* already a star chain, then the star chain of Lemma 2.3.4 is simply obtained by collapsing all leaves with the same one-step transition probability to state 0 into a single leaf. This is established as Proposition **??** in the Appendix, where it is also shown that the stationary probabilities collapse accordingly. For example, suppose the given chain is the star chain with transition matrix

$$P = \begin{pmatrix} 4/9 & 1/9 & 1/9 & 1/9 & 1/9 & 1/9 \\ 1/6 & 5/6 & 0 & 0 & 0 & 0 \\ 1/6 & 0 & 5/6 & 0 & 0 & 0 \\ 2/9 & 0 & 0 & 7/9 & 0 & 0 \\ 1/3 & 0 & 0 & 0 & 2/3 & 0 \\ 1/3 & 0 & 0 & 0 & 0 & 2/3 \end{pmatrix}.$$

We see that $\pi = \frac{1}{21}(6, 4, 4, 3, 2, 2)$ and that

$$\sigma(P) = \{1,\ 5/6,\ 0.8023,\ 0.7303,\ 2/3,\ 0.1896\},\ \ \sigma(P_0) = \{5/6,\ 5/6,\ 7/9,\ 2/3,\ 2/3\}.$$

The reduced set of eigenvalues of P is $\{1,\ 0.8023,\ 0.7303,\ 0.1896\}$ and the reduced set of eigenvalues of P_0 is $\{5/6,\ 7/9,\ 2/3\}$. The star chain constructed in Lemma 2.3.4 has three spokes with probabilities $1/6,\ 2/9,\ 1/3$ of moving to the hub in one step and respective stationary probabilities $8/21,\ 3/21,\ 4/21$ (with stationary probability $6/21$ at the hub).

The key to our construction will be the following "spoke-breaking" theorem.

Theorem 2.3.7. *For each $i = 1, \ldots, r$, the distribution $\pi_i^* \in \mathbb{R}^{r+1}$ can be represented as the mixture*

$$\pi_i^* = \rho_i \pi_{i-1}^* + (1 - \rho_i)\nu_i \tag{2.3.5}$$

of π_{i-1}^* and a probability distribution ν_i (regarded as a row vector in $\mathbb{R}^{r+1}$) satisfying

$$\nu_i P^* = \gamma_i \nu_i + (1 - \gamma_i)\pi_{i-1}^*. \tag{2.3.6}$$

Proof. Fix i. Clearly there is a unique row vector $\nu \equiv \nu_i$ satisfying (2.3.5), and it sums to unity because π_i^* and π_{i-1}^* each do. We will solve for ν and see immediately that ν has nonnegative entries; indeed, we will show that ν is given by

$$\nu(j) = \begin{cases} \frac{1-\gamma_i}{1-\gamma_i-\rho_i(1-\gamma_j)}\pi_i^*(j) & \text{if } 1 \le j \le i \\ 0 & \text{if } j = 0 \text{ or } j > i. \end{cases} \tag{2.3.7}$$

It will then be necessary only to prove that ν satisfies (2.3.6).

We begin by establishing (2.3.7) for $j = 1, \ldots, i-1$. (For $t = i-1$ and $t = i$, the notation $\prod_{k\neq j}^{t}$ will be shorthand for the product over values k satisfying both $1 \leq k \leq t$ and $k \neq j$.) In that case,

$$\begin{aligned}
(1-\rho_i)\nu(j) &= \pi_i^*(j) - \rho_i \pi_{i-1}^*(j) \\
&= (1-\rho_j)\prod_{k\neq j}^{i} \frac{1-\gamma_k-\rho_k(1-\gamma_j)}{\gamma_j-\gamma_k} - \rho_i(1-\rho_j)\prod_{k\neq j}^{i-1}\frac{1-\gamma_k-\rho_k(1-\gamma_j)}{\gamma_j-\gamma_k} \\
&= \pi_i^*(j)\left[1-\rho_i\frac{\gamma_j-\gamma_i}{1-\gamma_i-\rho_i(1-\gamma_j)}\right] \\
&= (1-\rho_i)\frac{1-\gamma_i}{1-\gamma_i-\rho_i(1-\gamma_j)}\pi_i^*(j),
\end{aligned}$$

as desired, where the first equality follows from (2.3.5), and the second and third employ the formula (2.3.2) both for π_i^* and for π_{i-1}^*.

For $j = i$ we calculate

$$(1-\rho_i)\nu(i) = \pi_i^*(i) - \rho_i\pi_{i-1}^*(i) = \pi_i^*(i),$$

i.e.,

$$\nu(i) = (1-\rho_i)^{-1}\pi_i^*(i) = \frac{1-\gamma_i}{1-\gamma_i-\rho_i(1-\gamma_i)}\pi_i^*(i),$$

again as desired.

For $j = 0$, (2.3.2) gives that

$$(1-\rho_i)\nu(0) = \pi_i^*(0) - \rho_i\pi_{i-1}^*(0) = \prod_{k=1}^{i}\rho_k - \rho_i\prod_{k=1}^{i-1}\rho_k = 0,$$

once again as desired. For $j > i$, (2.3.7) is clear because $\pi_i^*(j) = 0 = \pi_{i-1}^*(j)$.

It remains to check that ν satisfies (2.3.6). Since both sides are vectors summing to 1 (on the left because ν is a probability distribution and P^* is a transition kernel, and on the

right because both ν and π^*_{i-1} are probability distributions), we need only check $\mathrm{LHS}(j) = \mathrm{RHS}(j)$ for $j \neq 0$ (henceforth assumed). We begin by calculating the state-j entry of the LHS assuming $j \leq i$:

$$\begin{aligned}\mathrm{LHS}(j) &= \sum_{k=0}^{r} \nu(k)P^*(k,j) = \sum_{k=1}^{i} \nu(k)P^*(k,j) \\ &= \nu(j)P^*(j,j) = \left[\frac{1-\gamma_i}{1-\gamma_i-\rho_i(1-\gamma_j)}\pi^*_i(j)\right] \times (\gamma_j).\end{aligned}$$

On the other hand, using (2.3.5) we calculate

$$\begin{aligned}\mathrm{RHS} &= \pi^*_{i-1} + \gamma_i\left(\nu - \pi^*_{i-1}\right) \\ &= \pi^*_{i-1} + \gamma_i\left[\nu - \rho_i^{-1}\left(\pi^*_i - (1-\rho_i)\nu\right)\right] \\ &= \rho_i^{-1}\left(\pi^*_i - (1-\rho_i)\nu\right) + \frac{\gamma_i}{\rho_i}\left(\nu - \pi^*_i\right).\end{aligned}$$

Therefore, for $j \leq i$ the jth entry of the RHS is

$$\begin{aligned}\mathrm{RHS}(j) &= \rho_i^{-1}\pi^*_i(j)\left[1 - \frac{(1-\rho_i)(1-\gamma_i)}{1-\gamma_i-\rho_i(1-\gamma_j)} + \frac{\gamma_i\rho_i(1-\gamma_j)}{1-\gamma_i-\rho_i(1-\gamma_j)}\right] \\ &= \rho_i^{-1}\pi^*_i(j)\left[\frac{(1-\gamma_i)\rho_i\gamma_j}{1-\gamma_i-\rho_i(1-\gamma_j)}\right] \\ &= \mathrm{LHS}(j).\end{aligned}$$

If $j > i$, then $\mathrm{LHS}(j) = 0 = \mathrm{RHS}(j)$, finishing the proof that ν satisfies (2.3.6). □

The preceding Theorem 2.3.7 suggests the form for the chain $(\widehat{X}_t)_{t=0,1,2,\ldots}$ on $\{0,1,\ldots,r\}$, where the times spent in state $j = 0,1,2,\ldots,r$ in this chain are independent and distributed as the Y_j's in Theorem 2.3.2. Before proceeding to the construction in Lemma 2.3.10, the next lemma provides some preliminaries.

Lemma 2.3.8. *Let $0 < k \leq r$. Let $\hat{\pi}_k(j) := \rho_k \rho_{k-1} \dots \rho_{j+1}(1-\rho_j)$ for all $1 \leq j < k$, and let $\hat{\pi}_k(k) := 1 - \rho_k$. Then $\hat{\pi}_k(j) \geq 0$ for $1 \leq j \leq k$, and $\sum_{j=1}^{k} \hat{\pi}_k(j) = 1 - \prod_{i=1}^{k} \rho_i$. If we define $\hat{\pi}_k(0) := \prod_{i=1}^{k} \rho_i$, then $\hat{\pi}_k$ gives a probability distribution on $0, 1, \dots, k$.*

The proof of this lemma is very easy. Let us also adopt the convention $\hat{\pi}_0 := \delta_0$.

Remark 2.3.9. Paralleling (2.3.5) in Theorem 2.3.7, we have

$$\hat{\pi}_k = \rho_k \hat{\pi}_{k-1} + (1-\rho_k)\delta_k \quad \text{for } 1 \leq k \leq r.$$

We are now ready to construct $(\widehat{X}_t)$:

Lemma 2.3.10. *Let $(\widehat{X}_t)$ be the absorbing Markov chain with state space $\{0, \dots, r\}$ begun in distribution $\hat{\pi} := \hat{\pi}_r$, with transition matrix $\widehat{P}$ defined by*

$$\widehat{P}(i,j) = \begin{cases} 1 & \text{if } 0 = j = i \\ \gamma_i & \text{if } 0 < j = i \\ (1-\gamma_i) \cdot \hat{\pi}_{i-1}(j) & \text{if } j < i \\ 0 & \text{if } j > i. \end{cases}$$

Then

(a) *If Z_i is the time spent in state i (including time 0) by $(\widehat{X}_t)$ with initial distribution $\hat{\pi}$ prior to hitting 0, then $\mathcal{L}(Z_1, Z_2, \dots, Z_r) = \mathcal{L}(Y_1, Y_2, \dots, Y_r)$.*

(b) *If $\widehat{T}_0$ is the hitting time of state 0 for the chain $(\widehat{X}_t)$ with initial distribution $\hat{\pi}$, then $\widehat{T}_0 \stackrel{\mathcal{L}}{=} \sum_{i=1}^{r} Y_i$.*

Proof. (a) When viewed in the right light, the lemma is evident. The chain moves downward through the state space $\{0, 1, \ldots, r\}$, with ultimate absorption in state 0, and can be constructed by performing a sequence of r independent Bernoulli trials $W_r, \ldots, W_1$ with varying success probabilities $1-\rho_r, \ldots, 1-\rho_1$, respectively. If $W_i = 0$, then the chain does not visit state i, whereas if $W_i = 1$ then the amount of time spent in state i is Geom$(1-\gamma_i)$ independent of the amounts of time spent in the other states.

A formal proof of part (a) is not difficult but would obscure this simple construction and is therefore not included.

(b) This is immediate from part (a), since $\widehat{T}_0 = \sum_{i=1}^{r} Z_i$. □

As the culmination of this subsection we exhibit an intertwining between $(\pi^*, P^*_{\text{abs}})$ and $(\hat{\pi}, \widehat{P})$.

Theorem 2.3.11. *Let Λ_2 be defined as follows:*

$$\Lambda_2(0,:) := \delta_0, \qquad \Lambda_2(i,:) := \nu_i \text{ for } i = 1, \ldots, r.$$

Then $(\pi^*, P^*_{\text{abs}})$ and $(\hat{\pi}, \widehat{P})$ are intertwined by the link Λ_2, which satisfies (2.1.2)*; to wit,*

$$\Lambda_2 P^*_{\text{abs}} = \widehat{P}\Lambda_2, \tag{2.3.8}$$

$$\pi^* = \hat{\pi}\Lambda_2, \tag{2.3.9}$$

$$\Lambda_2 \delta_0^T = \delta_0^T. \tag{2.3.10}$$

Proof. We begin by noting that Λ_2 is stochastic because, as noted in Theorem 2.3.7, each ν_i is a probability distribution.

From Theorem 2.3.7 we have that $\pi_k^* = \rho_k \pi_{k-1}^* + (1-\rho_k)\nu_k$ for $1 \le k \le r$, and from Remark 2.3.9 we have the corresponding equations for $\hat{\pi}_k$, namely, $\hat{\pi}_k = \rho_k \hat{\pi}_{k-1} + (1-\rho_k)\delta_k$ for all $1 \le k \le r$. One can use these results to prove $\pi_k^* = \hat{\pi}_k \Lambda_2$ for $k = 0, 1, \ldots, r$ by induction on k; in particular, (2.3.9) follows by setting $k = r$.

To show (2.3.8), first observe $(\Lambda_2 P_{\text{abs}}^*)(0,:) = \delta_0 = (\widehat{P}\Lambda_2)(0,:)$. Comparing ith rows for $1 \le i \le r$, we see

$$(\Lambda_2 P_{\text{abs}}^*)(i,:) = \nu_i P_{\text{abs}}^* = \gamma_i \nu_i + (1-\gamma_i)\pi_{i-1}^* \tag{2.3.11}$$

by (2.3.6) and the fact that $\nu_i(0) = 0$ for all i. Iterating Theorem 2.3.7, we see for $i = 1, \ldots, r$ that

$$\begin{aligned}
\pi_i^* &= (1-\rho_i)\nu_i + \rho_i \pi_{i-1}^* \\
&= (1-\rho_i)\nu_i + \rho_i \left[(1-\rho_{i-1})\nu_{i-1} + \rho_{i-1}\pi_{i-2}^*\right] \\
&= (1-\rho_i)\nu_i + \rho_i(1-\rho_{i-1})\nu_{i-1} + \rho_i\rho_{i-1}\pi_{i-2}^* \\
&= \cdots = \hat{\pi}_i(i)\nu_i + \hat{\pi}_i(i-1)\nu_{i-1} + \cdots + \hat{\pi}_i(1)\nu_1 + \hat{\pi}_i(0)\delta_0.
\end{aligned}$$

So $\pi_i^* = \sum_{j=1}^{i} \hat{\pi}_i(j)\nu_j + \hat{\pi}_i(0)\delta_0$ for $i = 1, \ldots, r$, and the same equation holds for $i = 0$ because $\pi_0^* = \delta_0 = \hat{\pi}_0$. Applying this to equation (2.3.11) we find for $i = 1, \ldots, r$ that

$$\begin{aligned}
(\Lambda_2 P_{\text{abs}}^*)(i,:) &= \gamma_i \nu_i + \sum_{j=1}^{i-1} (1-\gamma_i)\hat{\pi}_{i-1}(j)\nu_j + (1-\gamma_i)\hat{\pi}_{i-1}(0)\delta_0 \\
&= (\widehat{P}\Lambda_2)(i,:),
\end{aligned}$$

as desired, where at the last equality we have recalled $\Lambda_2(0,:) = \delta_0$.

Finally, (2.3.10) asserts that the 0th column of Λ_2 is δ_0^T. This follows from the definition of Λ_2, since it has already been noted at (2.3.7) that $\nu_i(0) = 0$ for $i = 1, \ldots, r$. □

2.3.3 Quasi-link to the star chain

The main result of this subsection is Theorem 2.3.13, which provides a quasi-link between the *absorbing* transition matrices P_{abs} and P^*_{abs} corresponding to the given chain and the star chain, respectively. We begin with a linear-algebraic lemma.

Lemma 2.3.12. *The matrix P_{abs} has $n+1$ linearly independent left eigenvectors. Its multiset of $n+1$ eigenvalues is $\{1, \eta_1, \ldots, \eta_n\}$.*

Proof. Recall that $\sigma(P_0) = \{\eta_1, \ldots, \eta_n\}$. Recall also that $D_0 = \text{diag}(\pi_1, \ldots, \pi_n)$ and that $S_0 = D_0^{1/2} P_0 D_0^{-1/2}$ is a symmetric matrix. Let $\widetilde{U}$ be an n-by-n orthogonal matrix whose rows are orthonormal left eigenvectors of S_0, so that $\widetilde{U} S_0 \widetilde{U}^T = \text{diag}(\eta_1, \eta_2, \ldots, \eta_n)$. Then the rows (denoted $u_1, \ldots, u_n$) of the n-by-n matrix $U := \widetilde{U} D_0^{1/2}$ are left eigenvectors of P_0 with respective eigenvalues $\eta_1, \ldots, \eta_n$. For $i = 1, \ldots, n$, define the scalar

$$w_i := \frac{(0|u_i)P(:,0)}{\eta_i - 1};$$

then $(w_i|u_i)P_{\text{abs}} = \eta_i(w_i|u_i)$ and $\eta_i \in \sigma(P_{\text{abs}})$. Finally, $\delta_0 P_{\text{abs}} = \delta_0$. The $n+1$ eigenvectors δ_0 and $(w_i|u_i)$ for $i = 1, \ldots, n$ are clearly linearly independent, and our proof is complete. □

Note that

$$(w_i|u_i)\vec{1}^T = (w_i|u_i)P_{\text{abs}}\vec{1}^T = \eta_i(w_i|u_i)\vec{1}^T$$

and $\eta_i < 1$, implying that $(w_i|u_i)\vec{1}^T = 0$ and $w_i = -u_i\vec{1}^T$.

Let n_i denote the algebraic (also geometric) multiplicity of the eigenvalue γ_i as an eigenvalue of P_0 (here we are working with the reduced set of eigenvalues again). Relabel the eigenvectors corresponding to γ_i by $u_1^i, \ldots, u_{n_i}^i$. Note that, when viewed as an eigenvalue of P_{abs}, γ_i has algebraic (also geometric) multiplicity n_i, with corresponding eigenvectors $(-u_1^i\vec{1}^T|u_1^i), \ldots, (-u_{n_i}^i\vec{1}^T|u_{n_i}^i)$. In the next theorem we construct our $(r+1)$-by-$(n+1)$ quasi-link Λ_1 between (π, P_{abs}) and $(\pi^*, P_{\text{abs}}^*)$.

Theorem 2.3.13. *There exists a quasi-link Λ_1 providing a quasi-intertwining between (π, P_{abs}) and $(\pi^*, P_{\text{abs}}^*)$ and satisfying* (2.1.2)*, i.e., a matrix Λ_1 with rows summing to* 1 *such that*

$$\pi = \pi^*\Lambda_1, \tag{2.3.12}$$

$$\Lambda_1 P_{\text{abs}} = P_{\text{abs}}^*\Lambda_1, \tag{2.3.13}$$

$$\Lambda_1\delta_0^T = \delta_0^T. \tag{2.3.14}$$

Proof. If row i of Λ_1 is denoted by x_i for $i = 0, ..., r$, then for (2.3.13) we require

$$x_0P_{\text{abs}} = x_0; \quad x_iP_{\text{abs}} = (1-\gamma_i)x_0 + \gamma_i x_i \quad i = 1, \ldots, r.$$

This forces $x_0 = \delta_0$ and

$$x_i(P_{\text{abs}} - \gamma_i I) = (1-\gamma_i)\delta_0, \quad i = 1, \ldots, r.$$

Therefore, for $\Lambda_1 P_{\text{abs}} = P_{\text{abs}}^*\Lambda_1$ to hold, we necessarily set

$$x_i = \delta_0 + \sum_{j=1}^{n_i} c_j^i\,(-u_j^i\vec{1}^T|u_j^i), \quad i = 1 \ldots, r$$

where the c_j^i's are soon-to-be-determined real constants.

For *any* choices of c_j^i's above we have that the rows of Λ_1 sum to unity and $\Lambda_1 P_{\text{abs}} = P_{\text{abs}}^* \Lambda_1$, but it remains to be shown that we can define c_j^i's so that (2.3.12) holds. The difficulty is that there may exist values $\eta_i \in \sigma(P_0)$ such that $\eta_i \neq \gamma_j$ for any $j = 1, \ldots, r$. However, we will show in the next lemma that π is in the span of the eigenvectors corresponding to the remaining eigenvalues, and that will complete our proof of (2.3.12).

To prove (2.3.14), we use (2.3.12)–(2.3.13) to get $\pi P_{\text{abs}}^t = \pi^* \Lambda_1 P_{\text{abs}}^t = \pi^* P_{\text{abs}}^{*t} \Lambda_1$; we find [using $\Lambda_1(0,0) = 1$] that the 0th entry of this vector is

$$\begin{aligned}
\mathbf{P}_\pi(T_0 \leq t) &= \sum_i \pi^*(i) \sum_j P_{\text{abs}}^{*t}(i,j) \Lambda_1(j,0) \\
&= \pi^*(0) + \sum_{i \neq 0} \pi^*(i) [P_{\text{abs}}^{*t}(i,0) + P_{\text{abs}}^{*t}(i,i) \Lambda_1(i,0)] \\
&= \pi^*(0) + \sum_{i \neq 0} \pi^*(i) [1 + P_{\text{abs}}^{*t}(i,i)(\Lambda_1(i,0) - 1)] \\
&= \pi^*(0) + \sum_{i \neq 0} \pi^*(i) [1 + \gamma_i^t (\Lambda_1(i,0) - 1)] \\
&= 1 + \sum_{i=1}^r \pi^*(i) \gamma_i^t (\Lambda_1(i,0) - 1).
\end{aligned}$$

We also have from (2.3.15) in the proof of the next lemma that $\mathbf{P}_\pi(T_0 \leq t) = 1 - \sum_{i=1}^r \pi^*(i) \gamma_i^t$. Therefore $\Lambda_1(i,0) = 0$ for $i > 0$, and (2.3.14) follows. □

Lemma 2.3.14. *There exist real constants c_j^i such that $\pi = \pi^* \Lambda_1$.*

Proof. We will make use of the fact that

$$\mathbf{P}_\pi(T_0 > t) = \sum_{j=1}^r \pi^*(j) \gamma_j^t, \quad t = 0, 1, \ldots, \tag{2.3.15}$$

which follows from its continuous-time analogue, equation (1.1) in [9], using Lemma 2.3.3. [That analogue is established using the fact that the function ψ in our equations (2.3.3)–(2.3.4) is the Laplace transform of T_0 for the stationary continuized chain; see [9] for further details.] Define

$$\pi_{-0} := (\pi(1), \ldots, \pi(n)) \in \mathbb{R}^n;$$

we would use the notation π_0 to indicate this deletion of the 0th entry from π except that it conflicts with our notation for the initial distribution of the given chain. We then have that $P_\pi(T_0 > t) = \pi_{-0} P_0^t \vec{1}^T$. Using the spectral representation of P_0 we find for $t \geq 0$ that

$$\mathbf{P}_\pi(T_0 > t) = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \sqrt{\pi(i)\pi(j)} \widetilde{U}(k,i) \widetilde{U}(k,j) \eta_k^t = \sum_{k=1}^{n} q_k \eta_k^t. \tag{2.3.16}$$

Here $\mathbf{q} = (q_1, \ldots, q_n) = (\pi_{-0}^{1/2} \widetilde{U}^T)^2$, where both the nonnegative square root and the square are in the Hadamard sense. In particular, $q_k \geq 0$ for all $k = 1, \ldots, n$. Comparing (2.3.15) and (2.3.16), it is clear that if $\eta_i \neq \gamma_j$ for every $j = 1, \ldots, r$, then $q_i = 0$. Again comparing (2.3.15) and (2.3.16), for each γ_j there is an $\eta_k = \gamma_j$ such that the coefficient of η_k^t in (2.3.16), namely q_k, is strictly positive. Now $\mathbf{q} = (\pi_{-0}^{1/2} \widetilde{U}^T)^2$ equals the Hadamard square $(\pi_{-0} D_0^{-1/2} \widetilde{U}^T)^2$. We can therefore choose R, a diagonal matrix with ± 1 along the diagonal, such that $\pi_{-0} = \mathbf{q}^{1/2} R (\tilde{U} D_0^{1/2}) = \mathbf{q}^{1/2} R U$; here $\mathbf{q}^{1/2}$ is the Hadamard nonnegative square root of $\mathbf{q}$. Relabel the entries of the vector $\mathbf{q}$ (and of R) so that

$$\pi_{-0} = \sum_{i=1}^{r} \sum_{j=1}^{n_i} r_j^i (q_j^i)^{1/2} u_j^i.$$

Letting $c_j^i = r_j^i (q_j^i)^{1/2} / \pi^*(i)$ yields

$$\pi_{-0} = \sum_{i=1}^{r} \sum_{j=1}^{n_i} \pi^*(i)\, c_j^i\, u_j^i.$$

It remains only to show that for this choice of c_j^i's we have

$$\pi(0) = 1 + \sum_{i=1}^{r} \sum_{j=1}^{n_i} \pi^*(i) c_j^i (-u_j^i \vec{1}^T).$$

This is immediate from

$$1 - \pi(0) = \pi_{-0} \vec{1}^T = \mathbf{q}^{1/2} R U \vec{1}^T = \sum_{i=1}^{r} \sum_{j=1}^{n_i} r_j^i (q_j^i)^{1/2} (u_j^i \ \vec{1}^T). \qquad \square$$

Our construction of Λ_1 uses the eigenvectors of P_{abs}; the entries of these eigenvectors are not all nonnegative, and as a result neither (in general) are the entries of Λ_1. In the special case that the given chain is a star chain, the quasi-link Λ_1 is a *bona fide* link. For example, for the chain considered in Remark 2.3.6 the quasi-link Λ_1 is easily seen to be the link

$$\Lambda_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix}.$$

Remark 2.3.15. If $r = n$ (i.e., the reduced spectra are the same as the unreduced spectra), then it is not hard to show that the quasi-link Λ_1 of Theorem 2.3.13 is uniquely determined.

2.3.4 The big link Λ

Combining the quasi-link Λ_1 of Theorem 2.3.13 between (π, P_{abs}) and $(\pi^*, P^*_{\text{abs}})$ and the link Λ_2 of Theorem 2.3.11 between $(\pi^*, P^*_{\text{abs}})$ and $(\hat{\pi}, \widehat{P})$, we obtain the desired quasi-link $\Lambda = \Lambda_2 \Lambda_1$ between (π, P_{abs}) and $(\hat{\pi}, \widehat{P})$.

Theorem 2.3.16. *Let $\Lambda := \Lambda_2\Lambda_1$. Then Λ is a quasi-link providing a quasi-intertwining of (π, P_{abs}) and $(\hat{\pi}, \widehat{P})$, and therefore $\mathcal{L}_\pi T_0 = \mathcal{L}_{\hat{\pi}}\widehat{T}_0$.*

Proof. This follows from Remark 2.1.2 and the discussion in Section 2.1.3. □

If Λ is stochastic, then we have a link between P_{abs} and $\widehat{P}$ and we can use the discussion following Definition 2.1.1 to construct a sample path of $(\widehat{X}_t)$ given a realization of (X_t). However, it's easy to find examples showing that Λ is not nonnegative in general.

The discussion preceding Remark 2.3.15 shows that Λ is a link if the given chain X is a star chain. More generally, Λ is a link if the given chain is a "block star chain", defined as follows: Choose positive numbers $b_0, \ldots, b_k$ summing to unity and $0 < \pi_0 \leq 1$. For $i = 1, \ldots, k$, let $c_i := \pi_0 b_i$ and let Q_i be an ergodic and reversible Markov kernel with stationary probability mass function π_i. Let P be the following special case of (2.2.1):

$$P = \begin{pmatrix} b_0 & b_1\pi_1 & b_2\pi_2 & \ldots & b_k\pi_k \\ c_1\vec{1}^T & (1-c_1)Q_1 & 0 & \ldots & 0 \\ c_2\vec{1}^T & 0 & (1-c_2)Q_2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_k\vec{1}^T & 0 & 0 & \ldots & (1-c_k)Q_k \end{pmatrix};$$

it is easily checked that P is ergodic and reversible with stationary distribution equal to the concatenated row vector $(\pi_0 + k)^{-1}(\pi_0|\pi_1|\cdots|\pi_k)$, and that the reduction of spectra described in Section 2.3.1 results in $\{\gamma_1, \ldots, \gamma_r\}$ being some subset of distinct elements from $\{1 - c_1, \ldots, 1 - c_k\}$. If, for example, $r = k$, then Λ_1 is the matrix (2.2.2), where $\mu_0 = (1)$ is 1-by-1 and we recall for $1 \leq j \leq k$ that μ_j $(= \pi_j)$ is the quasi-stationary

distribution for the jth diagonal block $(1 - c_j)Q_j$ of P; hence Λ_1 is a link (and so, then, is $\Lambda = \Lambda_1\Lambda_2$). We are not aware of other interesting cases where Λ is guaranteed to be a link, but the key is to arrange, as for block star chains, for P_0 to have nonnegative eigenvectors corresponding to eigenvalues $\gamma_1 \ldots, \gamma_r$.

Remark 2.3.17. Is there a *unique* quasi-link Λ which, like the one constructed in Theorem 2.3.16, satisfies $\Lambda\delta_0^T = \delta_0^T$ and provides a quasi-intertwining of (π, P_{abs}) and $(\hat{\pi}, \widehat{P})$? We do not know the answer in general, but if $r = n$, then the answer is affirmative by Remark 2.3.15 and the invertibility of Λ_2.

2.4 Another representation for hitting times from stationarity

Our final application of the strategy outlined in Section 2.1.2 will provide a stochastic construction for an alternative characterization of the hitting-time distribution from stationarity first proved by Mark Brown [personal communication] in an unpublished technical report. A published version of a special case can be found in [8]. Our construction here is notable in that it will provide a generalization (to not necessarily reversible chains) of the discrete-time analogue of Brown's original result, and it is by applying our strategy that we discovered the generalization.

Brown's original theorem is the following, in which 0 is an arbitrary fixed state.

Theorem 2.4.1 (Mark Brown)**.** *Consider an ergodic time-reversible finite-state continuous-time Markov chain with stationary distribution π. Let V be a random variable with*

$$\mathbf{P}(V > t) = \frac{P_{00}(t) - \pi(0)}{1 - \pi(0)}, \quad 0 \le t < \infty.$$

Let $V_1, V_2, \ldots$ be iid copies of V, and let N be independent of the sequence (V_i) with $N+1$ distributed Geometric with success probability $\pi(0)$:

$$\mathbf{P}(N = k) = \pi(0)[1 - \pi(0)]^k, \quad k = 0, 1, \ldots.$$

Then the distribution $\mathcal{L}_\pi T_0$ of the nonnegative hitting time T_0 of 0 from a stationary start is the distribution of $\sum_{i=1}^{N} V_i$.

We will focus on the following discrete-time analogue. As in Section 2.3, analogues of all of our results can be established in the continuous-time setting as well, but we have chosen discrete time for convenience and ease of understanding.

Theorem 2.4.2. *Consider an ergodic time-reversible finite-state discrete-time Markov chain with stationary distribution π. Assume that $P^t(0,0)$ is nonincreasing in t. Let V be a random variable with*

$$\mathbf{P}(V > t) = \frac{P^t(0,0) - \pi(0)}{1 - \pi(0)}, \quad t = 0, 1, \ldots.$$

Let $V_1, V_2, \ldots$ be iid copies of V, and let N be independent of the sequence (V_i) with $N+1$ distributed Geometric with success probability $\pi(0)$:

$$\mathbf{P}(N = k) = \pi(0)[1 - \pi(0)]^k, \quad k = 0, 1, \ldots.$$

Then the distribution $\mathcal{L}_\pi T_0$ of the nonnegative hitting time T_0 of 0 from a stationary start is the distribution of $\sum_{i=1}^{N} V_i$.

The assumption in Theorem 2.4.2 that $P^t(0,0)$ is nonincreasing in t is met, for example, if the chain is time-reversible and all the eigenvalues of the one-step transition matrix P are nonnegative. **However, we do not need to assume reversibility to follow our approach**, so Theorem 2.4.2 (and likewise Theorem 2.4.1) is true without that assumption. For a non-reversible scenario in which the nonincreasingness assumption is satisfied, see Remark 2.4.7 and the paragraph preceding it.

Following our strategy, we aim to provide a sample-path intertwining of the given chain X in Theorem 2.4.2 with a chain $\widehat{X}$ (with, say, initial distribution $\hat{\pi}_0$ and transition matrix $\widehat{P}$) for which the hitting time $\widehat{T}_0$ has (for each sample path) a clear decomposition $\sum_{i=1}^{N} V_i$ as in the theorem. As in our earlier application, we can treat 0 as an absorbing state for the given chain, whose one-step transition matrix we then denote by P_{abs}. We thus wish to find $(\hat{\pi}_0, \widehat{P})$ and a link (or at least quasi-link) Λ such that $\pi = \hat{\pi}_0 \Lambda$ and $\Lambda P_{\text{abs}} = \widehat{P}\Lambda$. The chain $\widehat{X}$ we will construct has state space $\{0, 1, \dots\}$. Although the state space is infinite, this gives no difficulties as the needed intertwining results from [12] apply just as readily to Markov chains with countably infinite state spaces. First we construct our Λ.

Suppose the given chain has state space $\{0, 1, \dots, n\}$. We adopt notation that highlights the special role of state 0. Let $\pi = (\pi(0)|\pi_{-0}) \in \mathbb{R}^{n+1}$ with $\pi_{-0} \in \mathbb{R}^n$, and similarly let

$P^{i-1}(0,:) = (P^{i-1}(0,0)\,|P^{i-1}(0,:)_{-0}) \in \mathbb{R}^{n+1}$. For $i = 1, 2, 3, \ldots$, define

$$\mu_i := \left(0 \,\middle|\, \frac{P^{i-1}(0,0)\pi_{-0} - \pi(0)P^{i-1}(0,:)_{-0}}{P^{i-1}(0,0) - \pi(0)}\right) \in \mathbb{R}^{n+1}.$$

Lemma 2.4.3. *With μ_i defined above, we have for $i > 0$ that*

$$\mu_i P = q_i \pi + (1 - q_i)\mu_{i+1},$$

where

$$q_i := \frac{P^{i-1}(0,0) - P^i(0,0)}{P^{i-1}(0,0) - \pi(0)} \in [0, 1).$$

Proof. First note

$$\mu_i P = \frac{P^{i-1}(0,0)(0|\pi_{-0})P - \pi(0)(0|P^{i-1}(0,:)_{-0})P}{P^{i-1}(0,0) - \pi(0)}.$$

Now $(0|\pi_{-0})P = \pi - \pi(0)P(0,:)$, and similarly

$$\begin{aligned}(0|P^{i-1}(0,:)_{-0})P &= P^{i-1}(0,:)P - P^{i-1}(0,0)P(0,:) \\ &= P^i(0,:) - P^{i-1}(0,0)P(0,:);\end{aligned}$$

hence

$$\begin{aligned}&P^{i-1}(0,0)(0|\pi_{-0})P - \pi(0)(0|P^{i-1}(0,:)_{-0})P \\ &\quad = P^{i-1}(0,0)\pi - \pi(0)P^i(0,:) \\ &\quad = P^{i-1}(0,0)\pi - (P^i(0,0)\pi(0)|0) - (0|\pi(0)P^i(0,:)_{-0}) \\ &\quad = P^{i-1}(0,0)\pi - P^i(0,0)\pi + P^i(0,0)(0|\pi_{-0}) - (0|\pi(0)P^i(0,:)_{-0}) \\ &\quad = [P^{i-1}(0,0) - P^i(0,0)]\pi + (0|P^i(0,0)\pi_{-0} - \pi(0)P^i(0,:)_{-0}).\end{aligned}$$

Letting

$$q_i := \frac{P^{i-1}(0,0) - P^i(0,0)}{P^{i-1}(0,0) - \pi(0)},$$

it follows that $\mu_i P = q_i \pi + (1 - q_i)\mu_{i+1}$, as desired. □

This lemma suggests the form for $\widehat{P}$ and Λ. Let $\widehat{X}$ have state space $\{0, 1, 2, \ldots\}$. Define the transition kernel $\widehat{P}$ by setting $\widehat{P}(0,0) := 1$ and, for $i > 0$,

$$\widehat{P}(i,0) := \pi(0)q_i, \qquad \widehat{P}(i,1) := [1 - \pi(0)]q_i, \qquad \widehat{P}(i,i+1) := 1 - q_i;$$

we set $\widehat{P}(i,j) := 0$ for all other pairs (i,j). As the following lemma shows, the hitting time $\widehat{T}_0$ for this chain $\widehat{X}$ has a simple decomposition as a sum of Geometrically many iid copies of V.

Lemma 2.4.4. *Let $\widehat{X}$ have initial distribution $\hat{\pi}_0 := \pi(0)\delta_0 + [1 - \pi(0)]\delta_1$ and one-step transition matrix $\widehat{P}$. Then there exist random variables N and $V_1, V_2, \ldots$ with joint distribution as in Theorem 2.4.2 such that (for every sample path) $\widehat{T}_0 = \sum_{i=1}^{N} V_i$.*

Proof. Let $N \geq 0$ denote the number of visits to state 1; and for $i = 1, \ldots, N$, let V_i denote the highest state reached in the time interval $[\tau_i, \tau_{i+1})$, where τ_i denotes the epoch of ith visit to state 1. Then all of the assertions of the lemma are clear; it is perhaps worth noting only that for $t = 0, 1, \ldots$ we have

$$\mathbf{P}(V_1 > t) = \prod_{i=1}^{t}(1 - q_i) = \frac{P^t(0,0) - \pi(0)}{1 - \pi(0)}.$$

□

Define Λ by setting $\Lambda(0,:) := \delta_0$ and $\Lambda(i,:) := \mu_i$ for $i > 0$. Note that Λ has infinitely many rows, each of which is in $\mathbb{R}^{n+1}$. We then have the following theorem whose proof is almost immediate from the definitions and Lemma 2.4.3.

Theorem 2.4.5. *The quasi-link Λ provides a quasi-intertwining of (π, P_{abs}) and $(\hat{\pi}_0, \widehat{P})$ and satisfies* (2.1.2)*, and therefore $\mathcal{L}_\pi T_0 = \mathcal{L}_{\hat{\pi}_0} \widehat{T}_0$.*

Proof. It is easily checked that each row of Λ sums to unity. Further, for $i > 0$ using the observations that $\mu_i(0) \equiv 0$ and $\mu_1 = \left(0 \,\middle|\, \frac{\pi_{-0}}{1-\pi(0)}\right)$, one finds readily that the ith row of ΛP_{abs} is $\mu_i P = q_i \pi + (1 - q_i)\mu_{i+1}$, which is the ith row of $\widehat{P}\Lambda$. The 0th rows are both δ_0, so we conclude $\Lambda P_{\mathrm{abs}} = \widehat{P}\Lambda$. Similarly, $\pi = \hat{\pi}_0 \Lambda$. Finally, since $\mu_0(0) = 1$ and $\mu_i(0) = 0$ for $i > 0$, we have $\Lambda \delta_0^T = \delta_0^T$, which is (2.1.2). The equality of hitting-time laws then follows from the discussion in Section 2.1.3. □

Note that Λ is a link (in which case sample-path linking is possible) if and only if for every $t \geq 0$ the t-step transition probability $\tilde{P}^t(i, 0)$ is maximized when $i = 0$; here $\tilde{P}$ is the time-reversed transition matrix $\tilde{P}(i, j) := \pi(j)P(j, i)/\pi(i)$. A sufficient condition for this is that the state space is partially ordered, 0 is either a top element or a bottom element, and $\tilde{P}$ is stochastically monotone.

Remark 2.4.6. The intertwining constructed in Lemma 2.4.3 and Theorem 2.4.5 can be related to the fastest strong stationary time construction of [3] and the corresponding strong stationary dual constructed in Example 2.6 of [12]. In the interest of brevity, we omit an explanation of the connection.

Remark 2.4.7. We claim that if P is such that $\tilde{P}^t(i,0)$ is maximized for every t when $i = 0$, then P automatically satisfies the assumption in Theorem 2.4.2 that $P^t(0,0)$ is nonincreasing in t. To see this, consider the chain X with transition matrix P started in distribution $\mu_1 = [1-\pi(0)]^{-1}[\pi - \pi(0)\delta_0]$. Then, for any state i,

$$\frac{\mathbf{P}(X_t = i)}{\pi(i)} = \frac{\pi(i) - \pi(0)P^t(0,i)}{\pi(i)[1-\pi(0)]} = \frac{1-\tilde{P}^t(i,0)}{1-\pi(0)}.$$

If $s(t)$ is the separation of the chain at time t, then $1 - s(t)$ equals the minimum of this ratio over i, namely, $[1 - P^t(0,0)]/[1-\pi(0)]$. It is well known (e.g., [4, Chapter 9]) that separation is nonincreasing in t, so $P^t(0,0)$ is nonincreasing.

Chapter 3

Strong Stationary Duality for Diffusion Processes

In this chapter, we systematically develop the theory of strong stationary duality for diffusion processes on finite intervals. In Section 3.2.1 we define the strong stationary dual in the diffusion setting and in Section 3.2.2 we analytically derive the form of the dual diffusion's generator; in the process we explicitly derive the boundary behavior of the dual diffusion. In Section 3.3, we show that a suitably defined sequence of Markov chains and their strong stationary duals converge respectively to our primal diffusion and its strong stationary dual. In Section 3.4, we recover a central tenet of the classical theory by proving that the separation mixing time in the primal diffusion is equal in law to the absorption time in the dual diffusion. We exploit this connection in Section 3.5 to derive the analogue to the birth-and-death cut-off phenomenon theory of [14] in the present diffusion setting.

3.1 Background

A one-dimensional diffusion process X defined on the (possibly infinite) real interval I is a Markov process with sample paths in $C_I[0,\infty)$, where $C_I[0,\infty)$ is the space of continuous functions $x : [0,\infty) \to I$. In the sequel, we will consider only time-homogeneous diffusion processes. For a gentle (though occasionally non-rigorous) treatment of classical one-dimensional diffusion theory see [29, Chapter 15]. For a higher-level treatment see for example [27] or [35]. The generator of the diffusion is of the form

$$A = \frac{1}{2}b(x)\frac{d^2}{dx^2} + a(x)\frac{d}{dx} \tag{3.1.1}$$

where the infinitesimal drift $a(x)$ at a state x satisfies

$$\lim_{h\downarrow 0}\frac{1}{h}\mathbb{E}(X_{t+h} - X_t|X_t) = a(X_t) \text{ for all } t,$$

and the infinitesimal variance $b(x)$ satisfies

$$\lim_{h\downarrow 0}\frac{1}{h}\mathbb{E}[(X_{t+h} - X_t)^2|X_t] = b(X_t) \text{ for all } t.$$

Denoting the closure of I by $\bar{I}$ and the interior of I by I°, to avoid pathologies, we shall assume throughout that $a(\cdot)$, $b(\cdot) \in C(I^\circ)$ and $b > 0$ on I°.

We say that a diffusion X is *regular* if for all $x \in I^\circ$ and $y \in I$ we have $\mathbb{P}_x(T_y < \infty) > 0$, and we can then define a *scale function* of X to be a strictly increasing positive function $S : I \to \mathbb{R}$ such that for all $x < y < z$ we have

$$\mathbb{P}_y(T_z < T_x) = \frac{S(y) - S(x)}{S(z) - S(x)}.$$

If $S(x) = x$ is a scale function for X, then we say X is in *natural scale*. Note that S being a scale function for X implies $\alpha S + \beta$ is a scale function for X for any $\alpha > 0$ and β real, and it is easy to see that S is uniquely defined up to this affine transformation. We can find a scale function in terms of $a(\cdot)$ and $b(\cdot)$ via

$$S'(x) := s(x) = \exp\left\{-\int^x \frac{2a(y)}{b(y)}\,dy\right\}.$$

For a fixed but arbitrary $c \in I^\circ$, define the *speed function* $M : I^\circ \to \mathbb{R}$ of a diffusion via

$$M(x) = \int_{y=c}^{x} \frac{1}{s(y)b(y)}\,dy$$

and define the *speed measure* (also denoted M) via $M(x, y] = M(y) - M(x)$ to be the nonnegative measure on I° with density $m(x) := [s(x)b(x)]^{-1}$. Note that the speed measure is independent of our choice of c in the speed function. The speed measure derives its name from the following: Let X be in natural scale, and for $x \in I^\circ$ let $T_x(\varepsilon)$ be the first time the diffusion started in x exits $(x - \varepsilon, x + \varepsilon)$. Then

$$\lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon^2} \mathbb{E}_x T(\varepsilon) = m(x).$$

In words, $m(x)\varepsilon^2$ is the lead-order of the expected time to exit $(x - \varepsilon, x + \varepsilon)$ started in x. The speed measure M on I° can be extended to a measure on $\bar{I}$ by specifying arbitrary nonnegative (possibly infinite) mass at each boundary point of $\bar{I} \setminus I^\circ$.

For a diffusion X on I [with $I^\circ = (l, r)$] where $l \in I$, we classify the boundary behavior of X at l using the standard Feller boundary classification for one-dimensional diffusions (see [18, Section 8.1] for more details). Feller classified the boundary behavior of X at l

(analogous results holding at r if $r \in I$) by looking at the behavior of

$$N(l) := \int_{(l,x]} [S(x) - S(\eta)]\ M(d\eta), \quad \Sigma(l) := \int_{(l,x]} [M(x) - M(\eta)]\ dS(\eta)$$

for a fixed $x \in I^\circ$ and by calculating boundary conditions satisfied by elements of the domain of A, which we shall call $\mathcal{D}_A$. In classifying the boundary behavior of X, we are only concerned with the finiteness of $N(\cdot)$ and $\Sigma(\cdot)$, which is independent of x, and so the dependence of $N(\cdot)$ and $\Sigma(\cdot)$ on x has been supressed. The boundary l is said to be an *entrance boundary* if the diffusion cannot reach l from I° but can start in l and then immediately move to the interior of the state space. Entrance boundaries are characterized by

$$N(l) < \infty,\ \Sigma(l) = \infty.$$

Note that [29, Section 15.6] implies that to show l is entrance, it suffices to show that $N(l) < \infty$ and $S(l, x] = \lim_{y \downarrow l}[S(x) - S(y)] = \infty$. The boundary is said to be an *exit* boundary if it can be reached from the interior of the state space, but starting in l cannot reach I°. This behavior is characterized by

$$N(l) = \infty,\ \Sigma(l) < \infty.$$

Natural boundaries cannot be reached from I° and the diffusion cannot begin in a natural boundary. This behavior is characterized by $N(l) = \infty$ and $\Sigma(l) = \infty$. The boundary is *regular* if it can be reached from the interior of the state space and there exists a diffusion X with generator A (acting on a specified $\mathcal{D}_A$) that can enter I° starting at l. Regular

boundaries are characterized by

$$N(l) < \infty,\ \Sigma(l) < \infty.$$

The behavior of the diffusion at a regular boundary will be characterized by boundary conditions satisfied by elements $f \in \mathcal{D}_A$. In particular, we say that l is *instantaneously reflecting* if $f \in \mathcal{D}_A$ implies that

$$\frac{df}{dS}^{+}(l) = \lim_{x \downarrow l} \frac{df}{dS}(x) = \lim_{x \downarrow l} \frac{f'(x)}{s(x)} = 0.$$

We say l is *absorbing* if $f \in \mathcal{D}_A$ implies that $(Af)(l) = 0$. Instantaneously reflecting boundaries instantly reflect the diffusion back into the interior of the state space, while once a diffusion reaches an absorbing boundary, it remains at the boundary thereafter.

Presently and in the sequel, let X be a regular diffusion process on a finite closed interval I ($= [0, 1]$, without loss of generality) with initial distribution π_0 and generator A. Assume that 0 and 1 are instantaneously reflecting boundaries X. The boundary behavior of X guarantees that M is a finite measure on I°, and normalizing $M(dx)$ to a probability measure gives the unique invariant distribution of X, which we will denote by $\Pi(dx)$. As with M, for arbitrary $c \in I^\circ$, let us adopt the shorthand $\Pi(x) := \int_{y=c}^{x} \pi(y)\, dy$, where π is the density for Π with respect to Lebesgue measure, and note that regularity of X guarantees $\pi > 0$ on I°. The reflecting behavior at 0 guarantees $\lim_{c \downarrow 0} \int_{y=c}^{x} \pi(y)\, dy$ exists and is finite for all $x \in I^\circ$, and so to ease notation we may let $\Pi(x) = \int_{y=0}^{x} \pi(y)\, dy$ defined as an improper integral. Lastly, let $(P_t)_{t=0}^{\infty}$ be the Markov transition function associated with X and denote the corresponding transition densities with respect to Lebesgue measure

by $(p_t)_{t=0}^{\infty}$.

Based on the boundary behavior of X, we can completely specify the domain of A as

$$\mathcal{D}_A = \left\{ f \in C(I) \cap C^2(I^\circ) \mid Af \in C(I),\ \frac{df}{dS}^+(0) = \frac{df}{dS}^-(1) = 0 \right\} \tag{3.1.2}$$

(see [18, Section 8.1], especially (1.11) there, with $q_0 = 0 = q_1$ because both boundaries are instantaneously reflecting), where as above

$$\frac{df}{dS}^+(0) = \lim_{x \downarrow 0} \frac{df}{dS}(x) = \lim_{x \downarrow 0} \frac{f'(x)}{s(x)},$$

and

$$\frac{df}{dS}^-(1) = \lim_{x \uparrow 1} \frac{df}{dS}(x) = \lim_{x \uparrow 1} \frac{f'(x)}{s(x)}.$$

Let $F[0,1]$ be the space of bounded real valued measurable functions on $[0,1]$ equipped with its usual Borel σ-field $\mathcal{B}$. Let $M[0,1]$ be the space of signed measures on $([0,1], \mathcal{B})$. As in [36, Section 7.1], we note the natural bilinear functional on $F[0,1] \times M[0,1]$ defined by $(\mu, f) = \int_0^1 f(x)\, \mu(dx)$. We denote the adjoint of the operator T_t (with respect to this functional) by U_t, where $(T_t)_{t=0}^{\infty}$ is the one parameter Markov semigroup associated with $(P_t)_{t=0}^{\infty}$.

Note that $a(\cdot),\ b(\cdot),\ M(\cdot)$, and $\pi(\cdot)$ are defined only on I°. For notational convenience, any expressions involving these functions and ∂I are to be interpreted as the corresponding limiting expression (when such a limit exists!). For example, for $0 < x < 1$ we shall write the improper integral $\int_0^x f(y)\pi(y)\, dy$ rather than the equivalent $\lim_{z \downarrow 0} \int_z^x f(y)\pi(y)\, dy$.

3.2 Strong stationary duality for diffusions

3.2.1 Definition of the strong stationary dual

Let X^* be a second (Feller) diffusion process on I with initial distribution π_0^* and generator A^*. As in the continuous-time discrete-state Markov chain setting (see [19]), we define the notion of algebraic duality between X and X^*:

Definition 3.2.1. *Consider the integral operator Λ acting on $F[0,1]$ defined by*

$$(\Lambda f)(x) := \begin{cases} \int_0^x \pi^{(x)}(y) f(y)\, dy & \text{if } x > 0, \\ f(0) & \text{if } x = 0, \end{cases}$$

where we define the kernel

$$\pi^{(x)}(y) := \frac{\pi(y)}{\Pi(x)} \quad 0 < y \leq x < 1, \text{ and } \pi^{(1)} \equiv \pi.$$

We say that X^* is a strong stationary dual of X if

$$\Lambda \text{ maps } \mathcal{D}_A \text{ into } \mathcal{D}_{A^*} \tag{3.2.1}$$

and

$$\Lambda A = A^* \Lambda \text{ as operators defined on } \mathcal{D}_A \tag{3.2.2}$$

and

$$(\pi_0, f) = (\pi_0^*, \Lambda f) \text{ for all } f \in F[0,1]. \tag{3.2.3}$$

Remark 3.2.2. If $f \in C(I)$, then $\Lambda f \in C(I)$ as well. To show this, first note that $\pi \in C(I^\circ)$, $\Pi \in C(I)$, and for $x > 0$ we have $\Pi(x) > 0$. Clearly, then,

$$\Lambda f(x) = \frac{\int_0^x \pi(y) f(y)\, dy}{\Pi(x)}$$

is continuous at all $x > 0$. Continuity at zero is immediate as for any $\epsilon > 0$, we can choose x such that $|f(y) - f(0)| < \epsilon$ for all $y \leq x$, and so

$$|\Lambda f(0) - \Lambda f(x)| = \left| \int_0^x (f(0) - f(y)) \pi^{(x)}(y)\, dy \right| \leq \epsilon.$$

Remark 3.2.3. For $x < 1$, let $\Pi^{(x)}$ be the distribution Π conditioned to $(0, x]$, so that $\Pi^{(x)}$ has density $\pi^{(x)}$ when $x > 0$, and let $\Pi^{(0)} := \delta_0$ and $\Pi^{(1)} := \Pi$. If $\pi_0 = \Pi^{(x)}$ for some $x \in [0, 1)$, then (3.2.3) is uniquely satisfied by $\pi_0^* = \delta_x$. For $x \in (0, 1)$, this is easily seen via

$$\begin{aligned} \int_I f(y)\, \pi^{(x)}(y)\, dy &= (\pi_0, f) \\ &= (\pi_0^*, \Lambda f) = \int_I \int_{y \in (0,z]} \pi^{(z)}(y) f(y)\, dy\, \pi_0^*(dz) \\ &= \int_I \int_{z \in [y,1]} \pi^{(z)}(y)\, \pi_0^*(dz)\, f(y)\, dy. \end{aligned} \tag{3.2.4}$$

Letting $f(y) = \mathbb{1}(y > x)$ we see π_0^* must be concentrated on $(0, x]$. It also follows that for almost every y satisfying $0 < y \leq x$ we have

$$\pi^{(x)}(y) = \int_{z \in [y,1]} \pi^{(z)}(y)\, \pi_0^*(dz) = \int_{z \in [y,x]} \pi^{(z)}(y)\, \pi_0^*(dz),$$

or, equivalently,

$$\frac{1}{\Pi(x)} = \int_{z \in [y,x]} \frac{\pi_0^*(dz)}{\Pi(z)}.$$

Letting $y \uparrow x$ through such values, it follows that $\pi_0^* = \delta_x$ is the only possible initial distribution for X^*. To show that $\pi_0^* = \delta_x$ satisfies (3.2.3), note

$$(\delta_x, \Lambda f) = \Lambda f(x) = \int_0^x \pi^{(x)}(y) f(y)\, dy = (\Pi^{(x)}, f),$$

as desired. For $x = 0$, the argument goes as follows. For uniqueness, if $\pi_0 = \delta_0$, then letting $f(y) = \mathbb{1}(y \in (0, 1])$, the left side of (3.2.3) equals $f(0) = 0$, and the right side is strictly positive unless $\pi_0^* = \delta_0$. To see that $\pi_0^* = \delta_0$ satisfies (3.2.3) when $\pi_0 = \Pi^{(0)} = \delta_0$, we compute $(\delta_0, \Lambda f) = \Lambda f(0) = f(0) = (\delta_0, f)$.

3.2.2 The dual generator

From the definition of strong stationary duality, we derive the form of the dual generator:

Theorem 3.2.4. *With X as above, assume further that $b \in C^1(I^\circ)$. If X^* is a strong stationary dual of X, then the generator A^* of X^* has the form*

$$(A^* f)(x) = \left(\frac{1}{2} b'(x) - a(x) + b(x) \frac{\pi(x)}{\Pi(x)} \right) f'(x) + \frac{1}{2} b(x) f''(x)$$

for $x \in I^\circ$ and $f \in \mathcal{D}_{A^*}$. Also 0 is an entrance boundary for X^* and 1 is a regular absorbing boundary of X^*.

Proof. Let $f \in \mathcal{D}_A$. Then $Af \in C(I)$ and for $x > 0$ we have

$$\begin{aligned}(\Lambda A f)(x) &= \int_0^x a(y) f'(y) \pi^{(x)}(y)\, dy + \int_0^x \tfrac{1}{2} b(y) f''(y) \pi^{(x)}(y)\, dy \\ &= \frac{1}{\Pi(x)} \left(\int_0^x a(y) f'(y) \pi(y)\, dy + \int_0^x \tfrac{1}{2} b(y) f''(y) \pi(y)\, dy \right).\end{aligned}$$

We know that there exists a nonzero constant C such that $C \cdot \pi(x) = m(x)$, so that $\pi(x) = \frac{1}{Cb(x)s(x)}$. Also, $\frac{d}{dx}\frac{1}{s(x)} = \frac{1}{s(x)}\frac{2a(x)}{b(x)}$. The first term in $(\Lambda Af)(x)$ is then equal to

$$\frac{1}{\Pi(x)}\int_0^x \frac{1}{2C}\frac{2a(y)}{b(y)}\frac{1}{s(y)}f'(y)\,dy,$$

which by integration by parts equals

$$\frac{1}{\Pi(x)}\frac{1}{2C}\left[\frac{df}{dS}(x) - \frac{df}{dS}^{+}(0) - \int_0^x \frac{1}{s(y)}f''(y)\,dy\right].$$

The second term in $(\Lambda Af)(x)$ is equal to

$$\frac{1}{\Pi(x)}\frac{1}{2C}\int_0^x \frac{1}{s(y)}f''(y)\,dy,$$

and so

$$\begin{aligned}(\Lambda Af)(x) &= \frac{1}{\Pi(x)}\frac{1}{2C}\left[\frac{df}{dS}(x) - \frac{df}{dS}^{+}(0)\right] \\ &= \frac{1}{2}\frac{b(x)\pi(x)}{\Pi(x)}f'(x) - \frac{1}{\Pi(x)}\frac{1}{2C}\frac{df}{dS}^{+}(0).\end{aligned}$$

Since 0 is a reflecting boundary of X and $f \in \mathcal{D}_A$, we have $\frac{df}{dS}^{+}(0) = 0$ and thus

$$(\Lambda Af)(x) = \frac{1}{2}\frac{b(x)\pi(x)}{\Pi(x)}f'(x).$$

Let $g \in \mathcal{D}_{A^*}$. For $x \in (0,1)$, from equation (3.1.1) for A^* we can write

$$(A^*g)(x) = a^*(x)g'(x) + \tfrac{1}{2}b^*(x)g''(x).$$

for some a^*, $b^* \in C(I^\circ)$. If $f \in \mathcal{D}_A$ then by (3.2.1) we have $\Lambda f \in \mathcal{D}_{A^*}$, and so for $x \in (0,1)$ we know $(A^*\Lambda f)(x) = a^*(x)(\Lambda f)'(x) + \frac{1}{2}b^*(x)(\Lambda f)''(x)$. Note that $Af \in C(I)$

by assumption and so $\Lambda A f = A^* \Lambda f \in C(I)$ from Remark 3.2.2. Now

$$\begin{aligned}(\Lambda f)'(x) &= \frac{\Pi(x)\pi(x)f(x) - \pi(x)\int_0^x \pi(y)f(y)\,dy}{\Pi(x)^2} \\ &= \frac{\pi(x)}{\Pi(x)}[f(x) - (\Lambda f)(x)]\end{aligned}$$

and so

$$\begin{aligned}(\Lambda f)''(x) &= \frac{\Pi(x)\pi'(x) - \pi(x)^2}{\Pi(x)^2}[f(x) - (\Lambda f)(x)] + \frac{\pi(x)}{\Pi(x)}\left\{f'(x) - \frac{\pi(x)}{\Pi(x)}[f(x) - (\Lambda f)(x)]\right\} \\ &= \left[\frac{\pi'(x)}{\Pi(x)} - \frac{2\pi(x)^2}{\Pi(x)^2}\right][f(x) - (\Lambda f)x)] + \frac{\pi(x)}{\Pi(x)}f'(x).\end{aligned}$$

Now by (3.2.2), $\Lambda A = A^* \Lambda$ as operators on $\mathcal{D}_A$, which implies that for any $x \in (0,1)$ and $f \in \mathcal{D}_A$ we have

$$\begin{aligned}\frac{1}{2}\frac{b(x)\pi(x)}{\Pi(x)}f'(x) = &\left(a^*(x)\frac{\pi(x)}{\Pi(x)} + \frac{1}{2}b^*(x)\left[\frac{\pi'(x)}{\Pi(x)} - \frac{2\pi(x)^2}{\Pi(x)^2}\right]\right)[f(x) - (\Lambda f)(x)] \\ &+ \frac{1}{2}b^*(x)\frac{\pi(x)}{\Pi(x)}f'(x). \qquad (3.2.5)\end{aligned}$$

For any fixed $x \in I^\circ$, we can choose $f \in \mathcal{D}_A$ so that $f'(x) = 0$ and $f(x) \neq (\Lambda f)(x)$ [e.g., let f be a suitably smooth approximation of $\mathbb{1}(x/3, x/2)$], and for any such f, equation (3.2.5) yields

$$a^*(x)\frac{\pi(x)}{\Pi(x)} + \frac{1}{2}b^*(x)\left[\frac{\pi'(x)}{\Pi(x)} - \frac{2\pi(x)^2}{\Pi(x)^2}\right] = 0. \qquad (3.2.6)$$

We then find for $f \in \mathcal{D}_A$ and $x \in (0,1)$ that $(A^*\Lambda f)(x) = \frac{1}{2}\frac{b^*(x)\pi(x)}{\Pi(x)}f'(x)$, and by (3.2.2) this equals $(\Lambda A f)(x) = \frac{1}{2}\frac{b(x)\pi(x)}{\Pi(x)}f'(x)$. For each x in $(0,1)$, we can choose an $f \in \mathcal{D}_A$ such that $f'(x) \neq 0$, and using any such f we find that $b^*(x) = b(x)$.

Next, we have from $\pi(x) = \frac{1}{Cb(x)s(x)}$ that

$$\pi'(x) = \frac{-b'(x)s(x) - b(x)s'(x)}{Cb(x)^2 s(x)^2}.$$

Equation (3.2.6) and $b^* \equiv b$ then yields

$$\begin{aligned}
\frac{\pi(x)}{\Pi(x)} a^*(x) &= \frac{1}{2} b(x) \left[\frac{b'(x)s(x) + b(x)s'(x)}{C\Pi(x)b(x)^2 s(x)^2} + \frac{2\pi(x)^2}{\Pi(x)^2} \right] \\
&= \frac{1}{2C\Pi(x)} \left[\frac{b'(x)}{b(x)s(x)} + \frac{s'(x)}{s(x)^2} \right] + b(x) \frac{\pi(x)^2}{\Pi(x)^2} \\
&= \frac{1}{2C\Pi(x)} \left[Cb'(x)\pi(x) - \frac{2a(x)}{s(x)b(x)} \right] + b(x) \frac{\pi(x)^2}{\Pi(x)^2} \\
&= \frac{1}{2} b'(x) \frac{\pi(x)}{\Pi(x)} - a(x) \frac{\pi(x)}{\Pi(x)} + b(x) \frac{\pi(x)^2}{\Pi(x)^2},
\end{aligned}$$

so that $a^*(x) = \frac{1}{2}b'(x) - a(x) + b(x)\frac{\pi(x)}{\Pi(x)}$ on I°, as desired.

To find the boundary behavior of the dual diffusion at 0 and at 1, we calculate the dual scale function and the dual speed measure. First, note that

$$\begin{aligned}
s^*(x) &= \exp\left[-\int^x \frac{2a^*(y)}{b^*(y)}\, dy \right] \\
&= \exp\left[-\int^x \frac{b'(y)}{b(y)}\, dy + \int^x \frac{2a(y)}{b(y)}\, dy - \int^x \frac{2m(y)}{M(y)}\, dy \right] \\
&= \frac{1}{b(x)} \frac{1}{s(x)} \frac{1}{M(x)^2} \\
&= \frac{m(x)}{M(x)^2},
\end{aligned} \tag{3.2.7}$$

and a scale function for X^* is

$$S^*(x) = \frac{-1}{M(x)}. \tag{3.2.8}$$

Next, note

$$m^*(x) = \frac{1}{b^*(x)s^*(x)} = \frac{M(x)^2}{m(x)b(x)} = M(x)^2 s(x). \tag{3.2.9}$$

Now $M(x)$ is continuous on I and $M(0) = 0$, so there is a y such that $M(\zeta) \leq 1$ for all $\zeta \leq y$. For the dual scale measure S^* we then have

$$\begin{aligned} S^*(0, y]) &= \int_{(0,y]} s^*(\zeta)\, d\zeta \\ &= \lim_{z \downarrow 0} \int_z^y \frac{m(\zeta)}{M(\zeta)^2}\, d\zeta \\ &\geq \lim_{z \downarrow 0} \int_z^y \frac{m(\zeta)}{M(\zeta)}\, d\zeta \\ &= \lim_{z \downarrow 0} \big[\log M(y) - \log M(z) \big] = \infty. \end{aligned}$$

To show that 0 is an entrance boundary for X^*, it now suffices to show that $N^*(0) < \infty$. This is shown via

$$\begin{aligned} N^*(0) &= \lim_{z \downarrow 0} \int_z^x S^*[y, x]\, dM^*(y) \\ &= \lim_{z \downarrow 0} \int_z^x [S^*(x) - S^*(y)]\, m^*(y)\, dy \\ &= \lim_{z \downarrow 0} \int_z^x \left[\frac{-1}{M(x)} - \frac{-1}{M(y)} \right] M(y)^2\, s(y)\, dy \\ &\leq \frac{-1}{M(x)} \liminf_{z \downarrow 0} \int_z^x M(y)^2\, s(y)\, dy + \limsup_{z \downarrow 0} \int_z^x M(y)\, s(y)\, dy. \end{aligned}$$

It now clearly suffices to prove $\int_0^x M(y)\, s(y)\, dy < \infty$, which follows from the following calculation:

$$\int_0^x M(y)\, s(y)\, dy = \int_0^x M(y)\, dS(y) = \int_0^x S[y, x]\, dM(y) =: N(0) < \infty,$$

where we used the fact that 0 is a reflecting boundary for X to derive the final inequality.

To prove that 1 is a regular absorbing boundary for X^*, we use Proposition 3.2.5 below. From that proposition, for any $f \in C[0, 1]$ and $x \in [0, 1]$, we have $(\Lambda T_t f)(x) =$

$(T_t^*\Lambda f)(x)$. When $x > 0$, we then have

$$\int_0^1 \left[\int_0^x \pi^{(x)}(y)p_t(y,z)\,dy\right] f(z)\,dz = \int_{[0,1]}\int_0^y \pi^{(y)}(z)f(z)\,dz\,P_x^*\,(X_t^* \in dy)$$
$$= \int_0^1 \left[\int_{[z,1]} \pi^{(y)}(z)P_x^*\,(X_t^* \in dy)\right] f(z)\,dz.$$

In particular, letting $x = 1$ we find

$$\int_0^1 \pi(z)\,f(z)\,dz = \int_0^1 \left[\int_{[z,1]} \pi^{(y)}(z)\,P_1^*(X_t^* \in dy)\right] f(z)\,dz.$$

Since this holds for all $f \in C[0,1]$, and since both $\pi(z)$ and the expression in square brackets on the right are continuous functions of $z \in (0,1]$, it follows, for all $z \in (0,1]$, that

$$\pi(z) = \int_{[z,1]} \frac{\pi(z)}{\Pi(y)}\,P_1^*(X_t^* \in dy),$$

and hence $\int_{[z,1]} \frac{1}{\Pi(y)}\,P_1^*(X_t^* \in dy) = 1$. It now follows that $P_1^*(X_t^* = 1) = 1$ and hence that the boundary 1 is either regular absorbing or exit. To show that the boundary is absorbing, it suffices to show that $N^*(1) < \infty$. Indeed, for fixed x in I° we have [using (3.2.8)–(3.2.9)] that

$$N^*(1) = \int_{[x,1)} [S^*(y) - S^*(x)]m^*(y)\,dy = \int_{[x,1)} \left[\frac{1}{M(x)} - \frac{1}{M(y)}\right] s(y)M^2(y)\,dy$$
$$= \int_{[x,1)} s(y)\frac{M^2(y)}{M(x)}\,dy - \int_{[x,1)} s(y)M(y)\,dy$$
$$\leq \frac{M(1)}{M(x)} \int_{[x,1)} s(y)M(y)\,dy - \int_{[x,1)} s(y)M(y)\,dy < \infty$$

where the finiteness holds since 1 is reflecting for X [hence $\Sigma(1) < \infty$] and $M(\cdot)$ is increasing and bounded on I°. □

Proposition 3.2.5. *Let X^* be a strong stationary dual of X, and let the one-parameter Markov semigroups of operators for X^* and X be (T_t^*) and (T_t) respectively. Then for all t we have $\Lambda T_t = T_t^* \Lambda$ as operators on $C[0,1]$.*

Proof. For all λ we have $\Lambda(\lambda I - A) = (\lambda I - A^*)\Lambda$ and so the resolvent operators satisfy $\Lambda R_\lambda = R_\lambda^* \Lambda$. For $f \in C[0,1]$ and $x \in [0,1]$, note that

$$(R_\lambda^* \Lambda f)(x) = \int_0^\infty e^{-\lambda t} (T_t^* \Lambda f)(x)\, dt$$

and that

$$\begin{aligned}(\Lambda R_\lambda f)(x) &= \int_0^x \pi^{(x)}(y)\, (R_\lambda f)(y)\, dy \\ &= \int_0^x \int_0^\infty \pi^{(x)}(y)\, e^{-\lambda t}\, (T_t f)(y)\, dt\, dy \\ &= \int_0^\infty e^{-\lambda t}\, (\Lambda T_t f)(x)\, dt.\end{aligned}$$

Now, by the uniqueness of Laplace transforms of real valued functions, we have $(\Lambda T_t f)(x) = (T_t^* \Lambda f)(x)$ for all t, as desired. □

Remark 3.2.6. From Proposition 3.2.5, we have that $\Lambda T_t = T_t^* \Lambda$ as operators on $C[0,1]$ which implies that the equality also holds as operators on $F[0,1]$.

The choice of 0 and 1 as instantaneously reflecting boundaries was done to streamline exposition. However, we can establish analogues of Theorem 3.2.4 for more general boundary behaviors of X. If 0 and 1 are entrance boundaries for X, then the domain of A is

$$\mathcal{D}_A = \{f \in C(I) \cap C^2(I^\circ) \mid Af \in C(I)\}.$$

If 0 (resp., 1) is made reflecting then we impose the extra condition that $\frac{df}{dS}^+(0) = 0$ [resp., $\frac{df}{dS}^-(1) = 0$] for functions $f \in \mathcal{D}_A$. In the proof of Theorem 3.2.4, only the following properties of the boundary at 0 were needed:

$$\frac{df}{dS}^+(0) = 0 \text{ for } f \in C(I), \;\; N(0) < \infty,$$

and these properties also hold if 0 is an entrance boundary. Absorption of X^* at 1 is proven completely analogously to the reflecting case. If 1 is an entrance boundary for X, then 1 is an exit boundary for X^* since

$$\begin{aligned} N^*(1) &= \int_{[x,1)} [S^*(y) - S^*(x)] m^*(y)\, dy = \int_{[x,1)} \left[\frac{1}{M(x)} - \frac{1}{M(y)} \right] s(y) M^2(y)\, dy \\ &= \int_{[x,1)} s(y) \left[\frac{M^2(y)}{M(x)} - M(y) \right] dy \\ &\geq \int_x^1 s(y)[M(y) - M(x)]\, dy = \Sigma(1) = \infty \end{aligned}$$

and (twice utilizing integration by parts)

$$\begin{aligned} \Sigma^*(1) &= \int_x^1 [S^*(1) - S^*(y)] m^*(y)\, dy = \int_x^1 \left[\frac{1}{M(y)} - \frac{1}{M(1)} \right] s(y) M^2(y)\, dy \\ &= \int_x^1 s(y) \left[M(y) - \frac{M^2(y)}{M(1)} \right] dy \\ &\leq \int_x^1 s(y)[M(1) - M(y)]\, dy = N(1) < \infty. \end{aligned}$$

We thus arrive at the following generalization of Theorem 3.2.4.

Theorem 3.2.7. *Let X be a regular diffusion on I, and assume that each of the boundary points of I is either reflecting or entrance. Assume further that $b \in C^1(I^\circ)$. If X^* is a*

strong stationary dual of X, then the generator A^* of X^* has the form

$$(A^*f)(x) = \left(\frac{1}{2}b'(x) - a(x) + b(x)\frac{\pi(x)}{\Pi(x)}\right) f'(x) + \frac{1}{2}b(x)f''(x)$$

for $x \in I^\circ$ and $f \in \mathcal{D}_{A^*}$. Also 0 is an entrance boundary for X^*. If 1 is a reflecting boundary of X, then 1 is a regular absorbing boundary of X^*. If 1 is an entrance boundary of X, then 1 is an exit boundary of X^*.

***Example* 3.2.8.** For $\alpha \geq 0$, a diffusion X on $[0,1]$ is said to be a Bessel process with parameter α [written Bes(α)], reflected at 1, if the generator of X has the form

$$A = \frac{1}{2}\frac{d^2}{dx^2} + \frac{\alpha - 1}{2x}\frac{d}{dx},$$

and if for $f \in \mathcal{D}_A$ we have $\frac{df}{dS}^-(1) = 0$. The behavior at the boundary 0 is determined by the value of α. For $0 < \alpha < 2$, the boundary 0 is a regular reflecting boundary, and for $\alpha \geq 2$ the boundary 0 is an entrance boundary. For our discussion of duality, we do not consider the case $\alpha = 0$, for which 0 is an absorbing boundary. For $\alpha > 0$, a simple application of Theorem 3.2.7 gives that if X is a Bes(α) process on $[0,1]$ with instantaneously reflecting behavior at 1 begun in $\pi^{(x)}$, then X^* is a Bes(α+2) process begun in δ_x absorbed at 1. In particular, the dual of reflecting Brownian motion, i.e., the Bes(1) process reflected at 1, is the Bes(3) process reflected at 1. For an extensive background treatment of Bessel processes, see [34, Chapter4.3] or [42, Chapter V–VI].

***Example* 3.2.9.** For a second example, we turn to the Wright–Fisher gene frequency model from population genetics. The Wright–Fisher diffusion X is a diffusion on $[0,1]$

with generator of the form

$$A = \frac{1}{2}x(1-x)\frac{d^2}{dx^2} + [\alpha(1-x) - \beta x]\frac{d}{dx}.$$

The behavior at the boundaries is determined by the values of α and β. We have that

$$0 \text{ is a(n)} \begin{cases} \text{entrance boundary if } \alpha \geq 1/2, \\ \text{reflecting regular boundary if } 0 < \alpha < 1/2, \\ \text{exit boundary if } \alpha = 0, \end{cases}$$

and

$$1 \text{ is a(n)} \begin{cases} \text{entrance boundary if } \beta \geq 1/2, \\ \text{reflecting regular boundary if } 0 < \beta < 1/2, \\ \text{exit boundary if } \beta = 0. \end{cases}$$

If X is a Wright–Fisher diffusion with $\alpha = 1/2$ and $\beta > 0$, then from Theorem 3.2.7 we have that the strong stationary dual of X is a Wright–Fisher diffusion with $\alpha^* = \alpha + (1/2)$ and $\beta^* = 0$. For an extensive background on the Wright–Fisher model and its many applications, see [29, Section 15.8] or [18, Chapter 10].

Not surprisingly, we can also recover a partial converse to Proposition 3.2.5.

Lemma 3.2.10. *Let X and X^* be diffusions on $[0,1]$ and let 0 and 1 be either instantaneously reflecting or entrance boundaries for X. Then an intertwining*

$$\Lambda T_t = T_t^* \Lambda \text{ (for all } t \geq 0) \tag{3.2.10}$$

of the one-parameter semigroups by the link Λ *together with the initial condition* (3.2.3) *implies that* X^* *is a strong stationary dual of* X.

Proof. Let $f \in \mathcal{D}_A$. Then by definition $(Af)(x) = \lim_{t\downarrow 0} \frac{T_t f(x) - f(x)}{t}$. We have that $(Af)(x) = \frac{1}{2}b(x)f''(x) + a(x)f'(x)$ is continuous in x and so the convergence of $\frac{T_t f(x) - f(x)}{t}$ to $(Af)(x)$ is uniform (Theorem 7.7.3 in [36]). Now

$$\begin{aligned}
(\Lambda A f)(x) &= \int_0^x \pi^{(x)}(y) \left[\lim_{t\downarrow 0} \frac{T_t f(y) - f(y)}{t} \right] dy \\
&= \lim_{t\downarrow 0} \int_0^x \pi^{(x)}(y) \, \frac{T_t f(y) - f(y)}{t} \, dy \\
&= \lim_{t\downarrow 0} \frac{(\Lambda T_t f)(x) - (\Lambda f)(x)}{t} \\
&= \lim_{t\downarrow 0} \frac{(T_t^* \Lambda f)(x) - (\Lambda f)(x)}{t} \\
&= (A^* \Lambda f)(x),
\end{aligned}$$

where the last limit's existence is guaranteed by that of the first. This gives both that $\Lambda|_{\mathcal{D}_A} \subset \mathcal{D}_{A^*}$, and that on $\mathcal{D}_A$ we have $\Lambda A = A^* \Lambda$ as desired. □

Remark 3.2.11. Intertwinings of Markov semigroups have been well studied, appearing for example in [16], [41], etc. In the context of (3.2.10), the transition operator Λ is the following Markov kernel from $[0, 1]$ to $[0, 1]$: For $x \in [0, 1]$ and $A \in \mathcal{B}$ we have

$$\Lambda(x, A) = \Pi^{(x)}(A).$$

Remark 3.2.12. If (3.2.10) holds, then

$$(U_t \pi_0, f) = (\pi_0, T_t f) = (\pi_0^*, \Lambda T_t f) = (\pi_0^*, T_t^* \Lambda f) = (U_t^* \pi_0^*, \Lambda f),$$

mirroring the corresponding result that algebraic duality via link L of Markov chains yields $\pi_t = \pi_t^* L$.

3.3 Approximating duality via Markov chains

The purpose of the present section is twofold. Presently suppressing all details (which will be spelled out in full detail later in the section), we will show that a suitably defined sequence of Markov chains X^Δ and their corresponding strong stationary duals $\widehat{X}^\Delta$, as defined in [12], converge respectively to our primal diffusion $Y = S(X)$ (in natural scale) and its strong stationary dual Y^*. By establishing the newly defined diffusion strong stationary dual as a limit of an appropriately defined sequence of classical Markov chain strong stationary duals, we ground our definition and our present work in the classical theory.

In addition to tethering our duality to the classical theory, this has a number of interesting consequences. For example, we believe one of the great triumphs of strong stationary duality was its application in the perfect sampling algorithms of [20] and [23]. Via the work in the present section, for our primal diffusion Y we could approximately sample perfectly from Π_Y by using the theory of [20] to perfectly sample from the stationary distributions of the approximating sequence of chains. We could also use our approximating sequence of chains to study cut-off type behaviors of the dual hitting times of state $S(1)$, and hence of the primal diffusion's separation distance from stationarity. We are also able to recover the dual-hitting-time/primal-mixing-time duality of the classical Markov chain theory in

the diffusion setting by passing to appropriate limits; see Section 3.4 for full details.

This section is laid out as follows: First assuming instantaneously reflecting boundaries for our primal diffusion Y, in Sections 3.3.1–3.3.2 we explicitly spell out the one-dimensional convergence of our primal and dual sequences of Markov chains to the corresponding primal and dual diffusions. In Section 3.3.3, we prove the corresponding convergence theorems in the case when our primal diffusion has entrance boundaries at 0 and/or 1.

3.3.1 Primal convergence

Let $D_I[0,\infty)$ be the space of cadlag functions from $[0,\infty)$ into I. We can equip $D_I[0,\infty)$ with a metric d defined by

$$d(x,y) = \inf_{\lambda \in B} \left[\left(\sup_{s>t\geq 0} \left| \log \frac{\lambda(s)-\lambda(t)}{s-t} \right| \right) \vee \int_0^\infty e^{-u} d(x,y,\lambda,u)\, du \right]$$

where B is the set of strictly increasing Lipschitz continuous functions from $[0,\infty)$ to $[0,\infty)$ with the additional property that

$$\lambda \in B \text{ implies } \sup_{s>t\geq 0} \left| \log \frac{\lambda(s)-\lambda(t)}{s-t} \right| < \infty,$$

and

$$d(x,y,\lambda,u) := \sup_{t\geq 0} \left(|x(t\wedge u) - y(\lambda(t)\wedge u)| \wedge 1 \right).$$

The topology induced by d is known as the *Skorohod topology*, and under this topology $D_I[0,\infty)$ is both complete and separable (as I is both complete and separable). For more background on $D_I[0,\infty)$, see [18, Sections 3.5–3.10] or [7, Chapters 2–3].

We will consider stochastic processes with sample paths in $D_I[0,\infty)$ as $D_I[0,\infty)$-valued random variables and we will say that $X_n \Rightarrow X$ if we have convergence in law of the corresponding $D_I[0,\infty)$-valued random variables. Note that $X_n \Rightarrow X$ implies convergence of the associated finite-dimensional distributions of X_n to those of X (see [18, Theorem 3.7.8]), i.e., for all $\{t_1,\ldots,t_m\} \subset \{t \geq 0 \,|\, \mathbb{P}(X(t) = X(t-)) = 1\}$ we have

$$(X_n(t_1),\ldots,X_n(t_m)) \Rightarrow (X(t_1),\ldots,X(t_m)).$$

As in Section 3.1, let X be a regular diffusion on I with instantaneous reflection at the boundaries of I and scale function $S \equiv S_X$. To ease exposition, we will consider $Y = S_X(X)$, a regular diffusion in natural scale on $\mathcal{S} = [S_X(0), S_X(1)]$, and assume S_Y has been scaled to make $s_Y \equiv 1$. The speed function of Y is $M_Y = M_X \circ S_X^{-1} : \mathcal{S}^\circ \to \mathbb{R}$ (where M_X is the speed function of X). As with X, define the speed measure of Y as the nonnegative measure on $\mathcal{S}^\circ$, denoted $M_Y(\cdot)$, defined via $M_Y(x,y] = M_Y(y) - M_Y(x)$.

As X is regular and $\mathbb{P}_x(X_t = 0) = 0$ for all $t > 0$ and $x \in I$, it follows that Y is regular and $\mathbb{P}_x(X_t = 0) = \mathbb{P}_{S(x)}(Y_t = S(0)) = 0$ for all $t > 0$ and $S(x) \in \mathcal{S}$. Therefore $S(0)$ is an instantaneously reflecting reflecting boundary for Y. Analogous results hold at $S(1)$, and it follows that $S(1)$ is an instantaneously reflecting boundary for Y and $M_Y(\{S(1)\}) = 0$.

The generator of Y can be expressed as $(A_Y f)(y) = \frac{1}{2} b_Y(y) f''(y)$ with $b_Y(y) = b_X(x) s_X^2(x)$ where $y = S_X(x)$. Note that $M_Y(\mathcal{S}^\circ) = M_X(I^\circ) < \infty$ and so there exists a

unique invariant measure for Y which we will denote Π_Y. Observe

$$M_Y((c,d]) = M_Y(d) - M_Y(c) = M_X(S^{-1}(d)) - M_X(S^{-1}(c))$$
$$= \int_{S^{-1}(c)}^{S^{-1}(d)} m_X(z)\,dz = \int_c^d \frac{m_X(S_X^{-1}(w))}{s_X(S_X^{-1}(w))}\,dw$$

so that M_Y (resp., Π_Y) has density $m_Y(y) = m_X(S_X^{-1}(y))/s_X(S_X^{-1}(y))$ (resp., density $\pi_Y = \alpha\, m_Y$ for some constant α). On $\mathcal{S}^\circ$ we have $m_Y = b_Y^{-1}$ and so $\pi_Y b_Y$ is constant on $\mathcal{S}^\circ$. Assume that b_Y can be extended to a function in $C(\mathcal{S})$, so that $b_Y(S(0))$ and $b_Y(S(1))$ are well defined, and assume that $\lim_{y\to z}\ \pi_Y(y)b_Y(y) = \alpha$ for $z \in \{S(0), S(1)\}$.

For the remainder of the section, we shall be working with the diffusion Y rather than X, and so we will drop the Y subscript from b_Y, π_Y, Π_Y, etc.

Let $\Delta > 0$ be such that $S(1) - S(0) = n^\Delta \Delta$ for some integer n^Δ. As in [5, Chapter 6], define a birth-and-death transition matrix P^Δ on state space

$$\mathcal{S}^\Delta := \{S(0),\ S(0)+\Delta,\ S(0)+2\Delta, \ldots, S(1)-\Delta,\ S(1)\}$$

by setting (for ease of notation, we write i for $S(0) + i\Delta$ here):

$$P^\Delta(i, i+1) = P^\Delta(i, i-1) := \frac{b(i)h}{2\Delta^2} \quad \text{for } 0 < i < n^\Delta \text{ and}$$

$$P^\Delta(0,1) := \frac{b(0)h}{\Delta^2},\ \ P^\Delta(n^\Delta, n^\Delta - 1) := \frac{b(n^\Delta)h}{\Delta^2};$$

here

$$h \equiv h_\Delta := \frac{\Delta^2}{2\sup_y b(y)} \tag{3.3.1}$$

is chosen to make P^Δ monotone.

Note that for $i \in \{1, \ldots, n^\Delta - 1\}$ we have

$$\pi(i)P^\Delta(i, i+1) = \pi(i+1)P^\Delta(i+1, i),$$

and at the boundaries we have

$$\pi(0)P^\Delta(0,1) = 2\pi(1)P^\Delta(1,0),\ \pi(n^\Delta)P^\Delta(n^\Delta, n^\Delta - 1) = 2\pi(n^\Delta - 1)P^\Delta(n^\Delta - 1, n^\Delta).$$

It follows that there exists a constant C^Δ such that

$$\pi^\Delta(i) = \begin{cases} C^\Delta \pi(i), & i = 1, \ldots, n^\Delta - 1; \\ C^\Delta \pi(i)/2, & i = 0,\ n^\Delta \end{cases} \tag{3.3.2}$$

is the unique invariant probability distribution for P^Δ.

Let π_0^Δ be a probability measure on $\mathcal{S}^\Delta$, and let P^Δ be the transition matrix for a discrete-time birth-and-death chain X^Δ, begun in π_0^Δ, on state space $\mathcal{S}^\Delta$ [we write $X^\Delta \sim (\pi_0^\Delta, P^\Delta)$ as shorthand].

Theorem 3.3.1. *Assume there exists a constant $\delta > 0$ such that $b \geq \delta$ everywhere and that we can continuously extend b to the boundaries of $\mathcal{S}$. Consider a sequence of values $\Delta \downarrow 0$ such that for each Δ we have $S(1) - S(0) = n^\Delta \Delta$ for some integer n^Δ. Define the continuous-time stochastic process Y^Δ by setting $Y_t^\Delta := X_{\lfloor t/h \rfloor}^\Delta$ for $t \geq 0$. If $Y_0^\Delta \Rightarrow Y_0$, then $Y^\Delta \Rightarrow Y$.*

Our main proof tool will be the following theorem, adapted from [18, Corollary 4.8.9 and Theorem 1.6.5]:

Theorem 3.3.2. *Let A be the generator of a regular diffusion process Y with state space $\mathcal{S}$. Assume $h_\Delta > 0$ converges to 0 as $\Delta \downarrow 0$. Let $Y_t^\Delta := X^\Delta_{\lfloor t/h_\Delta \rfloor}$ where $X^\Delta \sim (\pi_0^\Delta, P^\Delta)$ is a Markov chain with some metric state space $\mathcal{S}^\Delta \subset \mathcal{S}$, and assume $Y_0^\Delta \Rightarrow Y_0$. Define $T^\Delta : B(\mathcal{S}^\Delta) \to B(\mathcal{S}^\Delta)$ via*

$$T^\Delta f(x) = \mathbb{E}_x f(X_1^\Delta)$$

for f in the space $B(\mathcal{S}^\Delta)$ of real-valued bounded measurable functions on $\mathcal{S}^\Delta$. Define $A^\Delta := h_\Delta^{-1}(T^\Delta - I)$. Suppose that $C(\mathcal{S})$ is convergence determining and that there is an algebra $B \subset C(\mathcal{S})$ that strongly separates points. Let $\rho_\Delta : C(\mathcal{S}) \to B(\mathcal{S}^\Delta)$ be defined via $\rho_\Delta f(\cdot) = f|_{\mathcal{S}^\Delta}(\cdot)$. If

$$\lim_{\Delta \to 0} \sup_{y \in \mathcal{S}^\Delta} \left| (A^\Delta \rho_\Delta f)(y) - (Af)(y) \right| = 0 \tag{3.3.3}$$

for all $f \in \mathcal{D}_A$, then $Y^\Delta \Rightarrow Y$.

The adaptation of Theorem 3.3.2 from [18, Corollary 4.8.9 and Theorem 1.6.5] is spelled out explicitly in Appendix B, as the notation between [18] and the present section differs considerably.

Proof of Theorem 3.3.1. Clearly $C(\mathcal{S})$ is convergence determining, and by considering suitably smooth uniform approximations to the indicator function of $\{x\}$ in $\mathcal{D}_A$ for each $x \in \mathcal{S}$, it follows that $\mathcal{D}_A \subset C(\mathcal{S})$ is an algebra that strongly separates points. Let $f \in \mathcal{D}_A$, so that $(Af)(y) = \frac{1}{2} b(y) f''(y)$. Using

$$\mathcal{D}_A = \{ f \in C(\mathcal{S}) \cap C^2(\mathcal{S}^\circ) \mid Af \in C(\mathcal{S}),\ f'(S(0)+) = f'(S(1)-) = 0 \},$$

we find that $bf'' \in C(\mathcal{S})$. As $b(y) \geq \delta > 0$ for all $y \in \mathcal{S}$, we have $1/b \in C(\mathcal{S})$ and therefore $f'' \in C(\mathcal{S})$. It follows that as $\Delta \downarrow 0$, uniformly for $y \in \mathcal{S}^\Delta \setminus \{S(0), S(1)\}$ we have

$$h^{-1}((T^\Delta - I)f)(y) = \frac{1}{2}b(y)\frac{f(y+\Delta) - 2f(y) + f(y-\Delta)}{\Delta^2} \to \frac{1}{2}b(y)f''(y).$$

Likewise,

$$h^{-1}((T^\Delta - I)f)(S(0)) = b(S(0))\frac{f(S(0)+\Delta) - f(S(0))}{\Delta^2} \to \frac{1}{2}b(S(0))f''(S(0)),$$
$$h^{-1}((T^\Delta - I)f)(S(1)) = b(S(1))\frac{f(S(1)-\Delta) - f(S(1))}{\Delta^2} \to \frac{1}{2}b(S(1))f''(S(1)).$$

Therefore

$$\sup_{y \in \mathcal{S}^\Delta} \left|(A^\Delta \rho_\Delta f)(y) - (Af)(y)\right| \to 0,$$

establishing (B.1). The result follows. □

Remark 3.3.3. For $x \in (S(0), S(1))$, let $i_{\Delta,x} := \lfloor [x - S(0)]/\Delta \rfloor$, and denote the invariant measure π^Δ truncated to $\{S(0), S(0)+\Delta, \ldots, S(0)+i_{\Delta,x}\Delta\}$ by $\pi^{\Delta, i_{\Delta,x}}$. If Y is begun with density $\pi^{(x)}$, then for $y \in (S(0)+k\Delta, S(0)+(k+1)\Delta)$ with $0 \leq k < i_{\Delta,x}$ we have

$$\mathbb{P}^\Delta(Y_0^\Delta \leq y) = \frac{\Delta \sum_{j=0}^{k} \pi^\Delta(S(0)+j\Delta)}{\Delta \sum_{j=0}^{i_{\Delta,x}} \pi^\Delta(S(0)+j\Delta)} \to \frac{\int_{S(0)}^{y} \pi(z)\,dz}{\int_{S(0)}^{x} \pi(z)\,dz} = \mathbb{P}(Y_0 \leq y).$$

If X^Δ is begun in $\pi^{\Delta, i_{\Delta,x}}$, it follows that $X_0^\Delta = Y_0^\Delta \Rightarrow Y_0$. If instead Y is begun deterministically at $S(0)$, then letting $X_0^\Delta = S(0)$ for all Δ again gives $X_0^\Delta = Y_0^\Delta \Rightarrow Y_0$.

Remark 3.3.4. For the sequence of birth-and-death chains (π_0^Δ, P^Δ), where either $\pi_0^\Delta = \delta_{S(0)}$ for each Δ or $x \in (S(0), S(1))$ is given and $\pi_0^\Delta = \pi^{\Delta, i_{\Delta,x}}$ for each Δ, we are guaranteed the existence of a sequence of birth-and-death strong stationary dual chains by [12, eqs. (4.16a)–(4.16b)] because of the following two observations.

(a) P^Δ is monotone. Indeed, for $i = 0, \ldots, n^\Delta - 1$ we easily see

$$P^\Delta(i, i+1) + P^\Delta(i+1, i) \leq 1.$$

(b) The ratio $\pi_0^\Delta / \pi^\Delta$ of probability mass functions (initial to stationary) is non-increasing.

3.3.2 Dual convergence

As in [12], construct on the same probability space as for X^Δ a strong stationary dual $\widehat{X}^\Delta \sim (\widehat{\pi}_0{}^\Delta, \widehat{P}^\Delta)$ of X^Δ using the link Λ of truncated stationary distributions [here, for ease of notation, i is again used as shorthand for $S(0) + i\Delta$]:

$$\Lambda^\Delta(i, j) := \mathbb{1}\{j \leq i\} \frac{\pi^\Delta(j)}{H^\Delta(i)};$$

we have used the shorthand $H^\Delta(i) := \sum_{j=0}^{i} \pi^\Delta(j)$. Note that $\widehat{X}^\Delta$ is also a birth-and-death chain on $\mathcal{S}^\Delta$. Assume either that $X_0^\Delta = S(0)$ for every Δ so that $\widehat{X}_0^\Delta = S(0)$ for every Δ as well, or that $X_0^\Delta \sim \pi^{\Delta, i_{\Delta,x}}$ for every Δ so that $\widehat{X}_0^\Delta = S(0) + i_{\Delta,x}\Delta$. The one-step transition matrix $\widehat{P}^\Delta$ for $\widehat{X}^\Delta$ is given by

$$\widehat{P}^\Delta(i, i-1) = \frac{H^\Delta(i-1)}{H^\Delta(i)} \frac{b(i)h}{2\Delta^2} = \frac{b(i)h}{2\Delta^2} - \frac{h \cdot \alpha \cdot C^\Delta}{H^\Delta(i) 2\Delta^2} \quad \text{for } 0 < i < n^\Delta, \tag{3.3.4}$$

$$\widehat{P}^\Delta(i, i+1) = \frac{H^\Delta(i+1)}{H^\Delta(i)} \frac{b(i+1)h}{2\Delta^2} = \frac{h \cdot \alpha \cdot C^\Delta}{H^\Delta(i) 2\Delta^2} + \frac{b(i+1)h}{2\Delta^2} \quad \text{for } 0 < i < n^\Delta, \tag{3.3.5}$$

$$\widehat{P}^\Delta(0, 1) = \frac{H^\Delta(1)}{H^\Delta(0)} \frac{b(1)h}{\Delta^2}, \tag{3.3.6}$$

$$\widehat{P}^\Delta(n_\Delta, n_\Delta) = 1, \tag{3.3.7}$$

with $\widehat{P}^\Delta(i, i)$ having values for $0 \leq i < n^\Delta$ so that the rows of $\widehat{P}^\Delta$ sum to unity. We next show the following theorem:

Theorem 3.3.5. *Assume that $b \in C^2(S(0), S(1)]$, and $\pi > 0$ on $\mathcal{S}$. Define the continuous-time processes $(\widehat{Y}_t^\Delta) := (\widehat{X}^\Delta_{\lfloor t/h \rfloor})$, and assume $\widehat{Y}_0^\Delta \Rightarrow Y_0^*$. Then*

$$\widehat{Y}^\Delta \Rightarrow Y^*,$$

where Y^* is a SSD of Y in the sense of Definition 3.2.1.

We will prove Theorem 3.3.5 after a series of preliminary results. We begin by putting Y^* into natural scale, i.e., consider the diffusion $Z^* = S^*(Y^*) = -1/M(Y^*)$ on state space $\mathcal{S}^* = (S^*(S(0)), S^*(S(1))]$. Note that the infinitesimal parameters of Z^* are given on $(-\infty, S^*(S(1)))$ [recalling (3.2.7)] by

$$a_{Z^*} \equiv 0, \;\; b_{Z^*}(S^*(y)) = b(y)s^*(y)^2 = \frac{b(y)m^2(y)}{M^4(y)} = \alpha^2 \frac{b(y)\pi^2(y)}{\Pi^4(y)}$$

(recall α is the constant such that $\pi = \alpha \cdot m = \alpha/b$). Note also that under the assumptions of Theorem 3.3.5, we have $b \in C^2(S(0), S(1)]$ [and so $\pi(\cdot) \propto b(\cdot)$ implies $\pi \in C^2(S(0), S(1)]$], $\pi > 0$ on $\mathcal{S}$. Note also that

$$(S^*)'(i) = \alpha \frac{\pi(i)}{\Pi^2(i)},$$

$$(S^*)''(i) = \alpha \frac{\Pi^2(i)\pi'(i) - 2\pi^2(i)\Pi(i)}{\Pi^4(i)},$$

$$(S^*)'''(i) = \alpha \frac{\Pi^2(i)\pi''(i) - 2\pi'(i)\Pi(i)\pi(i)}{\Pi^4(i)} - \alpha \frac{4\Pi^3(i)\pi(i)\pi'(i) - 6\pi^3(i)\Pi^2(i)}{\Pi^6(i)},$$

which implies that $S^* \in C^3[-S(0) + i_0\Delta, S(1)]$ for any $i_0 > 0$.

Define

$$\widehat{Z}_t^\Delta := S^* \left(\widehat{X}_t^\Delta \right),$$

and note that this is a birth and death chain on state space

$$\mathcal{S}^{*,\Delta} := \{S^*(S(0)),\ S^*(S(0)+\Delta), \ldots, S^*(S(1)-\Delta), S^*(S(1))\}.$$

Define

$$a_{\widehat{Z}^\Delta}(x) := \frac{1}{h_\Delta}\mathbb{E}_x\left(\widehat{Z}_1^\Delta - x\right), \tag{3.3.8}$$

where we require that $x = S^*(S(0)+i\Delta)$ for some nonnegative integer i.

Proposition 3.3.6. *With the assumptions of Theorem 3.3.5, letting $R < \infty$ be fixed, it follows that*

$$\lim_{\Delta\downarrow 0}\sup_{|x|<R}|a_{\widehat{Z}^\Delta}(x)| = 0.$$

Proof. Abbreviate $S(0)+i\Delta$ as i, and then x is of the form $x = S^*(i)$. Let

$$\mathrm{A} := \frac{S^*(i+1)-2S^*(i)+S^*(i-1)}{\Delta^2}\frac{b(i)}{2};\ \ \mathrm{B} := \frac{S^*(i+1)-S^*(i)}{\Delta}\frac{b(i+1)-b(i)}{2\Delta};$$
$$\mathrm{C} := \frac{\alpha\cdot C^\Delta}{\Delta H^\Delta(i)}\frac{S^*(i+1)-S^*(i-1)}{2\Delta}.$$

Then (3.3.4)–(3.3.5) allow us to rewrite (3.3.8) as

$$a_{\widehat{Z}^\Delta}(x) = \mathrm{A}+\mathrm{B}+\mathrm{C}$$

for $x \neq S^*(S(1))$. For all $|x| = S^*(i) < R$ we have $0 < \delta < i\Delta < \gamma < \infty$ uniformly in Δ for some γ and δ. A Taylor expansion of $S^*(\cdot)$ combined with $S^* \in C^3[\delta, S(1)]$ gives

$$\lim_{\Delta\downarrow 0}\sup_{|x|<R,\, x\neq S^*(S(1))}\left|\mathrm{A} - \alpha\frac{\Pi^2(i)\pi'(i)-2\pi^2(i)\Pi(i)}{\Pi^4(i)}\frac{b(i)}{2}\right| = 0,$$

or equivalently

$$\lim_{\Delta\downarrow 0}\sup_{|x|<R,\, x\neq S^*(S(1))}\left|\mathrm{A} - \frac{\alpha^2}{2}\left(\frac{\pi'(i)}{\pi(i)\Pi^2(i)} - \frac{2\pi(i)}{\Pi^3(i)}\right)\right| = 0;$$

and

$$\lim_{\Delta \downarrow 0} \sup_{|x|<R,\, x \neq S^*(S(1))} \left| \mathrm{B} - \alpha \frac{\pi(i)}{\Pi^2(i)} \frac{b'(i)}{2} \right| = 0,$$

or equivalently

$$\lim_{\Delta \downarrow 0} \sup_{|x|<R,\, x \neq S^*(S(1))} \left| \mathrm{B} + \alpha^2 \frac{\pi'(i)}{2\Pi^2(i)\pi(i)} \right| = 0.$$

Next, for C note that $(C^\Delta)^{-1}\Delta H^\Delta(i) > 0$ if $i\Delta > \delta$ (which is satisfied uniformly by all $|x| < R$). The function $\pi \circ S^*(\cdot)$ is uniformly continuous on $[-R, R]$, and, since the Riemann sum of a uniformly continuous function converges uniformly to the corresponding Riemann integral, we have for the sup-norm $\|\cdot\|_\infty$ on $[-R, R]$ that

$$\lim_{\Delta \downarrow 0} ||(C^\Delta)^{-1}\Delta H^\Delta - \Pi||_\infty = 0.$$

By regularity of the primal diffusion Y, we have $\Pi(i) > 0$ for $i\Delta > \delta > 0$, and therefore for such i we have $(C^\Delta)^{-1}\Delta H^\Delta(i)\Pi(i) > 0$. Note that $(C^\Delta)^{-1}\Delta H^\Delta(\cdot)\Pi(\cdot)$ is a bounded increasing function in i. All of this leads to

$$\lim_{\Delta \downarrow 0} \sup_{|x|<R,\, x \neq S^*(S(1))} \left| \frac{C^\Delta}{\Delta H^\Delta(i)} - \frac{1}{\Pi(i)} \right| =$$

$$\lim_{\Delta \downarrow 0} \sup_{|x|<R,\, x \neq S^*(S(1))} \left| \frac{C^\Delta}{\Delta H^\Delta(i)\Pi(i)} \right| \cdot \left| \Pi(i) - \frac{\Delta H^\Delta(i)}{C^\Delta} \right| = 0.$$

Therefore

$$\lim_{\Delta \downarrow 0} \sup_{|x|<R,\, x \neq S^*(S(1))} \left| \mathrm{C} - \frac{\alpha^2 \pi(i)}{\Pi(i)^3} \right| = 0.$$

Combining our results for A, B, and C with the observations that $a_{\widehat{Z}^\Delta}(S^*(S(1))) = 0$, we find

$$\lim_{\Delta \downarrow 0} \sup_{|x|<R} |a_{\widehat{Z}^\Delta}(x)| = 0,$$

as desired. □

Next, define

$$b_{\widehat{Z}^\Delta}(x) := \frac{1}{h_\Delta}\mathbb{E}_x\left(\widehat{Z}_1^\Delta - x\right)^2, \tag{3.3.9}$$

where again we require that $x = S^*(S(0) + i\Delta)$ for some nonnegative integer i.

Proposition 3.3.7. *With the assumptions of Theorem 3.3.5, letting $R < \infty$ be fixed, we have*

$$\lim_{\Delta\downarrow 0}\sup_{|x|<R}\left|b_{\widehat{Z}^\Delta}(x) - b_{Z^*}(x)\right| = 0,$$

where $b_{Z^*}(S^*(S(1))) = 0$ by the absorbing behavior of the boundary at $S^*(S(1))$.

Proof. We have

$$b_{Z^*}(x) = \alpha^2\frac{\pi^2((S^*)^{-1}(x))}{\Pi^4((S^*)^{-1}(x))}b((S^*)^{-1}(x))$$

for $x \neq S^*(S(1))$. There exists a constant $\delta > 0$ such that for all Δ and all x satisfying $|x| < R$, if we write $x = S^*(S(0) + i\Delta)$ then then $i\Delta \geq \delta$. Let

$$\mathrm{A} := (S^*(i+1) - S^*(i))^2\frac{\alpha\cdot C^\Delta}{2\Delta^2 H^\Delta(i)};\ \ \mathrm{B} := \frac{(S^*(i+1) - S^*(i))^2}{\Delta^2}\frac{b(i+1)}{2};$$

and

$$\mathrm{C} := \frac{(S^*(i-1) - S^*(i))^2}{\Delta^2}\frac{b(i)}{2};\ \ \mathrm{D} := -(S^*(i-1) - S^*(i))^2\frac{\alpha\cdot C^\Delta}{2\Delta^2 H^\Delta(i)}.$$

Note that $b_{\widehat{Z}^\Delta}(x) = \mathrm{A} + \mathrm{B} + \mathrm{C} + \mathrm{D}$ for $x \neq S^*(S(1))$.

As in the proof of Proposition 3.3.6,

$$\lim_{\Delta\downarrow 0}\sup_{|x|<R,\, x\neq S^*(S(1))}\left|\mathrm{B} - \alpha^2\frac{\pi^2(i)}{2\Pi^4(i)}b(i)\right| = 0;$$

$$\lim_{\Delta \downarrow 0} \sup_{|x|<R \; x \neq S^*(S(1))} \left| \mathrm{C} - \alpha^2 \frac{\pi^2(i)}{2\Pi^4(i)} b(i) \right| = 0.$$

Rewrite A (with analogous results holding for D) as

$$\mathrm{A} = (S^*(i+1) - S^*(i)) \frac{S^*(i+1) - S^*(i)}{\Delta} \frac{\alpha \cdot C^\Delta}{2\Delta H^\Delta(i)}.$$

From the uniform continuity of $(S^*)''(\cdot)$ on bounded intervals, $\frac{S^*(i+1)-S^*(i)}{\Delta}$ converges uniformly for $|x| = |S^*(i)| < R$ to $(S^*)'(i)$, which is uniformly bounded for $|x| = |S^*(i)| < R$. Also, $(C^\Delta)^{-1} 2\Delta H^\Delta(i)$ is bounded away from 0 for $|x| = |S^*(i)| < R$, and $S^*(\cdot)$ is uniformly continuous for $|x| = |S^*(i)| < R$. Hence

$$\lim_{\Delta \downarrow 0} \sup_{|x|<R,\, x \neq S^*(S(1))} |\mathrm{A}| = 0; \qquad \lim_{\Delta \downarrow 0} \sup_{|x|<R,\, x \neq S^*(S(1))} |\mathrm{D}| = 0.$$

Lastly, note that $b_{\widehat{Z}^\Delta}(S^*(S(1))) = 0$, which finishes the proof. □

We are now ready to prove Theorem 3.3.5:

Proof of Theorem 3.3.5. With x fixed so that $\mathbb{P}(Z^*)^{-1} = \delta_{S^*(x)}$, let R be such that $|S^*(x)| < R$ and $|S^*(S(1))| < R$, and define

$$\tau_R := \inf\{t \geq 0 \; : \; |Z_t^*| = R\}.$$

It follows that $Z^*(\cdot \wedge \tau_R)$ is equal in distribution to the diffusion process with state space $\mathcal{S}_R^* := [-R, S^*(S(1))]$ and generator

$$A_R^* := \frac{1}{2} b_{Z^*}(\cdot) \frac{\partial^2}{\partial x^2}$$

operating on the domain

$$\mathcal{D}_R := \left\{ f \in C(\mathcal{S}_R^*) \cap C^2[(\mathcal{S}_R^*)^\circ] \,\middle|\, A_R^* f \in C(\mathcal{S}_R^*),\; A_R^* f(-R) = A_R^* f(S^*(S(1))) = 0 \right\}.$$

Since $b_{Z^*} > 0$ on $\mathcal{S}^*$, it follows that if $f \in \mathcal{D}_R$ then $f \in C^2(\mathcal{S}_R^*)$. Let

$$\tau_R^\Delta := \inf\{t \geq 0 \ : \ |\widehat{Z}_t^\Delta| \geq R\},$$

and define the sequence of absorbing Markov chains $\widehat{V}^\Delta(\cdot) := \widehat{Z}^\Delta(\cdot \wedge \tau_R^\Delta)$ with state space

$$\mathcal{S}_R^{*,\Delta} := \{x \in \mathcal{S}^{*,\Delta} : x \leq \lceil R \rceil_\Delta\}$$

where $\lceil y \rceil_\Delta$ "rounds" y to the smallest element $\geq y$ in the grid $\{S^*(S(0)),\, S^*(S(0) + \Delta), \ldots, S^*(S(1) - \Delta),\, S^*(S(1))\}$. Lastly define $V^{*,\Delta}(t) := \widehat{V}^\Delta(\lfloor t/h \rfloor)$.

From [18, Corollary 4.8.9], to prove that $V^{*,\Delta}$ converges (as $\Delta \downarrow 0$) to $Z^*(\cdot \wedge \tau_R)$, it suffices to show that for each fixed $f \in \mathcal{D}_R$ we have

$$\lim_{\Delta \to 0} \sup_{x \in \mathcal{S}_R^{*,\Delta}} |\rho_\Delta f(x) - f(x)| = 0 = \lim_{\Delta \to 0} \sup_{x \in \mathcal{S}_R^{*,\Delta}} |\rho_\Delta A_R^* f(x) - \widehat{A}^\Delta f(x)| \tag{3.3.10}$$

where $\widehat{A}^\Delta f(x) := h_\Delta^{-1}[\mathbb{E}_x f(\widehat{V}_1^\Delta) - f(x)]$. The first equality in (3.3.10) is trivial. Consider the second equality. At $x = -R$ or $x = S^*(S(1))$, we have

$$|\rho_\Delta A_R^* f(x) - \widehat{A}^\Delta f(x)| = 0,$$

since both $\rho_\Delta A_R^* f(x)$ and $\widehat{A}^\Delta f(x)$ equal 0 for $x = -R$ or $x = S^*(S(1))$. For x in the interior of $\mathcal{S}_R^{*,\Delta}$, from a Taylor expansion of f with remainder in intermediate-point form we find

$$\left| \widehat{A}^\Delta f(x) - \left[f'(x) a_{\widehat{Z}^\Delta}(x) + \frac{f''(x)}{2} b_{\widehat{Z}^\Delta}(x) \right] \right| \leq \frac{c}{2} b_{\widehat{Z}^\Delta}(x), \tag{3.3.11}$$

where, with $x = S^*(S(0) + i\Delta)$, we take

$$c = \max\{|f''(S^*(S(0) + (i-1)\Delta)) - f''(x)|,\ |f''(S^*(S(0) + (i+1)\Delta) - f''(x)|\}.$$

From (3.3.11), Proposition 3.3.6, Proposition 3.3.7, and the fact that $f \in C^2(\mathcal{S}_R^*)$, we have that $\widehat{A}^\Delta f(x)$ converges uniformly to $A_R^* f(x)$, and so (3.3.10) is proven.

We have now established that $V^{*,\Delta}$ converges in distribution to $Z^*(\cdot \wedge \tau_R)$. The relative compactness of $V^{*,\Delta}$ follows as in the proof of [18, Theorem 7.4.1], and therefore [45, Theorem 11.1.1] implies $Z^{*,\Delta} \Rightarrow Z^*$. Lastly, noting that $(S^*)^{-1}(\cdot)$ is well-defined and measurable (indeed it is continuous!) we have that

$$\widehat{Y}^\Delta = (S^*)^{-1}(Z^{*,\Delta}) \Rightarrow (S^*)^{-1}(Z^*) = Y^*$$

by [18, Theorem 3.10.2]. □

3.3.3 Convergence extended to entrance boundary cases

For 0 an entrance boundary of X and 1 reflecting, again consider $Y = S(X)$, a regular diffusion in natural scale on $\mathcal{S} = [-\infty, S(1)]$ begun in $\pi_0 = \pi^{(x)}$, the stationary measure for Y truncated (conditioned) to $(-\infty, x)$ for some $x \in \mathcal{S}$. If $b(\cdot)$ is bounded away from both 0 and ∞ on $\mathcal{S}$, then the constructions of the approximating primal and dual chains are identical to the case where 0 is reflecting, and details are omitted. However, if $\lim_{x \to -\infty} b(x) = \infty$, then the approximating sequences of Markov chains need to be defined differently.

To this end, on $\mathcal{S}^\Delta := \{S(1) - i_\Delta \Delta, \ldots, S(1) - \Delta, S(1)\}$, with i_Δ chosen so that $i_\Delta \Delta \to \infty$, define a birth-and-death transition matrix P^Δ via (here using the shorthand i

for $S(1) - i\Delta$)

$$P^{\Delta}(i, i+1) = P^{\Delta}(i, i-1) := \frac{b(i)h_{\Delta}}{2\Delta^2} \text{ for } 0 < i < i_{\Delta}$$

$$P^{\Delta}(i_{\Delta}, i_{\Delta} - 1) := \frac{b(i_{\Delta})h_{\Delta}}{\Delta^2},$$

$$P^{\Delta}(0, 1) := \frac{b(0)h_{\Delta}}{\Delta^2},$$

with $P^{\Delta}(i, i)$ chosen to make the row sums of P^{Δ} equal to 1, and

$$h_{\Delta} := \frac{\Delta^2}{4 \cdot \sup_{i \leq i_{\Delta}} b(i)}$$

chosen again to ensure monotonicity. For an initial probability distribution π_0^{Δ} on $\mathcal{S}^{\Delta}$, consider a birth-and-death Markov chain $X^{\Delta} \sim (\pi_0^{\Delta}, P^{\Delta})$. Let the stationary distribution of X^{Δ} be denoted π^{Δ}. Let

$$i_{\Delta,x} := \lfloor [S(1) - x]/\Delta \rfloor,$$

and assume that $\pi_0^{\Delta} := \pi^{\Delta, i_{\Delta,x}}$ is π^{Δ} truncated (conditioned) to $\{S(1) - i_{\Delta}\Delta, \ldots, S(1) - i_{\Delta,x}\Delta\}$. Again note that $\pi_0^{\Delta} \Rightarrow \pi^{(x)}$.

The following theorem is proven in a similar fashion to Theorem 3.3.1, and so the proof will be sketched with some detail omitted (see Appendix B for notation).

Theorem 3.3.8. *Assume $b(\cdot)$ is continuous and bounded away from 0 over $(-\infty, S(1)]$. Let $\mathbb{P}Y^*(0)^{-1} = \pi^{(x)}$ and, as above, let $X^{\Delta} \sim (\pi_0^{\Delta}, P^{\Delta})$ with π_0^{Δ} equal to π^{Δ} truncated (conditioned) to $\{S(1) - i_{\Delta}\Delta, \ldots, S(1) - i_{\Delta,x}\Delta\}$. Define the continuous-time stochastic process Y^{Δ} by setting $Y_t^{\Delta} := X^{\Delta}_{\lfloor t/h_{\Delta} \rfloor}$ for $t \geq 0$. Then $Y^{\Delta} \Rightarrow Y$.*

Proof. Fix R such that $S(1) < R < \infty$. With

$$\tau_R^\Delta := \inf\{t \geq 0 : |Y_t^\Delta| \geq R\},$$

consider

$$Z^\Delta(\cdot) := Y^\Delta(\cdot \wedge \tau_R^\Delta) = X^\Delta\left(\left\lfloor \frac{\cdot \wedge \tau_R^\Delta}{h_\Delta} \right\rfloor\right).$$

With

$$\tau_R := \inf\{t \geq 0 : |Y_t| = R\},$$

let $Z(\cdot) := Y(\cdot \wedge \tau_R)$. Denote the generator of Z by A_Z, with domain $\mathcal{D}(A_Z)$. Writing (T_R^Δ) for the transition semigroup associated with the Markov chain X^Δ absorbed at absolute value R, let $A_R^\Delta := h_\Delta^{-1}(T_R^\Delta - I)$. Just as we showed (B.1) in the proof of Theorem 3.3.1, here we can show that

$$\lim_{\Delta \to \infty} \sup_{x \in [-R, S(1)]} \left|(A_R^\Delta \rho_\Delta f)(x) - (A_Z f)(x)\right| = 0 \tag{3.3.12}$$

for all $f \in \mathcal{D}(A_Z)$.

By [18, Corollary 4.8.9], we have (see Appendix B) that $Z^\Delta \Rightarrow Z$. The proof is finished by applying [45, Theorem 11.1.1] to see that $Y^\Delta \Rightarrow Y$ as desired. □

Let Y^* be a SSD of Y, and let Z^* be Y^* put into natural scale. Form the dual Markov chain to X^Δ, and denote the dual by $\widehat{X}^\Delta \sim (\delta_x, \widehat{P}^\Delta)$. The following proposition gives the dual-convergence theorem analogous to Theorem 3.3.5.

Theorem 3.3.9. *With the same assumptions as in Theorem 3.3.8, further assume* $b(\cdot) \in C^2(-\infty, S(1)]$ *and*

$$\inf_{y \in \mathcal{S}} y^4 m(M^{-1}(-1/y)) > 0. \tag{3.3.13}$$

For $t \geq 0$, define $\widehat{Y}^{\Delta}(t) := \widehat{X}^{\Delta}(\lfloor t/h_{\Delta} \rfloor)$. Then $\widehat{Y}^{\Delta} \Rightarrow Y^*$.

Proof. The proof follows along the same path as the proof of Theorem 3.3.5 and so details are omitted. The only wrinkle here is the assumption (3.3.13), which is a technical condition needed to make the infinitesimal variance of the dual diffusion in natural scale bounded away from 0, which we exploited in the proof of Theorem 3.3.5. □

Remark 3.3.10. Under some mild assumptions, the above theory can easily be extended to the case where both 0 and 1 are entrance boundaries for X. For example, if X is in natural scale, it is sufficient that b_X is bounded away from 0 and twice continuously differentiable on $\mathbb{R}$. The analogues of Theorem 3.3.1 and Theorem 3.3.5 can be easily recovered. Details are omitted.

3.4 Separation and hitting times

In the Markov chain setting, strong stationary duality gives that the separation mixing time in the primal chain is equal in law to a suitable absorption time in the dual chain. By studying and bounding the absorption time, which is sometimes more tractable than direct consideration of the mixing time, we can tightly bound the separation mixing time in our primal chain. See [12] for further detail. Spelling this out more fully, if $X \sim (\pi_0, P)$ is an ergodic discrete-time Markov chain with state space S, stationary distribution π, and with SSD (as defined in [12]) $X^* \sim (\pi_0^*, P^*)$ absorbing in m, then for every t we have

$$\operatorname{sep}(t) := \sup_{i \in S} \left(1 - \frac{\pi_t(i)}{\pi(i)}\right) \leq \mathbb{P}_{\pi_0^*}(T_m^* > t). \tag{3.4.1}$$

Under some monotonicity conditions, for example if the primal is a MLR chain on a linearly ordered state space, the inequality in (3.4.1) can be made to be an equality for every t by a suitable formation of the dual chain.

In our present diffusion setting, with X a regular diffusion on $[0, 1]$ with either reflecting or entrance behavior at the boundaries, we would like to recover a result similar to (3.4.1). Let Π be the invariant distribution for X, let $X_0 \sim \Pi_0$, and, given $t > 0$, let Π_t be the corresponding distribution of X_t. If $\Pi_t \ll \Pi$, define

$$a(t) := \operatorname{ess\,inf} R_t = \sup\left\{r \,\middle|\, \Pi(R_t < r) = 0\right\}$$

to be the essential infimum (with respect to Π) of (any version of) the Radon–Nikodym derivative $R_t := d\Pi_t/d\Pi$. We define the *separation* of the diffusion from Π at time t as follows:

$$\operatorname{sep}(\pi_t, \pi) := 1 - a(t). \tag{3.4.2}$$

To simplify the notation, we shall write $\operatorname{sep}(t)$ for $\operatorname{sep}(\pi_t, \pi)$ unless the full notation is needed to avoid confusion.

Claim 3.4.1. *Let* $\operatorname{sep}(t) = \operatorname{sep}(\pi_t, \pi)$ *be defined as above. Then*

(a) We have $0 \le \operatorname{sep}(t) \le 1$.

(b) For each t *we have* $\operatorname{sep}(t) = 0$ *if and only if* $\Pi_t = \Pi$.

(c) For any Π_0 *we have* $\Pi_t \ll \Pi$ *for all* $t > 0$.

(d) The separation $\operatorname{sep}(t)$ *is non-increasing in* t.

Proof. For (a), we show equivalently that $0 \leq a(t) \leq 1$. To this end, let R_t be (any version of) the Radon–Nikodym derivative $d\Pi_t/d\Pi$. Since $R_t(y) \geq 0$ for all y, we have $a(t) \geq 0$. But also

$$1 = \int_0^1 \Pi_t(dy) = \int_0^1 R_t(y)\,\Pi(dy) \geq a(t) \int_0^1 \Pi(dy) = a(t), \tag{3.4.3}$$

finishing the proof.

For (b), note that if $\Pi_t = \Pi$, we can take $R_t \equiv 1$ as a version of the Radon–Nikodym derivative $d\Pi_t/d\Pi$, and thence $\mathrm{sep}(t) = 0$. Conversely, if $\mathrm{sep}(t) = 0$, then $a(t) = 1$ and (3.4.3) is an equality; therefore $R_t = 1$ almost surely with respect to Π, and so $\Pi_t = \Pi$.

For (c), let $x \in (0, 1)$. When $\Pi_0 = \delta_x$, regularity of X guarantees the existence of a density for Π_t with respect to Π, call it $f_x(\cdot)$. For any Π_0, it follows that the Π_0-mixture of the densities $f_x(\cdot)$ is a density for Π_t with respect to Π [and so $\mathrm{sep}(t)$ is well defined].

For (d), for each $s > 0$ let $R_s = d\Pi_s/d\Pi$. Let $0 < t < u$ and note for any $A \in \mathcal{B}$, the Borel σ-field of $[0, 1]$, that

$$\begin{aligned}\int_A R_u(y)\,\Pi(dy) = \Pi_u(A) &= \int_0^1 P_{u-t}(x, A)\,\Pi_t(dx) = \int_0^1 R_t(x)\,P_{u-t}(x, A)\,\Pi(dx) \\ &\geq a(t) \int_0^1 P_{u-t}(x, A)\,\Pi(dx) = a(t)\,\Pi(A) = \int_A a(t)\,\Pi(dy).\end{aligned}$$

Hence $R_u \geq a(t)$ almost surely with respect to Π. Hence $a(u) \geq a(t)$, and therefore $\mathrm{sep}(u) \leq \mathrm{sep}(t)$, as desired. □

As in the discrete setting, we are able to bound $\mathrm{sep}(t)$ in our primal diffusion X using the absorption time in state 1 of our dual diffusion. In the diffusion setting, by virtue of diffusions being stochastically monotone, the inequality in (3.4.1) is an equality without

needing further assumptions. Spelling this out:

Lemma 3.4.2. *Let X be a regular diffusion on $[0,1]$ begun in Π_0, let X have either reflecting or entrance behavior at the boundaries, and let Π be the stationary measure for X. Let T_1^* be the hitting time of state 1 in the SSD diffusion X_t^* (as defined in Definition 3.2.1) begun in Π_0^* satisfying* (3.2.3)*. Then*

$$\operatorname{sep}(t) = \mathbb{P}_{\Pi_0^*}(T_1^* > t) = 1 - \mathbb{P}_{\Pi_0^*}(X_t^* = 1).$$

Proof. Let $f \in F[0,1]$. By Remark 3.2.12, we have for all $t > 0$ that $(\Pi_t, f) = (\Pi_t^*, \Lambda f)$. Therefore, writing $R_t = d\Pi_t / d\Pi$ as usual, we have

$$\begin{aligned}
\int_{[0,1]} \pi(x) R_t(x) f(x)\, dx &= \int_{[0,1]} \Pi(dx)\, R_t(x) f(x) \\
&= \int_{[0,1]} \Pi_t(dx)\, f(x) \\
&= \int_{[0,1]} \Pi_t^*(dx) \int_{[0,x]} \pi^{(x)}(y) f(y)\, dy \\
&= \int_{[0,1]} \int_{[y,1]} \Pi_t^*(dx)\, \pi^{(x)}(y) f(y)\, dy.
\end{aligned}$$

This holds for all $f \in F[0,1]$, and so

$$R_t(y) = \int_{[y,1]} \frac{\Pi_t^*(dx)}{\Pi(x)} \tag{3.4.4}$$

for Lebesgue-a.e. (i.e., for Π-a.e.) y. Thus $\Pi(R_t < r) = 0$ if and only if the right side of (3.4.4) is at least r for Π-a.e. y, or, equivalently, $\Pi_t^*(\{1\})/\Pi(1) = \Pi_t^*(\{1\}) \geq r$. Therefore $a(t) = \Pi_t^*(\{1\}) = \mathbb{P}_{\Pi_0^*}(X_t^* = 1)$ and so $\operatorname{sep}(t) = 1 - \mathbb{P}_{\Pi_0^*}(X_t^* = 1)$. □

Remark 3.4.3. We can also prove Lemma 3.4.2 by passing to the limit the corresponding discrete-time results for the Markov chains in Section 3.3. First, suppose that $Y_0 \sim \Pi^{(x)}$

for some $x > 0$ and hence $\Pi_0^* = \delta_x$ (see Remark 3.2.3). Adopting the notation of Section 3.3, the primal birth-and-death Markov chain $X^\Delta \sim (\pi_0^\Delta, P^\Delta)$ has

$$\mathrm{sep}^\Delta(t) = \sup_i \left(1 - \frac{\sum_j \pi_0^\Delta(j) P_t^\Delta(j,i)}{\pi^\Delta(i)}\right).$$

Now

$$\begin{aligned}\frac{\sum_j \pi_0^\Delta(j) P_t^\Delta(j,i)}{\pi^\Delta(i)} &= \frac{1}{H^\Delta(i_{\Delta,x})} \sum_{j \le i_{\Delta,x}} \frac{\pi^\Delta(j) P_t^\Delta(j,i)}{\pi^\Delta(i)} \\ &= \frac{1}{H^\Delta(i_{\Delta,x})} \mathbb{P}_i(X_t^\Delta \le i_{\Delta,x}).\end{aligned}$$

The monotonicity conditions outlined in Remark 3.3.4 and [12, Remark 4.15] imply that this last expression is minimized (for each $t = 0, 1, \ldots$) when $i = n^\Delta$, and that the minimum value is

$$\frac{1}{H^\Delta(i_{\Delta,x})} \mathbb{P}_{n^\Delta}(X_t^\Delta \le i_{\Delta,x}) = 1 - \mathrm{sep}^\Delta(t) = \mathbb{P}_{i_{\Delta,x}}(\widehat{T}_{n^\Delta} \le t), \tag{3.4.5}$$

where $\widehat{X}^\Delta$ is the strong stationary dual of X^Δ as defined at (3.3.4)–(3.3.7), with absorption time $\widehat{T}_{n^\Delta}$ in its largest state n^Δ. We now substitute $\lfloor t/h \rfloor$ for t, and recall that $h \equiv h_\Delta$ is a function of Δ and that $Y_t^\Delta := X_{\lfloor t/h \rfloor}^\Delta$ (and analogously for $\widehat{Y}_t^\Delta$), to find for real $t \ge 0$ that

$$\frac{1}{H^\Delta(i_{\Delta,x})} \mathbb{P}_{S(1)}(Y_t^\Delta \le S(0) + i_{\Delta,x}\Delta) = 1 - \mathrm{sep}^\Delta(t) = \mathbb{P}_{i_{\Delta,x}}\left(\widehat{Y}_t^\Delta = S(1)\right), \tag{3.4.6}$$

where $i_{\Delta,x}$ is short for $S(0) + i_{\Delta,x}\Delta$.

By Theorem 3.3.1, the left side of (3.4.6) converges to $\frac{1}{\Pi(x)} \mathbb{P}_{S(1)}(Y_t \le x)$ [where we note that the hypothesis of Theorem 3.3.1 is met for the deterministic initial conditions $Y_0^\Delta = Y_0 = S(1)$]. Theorem 3.3.5 implies that

$$\lim_{\Delta \downarrow 0} \mathbb{P}_{i_{\Delta,x}}(\widehat{Y}_t^\Delta > S(1) - \varepsilon) = \mathbb{P}_x(Y_t^* > S(1) - \varepsilon). \tag{3.4.7}$$

Let $\check{X}^\Delta$ be the Siegmund dual of (the time-reversal of) X^Δ; by definition, $\check{X}^\Delta$ is a Markov chain satisfying

$$\mathbb{P}_y(X_t^\Delta \le z) = \mathbb{P}_z(y \le \check{X}_t^\Delta)$$

for all $y,\ z \in \mathcal{S}^\Delta$ and $t = 0, 1, 2, \ldots$. Equation (5.3) in [12] gives, with $h = h_\Delta$ and with $\lceil x \rceil_\Delta$ (respectively, $\lfloor x \rfloor_\Delta$) being the smallest element $\ge x$ (resp., the largest element $\le x$) in the grid $\{S^*(S(0)),\ S^*(S(0)+\Delta), \ldots, S^*(S(1)-\Delta),\ S^*(S(1))\}$, that

$$\begin{aligned}
\mathbb{P}_{i_{\Delta,x}}(\widehat{Y}_t^\Delta > S(1)-\varepsilon) &= \mathbb{P}_{i_{\Delta,x}}(\widehat{X}^\Delta_{\lfloor t/h \rfloor} > S(1)-\varepsilon) \\
&= \sum_{j > S(1)-\varepsilon} \frac{H^\Delta(j)}{H^\Delta(i_{\Delta,x})} \mathbb{P}_{i_{\Delta,x}}(\check{X}^\Delta_{\lfloor t/h \rfloor} = j) \\
&\ge \frac{H^\Delta(\lfloor S(1)-\varepsilon \rfloor_\Delta)}{H^\Delta(i_{\Delta,x})} \mathbb{P}_{i_{\Delta,x}}(\check{X}^\Delta_{\lfloor t/h \rfloor} > \lceil S(1)-\varepsilon \rceil_\Delta) \\
&\ge \frac{H^\Delta(\lfloor S(1)-\varepsilon \rfloor_\Delta)}{H^\Delta(i_{\Delta,x})} \mathbb{P}_{i_{\Delta,x}}(\check{X}^\Delta_{\lfloor t/h \rfloor} \ge \lceil S(1)-2\varepsilon \rceil_\Delta) \\
&= \frac{H^\Delta(\lfloor S(1)-\varepsilon \rfloor_\Delta)}{H^\Delta(i_{\Delta,x})} \mathbb{P}_{\lceil S(1)-2\varepsilon \rceil_\Delta}(X^\Delta_{\lfloor t/h \rfloor} \le i_{\Delta,x}) \\
&\to \frac{\Pi(S(1)-\varepsilon)}{\Pi(x)} \mathbb{P}_{S(1)-2\varepsilon}(Y_t \le x) \text{ as } \Delta \downarrow 0,
\end{aligned}$$

and this last expression converges to $\frac{1}{\Pi(x)}\mathbb{P}_{S(1)}(Y_t \le x)$ as $\varepsilon \downarrow 0$. We can also get an upper bound on $\mathbb{P}_{i_{\Delta,x}}(\widehat{Y}_t^\Delta > S(1)-\varepsilon)$ using

$$\begin{aligned}
\sum_{j > S(1)-\varepsilon} \frac{H^\Delta(j)}{H^\Delta(i_{\Delta,x})} \mathbb{P}_{i_{\Delta,x}}(\check{X}^\Delta_{\lfloor t/h \rfloor} = j) &\le \frac{1}{H^\Delta(i_{\Delta,x})} \mathbb{P}_{i_{\Delta,x}}(\check{X}^\Delta_{\lfloor t/h \rfloor} \ge \lfloor S(1)-\varepsilon \rfloor_\Delta) \\
&= \frac{1}{H^\Delta(i_{\Delta,x})} \mathbb{P}_{\lfloor S(1)-\varepsilon \rfloor_\Delta}(X^\Delta_{\lfloor t/h \rfloor} \le i_{\Delta,x}) \\
&\to \frac{1}{\Pi(x)} \mathbb{P}_{\lfloor S(1)-\varepsilon \rfloor_\Delta}(Y_t \le x) \text{ as } \Delta \downarrow 0,
\end{aligned}$$

and this last expression converges to $\frac{1}{\Pi(x)}\mathbb{P}_{S(1)}(Y_t \le x)$ as $\varepsilon \downarrow 0$. Letting $\varepsilon \downarrow 0$ in (3.4.7)

gives

$$\frac{1}{\Pi(x)}\mathbb{P}_{S(1)}(Y_t \leq x) = \mathbb{P}_x(T^*_{S(1)} \leq t).$$

Now $\Pi_t \ll \Pi$ for all $t > 0$; let $R_t = d\Pi_t/d\Pi$, so that for any $A = [S(0), s)$ with $s \in \mathcal{S}$ we have

$$\mathbb{P}_{\Pi^{(x)}}(Y_t \in A) = \int_A R_t(y)\,\Pi(dy),$$

and also

$$\begin{aligned}\mathbb{P}_{\Pi^{(x)}}(Y_t \in A) &= \int_{[S(0),x]} \frac{\Pi(dy)}{\Pi(x)} P_t(y, A) \\ &= \int_{[S(0),x]} \frac{1}{\Pi(x)}\pi(y) \int_A p_t(y, z)\,dz\,dy.\end{aligned}$$

We will now appeal to the reversibility of Y. A diffusion process X with generator A and state space I is *reversible* with respect to the distribution μ if for all $f,\ g \in \mathcal{D}_A$ we have

$$\int f(y)\,(Ag)(y)\,\mu(dy) = \int (Af)(y)\,g(y)\,\mu(dy). \tag{3.4.8}$$

If Y satisfies the assumptions of Lemma 3.4.2, noting that $f,\ g \in \mathcal{D}_A$ implies that the derivatives of each function vanish at the boundary of the state space, integration by parts yields that (3.4.8) holds for $\mu = \Pi$, the stationary distribution of Y, and the primal diffusion is reversible with respect to Π. Also note that (3.4.8) is equivalent to the following (see [37, Section II.5]): for all $f, g \in C(\mathcal{S})$, and for all $t > 0$ we have

$$\int f(y)\,(T_t g)(y)\,\Pi(dy) = \int (T_t f)(y)\,g(y)\,\Pi(dy), \tag{3.4.9}$$

where (T_t) is the one parameter semigroup associated with Y.

Letting f and g be suitably continuous approximations of $\mathbb{1}([S(0), x])$ and $\mathbb{1}(A)$, and appealing to (3.4.9), we have

$$\begin{aligned}\int_{[S(0),x]} \frac{1}{\Pi(x)}\pi(y)\int_A p_t(y,z)\,dz\,dy &= \int_{[S(0),x]} \frac{1}{\Pi(x)}\int_A p_t(z,y)\pi(z)\,dz\,dy \\ &= \int_A \frac{1}{\Pi(x)}\mathbb{P}_z(Y_t \le x)\,\Pi(dz),\end{aligned}$$

and so $\frac{1}{\Pi(x)}\mathbb{P}_z(Y_t \le x)$ is a version of $R_t(z)$. By monotonicity of Y, we have that $\frac{1}{\Pi(x)}\mathbb{P}_z(Y_t \le x)$ is minimized when $z = S(1)$, and hence for Y we have

$$a(t) = 1 - \mathrm{sep}(t) = \frac{1}{\Pi(x)}\mathbb{P}_{S(1)}(Y_t \le x) = \mathbb{P}_x(T^*_{S(1)} \le t),$$

establishing Lemma 3.4.2 for Y.

Remark 3.4.4. In the Markov chain setting of [12] and [19], the authors were able to justify their "strong stationary duality" nomenclature by tying their then-new notion of duality to the more classical notions of duality in the stochastic process literature. Specifically, let $X \sim (\pi_0, P)$ be an ergodic Markov chain with stationary distribution π. If X satisfies specific monotnicity conditions, namely, that the time reversal $\tilde{P}$ is monotone and $\pi_0(x)/\pi(x)$ decreases in x, then with H be cumulative of π, they show that the SSD X^* of X is the Doob H-transform of the Siegmund dual of the time-reversal of X.

For a Markov process Y with transition operator $P_t(x, dy)$, the *Doob H-transform* of Y is the right-continuous Markov process with transition operator

$$Q_t(x, dy) := \frac{H(y)}{H(x)}P_t(x, dy).$$

It has played a central role in Markov process duality theory, especially in the context of processes conditioned to die in a given set or point. See [43, Chapter VII] for further detail. The *Siegmund dual* of a Markov process Y with state space $\mathcal{S}$ is a Markov process Z on $\mathcal{S}$ satisfying:

$$\mathbb{P}_y(Y_t \leq z) = \mathbb{P}_z(y \leq Z_t) \text{ for all } y, z \in \mathcal{S}.$$

It has played a prominent role in the study of birth-and-death chains and diffusion theory and in the study of interacting particle systems (see [37, Section II.3] for extensive background).

To justify the nomenclature in the present diffusion setting, consider the diffusion X as defined in Section 3.1, and let X^* be the strong stationary dual of X specified in Definition 3.2.1. Then, recalling from Remark 3.2.6 that for all $f \in F[0,1]$ we have $\Lambda T_t f = T_t^* \Lambda f$, a simple calculation yields

$$\int_{[0,x]} p_t(z,y)\, dy = \int_{[z,1]} \frac{\Pi(x)}{\Pi(y)} P_x^*(X_t^* \in dy),$$

giving us immediately that X^* is the Doob H-transform of the Siegmund dual of (the time reversal of) X, where H here is the cumulative stationary distribution Π.

A functional definition of duality generalizing Siegmund's definition was introduced in [25]. For extensive background see again [37, Section II.3]. Briefly, let X and Y be two Markov processes with state spaces $\mathcal{S}$ and $\mathcal{S}'$ and let f be a bounded measurable function on $\mathcal{S} \times \mathcal{S}'$. We define Y to be the *dual* of X with respect to the function f if

$$\mathbb{E}_x f(X_t, y) = \mathbb{E}_y f(x, Y_t), \quad \text{for all } x \in \mathcal{S},\ y \in \mathcal{S}'.$$

As in [12, Theorem 5.12], a simple calculation yields that, in the diffusion setting, X and its SSD X^* are dual with respect to the function

$$f(x, x^*) := \begin{cases} 1/\Pi(x^*), & \text{if } x \leq x^* \\ 0, & \text{otherwise,} \end{cases}$$

on $I \times I$, further justifying the duality name for X^*.

With [12, Definition 5.16], the authors generalized the classical notion of functional duality. Adapted to the present setting, let X and Y be two diffusions defined on a common probability space with state spaces $\mathcal{S}$ and $\mathcal{S}'$. We say Y is *dual* to X with respect to a function $f : \mathcal{S} \times \mathcal{S}' \to \mathbb{R}$ and distribution μ on $\mathcal{S} \times \mathcal{S}'$ if

$$\mathbb{E}_\mu f(X_t, Y_0) = \mathbb{E}_\mu f(X_0, Y_t).$$

In [12, Theorem 5.19], the authors were able to show that the strong stationary dual of an ergodic Markov chain $X \sim (\pi_0, P)$ with stationary distribution π, and with the additional properties that the time reversal $\tilde{P}$ is monotone and $\pi_0(x)/\pi(x)$ decreases in x, is dual to the primal chain with respect to this new functional definition, for suitable choices of f and μ. We are able to recover the analogue of their Theorem 5.19 here, as it is easy to see that X^* (the strong stationary dual of X) and X are dual with respect to the function $f(x^*, x) = \mathbf{1}(x \leq x^*)\pi(x)/\Pi(x^*)$ and μ equal to any mixture of the distributions $\delta_{x^*} \times \Pi^{(x^*)}$ with $x^* \in [0, 1]$.

3.5 Hitting times and eigenvalues

In the continuous-time birth-and-death chain setting, a famous theorem due to Karlin and MacGregor [28] asserts that the hitting time of state n for a birth-and-death chain X on $\{0, 1, \ldots, n\}$ started in state 0 is distributed as the sum of independent exponential random variables with parameters relating to the eigenvalues of the generator of X. Fill [19] used strong stationary duality to exploit Karlin and MacGregor's result to prove that the separation from stationarity for an ergodic continuous-time birth-and-death chain X at time t is equal to $\mathbb{P}(Y > t)$ where Y is a sum of independent exponential random variables with parameters depending on the eigenvalues of the generator of X. In [14], Diaconis and Saloff-Coste used Fill's result and tight concentration bounds on the tail probabilities of Y to prove the existence of a separation cutoff for a sequence (X_n) of birth-and-death chains under certain conditions on the eigenvalues of the generators of the chains X_n. In this section, we outline and recover the analogous theory in the diffusion setting.

To this effect, consider again a diffusion X on $[0, 1]$ with generator A, and with reflecting or entrance boundary behavior at each boundary, satisfying the assumptions of Theorem 3.2.4. Let X^* be a strong stationary dual of X according to Definition 3.2.1. For fixed λ, let $v_\lambda(x)$ be the solution to the eigenvalue problem associated with A (respectively, A^*):

$$Av + \lambda v = 0 \quad (A^*v + \lambda v = 0) \tag{3.5.1}$$

with boundary condition

$$B_0(v) = 0 \tag{3.5.2}$$

where B_0 represents the following boundary condition:

$$B_0(v) := \begin{cases} v(0), & \text{if } 0 \text{ is absorbing or exit;} \\ \frac{dv}{dS}^+(0), & \text{if } 0 \text{ is instantaneously reflecting or entrance.} \end{cases}$$

Let $T_{x,y}$ be the hitting time of y for X begun in x. From [27, Section 4.6], we have that $v_\lambda(x)$ is unique up to multiplicative constant and that the moment generating function of $T_{x,y}$, call it $\psi_{x,y}$, can be expressed as

$$\psi_{x,y}(\lambda) = v_\lambda(x)/v_\lambda(y). \tag{3.5.3}$$

A completely analogous set of results hold for A^*.

If we further add the relevant boundary condition at 1, namely that $B_1(v) = 0$ (where B_1 is defined analogously to B_0), then we have from Sturm–Liouville theory (see for example [30, Theorem 4.1]) that the eigenvalues of A^* (resp., nonzero eigenvalues of A) satisfying (3.5.1) with the two boundary conditions are countable, real, positive, and simple and can be ordered such that

$$0 < \lambda_1 < \lambda_2 < \cdots \uparrow \infty;$$

further, they satisfy $\sum_{k=1}^{\infty} \lambda_k^{-1} < \infty$. For extensive background on the relevant Sturm–Liouville theory, see for example [46]. The eigenfunctions and eigenvalues of A and A^* are connected by the following simple relationship:

Proposition 3.5.1. *Adopt the same assumptions as Theorem 3.2.4, and further assume that* $b(\cdot) > 0$ *on* $(0,1)$ *and that* 1 *is a reflecting boundary for* X *and* 0 *is either a reflecting or entrance boundary for* X*. Fix* $\lambda > 0$*.*

(a) *Suppose that* $v = f$ *is a solution of* (3.5.1) *for generator* A *with boundary conditions* $B_0(v) = B_1(v) = 0$*. Then* $v = \Lambda f$ *is a solution of* (3.5.1) *for generator* A^* *with boundary conditions* $B_0^*(v) = B_1^*(v) = 0$ *(and the same* λ*).*

(b) *Suppose that* $v = g$ *is a solution of* (3.5.1) *for generator* A^* *with boundary conditions* $B_0^*(v) = B_1^*(v) = 0$*. Then* $f(\cdot) = g(\cdot) + \frac{\Pi(\cdot)}{\pi(\cdot)} g'(\cdot)$ *is a solution of* (3.5.1) *for generator* A *with boundary conditions* $B_0(v) = B_1(v) = 0$ *(and the same* λ*).*

Proof. (a) If $f(\cdot)$ satisfies (3.5.1) for A and the boundary conditions $B_0(f) = B_1(f) = 0$, then $f \in \mathcal{D}_A$ and

$$\frac{df}{dS}^{+}(0) = \left(\frac{f'}{s}\right)^{+}(0) = 0 = \frac{df}{dS}^{-}(1) = \left(\frac{f'}{s}\right)^{-}(1).$$

From (3.2.1) we have $\Lambda f \in \mathcal{D}_{A^*}$, and from (3.2.2) we have

$$A^*\Lambda f(\cdot) = \Lambda A f(\cdot) = \Lambda(-\lambda f)(\cdot) = -\lambda \Lambda f(\cdot) \tag{3.5.4}$$

on I. Therefore, $\Lambda f(\cdot)$ satisfies (3.5.1) for A^*. Also $B_0^*(\Lambda f) = 0$ as 0 is an entrance boundary for the dual and $\Lambda f \in \mathcal{D}_{A^*}$, and similarly $B_1^*(\Lambda f) = 0$.

(b) Note that if $g(\cdot)$ satisfies (3.5.1) for A^*, then $g \in \mathcal{D}_{A^*}$, and hence $g \in C[0,1]$, and

$g(1) = 0$. Next, on $(0, 1)$ note

$$\begin{aligned} f' &= g' + \frac{\pi^2 - \Pi\pi'}{\pi^2} g' + \frac{\Pi}{\pi} g'' = 2g' - \frac{\Pi\pi'}{\pi^2} g' + \frac{\Pi}{\pi} g'' \\ &= 2g' + \frac{\Pi}{\pi}\left(g'' + \frac{b'}{b} g' - \frac{2a}{b} g' \right) = 2g' + \frac{\Pi}{\pi}\left(-\frac{2\lambda g}{b} - \frac{2\pi}{\Pi} g' \right) \\ &= -2\lambda g \frac{\Pi}{\pi}\frac{1}{b} = -2\lambda g M s \end{aligned}$$

where the fourth equality follows from (3.5.1). We have that $M^+(0) = 0 = g(1)$ and $M^-(1) < \infty$, and hence

$$\left(\frac{f'}{s}\right)^+ (0) = 0 = \left(\frac{f'}{s}\right)^- (1),$$

and therefore $B_0(f) = B_1(f) = 0$. Next, on $(0, 1)$ note

$$\begin{aligned} f'' &= \frac{-2\lambda}{b}\left(\frac{\Pi}{\pi} g' + g \frac{\pi^2 - \Pi\pi'}{\pi^2} - g \frac{\Pi b'}{\pi b} \right) \\ &= \frac{-2\lambda}{b}\left(\frac{\Pi}{\pi} g' + g + \frac{\Pi}{\pi} g \frac{s'}{s} \right) \in C(0, 1), \end{aligned}$$

and hence $f \in C^2(0, 1)$. Combining the above, on $(0, 1)$ we have

$$\begin{aligned} af' + \frac{1}{2} b f'' &= \lambda g \frac{\Pi}{\pi}\frac{s'}{s} - \lambda\left(\frac{\Pi}{\pi} g' + g + \frac{\Pi}{\pi} g \frac{s'}{s} \right) \\ &= -\lambda\left(\frac{\Pi}{\pi} g' + g \right) = -\lambda f. \end{aligned}$$

To show that $f \in \mathcal{D}_A$ and that f satisfies (3.5.1) for A with the relevant boundary conditions it remains only to show that $f \in C[0, 1]$. We have (by Theorem 3.2.4) that 0 is an entrance boundary for X^*, and hence for any fixed $\xi \in (0, 1)$ we have that $N(0) < \infty$ and hence

$$-\int_{(0,\xi]} S^*(\eta)\, M^*(d\eta) = \int_{(0,\xi]} \frac{1}{M(\eta)} s(\eta) M^2(\eta)\, d\eta = \int_{(0,\xi]} s(\eta) M(\eta)\, d\eta < \infty,$$

where the first equality holds by (3.2.8)–(3.2.9). Since $g \in C[0,1]$, it follows that

$$\int_{(0,\xi]} |-2\lambda g(\eta)s(\eta)M(\eta)|\,d\eta = \int_{(0,\xi]} |f'(\eta)|\,d\eta < \infty.$$

Hence, by the dominated convergence theorem,

$$\int_{(\omega,\xi]} f'(\eta)\,d\eta = f(\xi) - f(\omega)$$

has a finite limit as $\omega \downarrow 0$. We conclude that $f \in C[0,1)$. We have by assumption that 1 is a reflecting boundary for X and hence for any fixed $\xi \in (0,1)$ we have that $\Sigma(1) < \infty$ and hence

$$\int_{[\xi,1)} s(\eta)M(\eta)\,d\eta < \infty.$$

By the same argument that showed that f is continuous at 0, we find that f is also continuous at 1. The proof is finished, as we have established that $f \in C[0,1]$. □

In the diffusions setting, we have an analogue (namely [30, Theorem 5.1]) of Karlin and MacGregor's famous result on the eigenvalue expansion on birth-and-death hitting times. Adapted to the present setting, we state the analogue as follows:

Theorem 3.5.2. *Let X be a regular diffusion process on $[0,1]$ and assume 0 is either instantaneously reflecting or entrance. Then*

$$\lim_{x\to 0} \psi_{x,1}(\lambda) = \prod_{k=1}^{\infty} \left(1 - \frac{\lambda}{\lambda_k}\right)^{-1}, \tag{3.5.5}$$

which is the moment generating function of an infinite sum of independent exponential random variables with parameters λ_k.

Combining this with Proposition 3.5.1 and Lemma 3.4.2, we arrive at

Theorem 3.5.3. *Let X be a diffusion on $[0,1]$ with $X_0 = 0$, with generator A, and with either reflecting or entrance behavior at the bounday 0 and reflecting behavior at the boundary 1. Let the eigenpairs $(\lambda_i, v_{\lambda_i})$, $i = 1, 2, \ldots$, of A with $\lambda_i > 0$ satisfying (3.5.1) and boundary conditions $B_0(v_{\lambda_i}) = 0 = B_1(v_{\lambda_i})$ be labeled so that $0 < \lambda_1 < \lambda_2 < \cdots$. Let X^* be a strong stationary dual of X with generator A^*, and note that $X_0^* = 0$ by Remark 3.2.3. Let $W_1, W_2, \ldots$ be independent random variables with $W_i \sim \text{Exp}(\lambda_i)$. Then*

$$\text{sep}(t) = \mathbb{P}_0(T_1^* > t) = \mathbb{P}(W > t) \text{ where } W \overset{\mathcal{L}}{=} \sum_{i=1}^{\infty} W_i.$$

This mirrors the corresponding result for birth-and-death Markov chains given by [12, Theorem 4.20] in discrete time and by [19, Theorem 5] in continuous time.

In [14], the authors used [12, Theorem 4.20] to determine conditions for a separation cut-off to occur in a sequence of birth-and-death Markov chains. We shall presently derive analogous results for diffusions using Theorem 3.5.3. Consider now a sequence of diffusion generators $(A_n)_{n=1}^{\infty}$ defining a sequence of diffusions $(X^n)_{n=1}^{\infty}$ with $X_0^n \sim \nu^n$, on finite intervals $[l_1, r_1] = I_1$, $[l_2, r_2] = I_2$, $\ldots$ where all left boundary points, l_n are assumed to be reflecting or entrance and all right boundary points r_n are assumed to be reflecting. Note that without loss of generality we can take $I_n = [0, r_n]$ for all $n \geq 1$. We write π^n for the stationary distribution for X^n, and we write ν_t^n for the distribution of X^n at time t. This sequence of diffusions exhibits a *separation cut-off at* (t_n) if the sequence (t_n) is such that

for any $\epsilon \in (0,1)$ we have

$$\text{(i)} \lim_{n\to\infty} \text{sep}(\nu^n_{(1+\varepsilon)t_n}, \pi^n) = 0, \text{ and}$$

$$\text{(ii)} \lim_{n\to\infty} \text{sep}(\nu^n_{(1-\varepsilon)t_n}, \pi^n) = 1.$$

To apply Theorem 3.5.3 here, let the nonzero eigenvalues of A_n be labeled $0 < \lambda_{n,1} < \lambda_{n,2} < \cdots$, and let $\nu^n = \delta_0$ for all $n \geq 1$. We further assume that each A_n satisfies the assumptions of Theorem 3.2.4, and let $(A^*_n)_{n=1}^\infty$ be the sequence of generators of the strong stationary duals of $(X^n)_{n=1}^\infty$ as defined by Definition 3.2.1 For each $n \geq 1$, let $W_{n,j} \sim \text{Exp}(\lambda_{n,j})$ be independent random variables, and let $W_n \stackrel{\mathcal{L}}{=} \sum_{j=1}^\infty W_{n,j}$. From Theorem 3.5.3, we have $\text{sep}^n(t) = \mathbb{P}(W_n > t)$. We can therefore get sharp bounds on separation by deriving sharp bounds for the tail probabilities of W_n. To this end, note that we have

$$\mathbb{E}\, W_n = \sum_{j=1}^\infty \lambda_{n,j}^{-1} < \infty, \quad \mathbb{V}\text{ar}\, W_n = \sum_{j=1}^\infty \lambda_{n,j}^{-2} < \infty.$$

An application of the one-sided Chebyshev's inequality gives the analogue to the separation cut-off result [14, Theorem 5.1]:

Theorem 3.5.4. *Let $(A_n)_{n=1}^\infty$ be a sequence of diffusion generators defining diffusions $(X^n)_{n=1}^\infty$, with $X_0^n \sim \nu^n$, on finite intervals $[0, r_1] = I_1$, $[0, r_2] = I_2$, ..., where 0 is assumed to be reflecting or entrance for all n, and all right boundary points r_n are assumed to be reflecting. With the eigenvalues $\lambda_{n,i}$ defined as above, this sequence of diffusions exhibits a separation cut-off if and only if*

$$\lim_{n\to\infty} \lambda_{n,1}\, \mathbb{E}W_n = \infty,$$

in which case there is a separation cut-off at (t_n) *with* $t_n := \mathbb{E}W_n$. *Further, for any* $c > 0$ *the following separation bounds hold for* any *sequence* (t_n), *where we restrict to* $c \leq 1$ *in the second bound:*

$$\mathrm{sep}(\nu^n_{(1+c)t_n}, \pi^n) \leq \frac{1}{1 + c^2\lambda_{n,1}t_n}, \;\; \mathrm{sep}(\nu^n_{(1-c)t_n}, \pi^n) \geq 1 - \frac{1}{1 + c^2\lambda_{n,1}t_n}.$$

The proof is completely analogous to the proof of Theorem 5.1 in [14], and so is omitted.

***Example* 3.5.5.** Let $0 < 1 = r_1 \leq r_2 \leq r_3 \leq \cdots$ be an arbitrary increasing sequence of positive real numbers, and let A_n be the generator of reflecting Brownian motion on $I_n = [0, r_n]$. Then (see [30, Section 6]), we know that

$$\lambda_{n,k} = \frac{j_k^2}{2r_n^2}$$

where $(j_k)_{k=1}^{\infty}$ are the positive zeros of the usual Bessel function $J_{1/2}$. Note

$$\lambda_{n,1}\mathbb{E}W_n = \sum_{k=1}^{\infty} \frac{j_1^2}{j_k^2}$$

is constant in n, and therefore there is no separation cut-off.

***Example* 3.5.6.** Let (η_n) be a sequence of positive real numbers diverging monotonically to infinity. Let A_n be the generator for a Bes($2\eta_n$+2) process on $[0, 1]$ with 1 a reflecting boundary. Again from [30, Section 6], we have that

$$\lambda_{n,k} = \frac{j_{n,k}^2}{2}$$

where $(j_{n,k})_{k=1}^{\infty}$ are the positive zeros of the Bessel function J_{η_n+1}. Then

$$\lambda_{n,1}\mathbb{E}W_n = \sum_{k=1}^{\infty} \frac{j_{n,1}^2}{j_{n,k}^2}.$$

It is well known (see for instance [44, equations (1) and (40)]) that

$$\sum_{k=1}^{\infty} \frac{1}{j_{n,k}^2} = \frac{1}{4(\eta_n + 2)}$$

and (see [1, pg. 371]) that

$$j_{n,1}^2 = \left[\eta_i + 1 + O\left(\eta_i^{1/3}\right)\right]^2;$$

so there is a separation cut-off for this sequence of diffusions at (t_n), with $t_n = 2\sum_{k=1}^{\infty} j_{n,k}^{-2} = (2\eta_n + 4)^{-1}$.

This is, perhaps, not a surprising result in light of the interpretation of the Bes(m) process as the radial part of m-dimensional Brownian motion for integer m. As the strong stationary dual of a Bes(α) process is a Bes($\alpha + 2$) process (recall Example 3.2.8), for integer sequences $\eta_n = m_n$, a separation cut-off is equivalent to a sharp concentration in the hitting time of 1 of the dual Bes($2m_n + 4$) sequence, i.e., a sharp concentration in the hitting time of the unit sphere for $(2m_n + 4)$-dimensional Brownian motion started in $\vec{0}$. For large m_n, at time t the ratio of the square of the radial part of $(2m_n + 4)$-dimensional Brownian motion to t has a distribution which doesn't depend on t and (by the central limit theorem) is approximately normal with mean $2m_n + 4$ and variance $2(2m_n + 4)$. We therefore expect to have a sharp concentration of the hitting time of the unit sphere at $t = (2m_n + 4)^{-1}$, and indeed we found that the cut-off occurs there.

Appendix A

P^* when P is a star chain

In Remark 2.3.6 it is claimed that if the given chain P is a star chain, then the star chain of Lemma 2.3.4 is simply obtained by collapsing all leaves with the same one-step transition probability to state 0 into a single leaf. More precisely, we establish the following:

Proposition A.1 *Let P be the transition matrix of an ergodic star chain with hub at 0. If for each γ_i in the reduced set of eigenvalues of P_0 we define*

$$m(i) := \{j \in [n] : \eta_j = \gamma_i\},$$

then $P^*(0,i) = \sum_{j \in m(i)} P(0,j)$.

Proof. Define $H := \text{diag}(\eta_1, \ldots, \eta_n)$ and

$$x := (P(0,1), \ldots, P(0,n)),$$

$$y := (1-\eta_1, \ldots, 1-\eta_n),$$

so that

$$P = \left(\begin{array}{c|c} P(0,0) & x \\ \hline y^T & H \end{array}\right).$$

By the standard formula for the determinant of a partitioned matrix (e.g., [26, Section 0.8.5]), if t is not in the spectrum $\{\eta_1, \ldots, \eta_n\}$ of H then we find

$$\det(tI - P) = [t - P(0,0) - x(tI - H)^{-1}y^T] \det(tI - H) \tag{A.1}$$

for the characteristic polynomial of P. Analogously, define $\Gamma := \text{diag}(\gamma_1, \ldots, \gamma_r)$ and

$$x^* := (P^*(0,1), \ldots, P^*(0,r)),$$

$$y^* := (1 - \gamma_1, \ldots, 1 - \gamma_r);$$

if t is not in the spectrum $\{\gamma_1, \ldots, \gamma_r\}$ of Γ, then we find

$$\det(tI - P^*) = [t - P^*(0,0) - x^*(tI - \Gamma)^{-1}y^{*T}] \det(tI - \Gamma) \tag{A.2}$$

for the characteristic polynomial of P^*.

Note that

$$\begin{aligned} P(0,0) = \operatorname{tr} P - \operatorname{tr} H &= \sum_{i=0}^{n} \theta_i - \sum_{i=1}^{n} \eta_i \\ &= \sum_{i=0}^{r} \lambda_i - \sum_{i=1}^{r} \gamma_i = \operatorname{tr} P^* - \operatorname{tr} \Gamma = P^*(0,0), \end{aligned} \tag{A.3}$$

where the third equality is a result of the eigenvalue reduction procedure discussed in Section 2.3.1 and the fourth equality is from Lemma 2.6 in [9]. Similarly, for all $t \notin \{\eta_1, \ldots, \eta_n\}$ we have

$$\frac{\det(tI - P)}{\det(tI - H)} = \frac{\det(tI - P^*)}{\det(tI - \Gamma)}. \tag{A.4}$$

Therefore, for all $t \notin \{\eta_1, \ldots, \eta_n\}$ we have

$$\sum_{i=1}^{n} P(0,i)\frac{1-\eta_i}{t-\eta_i} = \sum_{i=1}^{r} P^*(0,i)\frac{1-\gamma_i}{t-\gamma_i}, \tag{A.5}$$

because using definitions of $H, x, y, \Gamma, x^*, y^*$ and equations (A.1)–(A.4) we find

$$\begin{aligned}\sum_{i=1}^{n} P(0,i)\frac{1-\eta_i}{t-\eta_i} &= x(tI-H)^{-1}y^T = t - P(0,0) - \frac{\det(tI-P)}{\det(tI-H)} \\ &= t - P^*(0,0) - \frac{\det(tI-P^*)}{\det(tI-\Gamma)} = x^*(tI-\Gamma)^{-1}y^{*T} = \sum_{i=1}^{r} P^*(0,i)\frac{1-\gamma_i}{t-\gamma_i}.\end{aligned}$$

Rewrite (A.5) as

$$\sum_{i=1}^{r} P^*(0,i)\frac{1-\gamma_i}{t-\gamma_i} = \sum_{i=1}^{r}\left(\sum_{j\in m(i)} P(0,j)\right)\frac{1-\gamma_i}{t-\gamma_i}.$$

Since $\gamma_1, \ldots, \gamma_r$ are distinct, it follow easily that $P^*(0,i) = \sum_{j\in m(i)} P(0,j)$ for $i = 1, \ldots, r$, as desired. □

Let π be the stationary distribution for P. Using the formula for $P^*(0,i)$ provided by Proposition A.1, it is a simple matter to check that the probability mass function π^* defined by $\pi^*(0) := \pi(0)$ and $\pi^*(i) = \sum_{j\in m(i)} \pi(j)$ for $i \neq 0$ satisfies the detailed balance condition and is therefore the stationary distribution for P^*; indeed, using the reversibility of P with respect to π we have

$$\begin{aligned}\pi^*(0)P^*(0,i) = \pi(0)\sum_{j\in m(i)} P(0,j) &= \sum_{j\in m(i)} \pi(j)P(j,0) \\ &= \sum_{j\in m(i)} \pi(j)(1-\gamma_i) = \pi^*(i)P^*(i,0).\end{aligned}$$

Appendix B

Details of Theorem 3.3.2

In proving Theorem 3.3.1, we made use of Theorem 3.3.2 adapted from [18, Theorems 4.8.2 and 1.6.5 and Corollary 4.8.9]]. We restate the theorem here:

Theorem 3.3.2 *Let A be the generator of a regular diffusion process Y with state space $\mathcal{S}$. Assume $h_\Delta > 0$ converges to 0 as $\Delta \downarrow 0$. Let $Y_t^\Delta := X^\Delta_{\lfloor t/h_\Delta \rfloor}$ where $X^\Delta \sim (\pi_0^\Delta, P^\Delta)$ is a Markov chain with some metric state space $\mathcal{S}^\Delta \subset \mathcal{S}$, and assume $Y_0^\Delta \Rightarrow Y_0$. Define $T^\Delta : B(\mathcal{S}^\Delta) \to B(\mathcal{S}^\Delta)$ via*

$$T^\Delta f(x) = \mathbb{E}_x f(X_1^\Delta)$$

for f in the space $B(\mathcal{S}^\Delta)$ of real-valued bounded measurable functions on $\mathcal{S}^\Delta$. Define $A^\Delta := h_\Delta^{-1}(T^\Delta - I)$. Suppose that $C(\mathcal{S})$ is convergence determining and that there is an algebra $B \subset C(\mathcal{S})$ that strongly separates points. Let $\rho_\Delta : C(\mathcal{S}) \to B(\mathcal{S}^\Delta)$ be defined via

$\rho_\Delta f(\cdot) = f|_{\mathcal{S}^\Delta}(\cdot)$. *If*

$$\lim_{\Delta \to 0} \sup_{y \in \mathcal{S}^\Delta} \left| (A^\Delta \rho_\Delta f)(y) - (Af)(y) \right| = 0 \tag{B.1}$$

for all $f \in \mathcal{D}_A$*, then* $Y^\Delta \Rightarrow Y$.

The purpose of this appendix is to carefully spell out the proof of the above theorem, as the notation in [18] differs considerably from the notation we have adopted. The following chart gives the notational equivalences between the present work and [18]; in connection with $\mu_n(x, \cdot)$, see Corollary 4.8.5 in [18].

Notation in present work:	Notation in [18]:
$\mathcal{S} = [S(0), S(1)]$ with the Euclidean metric	(E, r)
$\mathcal{S}^\Delta,\ A^\Delta,\ T^\Delta$	$E_n,\ A_n,\ T_n$
$P^\Delta(x, \cdot)$	$\mu_n(x, \cdot)$
$\{(f, Af) \mid f \in \mathcal{D}_A\}$	A
$\mathcal{D}_A$	$\mathcal{D}_A$
$C(\mathcal{S})$	L
$1/h^\Delta$	α_n
id	η_n
ρ_Δ	π_n

Here $\rho_\Delta : C(\mathcal{S}) \to B(\mathcal{S}^\Delta)$ is defined via $\rho_\Delta f(\cdot) = f|_{\mathcal{S}^\Delta}(\cdot)$, and id : $\mathcal{S}^\Delta \to \mathcal{S}$ is the inclusion function embedding $\mathcal{S}^\Delta$ into $\mathcal{S}$.

Proof of Theorem 3.3.2. As in [18, Section 3.4], define a set $B \subset C(\mathcal{S})$ to be *convergence*

determining if

$$\lim_{n\to\infty} \int f \, d\mathbb{P}_n = \int f \, d\mathbb{P} \text{ for all } f \in B$$

implies $\mathbb{P}_n \Rightarrow \mathbb{P}$. We say that $B \subset C(\mathcal{S})$ *strongly separates points* if for every $y \in \mathcal{S}$ and $\varepsilon > 0$ there exists a finite set $\{f_1, \ldots, f_k\} \subset B$ such that

$$\inf_{z:|z-y|\geq\varepsilon} \max_{1\leq i\leq k} |f_i(z) - f_i(y)| > 0.$$

Clearly $C(\mathcal{S})$ is convergence determining, and by considering suitably smooth uniform approximations in $\mathcal{D}_A$ to the indicator function of $\{x\}$ for each $x \in \mathcal{S}$, it follows that $\mathcal{D}_A \subset C(\mathcal{S})$ is an algebra that strongly separates points. In the notation of [18, Corollary 4.8.9], we have $G_n = E_n = \mathcal{S}^\Delta$, and so to prove $Y^\Delta \Rightarrow Y$, it suffices to prove that for each $T > 0$ and $f \in C(\mathcal{S})$ we have

$$\lim_{\Delta\to 0} \sup_{y\in\mathcal{S}^\Delta} \left|(T^\Delta)^{\lfloor t/h\rfloor}\rho_\Delta f(y) - \rho_\Delta T_t f(y)\right|, \qquad 0 \leq t \leq T. \tag{B.2}$$

From [18, Theorem 1.6.5], to prove (B.2) it suffices to establish that for all $f \in \mathcal{D}_A$ we have that $\rho_\Delta f \in B(\mathcal{S}^\Delta)(= L_n$ in the notation of [18, Theorem 1.6.5]) satisfies

$$\lim_{\Delta\to 0} \sup_{y\in\mathcal{S}^\Delta} |\rho_\Delta f(y) - f(y)| = 0 \tag{B.3}$$

and

$$\lim_{\Delta\to 0} \sup_{y\in\mathcal{S}^\Delta} \left|(A^\Delta\rho_\Delta f)(y) - (Af)(y)\right| = 0. \tag{B.4}$$

But (B.3) is clearly true, and (B.4) is assumed (for all $f \in \mathcal{D}_A$) in the statement of Theorem 3.3.2. □

Bibliography

[1] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, ninth Dover printing, tenth GPO printing edition, 1964.

[2] D. J. Aldous and M. Brown. Inequalities for rare events in time-reversible Markov chains. I. In *Stochastic inequalities (Seattle, WA, 1991)*, volume 22 of *IMS Lecture Notes Monogr. Ser.*, pages 1–16. Inst. Math. Statist., Hayward, CA, 1992.

[3] D. J. Aldous and P. Diaconis. Strong uniform times and finite random walks. *Adv. in Appl. Math.*, 8(1):69–97, 1987.

[4] D. J. Aldous and J. A. Fill. Reversible Markov chains and random walks on graphs. Chapter drafts available from http://www.stat.Berkeley.EDU/users/aldous/book.html.

[5] R. N. Bhattacharya and E. C. Waymire. *Stochastic processes with applications*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Inc., New York, 1990. A Wiley-Interscience Publication.

BIBLIOGRAPHY

[6] P. Biane. Intertwining of Markov semi-groups, some examples. In *Séminaire de Probabilités, XXIX*, volume 1613 of *Lecture Notes in Math.*, pages 30–36. Springer, Berlin, 1995.

[7] P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[8] M. Brown. The first passage time distribution for a parallel exponential system with repair. In *Reliability and fault tree analysis (Conf., Univ. California, Berkeley, Calif., 1974)*, pages 365–396. Soc. Indust. Appl. Math., Philadelphia, Pa., 1975.

[9] M. Brown. Interlacing eigenvalues in time reversible Markov chains. *Math. Oper. Res.*, 24(4):847–864, 1999.

[10] M. Brown and Y. S. Shao. Identifying coefficients in the spectral representation for first passage time distributions. *Probab. Engrg. Inform. Sci.*, 1:69–74, 1987.

[11] P. Carmona, F. Petit, and M. Yor. Beta-gamma random variables and intertwining relations between certain Markov processes. *Rev. Mat. Iberoamericana*, 14(2):311–367, 1998.

[12] P. Diaconis and J. A. Fill. Strong stationary times via a new form of duality. *Ann. Probab.*, 18(4):1483–1522, 1990.

[13] P. Diaconis and L. Miclo. On times to quasi-stationarity for birth and death processes. *J. Theoret. Probab.*, 22(3):558–586, 2009.

[14] P. Diaconis and L. Saloff-Coste. Separation cut-offs for birth and death chains. *Ann. Appl. Probab.*, 16(4):2098–2122, 2006.

[15] R. Durrett. *Probability models for DNA sequence evolution*. Probability and its Applications (New York). Springer, New York, second edition, 2008.

[16] E. B. Dynkin. *Markov processes. Vols. I, II*, volume 122 of *Translated with the authorization and assistance of the author by J. Fabius, V. Greenberg, A. Maitra, G. Majone. Die Grundlehren der Mathematischen Wissenschaften, Bände 121*. Academic Press Inc., Publishers, New York, 1965.

[17] A. Etheridge. *Some mathematical models from population genetics*, volume 2012 of *Lecture Notes in Mathematics*. Springer, Heidelberg, 2011. Lectures from the 39th Probability Summer School held in Saint-Flour, 2009.

[18] S. N. Ethier and T. G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986. Characterization and convergence.

[19] J. A. Fill. Strong stationary duality for continuous-time Markov chains. Part I: Theory. *Journal of Theoretical Probability*, 5:45–70, 1992. 10.1007/BF01046778.

[20] J. A. Fill. An interruptible algorithm for perfect sampling via Markov chains. *Ann. Appl. Probab.*, 8(1):131–162, 1998.

[21] J. A. Fill. On hitting times and fastest strong stationary times for skip-free and more general chains. *J. Theoret. Probab.*, 22(3):587–600, 2009.

[22] J. A. Fill. The passage time distribution for a birth-and-death chain: strong stationary duality gives a first stochastic proof. *J. Theoret. Probab.*, 22(3):543–557, 2009.

[23] J. A. Fill, M. Machida, D. J. Murdoch, and J. S. Rosenthal. Extension of Fill's perfect rejection sampling algorithm to general chains. *Random Structures Algorithms*, 17(3-4):290–316, 2000. Special issue: Proceedings of the Ninth International Conference "Random Structures and Algorithms" (Poznan, 1999).

[24] JamesAllen Fill and Vince Lyzinski. Hitting times and interlacing eigenvalues: A stochastic approach using intertwinings. *Journal of Theoretical Probability*, pages 1–28, 2012.

[25] R. Holley and D. Stroock. Dual processes and their application to infinite interacting systems. *Adv. in Math.*, 32(2):149–174, 1979.

[26] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, 1985.

[27] K. Itô and H. P. McKean, Jr. *Diffusion processes and their sample paths*. Springer-

Verlag, Berlin, 1974. Second printing, corrected, Die Grundlehren der mathematischen Wissenschaften, Band 125.

[28] S. Karlin and J. McGregor. Coincidence properties of birth and death properties. *Pacific J. Math.*, 9:1109–1140, 1959.

[29] S. Karlin and H. M. Taylor. *A second course in stochastic processes*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1981.

[30] J. T. Kent. Eigenvalue expansions for diffusion hitting times. *Z. Wahrsch. Verw. Gebiete*, 52(3):309–319, 1980.

[31] J. F. C. Kingman. Random partitions in population genetics. *Proc. Roy. Soc. London Ser. A*, 361(1704):1–20, 1978.

[32] J. F. C. Kingman. The representation of partition structures. *J. London Math. Soc. (2)*, 18(2):374–380, 1978.

[33] J.F.C. Kingman. The coalescent. *Stochastic Processes and their Applications*, 13(3):235 – 248, 1982.

[34] F. B. Knight. *Essentials of Brownian motion and diffusion*, volume 18 of *Mathematical Surveys*. American Mathematical Society, Providence, R.I., 1981.

[35] V. N. Kolokoltsov. *Markov processes, semigroups and generators*, volume 38 of *de Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 2011.

[36] J. Lamperti. *Stochastic processes*. Springer-Verlag, New York, 1977. A survey of the mathematical theory, Applied Mathematical Sciences, Vol. 23.

[37] T. M. Liggett. *Interacting Particle Systems*. Springer Berlin Heidelberg, 1985.

[38] R. Mena and W. Murray. Markov chains for collaboration. *Math. Mag.*, 84(1):3–15, 2011.

[39] L. Miclo. On absorption times and Dirichlet eigenvalues. *ESAIM Probab. Stat.*, 14:117–150, 2010.

[40] B. Morris and Y. Peres. Evolving sets, mixing and heat kernel bounds. *Probab. Theory Related Fields*, 133(2):245–266, 2005.

[41] L. C. G. Rogers and J. W. Pitman. Markov functions. *Ann. Probab.*, 9(4):573–582, 1981.

[42] L. C. G. Rogers and D. Williams. *Diffusions, Markov processes, and martingales. Vol. 2*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000. Itô calculus, Reprint of the second (1994) edition.

[43] M. Sharpe. *General theory of Markov processes*, volume 133 of *Pure and Applied Mathematics*. Academic Press Inc., Boston, MA, 1988.

[44] I. N. Sneddon. On some infinite series involving the zeros of Bessel functions of the first kind. *Proc. Glasgow Math. Assoc.*, 4:144–156 (1960), 1960.

[45] D. W. Stroock and S. R. S. Varadhan. *Multidimensional diffusion processes*, volume 233 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1979. Reprinted in 2006.

[46] E. C. Titchmarsh. *Eigenfunction expansions associated with second-order differential equations. Part I*. Second Edition. Clarendon Press, Oxford, 1962.

Vita

Vince Lyzinski

PERSONAL INFORMATION

Birth: December 8, 1983, Pittsburgh, PA

Home: 3503 N. Charles St. Apt. 3A, Baltimore, MD 21218

EDUCATION

Johns Hopkins University, Baltimore, Maryland

- Ph.D. candidate in Applied Mathematics and Statistics, August 2007–present

  Dissertation: Intertwinings, Interlacing Eigenvalues, and Strong Stationary Duality for Diffusions

- Master of Science in Engineering, Department of Applied Mathematics and Statistics, 2011

- Master of Science, Department of Mathematics, 2007

VITA

University of Notre Dame

- Bachelor of Science, Department of Mathematics, 2006

FIELDS OF INTEREST

Probability theory, stochastic processes, probabilistic number theory, graph theory, combinatorics, statistics

RELEVANT COURSEWORK

110.601 Algebra	110.602 Algebra
110.615 Algebraic Topology	550.671 Combinatorial Analysis
110.607 Complex Variables	550.661 Foundations of Optimization
550.672 Graph Theory	110.620 LIE Groups and LIE Algebra
550.723 Markov Chains	550.692 Matrix Analysis
550.892 Matrix Analysis II Seminar	550.681 Numerical Analysis
550.722 Poisson Convergence	550.620 Probability Theory I
550.621 Probability Theory II	110.605 Real Variables
550.480 Shape and Differential Geometry	550.630 Statistical Theory
550.631 Statistical Theory II	550.426 Stochastic Processes
550.626 Stochastic Processes II	550.735 Topics in Stat. Pattern Recognition
550.730 Topics in Statistics	550.484 Wavelets & Fourier Analysis

TEACHING EXPERIENCE

At Johns Hopkins, I was the instructor for the following courses:

- 550.310 Probability and Statistics for the Physical and Information Sciences and Engineering, Taught Fall 2012, Fall 2011, Spring 2011, Spring 2010
- 550.230 Introduction to Biostatistics, Taught Summer 2012, Summer 2011
- 550.171 Discrete Mathematics, Taught Summer 2009
- 550.111 Statistical Analysis I, Taught Summer 2008

At Johns Hopkins, I was the teaching assistant for the following courses:

- 110.109 Calculus II
- 110.202 Calculus III
- 550.361 Introduction to Optimization
- 550.362 Introduction to Optimization II
- 550.291 Linear Algebra and Differential Equations
- 550.692 Matrix Analysis and Linear Algebra
- 550.621 Probability Theory II
- 550.386 Scientific Computing: Differential Equations
- 550.385 Scientific Computing: Linear Algebra
- 550.111 Statistical Analysis I

VITA

AWARDS AND FELLOWSHIPS

- GAANN Fellowship, U.S. Department of Education, 2009–2012
- Counselman Fellowship, Johns Hopkins Department of Applied Mathematics and Statistics, 2008–2012
- Naddor Teaching Fellowship, Johns Hopkins Department of Applied Mathematics and Statistics, 2007–2009
- Magna Cum Laude, University of Notre Dame, 2006

PUBLICATIONS

- J. A. Fill, V. Lyzinski (2012). Hitting times and interlacing eigenvalues: A stochastic approach using intertwinings. *Journal of Theoretical Probability*, 2012, DOI 10.1007/s10959-012-0457-9.
- A. Godbole, C. M. Lim, V. Lyzinski, N. Triantafillou (2012). Sharp threshold asymptotics for the emergence of additive bases. Submitted for publication.

CONFERENCE PRESENTATIONS AND TALKS

- "Strong Stationary Duality for Diffusion Processes," Session on Computational and Discrete Mathematics, CAIMS, Toronto, Canada, June 2012.

VITA

- "Sharp Threshold Asymptotics for the Emergence of Additive Bases," AMS Session on Probability Theory, Stochastic Processes and Statistics, AMS/MAA Joint Mathematics Meeting, Boston, MA, January 4, 2012.

- "Sharp Threshold Asymptotics for the Emergence of Additive Bases," Johns Hopkins University Department of Applied Mathematics and Statistics Student Seminar Series, October 18, 2011.

- "Multivariate Records," George Washington University Department of Statistics Student Seminar Series, October 15, 2010.

- "Multivariate Records," Johns Hopkins University Department of Applied Mathematics and Statistics Student Seminar Series, October 12, 2010.

- "A Stochastic Interpretation of a Hitting Time Result of Mark Brown," Johns Hopkins University Department of Applied Mathematics and Statistics Student Seminar Series, April 13, 2010.

- "The Comparison Method for the Relaxation Time of Ergodic Markov Chains," Fields–MITACS Summer School in Applied Probability, May 2009.

- "The Comparison Method for the Relaxation Time of Ergodic Markov Chains," Johns Hopkins University Department of Applied Mathematics and Statistics Student Seminar Series, March 24, 2009.

VITA

REFEREEING

Annals of Applied Probability, Involve-A Mathematics Journal, Electronic Journal of Probability

PROFESSIONAL SOCIETY MEMBERSHIPS

IMS

Phi Beta Kappa