ANALYSIS OF NON-REVERSIBLE MARKOV CHAINS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

(Michael) Chek Hin Choi

August 2017

ANALYSIS OF NON-REVERSIBLE MARKOV CHAINS

(Michael) Chek Hin Choi, Ph.D.

Cornell University 2017

The analysis of non-reversible Markov chains is of great theoretical and applied interest. In this thesis, we summarize our contributions in this direction into four parts.

In the first part, titled "Skip-free Markov chains", we aim at developing a general theory for the class of skip-free Markov chains on denumerable state space. This encompasses their potential theory via an explicit characterization of their potential kernel expressed in terms of family of fundamental excessive functions, which are defined by means of the theory of Martin boundary. We also describe their fluctuation theory generalizing the celebrated fluctuations identities that were obtained by using the Wiener-Hopf factorization for the specific skip-free random walks. We proceed by resorting to the concept of similarity to identify the class of skip-free Markov chains whose transition operator has only real and simple eigenvalues. We manage to find a set of sufficient and easy-to-check conditions on the one-step transition probability for a Markov chain to belong to this class. We also study several properties of this class including their spectral expansions given in terms of Riesz basis, derive a necessary and sufficient condition for this class to exhibit a separation cutoff, and give a tighter bound on its convergence rate to stationarity than existing results.

In the second part, titled "Analysis of non-reversible Markov chains via similarity orbit", we examine the spectral theory of Markov chains on a denumerable state space from a similarity orbit perspective. In particular, we study the class of Markov chains that are in the similarity orbit of Markov chains with normal transition kernels such as birth-death chains or reversible Markov chains. This allows us to derive spectral

expansions and offer a detailed analysis on the convergence rate, separation cutoff and L^2-cutoff of this class of non-reversible Markov chains. We also look into the problem of estimating the integral functionals from discrete observations for this class. In the last part of this Chapter, we investigate three particular similarity orbits of reversible Markov kernels, that we call the permutation, pure birth and random walk orbit, and analyze various possibly non-reversible variants of classical birth-death processes in these orbits.

In the third part, titled "Metropolis-Hastings reversiblizations of non-reversible Markov chains", we study two types of Metropolis-Hastings (MH) reversiblizations for non-reversible Markov chains with transition kernel P. While the first type is the classical Metropolised version of P, we introduce a new self-adjoint kernel which captures the opposite transition effect of the first type, that we call the second MH kernel. We investigate the spectral relationship between P and the two MH kernels. Along the way, we state a version of Weyl's inequality for the spectral gap of P (and hence its additive reversiblization), as well as an expansion of P. Both results are expressed in terms of the spectrum of the two MH kernels. In the spirit of Fill [42] and Paulin [89], we define a new pseudo-spectral gap based on the two MH kernels, and show that the total variation distance from stationarity can be bounded by this gap. We give variance bounds of the Markov chain in terms of the proposed gap, and offer spectral bounds in metastability and Cheeger's inequality in terms of the two MH kernels by comparison of Dirichlet form and Peskun ordering.

Finally, in the fourth part, titled "Estimation of the log-Sobolev constant and Eigenspace of reversible Markov chain via a single sample path", we build upon the results of Hsu et al. [50] to consider the problem of estimating the log-Sobolev constant and the eigenspace of transition matrix via a single sample path from a reversible and ergodic Markov chain. This allows us to have a tighter mixing time estimate by Wilson's method and the log-Sobolev upper bound. We prove statistical guarantees for our proposed estimators, and demonstrate that the idea can be used to estimate other functional constants such as the modified log-Sobolev constant and the Cheeger constant of the Markov chain.

BIOGRAPHICAL SKETCH

Michael Choi was born on March 22, 1990. He grew up in Hong Kong and graduated from St. Paul's Co-educational College in 2009. Before he came to the United States, he received a Bachelor of Science in Actuarial Science with First Class Honours at The University of Hong Kong in 2013.

In Fall 2013, he came to Cornell to pursue a Ph.D. degree in the School of Operations Research and Information Engineering, with a concentration in Applied Probability and Statistics.

To my family. In memory of Francis and Kelvin.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Carole Bernard, Eric Cheung and Hailiang Yang for opening the door to research for me. Without their kind help, support and mentorship at the beginning, this journey will not start at the very first place.

Words cannot express my sincerest gratitude to Pierre Patie, who is simply the best advisor one can hope for. His unparallel knowledge, enthusiasm, creativity, curiosity, diligence together with his foresight never cease to amaze and impress me. He provided an exemplary model and aspiration for me as a teacher, scholar, researcher and friend. This thesis cannot possibly be done without his advice and guidance in which I am indebted to. On a personal level, he is one of the most charming gentleman that I have met, and hopefully I have inherited a bit of his charisma and intellectual traits throughout the years.

I am also extremely grateful to Jim Dai and Laurent Saloff-Coste for being my thesis committees and offering constructive comments that greatly improve various parts of this thesis.

I am fortunate enough to collaborate with many amazing teachers during my sojourn at Cornell. I am grateful to Yudong Chen, Jim Dai, Shane Henderson, Peter Jackson, Mark Lewis, Ronnie Loeffen, Sidney Resnick and Anna Srapionyan for being excellent role models in teaching and for their complete trust on my teaching duties. I also thank Irina Gaynanova and Ravi Kumar for our joint recitation sessions, and all of my students for the pleasant and enjoyable experience throughout the years.

Researchwise, I am extremely indebted to the impressive group of researchers that I have cited. Their masterpieces have fundamentally shaped various research directions in the field and greatly enhanced my understanding of the subject. This thesis certainly stands on no ground without their foundational and pioneering work, and I will continue to be an admirer of their intellectual achievements.

I have benefited enormously from the research diversity at ORIE. I am always impressed by the foresight of the founders of ORIE to create a platform that combine, in *alphabetical* order, operations researchers, optimizers, probabilists and statisticians together fifty one years ago at the time of writing. This is particularly evident nowadays when we look at the cutting-edge development in a wide variety of disciplines or applications that are often at the core and the intersection of these subjects. In this direction, I thank David Shmoys for steering the wheel and being a great leader of the school, and Mark Lewis and David Williamson for being the DGS. I would like to take this opportunity to thank the entire ORIE for their warm hospitality and willingness to let me to be part of the family, and to all the staff for their professionalism. To Ravi Kumar, Calvin Wylie and James Dong for spoiling me with spare phone batteries before Christmas. To Bangrui Chen, for his insights, kindness, sharing and simply being the best officemate I can ever hope for in the past four years. To Tom Fei, for instilling me a different perspective on applications and teaching me the state-of-the-art research in SBM, deep learning and computer vision. To Tom Chen, for introducing me to Cornell and the memorable time we had in learning functional analysis together back in my first year. To Kenneth Chong, for his professional baking and cheerful smile. To Weici Hu, for her kindness, thoughtfulness and invitations to various events. To Chaoxu Tong, for our funny moments on Agar.io and teaching me so much about coding. To Venus Lo, for her positive energy that always shines through the darkest days and coldest winter, for offering me the chance to practice my gradually rusty Cantonese, and for our friendship since we met in Waterloo. To Cory Girard, for being my roommate in my second year and his contagious laughters that always brighten my days. To Julian Sun, for being the first person I know at Cornell since our Ph.D. visit day, for the fun times we had in learning probability and ergodic theory as well as being picnic organizers, and for his long emails. To Saul Toscano-Palmerin, for always being a great listener on different topics and being with me through various ups and downs. To Alberto Vera, for our Monday post-episode discussion on The Walking Dead and The Westworld. To Yixuan Zhao, for being my academic big brother and his generous and heartfelt advice. To James Dong, Angela Gao, Weilong Guo and Yuhang Ma, for being very flexible officemates and for the laughters we once shared in Rhodes 287 and 296.

Last, but not very least, I sincerely thank my family Connie, Andrew, Catherine and Jacqueline for their love, support, patience, encouragement and sacrifices throughout the years. I am obliged to the ultimate freedom that they have enchanted me in my intellectual quest. This thesis is dedicated to them and to the memory of Francis, who planted the scholastic seeds in my mind since high school that have finally came to blossom, and Kelvin, for everything he has done for me and for my family.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Consider a Markov chain with transition kernel P and stationary distribution π with its time-reversal P^* on a general state space $\mathcal{X}$. In the reversible case, that is, when P is viewed as a linear self-adjoint operator in the weighted Hilbert space $L^2(\pi)$, the quantitative rate of convergence to equilibrium, as measured in the Hilbert space topology, total variation distance or separation distance, is well-known to be closely connected to the spectrum or the spectral gap of P, see for instance Aldous and Fill [1], Levin et al. [74], Meyn and Tweedie [78], Miclo [81], Montenegro and Tetali [84], Saloff-Coste [97] and the references therein. Roberts and Rosenthal [93] shows that the existence of a L^2-spectral gap is equivalent to P being geometrically ergodic. The main technical insight relies heavily on the spectral theory of self-adjoint operators, which facilitates the analysis of the spectrum of P. From an application point of view, reversible Markov chains form the backbone in many diverse areas ranging from Markov chain Monte Carlo (see e.g. Roberts and Rosenthal [94] and Diaconis et al. [34]) or in the study of population genetics and community ecology (see e.g. Khare and Zhou [65] and Griffiths [46]). However, P need not be reversible in general. If P is non-reversible, the analysis on the rate of convergence is fragmentally understood, possibly due to a much less developed spectral theory for non-self-adjoint operators. We refer interested readers to the work of Montenegro and Tetali [84], Wilmer [112] and Lorek [75] for excellent discussion on the convergence rate of some particular non-reversible Markov chains.

On the other hand, in addition to the theoretical interest as described above, the study of non-reversible Markov chains is also fascinating from an application perspective. It has been shown in Hwang et al. [53, 54] and Chen and Hwang [18] that one can accelerate the convergence of Markov processes by adding anti-symmetric drift to a common underlying equilibrium. Diaconis et al. [33] proposes a non-reversible Markov chain sampler and analyzes the convergence rate in total variation distance, and more recently Sun et al. [106] and Bierkens [6] propose a non-reversible Metropolis-Hastings method by inserting vortices or perturbations to improve convergence rate, while Duncan et al. [35] studies non-reversible Langevin samplers and demonstrates a faster convergence rate with smaller asymptotic variance both theoretically and empirically. In this viewpoint, it is interesting to see if one can derive new algorithms to speed up convergence and offer an unifying framework for the analysis of non-reversible Markov chains.

We now give a historical account of the theoretical development and describe three different approaches that have been elaborated to deal with non-reversibility.

The first approach, suggested by Kendall [62], shows that non-reversible Markov kernel can be "dilated" to a semigroup of unitary operators in a larger Hilbert space by Sz.-Nagy's dilation theorem. Yet, this approach seems to be dedicated enough to yield useful information on the rate of convergence. In this spirit of modifying the functional space, it is recently proposed by Kontoyiannis and Meyn [67] to cast P in a weighted Banach space L_V^∞ instead of the classical $L^2(\pi)$ framework, where V is the Lyapunov function associated with P. In particular, they show that for a ϕ-irreducible and aperiodic Markov chain, P is geometrically ergodic if and only if P admits a spectral gap in the space L_V^∞ equipped with the V-norm. They also give an example in which a non-reversible Markov chain is geometrically ergodic yet it fails to have a $L^2(\pi)$ spectral gap.

The second approach, initiated by Fill [42], is to resort to an appropriate *reversiblized* version of P and analyze how its spectrum can be related to the chi-squared distance to stationarity of the original chain. Two reversiblizations are proposed, namely the multiplicative reversiblization PP^* and the additive reversiblization $(P + P^*)/2$. In the discrete-time setting, it is shown that the second largest eigenvalue of PP^* can be used to upper bound the distance from stationarity, while the spectral gap of $(P+P^*)/2$ is used for continuous-time Markov chain. More recently, Paulin [89] generalizes this approach and defines a *pseudo*-spectral gap, based upon the maximum spectral gap of $P^{*k}P^k$ for $k \geqslant 1$. He demonstrates that the proposed gap plays a similar role as that of spectral gap in the reversible case. He proves variance bounds and Bernstein inequality based on his proposed gap. In this fashion, we will explain an original approach that involves two types of Metropolis-Hastings reversibilizations in Chapter 4.

The third approach, presented by Patie and Savov [85], is to resort to the concept of intertwining to build a link between non-reversible and reversible operators. More precisely, they investigate the rate of convergence to equilibrium of the so-called generalized Laguerre semigroups which are associated to non-self-adjoint integro-differential operators, and obtained explicit spectral estimate for the hypocoercivity phenomenon of this class. We will illustrate the spectral theory of skip-free Markov chains and more general non-reversible Markov chains via the concept of similarity, that is, an intertwining relationship with the link operator having a bounded inverse, in Chapter 2 and Chapter 3 respectively.

1.1 Organization and outline of the Thesis

This thesis consists of four self-contained chapters, with a recurrent theme of analyzing non-reversible Markov chains from various perspectives. We now summarize the contributions and give a high-level discussion of the content in each chapter. Note that each chapter has its own introduction and a more detailed outline of the work.

Chapter 2. In the first Chapter, we develop a general theory for the class of skip-free

Markov chains on denumerable state space. This encompasses their potential theory via an explicit characterization of their potential kernel expressed in terms of family of fundamental excessive functions, which are defined by means of the theory of Martin boundary. We also describe their fluctuation theory generalizing the celebrated fluctuations identities that were obtained by using the Wiener-Hopf factorization for the specific skip-free random walks. We proceed by resorting to the concept of similarity to identify the class of skip-free Markov chains whose transition operator has only real and simple eigenvalues. We manage to find a set of sufficient and easy-to-check conditions on the one-step transition probability for a Markov chain to belong to this class by stochastic monotonicity as introduced by Siegmund [102] and Clifford and Sudbury [22]. We also study several properties of this class including their spectral expansions given in terms of Riesz basis, derive a necessary and sufficient condition for this class to exhibit a separation cutoff, and give a tighter bound on its convergence rate to stationarity than existing results.

Chapter 3. Building upon the work of Chapter 2, we further examine the spectral theory of Markov chains on a denumerable state space from a similarity orbit perspective in this Chapter, expanding along the intertwining approach proposed by Patie and Savov [85] and Miclo [83]. In particular, we study the class of Markov chains that are in the similarity orbit of normal Markov chains such as birth-death chains or reversible Markov chains, which allows us to derive spectral expansions and offer a detailed analysis on the convergence rate, separation cutoff (Chen and Saloff-Coste [13], Diaconis and Saloff-Coste [31], Mao et al. [76]) and L^2-cutoff (Chen and Saloff-Coste [12], Chen et al. [15]) of this class of non-reversible Markov chains that is otherwise inaccessible. We also look into the problem of estimating the integral functionals from discrete observations for this class, extending the work of Altmeyer and Chorowski [2] in this direction. Finally, as illustrations, we investigate three particular similarity orbits of reversible Markov kernels, that we call the permutation, pure birth and random walk orbit, and analyze various variants of classical birth-death processes, such as the Ehrenfest model, linear birth-death process, quadratic birth-death process and $M/M/\infty$ model and their relationships with orthogonal polynomials in these orbits. Another highlight of this approach is that the Markov chain and all other chains in its orbits share the same eigentime identity as studied by Aldous and Fill [1], Cui and Mao [25], Miclo [82].

Chapter 4. In this Chapter, along the reversiblization approach suggested by Fill [42] and Paulin [89], we study two types of Metropolis-Hastings (MH) reversiblizations for non-reversible Markov chains with transition kernel P. While the first MH kernel is the classical Metropolis chain of P as introduced by Hastings [48], Metropolis et al. [77] and Roberts and Rosenthal [94], we identify a new self-adjoint yet possibly non-Markovian operator that we call the second MH kernel to capture the opposite transition effect of the first kernel. We state a version of Weyl's inequality for the spectral gap of P and its additive reversiblization in the finite state space case, and illustrate the sharpness of the inequality by investigating in details the asymmetric simple random walk on the n-cycle and on discrete torus. We proceed by defining a *pseudo*-spectral gap, that we call the MH-spectral gap, based on the spectrum of the two MH kernels. We show that the existence of a MH-spectral gap implies that P is geometrically ergodic, and by Weyl's inequality the second MH kernel is a contraction operator whenever P is lazy and ergodic in the finite state space case. We carry out some numerical examples that reveal that our MH-spectral gap is, for non-reversible chains, a better estimate than the existing bounds found in the literature. Variance bounds are also proved in terms of the proposed gap. Finally, we revisit the notion of metastability and the Cheeger inequality, and offer a variant of these celebrated inequalities by means of comparison of the Dirichlet form between the non-reversible chain and the two MH kernels.

Chapter 5. In the final Chapter, motivated by the work of Hsu et al. [50], we consider the problem of estimating the log-Sobolev constant and the eigenspace of transition matrix via a single sample path from a reversible and ergodic Markov chain, in order to yield a tighter mixing time estimate by the log-Sobolev upper bound (see e.g. Diaconis and Saloff-Coste [30]) and the Wilson's method for the lower bound (see e.g. Saloff-Coste [98]). We rely on the well-known relationship between the spectral gap and the log-Sobolev constant to provide an interval estimator of the latter, and we invoke a variant of Davis-Kahan $\sin\theta$ theorem to prove statistical guarantee for our proposed estimator of the eigenspace. We then proceed to illustrate that the same idea can be used to estimate other functional constants such as the modified log-Sobolev constant and the Cheeger constant of the Markov chain.

CHAPTER 2

SKIP-FREE MARKOV CHAINS

2.1 Introduction

Let $X = (X_n)_{n\in\mathbb{N}}$ be a Markov chain on the countable state space $E = [[\mathfrak{l}, \mathfrak{r}]] \subseteq \mathbb{Z}$, where we use the notation [[(resp.]]) to denote that $\mathfrak{l}$ (resp. $\mathfrak{r}$) may or may not be in E, defined on the filtered probability space $(\Omega, (\mathcal{F}_n)_{n\in\mathbb{N}}, \mathbb{P} = (\mathbb{P})_{x\in\mathbb{E}})$. We denote its transition matrix by $P = (p(x,y))_{x,y\in E}$. We assume further that X is irreducible, i.e. for all $x, y \in \dot{E} =]\mathfrak{l}, \mathfrak{r}[$, $p^{(n)}(x,y) = \mathbb{P}_x(X_n = y) > 0$ for some $n \in \mathbb{N}$, and *upward skip-free*, i.e. for all $x \in E$, $p(x, x+1) > 0$ and $p(x, x+y) = 0, y \geqslant 2$. We denote by $\mathcal{SF}$ be the set of such upward skip-free Markov chains (or transition operators) on E.

The aim of this Chapter to develop a comprehensive theory for the set $\mathcal{SF}$, including *the potential theory, the fluctuation theory* and, resorting to the algebraic concept of similarity, *the spectral theory*. As a by-product, we also provide, for ergodic chains, a detailed analysis of the speed of convergence to stationarity and investigate the separation cutoff phenomena.

We recall that under the additional condition that $p(x, x-y) = 0, y \geqslant 2$, that is, it is also skip-free to the left, $X \in \mathcal{SF}$ becomes a birth-death chain. These chains have been and are still the object of intensive and fascinating studies. This probably originates from the seminal work of Karlin and McGregor [59], see also Lederman and Reuter [69], on the diagonalization of their transition operator that provide deep insights into fine distributional properties of these chains. Note that the spectral analysis of these operators has also revealed fascinating links with the theory of orthogonal polynomials and Stieljes moment problem. A review of birth-death chains including their potential theory, is given below in Section 2.2.2.

Our contributions to the theory of the class $\mathcal{SF}$ can be summarized as follows.

(i) *Potential theory*: We shall start our program by studying the potential theory of chains in $\mathcal{SF}$. More specifically, we implement an original approach based on the theory of Martin boundary for Markov chains as developed by Dynkin [39] to express their q-potential kernel. In this vein, we recall that in the specific case of birth-death chains, this q-potential is given in terms of the two fundamental solutions (the decreasing and increasing one) of a three-term recurrence equations (discrete analogue of a second order differential equation), see Section 2.2.2 for more details regarding this expression. In our context, the situation is more delicate as one has to solve an infinite recurrence equation whose set of solutions does not seem to have been clearly identified in the literature. Although the issue of solving this equation is of algebraic nature, we shall elaborate a strategy based mostly on a combination of techniques from probability theory and potential theory. More specifically, instead of trying to identify directly the convex cone of q-excessive functions for the chain $X \in \mathcal{SF}$, we take an alternative route which consists in characterizing the q-potential kernel in terms of the so-called fundamental q-excessive (for short FqE) functions of three different chains: $(X, \mathbb{P})$, $(X, \widehat{\mathbb{P}})$ the dual chain as defined in (2.2.2) below and $(X, \mathbb{P}^{y]}), y \in E$, where $(X, \mathbb{P}^{y]})$ is the Markov chain $(X, \mathbb{P})$ killed upon entering the half-line $[[\mathfrak{l}, y]$, which is plainly an upward skip-free Markov chain on the state space $E^{y]} = (y, \mathfrak{r}]]$.

(ii) *Fluctuation identities*: From this representation of the q-potential kernel, we derive the main fluctuations identities for chains in $\mathcal{SF}$. We recall that the fluctuation theory, which is concerned with the distribution of the first visit of the chain to some (finite or infinite) intervals is of great importance in many applications such as biology, epidemiology and also in ruin theory where the upward skip-free property is required. This theory is well established for the case of birth-death chains (skip-free on both sides), see e.g. Karlin and McGregor [58], Keilson [60]. On the other hand the famous Wiener-Hopf factorization technique has proven to be useful to characterize the law of the first exit times for the class of random walks, that is for Markov chains with stationary and independent increments, see Spitzer [104] for further details on these identities. We also mention the interesting work of Fill [43] where he characterizes the upward hitting time distribution of upward skip-free chain via establishing an intertwining with pure-birth chain. Our original approach goes decisively beyond these frameworks as it allows to treat in an unified way, not only the first upward passage time, but also the first downward passage time, including the possibility of undershoot, of the irreducible skip-free Markov chain. To implement our methodology on some specific examples, one merely needs to have access to a transformation (Laplace transform, Fourier transform, moment generating function...) that determines the one-step transition kernel of the chain. We shall illustrate this idea by recovering, in a simple manner, Spitzer's identities for skip-free random walks, and, by studying the first passage times of branching Galton-Watson processes with immigration, whose details will be provided in a subsequent paper [20]. We also mention that the continuous analogue of these results for skip-free continuous-time Markov processes on the real line has been detailed in [87] and applications to generalized Ornstein-Uhlenbeck processes and continuous branching processes with immigration are carried out in [72] and [86] respectively.

(iii) *Spectral theory and its applications*: We also aim at providing some insights into the spectral theory of the (transition operator of) Markov chains that belong to the class $\mathcal{SF}$. This is a challenging issue as the transition operator P of such a chain is non-self-adjoint (non-reversible) in the weighted Hilbert space $\ell^2(\pi)$, where π is the reference measure, implying that there is no spectral theorem available for such bounded linear operator. To overcome this difficulty, we propose an original approach based on the concept of similarity. More specifically, we introduce a subclass of $\mathcal{SF}$, denoted by $\mathcal{S}_{sf}$, whose each element is related to a (diagonalizable) transition matrix of a birth-death Markov chain via a certain commutation relation. Using this identity, we resort to some techniques from non-harmonic analysis to investigate how to transfer the known spectral information of the reversible birth-death transition operator to the non-reversible one in order to obtain its spectral decomposition. By means of the inverse spectral theorem we manage to show that the class $\mathcal{S}_{sf}$ characterizes completely the family of transition operators in $\mathcal{SF}$ that have real and distinct eigenvalues. On the other hand, we shall provide a set of sufficient and easy-to-check conditions on the one-step transition probability for a Markov chain to belong to $\mathcal{S}_{sf}$. It is worth mentioning that, to the best of our knowledge, this is the first identification of a class of non-reversible chains with such a spectrum. This is a useful fact which answers an open question raised by Fill [43] on understanding the class of skip-free chains that have real and non-negative eigenvalues.

We believe that this new way of classifying Markov chains based on similarity orbit is powerful enough to tackle many substantial and delicate problems arising in the analysis of such chains. To illustrate this fact, we provide the spectral decomposition of their non-symmetric transition matrices, the law of their first passage times, including the case with possible overshoot. We also study for ergodic chains in $\mathcal{S}_{sf}$ the speed of convergence to equilibrium in both the $\ell^2(\pi)$-topology and the total variation distance. In this line of work, we indicate that Miclo [82] has recently introduced a new notion known as Markov similarity and compares the mixing speed of Markov similar generators. As for the speed of convergence to equilibrium, there is a vast literature devoted to this important topic in various settings. For instance, in the case when P is a linear self-adjoint operator in $\ell^2(\pi)$ (reversible), the chain satisfies the spectral gap, which according to Roberts and Rosenthal [93], is equivalent to P being geometrically ergodic. In the non-reversible case, to overcome the lack of a spectral theory, many interesting reversibilization techniques have been implemented to obtain a quantitative rate of convergence and we refer the interested readers to Fill [42] and Saloff-Coste [97]. We propose an alternative approach based on the concept of similarity which seems to be a natural extension of the spectral gap estimate developed in the reversible case. Moreover, our technique enables us to provide an explicit and a spectral interpretation as a perturbed spectral gap estimate of the (discrete) hypocoercivity phenomena that was introduced by Villani [109] for general non-self-adjoint semigroups. We also manage to obtain generalization to the class $\mathcal{S}_{sf}^M$, a subclass of $\mathcal{S}_{sf}$ with stochastic monotone time-reversal to be formally introduced in Section 2.6, of the remarkable *spectral gap times mixing time going to infinity* separation cutoff criteria established by Diaconis and Saloff-Coste [31] for reversible birth-death chains, and recently by Mao et al. [76] for continuous-time skip-free chain with stochastic monotone time-reversal.

We point out that these three topics (potential, fluctuation and spectral theory) are intertwined. Indeed, there is of course a fundamental connection between the resolvent operator and the spectral theory of P as the spectrum of P is, by definition, the set of complex numbers z such that the resolvent operator $R_z = (zI - P)^{-1}$ does not exist or is unbounded. Note that since P is a contraction operator, this set is included in the unit disc. Moreover, the development of the fluctuation identities is based on the expression of the q-potential kernels whereas the spectral theory allow us to get an explicit representation of the transition kernel (and hence by integration of the q-resolvent operator) and, also, of the distribution of the first passage times.

The rest of the Chapter is organized as follows. In Section 2.2, we fix our notations and provide a review of birth-death chains and Martin boundary theory. The definition and study of the fundamental q-excessive functions as well as the expression of the q-potential kernel are discussed in Section 2.3. In Section 2.4, we present the fluctuations identities, that is the explicit characterization (via the probability generating function) of the law of the first exit times of skip-free chains, and, following the line of work by Feller [41], Kent and Longford [63] and Viskov [110], a characterization of these stopping times as discrete infinitely divisible variables in Section 2.5. We proceed in the Sections 2.6-2.8 by introducing the subclass of skip-free Markov chains that are similar to a diagonalizable birth-death chain and discuss their spectral properties as well as its applications to the speed of convergence to equilibrium and the study of the cutoff phenomena.

2.2 Preliminaries

In this Section, we review some classical concepts on Markov chains that will play a central role throughout the chapter. This include some facts of the potential theory and the Martin boundary theory of Markov chains.

2.2.1 Basic facts on Markov chains

We recall that X is said to be *upward skip-free* if the only upward transition is of unit size, yet it can have downward jump of any arbitrary magnitude, i.e. for all $x \in E$ and

$y \geqslant x+2$, $\mathbb{P}_x(X_1 = y) = 0$. We consider an upward skip-free irreducible Markov chain $X = (X_n)_{n\in\mathbb{N}_0}$ on a denumerable state space $E \subseteq \mathbb{Z}$ with left endpoint $\mathfrak{l}$ and right endpoint $\mathfrak{r}$. We use the convention that $\mathfrak{r} \in E$ if the boundary point $\mathfrak{r}$ is not absorbing. Otherwise, if $\mathfrak{r}$ is absorbing or $\mathfrak{r} = \infty$, we say that $X \in \mathcal{SF}_\infty$. We assume now that $X \in \mathcal{SF}_\infty$ and postpone to Section 2.3.3 the study of the Markov chain $X \notin \mathcal{SF}_\infty$.

Since X is irreducible, there exists π a positive and finite-valued excessive measure for P, that is, $\pi P \leqslant \pi$, see [61, Section 5.2 and 6.8], and that will serve as a reference measure. Let G_q be the q-potential kernel of X (or its Green function) with respect to the reference measure π. That is, for $0 < q < 1$,

$$G_q(x,y) \quad = \quad \sum_{n=0}^{\infty} q^n \frac{\mathbb{P}_x(X_n = y)}{\pi(y)} = \sum_{n=0}^{\infty} q^n \frac{p^{(n)}(x,y)}{\pi(y)}, \; x,y \in E. \qquad (2.2.1)$$

When $q = 1$, we write G, rather than G_1, to denote the 1-potential kernel. The π-dual matrix $\widehat{P} = (\widehat{p}(y,x))_{y,x\in E}$ is defined to be

$$\widehat{p}(y,x)\pi(y) \quad = \quad p(x,y)\pi(x). \qquad (2.2.2)$$

Denote ζ to be the lifetime of X, that is, $\zeta(\omega) = k$ if k is the last time that X is in the state space E, and $\zeta(\omega) = \infty$ otherwise.

Let us now define the hitting times associated with X. Denote T_A to be the first hitting time of the set A, that is,

$$T_A = \inf\{n \geqslant 0; X_n \in A\},$$

with the usual convention that $\inf\{\emptyset\} = \infty$. If $A = \{a\}$, we write $T_a = T_A$. Similarly, if $A = (\mathfrak{l}, b]$, we use $T_{b]} = T_A$. Denote the first return time to be $T_a^+ = \inf\{n \geqslant 1; X_n = a\}$.

We also use, for a real-valued function f and a measure μ on E, the following notation, for any $x, y \in E$,

$$Pf(x) = \sum_{y \in E} p(x,y) f(y),$$
$$\mu P(y) = \sum_{x \in E} \mu(x) p(x,y).$$

2.2.2 A review of birth-death chains

Let Y be a birth-death chain on E with transition operator $Q \in \mathcal{B}$, that is, $Q \in \mathcal{SF}$ with the additional requirements that $Q(x, x-1) > 0$ and $Q(x, x-y) = 0$ for $y \geqslant 2$. Q is a bounded self-adjoint operator in the Hilbert space $\ell^2(\pi_Q)$ where π_Q is the reversible measure, i.e. for all $x, y \in E$, $Q(x,y)\pi_Q(x) = Q(y,x)\pi_Q(y)$. The q-potential of Q, denoted by U_q, is well-known to take the form, for any $x, y \in E$,

$$U_q(x,y) = C_q F_q(x \wedge y) \widehat{F}_q(x \vee y), \tag{2.2.3}$$

where F_q (resp. $\widehat{F}_q$) is the unique increasing (resp. decreasing) solution to the equation $qQF = F$ satisfying appropriate boundary conditions, and C_q is their Wronskrian. Note that this equation boils down to a three-term recurrence equations, see [38, Ex. 5.3 p.150 and Section 5.4] for details regarding this expression. This is reminiscent to the expression of the potential kernel of the continuous time-space analogue, whose generator is a second order differential operator. We recall that a systematic and thorough study of one-dimensional diffusion has been undertaken by Feller [40]. Moreover, the moment generating function of the first hitting time $T_a^Y = \inf\{n \geqslant 0;\ Y_n = a\}$ of Y to a fixed level $a \in E$ is given by

$$\mathbb{E}_x[q^{T_a^Y}] = \frac{F_q(x)}{F_q(a)} \mathbb{1}_{\{x \leqslant a\}} + \frac{\widehat{F}_q(x)}{\widehat{F}_q(a)} \mathbb{1}_{\{x > a\}}. \tag{2.2.4}$$

Let us now describe a link between this expression and the diagonalization of the operator Q in $\mathcal{B}$. We assume, for sake of simplicity, that $\mathfrak{l} = 0$ and thus E is countable subset of $\mathbb{N}$. It is worth mentioning that the similarity transform $D_{\sqrt{\pi}} Q D_{\frac{1}{\sqrt{\pi}}}$, where D_a is the diagonal matrix of a yields to a symmetric tridiagonal matrix which belongs to the well-studied class of Jacobi matrices. We also recall that a function ϕ is a Pick function if ϕ admits an analytical continuation to the cut complex plane $\mathbb{C} \setminus [0, \infty)$ such that $\Im\phi(z)\Im(z) \geqslant 0$, and we denote by $\mathcal{P}$ the subset of such Pick functions that admit the representation, for $\Im(z) > 0$,

$$\phi(z) = \int_{-1}^{1} \frac{d\Delta(r)}{r - z}, \tag{2.2.5}$$

where Δ is a probability measure on $[-1, 1]$. From the work of Karlin and McGregor [59], we know that there exists the spectral mapping $\mathrm{K} : \mathcal{B} \to \mathcal{P}$ which is one-to-one and Δ is the spectral measure of Q. More specifically, we have for any $f \in \ell^2(\pi_Q)$ and $n \in \mathbb{N}$, Q^n can be diagonalizable as follows

$$Q^n f = \int_{-1}^{1} r^n \langle f, \mathrm{F}_r \rangle_{\pi_Q} \mathrm{F}_r \, d\Delta(r), \tag{2.2.6}$$

where for $r \in supp(\Delta)$, $Q\mathrm{F}_r = r\mathrm{F}_r$. Note that $S(Q)$ the spectrum of Q is such that $S(Q) = supp(\Delta) \subset [-||Q||, ||Q||] \subseteq [-1, 1]$. Another remarkable and deep result, which is due to Krein, is the onto property of the spectral map K. Indeed for any $\phi \in \mathcal{P}$, i.e. a Pick function of the form (2.2.5), there exists an unique $Q \in \mathcal{B}$ on E with $\mathfrak{l}$ a non-killing boundary with a spectral representation of the form (2.2.6). The first proof of this conjecture was given in [38, Chapter 6]; it uses the theory of Hilbert spaces of entire functions, and a deep uniqueness theorem due to de Branges. Note also that such a spectral representation reveals that when the spectrum of Q is composed of isolated eigenvalues then they are necessarily simple (see e.g. Karlin and McGregor [59] and [38]).

2.2.3 Martin boundary theory of denumerable Markov chains

In this subsection, we review some essential results of Martin boundary theory of denumerable Markov chains based on [39, 61, 113]. First, we recall the definition of q-excessive, q-harmonic and q-potential functions. A non-negative function f on E is q-excessive (resp. q-harmonic) if $qPf(x) \leqslant f(x)$ (resp. $qPf(x) = f(x)$) for all $x \in E$, where $0 < q \leqslant 1$. A non-negative function f on E is a q-potential if f is q-excessive and $\lim_{n\to\infty}(qP)^n f(x) = 0$ for all $x \in E$. We say that f is harmonic (resp. excessive) if f is 1-harmonic (resp. 1-excessive). Note that the irreducibility property implies that $f \in \mathcal{E}_q$ is positive, since if there exists $y \in E$ such that $f(y) > 0$, then by irreducibility there exists n such that for $x \in E$ $f(x) \geqslant qp^n(x,y)f(y) > 0$. We write

$$\mathcal{E}_q = \{f : E \to \mathbb{R}^+;\ qPf \leqslant f\}\ (\text{resp. } \mathcal{H}_q, \mathcal{P}_q)$$

to be the set of q-excessive (resp. q-harmonic, q-potential) functions on E. We simply write $\mathcal{E}$ (resp. $\mathcal{H}$, $\mathcal{P}$) to denote the set of excessive (resp. harmonic, potential) functions. We will also use the notation $\widehat{\mathcal{E}}_q$ (resp. $\widehat{\mathcal{H}}_q, \widehat{\mathcal{P}}_q$) to denote the set of q-excessive (resp. q-harmonic, q-potential) functions associated with $\widehat{P}$, which was defined in (2.2.2). We point out that if $\hat{h} \in \widehat{\mathcal{E}}_q$, then $\hat{h}\pi$ is a q-excessive measure for P in the sense that $q\,\hat{h}\pi P f \leqslant \hat{h}\pi f$. Another commonly used terminology for excessive (resp. harmonic, potential) function is superharmonic (resp. invariant, purely-excessive) function.

We further recall the definition of minimal q-harmonic function. A non-zero function f on E is minimal q-excessive if $f = f_1 + f_2$ implies $f = c_i f_i$ for $i = 1, 2$, where $f_1, f_2 \in \mathcal{E}_q$, c_1 and c_2 are constants and $0 < q \leqslant 1$. We say that f is minimal excessive if f is minimal 1-excessive. We write $\mathcal{E}_q^{min}$ (resp. $\mathcal{H}_q^{min}$) to be the set of minimal q-excessive (resp. minimal q-harmonic) functions on E. Next, we state the classical Riesz representation theorem, which gives an unique decomposition of excessive function.

Theorem 2.2.1 (Riesz representation theorem). *For $0 < q \leqslant 1$, every q-excessive function can be written uniquely as the sum of a q-potential and a q-harmonic function. That is, if $f \in \mathcal{E}_q$, then*

$$f(x) = \sum_{y \in E} G_q(x,y) k_q(y) \pi(y) + h_q(x),$$

where $k_q(x) = f(x) - qPf(x)$ and $h_q(x) = \lim_{n\to\infty} (qP)^n f(x) \in \mathcal{H}_q$.

Suppose $(X, \mathbb{P})$ is a transient Markov chain with transition matrix P and reference measure π on a denumerable state space E. For a measure μ on E, define $E_\mu = \{y \in E;\ \mu G(y) > 0\}$, where $\mu G(y) = \sum_{x \in E} \mu(x) G(x,y)$. We say that the measure μ is a *standard measure* if $E_\mu = E$. For $x, y \in E$ and $0 < q \leqslant 1$, the q-Martin kernel associated to a standard measure μ is defined to be

$$_{\mu}K_q(x,y) = \frac{G_q(x,y)}{\mu G_q(y)}, \tag{2.2.7}$$

where the denominator is positive since μ is standard. Since X is transient, $_{\mu}K_q(x,y)$ is finite for any $x, y \in E$. Next, define a metric $_{\mu}d_q$ on the space E by

$$_{\mu}d_q(y,z) = \sum_{x \in E} w_x \frac{|_{\mu}K_q(x,y) - {}_{\mu}K_q(x,z)| + |\mathbb{1}_{\{x=y\}} - \mathbb{1}_{\{x=z\}}|}{C_x^\mu + 1},$$

where the weights $(w_x)_{x \in E}$, with $w_x > 0$, are chosen such that $\sum_{x \in E} w_x < \infty$, and C_x^μ is a function that depends only on x and satisfies $_{\mu}K_q(x,y) \leqslant C_x^\mu$. We can obtain the completion of E with respect to the metric $_{\mu}d_q$, namely $\overline{E}$, and the boundary of E in $\overline{E}$ is denoted as $\partial E = \overline{E} - E$. $\overline{E}$ is the Martin compactification of E and ∂E is the Martin boundary. The set

$$\partial_{\mathbb{P}} E = \{y \in \partial E;\ {}_{\mu}K_q(x,y) \text{ is minimal } q\text{-harmonic in } x\}$$

is known as the *minimal Martin boundary*. When there is no ambiguity in the probability measure, we write $\partial_m E = \partial_{\mathbb{P}} E$. The inclusion of the indicator terms $\mathbb{1}_{\{x=y\}}$ and $\mathbb{1}_{\{x=z\}}$ in the metric $_{\mu}d_q$ ensures that E is an open set in the Martin compactification $\overline{E}$, and the

Martin boundary ∂E is closed. Next, fix a reference point $\mathfrak{o} < \mathfrak{r} - 1$, and we write the q-Martin kernel ${}_{\delta_\mathfrak{o}}K_q(x, y)$ as

$$ {}_{\delta_\mathfrak{o}}K_q(x, y) = \frac{\mu G_q(y)}{G_q(\mathfrak{o}, y)} {}_\mu K_q(x, y) = \frac{{}_\mu K_q(x, y)}{{}_\mu K_q(\mathfrak{o}, y)}. $$

Similarly, we have

$$ {}_\mu K_q(x, y) = \frac{G_q(\mathfrak{o}, y)}{\mu G_q(y)} {}_{\delta_\mathfrak{o}}K_q(x, y) = \frac{{}_{\delta_\mathfrak{o}}K_q(x, y)}{{}_\mu K_q(y)}. $$

A sequence (y_n) is a Cauchy sequence in the metric space $(E, {}_\mu d_q)$ if and only if $({}_\mu K_q(x, y_n))$ is a Cauchy sequence of real numbers for every x if and only if $({}_{\delta_\mathfrak{o}}K_q(x, y_n))$ is a Cauchy sequence of real numbers for every x if and only if (y_n) is a Cauchy sequence in the metric space $(E, {}_{\delta_\mathfrak{o}}d_q)$. Thus, up to homeomorphism, $\overline{E}$ is independent of the choice of the reference point $\mathfrak{o}$. It can also be shown that $\overline{E}$ is independent of the choice of the weights $(w_x)_{x \in E}$ (see e.g. Proposition 10.13 in [61]). From now on, we fix the reference point $\mathfrak{o}$ and write

$$ K_q(x, y) = {}_{\delta_\mathfrak{o}}K_q(x, y) = \frac{G_q(x, y)}{G_q(\mathfrak{o}, y)}. $$

Let $\Omega_\infty = \{\omega;\ \text{there is}\, x_\infty \in \partial E \,\text{such that}\, x_n \to x_\infty \,\text{in the Martin topology}\}$. Ω_∞ can be interpreted as the set of *non-terminating* trajectories of X that converges to ∂E. If (X, P) is transient and P is a stochastic matrix then there is a random variable X_∞ taking values in ∂E such that for each $x \in E$, $\mathbb{P}_x\left(\lim_{n\to\infty} X_n = X_\infty\right) = 1$. In terms of trajectory space, we have $\mathbb{P}_x\left(\Omega_\infty\right) = 1$. If P is strictly substochastic at some states, we should extend the trajectories to $E \cup \{\nabla\}$, where ∇ is the graveyard state. Define

$$ \Omega_\nabla = \{\omega;\ \text{there is}\, k \geqslant 1 \,\text{such that}\, x_n \in E,\, \forall n \leqslant k \,\text{and}\, x_n = \nabla, \forall n \geqslant k+1\}. $$

Ω_∇ is the set of trajectories that eventually reach ∇. Denote ζ to be the lifetime of X, that is, $\zeta(\omega) = \mathfrak{K}$ for $\omega \in \Omega_\nabla$, where $\mathfrak{K}$ is the last time that X is in the state space E (as defined in Ω_∇), and $\zeta(\omega) = \infty$ otherwise. Define $\Omega_\zeta = \Omega_\nabla \cup \Omega_\infty$. If $(X, \mathbb{P})$ is transient then there is a random variable X_ζ taking values in $\overline{E}$ such that for each $x \in E$, $\mathbb{P}_x\left(\lim_{n\to\zeta} X_n = X_\zeta\right) = 1$. In terms of trajectory space, this means that $\mathbb{P}_x\left(\Omega_\zeta\right) = 1$.

Next, suppose now that $h \in \mathcal{E}_q$. The h-process of X is defined to be the Markov chain on $E^h = \{y \in E;\ h(y) > 0\} = E$, by irreducibility, with the canonical measure $\mathbb{P}_x^h$ such that $\mathbb{P}_x^h(X_1 = y) = \dfrac{p(x,y)h(y)}{h(x)}$. Recalling that $\mathfrak{o}$ is the fixed reference point, the q-potential and q-Martin kernels associated with the h-process take respectively the form

$$G_q^h(x,y) = \frac{G_q(x,y)h(y)}{h(x)}, \tag{2.2.8}$$

$$K_q^h(x,y) = \frac{K_q(x,y)h(\mathfrak{o})}{h(x)}. \tag{2.2.9}$$

From (2.2.9) and the definition of the metric ${}_\mu d_q$, we observe that the Martin compactification $\overline{E}$ is homeomorphic to the Martin compactification of the h-process. Next, we state the following which are the main claims of Theorem 6 and 7 in [39].

Theorem 2.2.2 (Uniqueness of the representation)**.** *Let $q \in (0,1]$. If $h \in \mathcal{E}_q$ such that $h(\mathfrak{o}) = 1$ then h has a unique representation of the form*

$$h(x) = \int_{E\cup\partial_m E} K_q(x,y)\,\mu_h(dy) = K_q\mu_h(x),$$

where $\mu_h(\cdot) = \mathbb{P}^h(X_\zeta \in \cdot)$ defines a probability measure. Conversely, for any finite measure μ, the mapping $x \mapsto \int_{E\cup\partial_m E} K_q(x,y)\,d\mu(y)$ defines an q-excessive function, which is q-harmonic if and only if $\mu(E) = 0$. Finally, for all $y \in E\cup\partial_m E$, let $h^y(\cdot) = K(\cdot,y)$. Then $y \in E\cup\partial_m E$ if and only if $\mathbb{P}^{h^y}(X_\zeta = y) = 1$. Moreover we have $\mathbb{P}^{h^y}(\zeta = \infty) = 1$ if and only if $y \in \partial_m E$.

The previous claim means that X is forced to terminate at the point $y \in E\cup\partial_m E$ $\mathbb{P}_x^{h^y}$-almost surely.

Theorem 2.2.3. *We have $\mathcal{E}_q^{min} = \{h_q;\ h_q(x) = CK_q(x,y), C > 0$ and $y \in E\cup\partial_m E\}$.*

Finally, we recall the following useful result whose proof follows readily from [21, Theorem 11.9].

Lemma 2.2.1. *Suppose that $h_q \in \mathcal{E}_q$, and T is a stopping time with respect to $(\mathcal{F}_n)_{n\geqslant 1}$. For $x \in E^{h_q}$,*

$$\mathbb{P}_x^{h_q}(T < \infty) = \frac{1}{h_q(x)} \mathbb{E}_x \left[q^T h_q(X_T) \mathbb{1}_{\{T<\infty\}} \right].$$

2.3 Potential Theory

In this Section, we provide an explicit representation of the q-potential kernel G_q, as defined in (2.2.1), of an upward skip-free Markov chain $X \in \mathcal{SF}$. We recall that in the simpler case when X boils down to a birth-death Markov chain, i.e. skip-free in both directions, then its q-potential kernel takes the form

$$U_q(x,y) = C_q F_q(x \wedge y) \widehat{F}_q(x \vee y), \tag{2.3.1}$$

where F_q and $\widehat{F}_q$ are the two fundamental solutions of a three-term recurrence equations (discrete analogue of a second order differential equation), see Section 2.2.2 for more details regarding this expression. In our context, the situation is more delicate as one has to solve an infinite recurrence equation whose set of solutions does not seem to have been clearly identified in the literature. Although the issue of solving this equation is of algebraic nature, we shall elaborate a strategy based mostly on a combination of techniques from probability theory and potential theory. We start by expressing the q-potential kernel in terms of the so-called fundamental q-excessive (for short FqE) functions of the following three processes: $(X, \mathbb{P})$, $(X, \widehat{\mathbb{P}})$ and $(X, \mathbb{P}^{y]})$, where $(X, \mathbb{P}^{y]})$ is the Markov chain $(X, \mathbb{P})$ killed upon entering the half-line $[[\mathfrak{l}, y]$, which is plainly an upward skip-free Markov chain on the state space $E^{y]} = (y, \mathfrak{r}]]$, with transition kernel denoted by $P^{y]}$. We are now ready to state the main result of this Section.

Theorem 2.3.1. *1. Writing, for any $x \in E$ and $0 < q < 1$,*

$$H_q(x) = K_q \delta_{\mathfrak{r}}(x), \tag{2.3.2}$$

(resp. $\widehat{H}_q(x) = \widehat{K}_q \delta_{\mathfrak{l}}(x)$) with $\delta_{\mathfrak{r}}$ is the Dirac mass at $\mathfrak{r}$, we have, for all $0 < q < 1$, that

$$H_q \in \mathcal{E}_q^{min}$$

(resp. $\widehat{H}_q \in \widehat{\mathcal{E}}_q^{min}$) and it is the unique minimal increasing (resp. decreasing) q-excessive for P (resp. $\widehat{P}$) such that $H_q(\mathfrak{o}) = 1$ (resp. $\widehat{H}_q(\mathfrak{o}) = 1$). Moreover, if $X \in \mathcal{SF}_\infty$ (resp. $X \in \mathcal{SF} \setminus \mathcal{SF}_\infty$) then H_q is the unique increasing function in $\mathcal{H}_q$ (resp. $H_q \in \mathcal{P}_q$ with $H_q(\mathfrak{r}) < \infty$).

2. *For any $y < \mathfrak{r}$, $0 < \kappa_q^{y]} = \lim_{x \to \mathfrak{r}} \frac{K_q \delta_{\mathfrak{r}}(x)}{K_q^{y]} \delta_{\mathfrak{r}}(x)} < \infty$. Then the function $\mathbf{H}_q^{y]}$ defined, for any $x \in E^{y]}$, by*

$$\mathbf{H}_q^{y]}(x) = \kappa_q^{y]} K_q^{y]} \delta_{\mathfrak{r}}(x) \tag{2.3.3}$$

has (with respect to $P^{y]}$) the same property than H_q but with the normalization

$$\lim_{x \to \mathfrak{r}} \frac{\mathbf{H}_q^{y]}(x)}{H_q(x)} = 1$$

(recall that by convention $\mathbf{H}_q^{y]}(x) = 0$ for any $x \leqslant y$).

3. *Finally, let $X \in \mathcal{SF}_\infty$ and set $C_q = G_q(\mathfrak{o}, \mathfrak{o}) > 0$. Then, we have, for all $x, y \in E$,*

$$G_q(x, y) = C_q \, \widehat{H}_q(y) \left(H_q(x) - \mathbf{H}_q^{y]}(x) \right). \tag{2.3.4}$$

Remark 2.3.1. Up to minor modifications all statements hold for $q = 1$ when $(X, \mathbb{P})$ is transient.

Remark 2.3.2. The terminology FqE comes from the birth-death case where these functions are usually called the fundamental solutions of the associated three-term recurrence equation and are q-excessive.

Remark 2.3.3. In comparison to the expression of the q-potential of the birth-death chain given in (2.3.1), there is in addition to the (unique) FqE functions for $(X, \mathbb{P})$ and its dual $(X, \widehat{\mathbb{P}})$, the sequence of F$q$E functions associated to the family of killed Markov chains $(X, \mathbb{P}^{y]})_{y \in E}$.

Remark 2.3.4. Additional properties of H_q and its relation with infinite divisibility are studied in Section 2.5 Corollary 2.5.1.

We proceed with the proof of these statements which is split into several intermediate results.

2.3.1 Proof of Theorem 2.3.1

We start with the following result which relates the Martin kernel to the hitting time distribution.

Lemma 2.3.1. *For any $x, y \in E$,*

$$\mathbb{E}_x(q^{T_y}) = \frac{K_q\delta_y(x)}{K_q\delta_y(y)}.$$

Proof. By Theorem 2.2.1, for any $y \in E$, $K_q\delta_y \in \mathcal{P}_q$ which leads, by Theorem 2.2.2, to $\mathbb{P}^{K_q\delta_y}(X_\zeta = y, \zeta < \infty) = 1$, that is, for any $x, y \in E$,

$$\mathbb{P}_x^{K_q\delta_y}(T_y < +\infty) = 1.$$

Since, on the other hand, by Lemma 2.2.1, we have, for any $x, y \in E$,

$$\mathbb{P}_x^{K_q\delta_y}(T_y < +\infty) = \mathbb{E}_x(q^{T_y})\frac{K_q\delta_y(y)}{K_q\delta_y(x)}$$

we complete the proof. $\square$

Proof of Theorem 2.3.1(1)

Suppose that $x \vee \mathfrak{o} \leqslant y \leqslant a$, where $x \vee \mathfrak{o} = \max\{x, \mathfrak{o}\}$. By means of Lemma 2.3.1, the upward skip-free property and the strong Markov property, we obtain that

$$K_q\delta_a(x) = \frac{\mathbb{E}_x(q^{T_a})}{\mathbb{E}_{\mathfrak{o}}(q^{T_a})} = \frac{\mathbb{E}_x(q^{T_y})\mathbb{E}_y(q^{T_a})}{\mathbb{E}_{\mathfrak{o}}(q^{T_y})\mathbb{E}_y(q^{T_a})} = K_q\delta_y(x). \tag{2.3.5}$$

Thus, for any $y \geqslant x \vee \mathfrak{o},\ K_q\delta_y(x) = K_q(x, x \vee \mathfrak{o})$. Hence, one can trivially define the function H_q as the extended Martin kernel at $\mathfrak{r}$, that is, for $x \in E$,

$$H_q(x) = \lim_{y \to \mathfrak{r}} K_q(x, y) = \int_{E \cup \partial_m E} K_q(x, y)\delta_{\mathfrak{r}}(dy).$$

Hence if $X \in \mathcal{SF}_\infty$ (resp. $X \in \mathcal{SF} \setminus \mathcal{SF}_\infty$) then $\mathfrak{r} \in \partial_{\mathbb{P}} E$ (resp. $\mathfrak{r} \in E$) and thus by theorems 2.2.2 and 2.2.3 (resp. and Theorem 2.2.1), we get that $H_q \in \mathcal{H}_q \cap \mathcal{E}_q^{min}$ (resp. $H_q \in \mathcal{P}_q \cap \mathcal{E}_q^{min}$). Next note, from the first identity in (2.3.5), that $H_q(\mathfrak{o}) = 1$, and, for $x \leqslant y$, $H_q(x) = K_q\delta_y(x)$. Hence, by Lemma 2.3.1 we have, for any $x \leqslant y$,

$$\mathbb{E}_x(q^{T_y}) = \frac{H_q(x)}{H_q(y)}, \tag{2.3.6}$$

which entails, by the irreducibility of X, that H_q is positive everywhere since for any $x \in E$, the ratio $H_q(x) = \frac{\mathbb{E}_x(q^{T_a})}{\mathbb{E}_{\mathfrak{o}}(q^{T_a})} > 0$. To see that the mapping $x \mapsto H_q(x)$ is increasing, one observes from again the strong Markov property and the upward skip-free property that (recall that $x \leqslant y \leqslant a$)

$$H_q(x) = \frac{\mathbb{E}_x(q^{T_y})\mathbb{E}_y(q^{T_a})}{\mathbb{E}_{\mathfrak{o}}(q^{T_a})} = \mathbb{E}_x(q^{T_y})H_q(y) < H_q(y).$$

To prove the uniqueness, we proceed by contradiction and thus assume that there exists a positive function $h_q \in \mathcal{E}_q^{min}$ (resp. in $\mathcal{H}_q$ when $X \in \mathcal{SF}_\infty$) which differs from H_q and which is also an increasing function on E. Then, according to Theorem 2.2.3, there exists $y_0 \in E$ (resp. $y_0 = \mathfrak{l}$ or $y_0 = \mathfrak{r}$) such that for all $x \in E$, $h_q(x) = \frac{K_q(x,y_0)}{K_q(o,y_0)}$. Thus, on the one hand, from Lemma 2.3.1 we deduce that for any $x \leqslant y_0$,

$$\mathbb{E}_x(q^{T_y}) = \frac{K_q\delta_{y_0}(x)}{K_q\delta_{y_0}(y_0)} = \frac{h_q(x)}{h_q(y_0)}.$$

This combines with (2.3.6) yield that $h_q(x) = H_q(x)$ for any $x \leqslant y_0$, which proves the claim when $H_q \in \mathcal{H}_q$ and $y_0 = \mathfrak{r}$. In the other cases, choose $x > y_0$ such that $h_q(x) \neq H_q(x)$. Then, observe from Theorem 2.2.2 that $\mathbb{P}_x^{h_q}(T_{y_0} < +\infty) = 1$. As $\mathbb{P}_x^{h_q}(T_{y_0} < +\infty) = \frac{h_q(y_0)}{h_q(x)}\mathbb{E}_x(q^{T_{y_0}})$ and h_q is increasing and $x > y_0$, we get that $\mathbb{E}_x(q^{T_{y_0}}) > 1$ which is impossible. This completes the uniqueness property of H_q. To complete the proof of Theorem 2.3.1(1), we use similar arguments for deriving the stated properties of $\widehat{H}_q$ after recalling that the dual chain $(X, \widehat{\mathbb{P}})$ is downward skip-free.

Proof of Theorem 2.3.1(2)

We start with the following claim which is a straightforward reformulation of Theorem 2.3.1(1) for the killed chains.

Lemma 2.3.2. *Let $b \in E$ and choose a reference point $\mathfrak{o}^{b]} \in E^{b]}$. Then, for all $0 < q < 1$, the function $H_q^{b]}(x) = K_q^{b]}\delta_{\mathfrak{r}}(x)$ defined on $E^{b]}$ is positive on $E^{b]}$, minimal, increasing q-harmonic for $P^{b]}$ with $H_q^{b]}(\mathfrak{o}^{b]}) = 1$. Moreover, for any $b \leqslant x \leqslant a$,*

$$\mathbb{E}_x^{b]}(q^{T_a}) = \mathbb{E}_x(q^{T_a}\mathbb{1}_{\{T_a < T_{b]}\}}) = \frac{H_q^{b]}(x)}{H_q^{b]}(a)}.$$

Proof. Under $\mathbb{P}^{b]}$, X is an upward skip-free Markov chain on $E^{b]}$, the results follows from Theorem 2.3.1(1) and the identity (2.3.6). □

The following lemma expresses the pgf of the *downward* hitting times $(T_b, \mathbb{P}_x^{[a})$, where $a > x > b$, in terms of the FqE functions of $(X, \mathbb{P})$ and $(X, \mathbb{P}^{b]})$.

Lemma 2.3.3. *For any $b < x < a$ and $0 < q < 1$, we have*

$$\mathbb{E}_x(q^{T_b}\mathbb{1}_{\{T_b < T_a\}}) = \frac{H_q(x)}{H_q(b)} - \frac{H_q(a)}{H_q(b)}\frac{H_q^{b]}(x)}{H_q^{b]}(a)}.$$

Proof. Since by definition $H_q = K_q\delta_{\mathfrak{r}}$, we have, from Theorem 2.2.2 that,

$$\mathbb{P}^{H_q}(X_\zeta = \mathfrak{r}) = 1. \tag{2.3.7}$$

This combines with the upward skip-free property, see Lemma 2.2.1, yields that under $\mathbb{P}^{H_q}_x$ the sample paths of X that drop below b before hitting a must reaches b before reaching a. That is

$$\mathbb{P}^{H_q}_x(T_b < T_a) = \mathbb{P}^{H_q}_x(T_{b]} < T_a) = 1 - \mathbb{P}^{H_q}_x(T_a < T_{b]}), \tag{2.3.8}$$

where we used again (2.3.7) for the second identity. Hence, an application of Lemma 2.2.1 gives

$$\mathbb{P}^{H_q}_x(T_b < T_a) = \frac{H_q(b)}{H_q(x)}\mathbb{E}_x(q^{T_b}\mathbb{1}_{\{T_b<T_a\}}) = 1 - \frac{H_q(a)}{H_q(x)}\mathbb{E}_x(q^{T_a}\mathbb{1}_{\{T_a<T_{b]}\}}) = 1 - \frac{H_q(a)}{H_q(x)}\frac{H_q^{b]}(x)}{H_q^{b]}(a)},$$

where the last equality follows from Lemma 2.3.2. Rearranging the terms provides the desired result. □

The proof of Theorem 2.3.1(2) follows readily after the following claim.

Lemma 2.3.4. *For $x \in E^{b]}$, define*

$$\kappa_q^{b]}(x) = \frac{H_q(x)}{H_q^{b]}(x)}.$$

Then the mapping $x \mapsto \kappa_q^{b]}(x)$ is non-increasing on $E^{b]}$ with $0 < \kappa_q^{b]}(x) < \infty$. Furthermore, $0 < \kappa_q^{b]} = \lim_{x\to\mathfrak{r}} \kappa_q^{b]}(x) < \infty$.

Proof. It is clear that, for all $x \in E^{b]}$, $0 < \kappa_q^{b]}(x) < \infty$, since both H_q and $H_q^{b]}$ are positive and finite on $E^{b]}$. Next, for any $x \in E^{b]}$,

$$qP^{b]}H_q(x) \leqslant qPH_q(x) \leqslant H_q(x),$$

where the first inequality follows from the fact that $P^{b]}$ is the restriction of P to $E^{b]}$, and we use that $H_q \in \mathcal{E}_q$ in the second inequality. Therefore, H_q (restricted on $E^{b]}$) is q-excessive for $P^{b]}$. Thus, one may define the H_q-transform of $qP^{b]}$ by

$$^{H_q}P^{b]}(x,y) = \frac{H_q(y)}{H_q(x)} qP^{b]}(x,y),$$

where $x, y \in \{x \in E^{b]} : H_q(x) > 0\} = E^{b]}$ by Theorem 2.3.1. Using Lemma 2.2.1 and Lemma 2.3.2, we have, for any $x \leqslant a$,

$$^{H_q}\mathbb{P}_x^{b]}(T_a < \infty) = \frac{H_q(a)}{H_q(x)} \mathbb{E}_x^{b]}(q^{T_a}) = \frac{H_q^{b]}(x)}{H_q(x)} \frac{H_q(a)}{H_q^{b]}(a)} = \frac{\kappa_q^{b]}(a)}{\kappa_q^{b]}(x)}. \tag{2.3.9}$$

Since $(X, {}^{H_q}\mathbb{P}_x^{b]})$ is a transient upward skip-free Markov chain, using Lemma 2.3.1 and Theorem 2.3.1(1) for $q = 1$, one easily deduces, with the obvious notation, that $\frac{1}{C\kappa_q^{b]}(x)} = {}^{H_q}K^{b]}\delta_{\mathfrak{r}}(x)$, for some $C > 0$. Thus the mapping $x \mapsto \kappa_q^{b]}(x)$ is non-increasing. Henceforth $\kappa_q^{b]} = \lim_{x \to \mathfrak{r}} \kappa_q^{b]}(x) \leqslant \kappa_q^{b]}(b+1) < \infty$. Observing that both $H_q(\mathfrak{r})$ and $H_q^{b]}(\mathfrak{r})$ are finite when $X \in \mathcal{SF} \setminus \mathcal{SF}_\infty$, we readily get that $\kappa_q^{b]} > 0$ which completes the proof in this case. It remains to show that

$$\lim_{a \to \mathfrak{r}} \frac{\kappa_q^{b]}(a)}{\kappa_q^{b]}(x)} = \lim_{a \to \mathfrak{r}} {}^{H_q}\mathbb{P}_x^{b]}(T_a < \infty) > 0, \tag{2.3.10}$$

when $X \in \mathcal{SF}_\infty$. To this end, we assume the contrary, that is, $\lim_{a \to \mathfrak{r}} {}^{H_q}\mathbb{P}_x^{b]}(T_a < \infty) = 0$. Since ${}^{H_q}\mathbb{P}_x^{b]}(T_a < \infty) = \mathbb{P}_x^{H_q}(T_a < T_{b]})$, the assumption becomes $\lim_{a \to \mathfrak{r}} \mathbb{P}_x^{H_q}(T_a < T_{b]}) = 0$ and (2.3.8) leads to

$$\lim_{a \to \mathfrak{r}} \mathbb{P}_x^{H_q}(T_a < T_b) = 0. \tag{2.3.11}$$

Next, by means of a first step analysis and the upward skip-free property, we obtain

$$\mathbb{P}_b^{H_q}(T_b^+ < T_a) = p^{H_q}(b, b+1)\mathbb{P}_{b+1}^{H_q}(T_b^+ < T_a) + p^{H_q}(b,b) + \sum_{y<b} p^{H_q}(b,y)\mathbb{P}_y^{H_q}(T_b^+ < T_a). \tag{2.3.12}$$

Taking $a \to \mathfrak{r}$, the left-hand side converges to $\mathbb{P}_b^{H_q}(T_b^+ < \infty)$ (recall that $T_b^+ = \inf\{n \geqslant 1;\ X_n = b\}$) due to the monotone convergence theorem, the upward skip-free property and the fact that $X \in \mathcal{SF}_\infty$. On the right-hand side of (2.3.12), the first term converges to $p^{H_q}(b, b+1)$ as a result of (2.3.11), while the third term converges to $\sum_{y<b} p^{H_q}(b,y)\mathbb{P}_y^{H_q}(T_b^+ < \infty)$ by invoking the dominated convergence theorem. Therefore, we arrive at

$$\begin{aligned}\mathbb{P}_b^{H_q}(T_b^+ < \infty) &= p^{H_q}(b, b+1) + p^{H_q}(b,b) + \sum_{y<b} p^{H_q}(b,y)\mathbb{P}_y^{H_q}(T_b^+ < \infty) \\ &= p^{H_q}(b, b+1) + \sum_{y \leqslant b} p^{H_q}(b,y) = \mathbb{P}_b^{H_q}(\zeta > 1) = 1,\end{aligned}$$

where the second equality comes from the identity $\mathbb{P}_y^{H_q}(T_b^+ < \infty) = 1$ which holds since $\mathbb{P}_y^{H_q}(X_\zeta = \mathfrak{r}) = 1$ and $y < b$, while the last equality is due to Theorem 2.2.2 with the fact that $H_q \in \mathcal{H}_q$ since $X \in \mathcal{SF}_\infty$. This is not possible since X is transient. Therefore, we conclude that $\kappa_q^{b]} > 0$. □

2.3.2 Proof of Theorem 2.3.1(3)

We start with the following extension of Lemma 2.3.3.

Lemma 2.3.5. *For any $x, y \in E$ and $0 < q < 1$, we have*

$$\mathbb{E}_x\left(q^{T_y}\right) = \frac{H_q(x) - \mathbf{H}_q^{y]}(x)}{H_q(y)}. \tag{2.3.13}$$

Proof. The case when $x \leqslant y$ is proved in (2.3.6). Next assume that $x > y$. Thanks to the upward skip-free property of X, for any $b \in E$ the mapping $\mathbb{1}_{\{T_b < T_a\}}$ is increasing in a large enough. Then, the monotone convergence theorem and the fact that $X \in \mathcal{SF}_\infty$ give $\lim_{a\to\mathfrak{r}} \mathbb{E}_x(q^{T_b}\mathbb{1}_{\{T_b<T_a\}}) = \mathbb{E}_x(q^{T_b})$. The sought result follows immediately from Lemma 2.3.3 and Lemma 2.3.4. □

We are now ready to prove the expression (2.3.4). First using (2.3.6), Lemma 2.3.1 and the definition of the Martin kernel in (2.2.7), we obtain, for any $x \leqslant \mathfrak{o}$,

$$\mathbb{E}_x(q^{T_{\mathfrak{o}}}) = H_q(x) = \frac{G_q(x,\mathfrak{o})}{G_q(\mathfrak{o},\mathfrak{o})} = \frac{G_q(x,\mathfrak{o})}{C_q}. \tag{2.3.14}$$

Next, for sake of clarity we state the analogue of the identity (2.3.6) for the dual chain $(X, \widehat{\mathbb{P}})$.

Lemma 2.3.6. *For all $0 < q < 1$ and any $x \leqslant y$,*

$$\widehat{\mathbb{E}}_y(q^{T_x}) = \frac{\widehat{H}_q(y)}{\widehat{H}_q(x)}. \tag{2.3.15}$$

A specific application of the previous result yields, for any $x \leqslant \mathfrak{o}$, that

$$\widehat{\mathbb{E}}_{\mathfrak{o}}(q^{T_x}) = \frac{1}{\widehat{H}_q(x)} = \frac{\widehat{G}_q(\mathfrak{o},x)}{\widehat{G}_q(x,x)} = \frac{G_q(x,\mathfrak{o})}{G_q(x,x)}, \tag{2.3.16}$$

where we use the identity $\widehat{H}_q(\mathfrak{o}) = 1$ in the first equality, the dual version of (2.3.14) for the second one and the integrated version of the dual identity (2.2.2) for the last one. Combining (2.3.14) and (2.3.16), we get, for any $x \leqslant \mathfrak{o}$,

$$G_q(x,x) = C_q H_q(x) \widehat{H}_q(x). \tag{2.3.17}$$

For any $x \geqslant \mathfrak{o}$, we reverse the role of x and $\mathfrak{o}$ to obtain, respectively,

$$\mathbb{E}_{\mathfrak{o}}(q^{T_x}) = \frac{1}{H_q(x)} = \frac{G_q(\mathfrak{o},x)}{G_q(x,x)}, \tag{2.3.18}$$

$$\widehat{\mathbb{E}}_x(q^{T_{\mathfrak{o}}}) = \widehat{H}_q(x) = \frac{\widehat{G}_q(x,\mathfrak{o})}{\widehat{G}_q(\mathfrak{o},\mathfrak{o})} = \frac{G_q(\mathfrak{o},x)}{G_q(\mathfrak{o},\mathfrak{o})}. \tag{2.3.19}$$

Combining (2.3.18) and (2.3.19), we again arrive at (2.3.17), which shows that (2.3.17) holds for all $x \in E$. Note that (2.3.17) holds regardless of the boundary condition at $\mathfrak{r}$. In particular, when $X \in \mathcal{SF}_\infty$, (2.3.13), (2.3.14) (replacing y by $\mathfrak{o}$), (2.3.17) and Lemma 2.3.1 give (2.3.4) which complete the proof of Theorem 2.3.1.

2.3.3 $X \in \mathcal{SF} \setminus \mathcal{SF}_\infty$

We suppose, in this Section, that $X \in \mathcal{SF} \setminus \mathcal{SF}_\infty$, that is, $\mathfrak{r}$ is a regular boundary. The state space E includes the point $\mathfrak{r}$, that is, $E = [[\mathfrak{l}, \mathfrak{r}]$. Under $\mathbb{P}^{[\mathfrak{r}}$, X is a skip-free Markov chain on the state space $E^{[\mathfrak{r}} = [[\mathfrak{l}, \mathfrak{r})$, which is killed whenever the process hits the state $\mathfrak{r}$. We are ready to state the following.

Proposition 2.3.1. *Suppose $\mathfrak{r}$ is a regular boundary. Then for any $x > b$,*

$$\mathbb{E}_x(q^{T_b}) = \begin{cases} \dfrac{H_q(x) - \bar{K}_q^{b]} H_q^{b]}(x)}{H_q(b)} & \text{if } \mathfrak{r} > x > b, \\ \dfrac{G_q(\mathfrak{r}, b)}{G_q(b,b)} & \text{if } \mathfrak{r} = x > b, \end{cases}$$

where $\bar{K}_q^{b]} = \dfrac{G_q^{[\mathfrak{r}}(\mathfrak{o},\mathfrak{o})\widehat{H}_q^{[\mathfrak{r}}(b)}{G_q(\mathfrak{o},\mathfrak{o})\widehat{H}_q(b)}$ *and* $G_q(\mathfrak{r}, b) = \dfrac{\eta_b(q)}{c + (1-q)p + \sum_{j\in E}(1-q)\eta_j(q)}$, *where* $c \geqslant 0$, $p > 0$, *and* $\eta(q) = (\eta_j(q))_{j\in E}$ *is a family of non-negative numbers that satisfies, for any* $0 < s, q < 1$,

$$s\eta(s) - q\eta(q) = (s-q)\eta(q)G_q^{[\mathfrak{r}},$$

where $G_q^{[\mathfrak{r}} = (G_q^{[\mathfrak{r}}(i,j))_{i,j\in E}$.

In order to prove this Proposition, we need the following classical result that enables to connect the q-potential of G_q and $G_q^{[a}$ (resp. $\widehat{G}_q$ and $\widehat{G}_q^{b]}$).

Lemma 2.3.7. *We have, for any* $0 < q < 1$,

$$G_q(x,y) = G_q^{[a}(x,y) + \mathbb{E}_x(q^{T_a})G_q(a,y), \quad x, y \in E^{[a}, \tag{2.3.20}$$

$$\widehat{G}_q(x,y) = \widehat{G}_q^{b]}(x,y) + \widehat{\mathbb{E}}_x(q^{T_b})\widehat{G}_q(b,y), \quad x, y \in E^{b]}. \tag{2.3.21}$$

Proof. Since X is upward skip-free, $T_{[a} = T_a$ and for any $n \in \mathbb{N}$, $a \in E$, $x, y \in E^{[a}$, we have

$$\mathbb{P}_x(X_n = y) = \mathbb{P}_x^{[a}(X_n = y) + \sum_{k=1}^{n-1} \mathbb{P}_x(X_n = y | T_a = k)\mathbb{P}_x(T_a = k)$$
$$= \mathbb{P}_x^{[a}(X_n = y) + \sum_{k=1}^{n-1} \mathbb{P}_a(X_{n-k} = y)\mathbb{P}_x(T_a = k),$$

where the second equality follows from strong Markov property. Next, multiplying by q^n, dividing by $\pi(y)$, which is positive by irreducibility, and summing over n, we obtain

$$G_q(x, y) = G_q^{[a}(x, y) + \sum_{n=1}^{\infty}\sum_{k=1}^{n-1} q^{n-k}\frac{\mathbb{P}_a(X_{n-k} = y)}{\pi(y)} q^k \mathbb{P}_x(T_a = k)$$
$$= G_q^{[a}(x, y) + \sum_{k=1}^{\infty}\sum_{n=k+1}^{\infty} q^{n-k}\frac{\mathbb{P}_a(X_{n-k} = y)}{\pi(y)} q^k \mathbb{P}_x(T_a = k)$$
$$= G_q^{[a}(x, y) + \mathbb{E}_x(q^{T_a})G_q(a, y)$$

which proves (2.3.20). (2.3.21) is the dual statement of (2.3.20). □

We are now ready to complete the proof of Proposition 2.3.1. First, observe that $(X, \mathbb{P}^{[\mathfrak{r}}) \in \mathcal{SF}_\infty$ and thus by means of Lemma 2.3.7 and (2.3.17), one gets

$$G_q(b, b) = G_q^{[\mathfrak{r}}(\mathfrak{o}, \mathfrak{o})H_q^{[\mathfrak{r}}(b)\widehat{H}_q^{[\mathfrak{r}}(b) + \frac{H_q(b)}{H_q(\mathfrak{r})}G_q(\mathfrak{r}, b). \tag{2.3.22}$$

If one chooses the same reference point $\mathfrak{o}$ for $(X, \mathbb{P})$ and $(X, \mathbb{P}^{[\mathfrak{r}})$, then plainly, for any $x \in E^{[\mathfrak{r}}$, $H_q(x) = H_q^{[\mathfrak{r}}(x)$, this leads, using again (2.3.17), to

$$G_q(\mathfrak{o}, \mathfrak{o})\widehat{H}_q(b) = G_q^{[\mathfrak{r}}(\mathfrak{o}, \mathfrak{o})\widehat{H}_q^{[\mathfrak{r}}(b) + \frac{G_q(\mathfrak{r}, b)}{H_q(\mathfrak{r})}. \tag{2.3.23}$$

For $\mathfrak{r} > x > b$, using the lemmas 2.3.1 and 2.3.7, we have

$$\mathbb{E}_x(q^{T_b}) = \frac{G_q(x, b)}{G_q(b, b)} = \frac{G_q^{[\mathfrak{r}}(x, b) + \mathbb{E}_x(q^{T_\mathfrak{r}})G_q(\mathfrak{r}, b)}{G_q(\mathfrak{o}, \mathfrak{o})H_q(b)\widehat{H}_q(b)}$$
$$= \frac{G_q^{[\mathfrak{r}}(\mathfrak{o}, \mathfrak{o})(H_q^{[\mathfrak{r}}(x) - K_q^{b]}H_q^{b]}(x))\widehat{H}_q^{[\mathfrak{r}}(b) + \frac{H_q(x)}{H_q(\mathfrak{r})}G_q(\mathfrak{r}, b)}{G_q(\mathfrak{o}, \mathfrak{o})H_q(b)\widehat{H}_q(b)}$$
$$= \frac{G_q(\mathfrak{o}, \mathfrak{o})H_q(x)\widehat{H}_q(b) - G_q^{[\mathfrak{r}}(\mathfrak{o}, \mathfrak{o})H_q^{b]}(x)\widehat{H}_q^{[\mathfrak{r}}(b)}{G_q(\mathfrak{o}, \mathfrak{o})H_q(b)\widehat{H}_q(b)}$$
$$= \frac{H_q(x) - \bar{K}_q^{b]}H_q^{b]}(x)}{H_q(b)},$$

where we used (2.3.23) in the fourth equality. For $x = \mathfrak{r} > b$, it follows from (2.3.17) that

$$\mathbb{E}_{\mathfrak{r}}(q^{T_b}) = \frac{G_q(\mathfrak{r}, b)}{G_q(b, b)} = \frac{G_q(\mathfrak{r}, b)}{G_q(\mathfrak{o}, \mathfrak{o}) H_q(b) \widehat{H}_q(b)}.$$

To complete the proof one appeals to Theorem 3.1 of [92] that enable to determine $G_q(\mathfrak{r}, b)$.

2.4 Fluctuation identities

We pursue our program by exploiting the potential theoretic results of the previous Section to derive the fluctuation identities of a general skip-free Markov chain. This consists in determining the law, through the pgf, of the first exit time of X to a (in)finite interval. These quantities are critical in many applications in various settings including insurance mathematics, biology and epidemiology to name but a few. For instance, the time $T_{0]}$ corresponds to the time of ruin for a risk process having only negative jumps, which is a natural assumption in risk theory as they (the jumps) model the size of claims. We emphasize that our methodology enables to get the pgf of this time of ruin and hence allows to solve this problem for general Markovian risk processes. We also recall that for a birth-death chain the pgf of exit times is given as a linear combination of the two FqE functions, the one for $(X, \mathbb{P})$ and its dual $(X, \widehat{\mathbb{P}})$. There exists also a theoretical characterization of the laws of these variables for random walks through the celebrated Wiener-Hopf factorization. This technique, which is of complex analytical nature, seems to be limited only to the specific class of random walks. Our original approach appears to be more comprehensive for this issue in the context of general skip-free Markov chains. We have already mentioned that in order to illustrate its applicability, we will provide in the next subsection an alternative way to recover the fluctuation identities for skip-free random walks. The analogue of these results for skip-free continuous-time Markov processes on the real line can be found in [87] and applications to generalized Ornstein-Uhlenbeck processes and continuous branching processes with immigration are carried out in [72] and [86] respectively. We focus below on the case $X \in \mathcal{SF}_\infty$, the other case can be dealt with similarly by means of Proposition 2.3.1.

Theorem 2.4.1. *Suppose that $X \in \mathcal{SF}_\infty$. Then for any $b \in E$, $x \in E^{b]}$ and $0 < q \leqslant 1$, we have*

$$\mathbb{E}_x(q^{T_{b]}}) = 1 + (q-1)C_q \sum_{y \in E^{b]}} \widehat{H}_q(y) \left(\mathbf{H}_q^{b]}(x) - \mathbf{H}_q^{y]}(x)\right) \pi(y), \tag{2.4.1}$$

and, for any $a > b$, $x \in E^{(b,a)^c}$,

$$\mathbb{E}_x(q^{T_{(b,a)^c}}) = 1 + (q-1)C_q \sum_{y \in E^{(b,a)^c}} \widehat{H}_q(y) \left(\frac{\mathbf{H}_q^{y]}(a)}{\mathbf{H}_q^{b]}(a)} \mathbf{H}_q^{b]}(x) - \mathbf{H}_q^{y]}(x)\right) \pi(y). \tag{2.4.2}$$

Remark 2.4.1. We emphasize that the expressions (2.4.1) and (2.4.2) are comprehensive and are very useful to solve the first exit times problems when applied to some specific instances. Indeed they reveal that the characterization of the (probability generating function of the) distribution of first passage times of skip-free Markov chains boils down to determine the positive constant C_q and the functions $\widehat{H}_q$ and $\mathbf{H}_q^{b]}$. In practice, it turns out that the knowledge of a transformation, such as the Laplace transform, Fourier transform or moment generating function, of the one-step transition (and hence by integration of the q-potential) of the chain enables to identify H_q and $\widehat{H}_q$. The constant C_q can be determined by an argument of analytical continuation applied to the transform of the q-resolvent. Finally, (a transformation of) the function $\mathbf{H}_q^{b]}$ can be obtained from the previous identifications combined with the following relations

$$\mathbf{H}_q^{y]}(x) = \frac{1}{C_q\, \widehat{H}_q(y)} \left(H_q(x) - G_q(x,y)\right).$$

that can be easily derived from (2.3.4). This procedure will be illustrated in Section 2.4.1 below to the class of skip-free random walks and, in the subsequent paper [20], to the branching Galton-Watson processes with immigration.

Remark 2.4.2. The analysis in Theorem 2.4.1 can be extended to study the joint law of $X_{T_{b]}-1}$ and $X_{T_{b]}}$. Indeed, for $x, y \in E$ and $k \leqslant b$,

$$\begin{aligned}\mathbb{E}_x(q^{T_{b]}}\mathbb{1}_{\{X_{T_{b]}-1}=y, X_{T_{b]}}=k\}}) &= \sum_{n=1}^{\infty} q^n \mathbb{P}_x(X_{n-1}=y, X_n=k, T_{b]}=n) \\ &= q\, G_q^{b]}(x,y)p(y,k),\end{aligned}$$

where the last equality follows from the Markov property. Taking $q \to 1^-$ and using the monotone convergence theorem, one gets

$$\mathbb{P}_x(X_{T_{b]}-1}=y, X_{T_{b]}}=k) = G^{b]}(x,y)p(y,k).$$

Summing over $y \in E$ (or $y \in E^{b]}$, since $G^{b]}(x,y)=0$ for $y \leqslant b$), we have

$$\mathbb{P}_x(X_{T_{b]}}=k) = \sum_{y \in E} G^{b]}(x,y)p(y,k).$$

We proceed with the proof of Theorem 2.4.1. First, let $B \subset E$ and denote G_q^B (resp. $\widehat{G}_q^B$) to be the q-potential of the X (resp. its dual $\widehat{X}$) killed upon entering into the set B. We recall the Hunt's switching identity for Markov chains, which can be found in [61, page 140], and says that, for any $x, y \in E \backslash B$,

$$G_q^B(x,y) = \widehat{G}_q^B(y,x). \tag{2.4.3}$$

For sake of simplicity, we simply write $G_q^B = G_q^{b]}$ (resp. $G_q^A = G_q^{[a}$) if $B = (\mathfrak{l}, b]$ (resp. if $B = [a, \mathfrak{r})$). With this notation in mind, we express the q-potential kernels of $(X, \mathbb{P}^{[a})$ and $(X, \mathbb{P}^{b]})$ in terms of FqE functions of the three processes $(X, \mathbb{P})$, $(X, \widehat{\mathbb{P}})$ and $(X, \mathbb{P}^{y]})$.

Lemma 2.4.1. *Suppose that $X \in \mathcal{SF}_\infty$.*

$$G_q^{[a}(x,y) = C_q \widehat{H}_q(y)\left(\frac{\mathbf{H}_q^{y]}(a)}{H_q(a)}H_q(x) - \mathbf{H}_q^{y]}(x)\right), \quad x, y \in E^{[a}, \tag{2.4.4}$$

$$G_q^{b]}(x,y) = C_q \widehat{H}_q(y)\left(\mathbf{H}_q^{b]}(x) - \mathbf{H}_q^{y]}(x)\right), \quad x, y \in E^{b]}, \tag{2.4.5}$$

$$G_q^{(b,a)^c}(x,y) = C_q \widehat{H}_q(y)\left(\frac{\mathbf{H}_q^{y]}(a)}{\mathbf{H}_q^{b]}(a)}\mathbf{H}_q^{b]}(x) - \mathbf{H}_q^{y]}(x)\right), \quad x, y \in E^{(b,a)^c}. \tag{2.4.6}$$

Proof. To show (2.4.4), we use (2.3.20) and (2.3.4) to get

$$\begin{aligned}
G_q^{[a}(x,y) &= G_q(x,y) - \mathbb{E}_x(q^{T_a})G_q(a,y) \\
&= C_q(H_q(x) - \mathbf{H}_q^{y]}(x))\widehat{H}_q(y) - \frac{H_q(x)}{H_q(a)}C_q(H_q(a) - \mathbf{H}_q^{y]}(a))\widehat{H}_q(y) \\
&= C_q\widehat{H}_q(y)\left(\frac{H_q(x)\mathbf{H}_q^{y]}(a)}{H_q(a)} - \mathbf{H}_q^{y]}(x)\right).
\end{aligned}$$

Next, using the Hunt's switching identity (2.4.3) and (2.3.21), we have

$$\begin{aligned}
G_q^{b]}(x,y) = \widehat{G}_q^{b]}(y,x) &= \widehat{G}_q(y,x) - \widehat{\mathbb{E}}_y(q^{T_b})\widehat{G}_q(b,x) = G_q(x,y) - \frac{\widehat{H}_q(y)}{\widehat{H}_q(b)}G_q(x,b) \\
&= C_q(H_q(x) - \mathbf{H}_q^{y]}(x))\widehat{H}_q(y) - \frac{\widehat{H}_q(y)}{\widehat{H}_q(b)}C_q(H_q(x) - \mathbf{H}_q^{b]}(x))\widehat{H}_q(b) \\
&= C_q\widehat{H}_q(y)\left(\mathbf{H}_q^{b]}(x) - \mathbf{H}_q^{y]}(x)\right),
\end{aligned}$$

which proves (2.4.5). Finally, to get (2.4.6), we use the identity $G_q^{b]}(x,y) = G_q^{(b,a)^c}(x,y) + \mathbb{E}_x^{b]}(q^{T_a})G_q^{b]}(a,y)$ and (2.4.5) combined with Lemma 2.3.2. $\square$

With Lemma 2.4.1 in mind, the proof of Theorem 2.4.1 follows readily by applying the second claim stated in the following classical results which we prove for sake of completeness.

Lemma 2.4.2. *For any $b < x$ and $n \in \mathbb{N}$, we have*

$$\mathbb{P}_x(T_{b]} > n) = P_n^{b]}\mathbb{1}(x) = \frac{\pi\widehat{P}_n^{b]}\delta_x}{\pi(x)} \tag{2.4.7}$$

and thus, for any $0 < q < 1$,

$$\mathbb{E}_x(q^{T_{b]}}) = 1 + (q-1)\mathbf{G}_q^{b]}\mathbb{1}(x) = 1 + (q-1)\frac{\pi\widehat{\mathbf{G}}_q^{b]}\delta_x}{\pi(x)}, \tag{2.4.8}$$

where $\mathbf{G}_q^{b]}f(x) = \sum_{y\in E^{b]}} f(y)G_q^{b]}(x,y)\pi(y)$.

Proof. First, an application of Fubini's theorem yields, that for any $b < x$ and $0 < q \leqslant 1$,

$$\begin{aligned}\mathbf{G}_q^{b]}\mathbb{1}(x) &= \sum_{y \in E^{b]}} G_q^{b]}(x,y)\pi(y) = \sum_{y \in E^{b]}} \sum_{n=0}^{\infty} q^n \mathbb{P}_x(X_n = y, n < T_{b]}) \\ &= \sum_{n=0}^{\infty} \sum_{y \in E^{b]}} q^n \mathbb{P}_x(X_n = y, n < T_{b]}) = \sum_{n=0}^{\infty} q^n \mathbb{P}_x(T_{b]} > n).\end{aligned}$$

Moreover, from the Hunt's switching identity (2.4.3), we observe that

$$\mathbf{G}_q^{b]}\mathbb{1}(x) = \sum_{y \in E^{b]}} G_q^{b]}(x,y)\pi(y) = \sum_{y \in E^{b]}} \pi(y)\widehat{G}_q^{b]}(y,x) = \frac{\pi \widehat{\mathbf{G}}_q^{b]}\delta_x}{\pi(x)}$$

which completes the proof of the first claim. Finally, the lemma is proved after observing that, for any $x \in E^{b]}$,

$$\begin{aligned}\mathbb{E}_x(q^{T_{b]}}) &= \sum_{n=1}^{\infty} q^n \mathbb{P}_x(T_{b]} = n) = \sum_{n=1}^{\infty} q^n \left(\mathbb{P}_x(T_{b]} > n-1) - \mathbb{P}_x(T_{b]} > n)\right) \\ &= q\sum_{n=0}^{\infty} q^n \mathbb{P}_x(T_{b]} > n) - \sum_{n=1}^{\infty} q^n \mathbb{P}_x(T_{b]} > n) \\ &= 1 + (q-1)\sum_{n=0}^{\infty} q^n \mathbb{P}_x(T_{b]} > n).\end{aligned}$$

□

Remark 2.4.3. Another method to study the first passage time of b is by collapsing and combining all the states at or below b, and consider the hitting time to the glued state for the chain. Precisely, let $\widetilde{P}$ be the transition matrix of the glued chain with an absorbing boundary at b, that is,

$$\widetilde{P} = \begin{array}{c c} & \begin{array}{cccc} b & b+1 & b+2 & \ldots \end{array} \\ \begin{array}{c} b \\ b+1 \\ b+2 \\ \vdots \end{array} & \begin{bmatrix} 1 & 0 & \ldots & \ldots \\ \sum_{j=0}^{b} p(b+1,j) & p(b+1,b+1) & \ldots & \ldots \\ \sum_{j=0}^{b} p(b+2,j) & p(b+2,b+1) & \ldots & \ldots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \end{array}$$

By construction,

$$\mathbb{P}_x(T_{b]} \leqslant n) = \widetilde{\mathbb{P}}_x(T_b \leqslant n).$$

Therefore, by (2.3.13), using the obvious notation,

$$\mathbb{E}_x(q^{T_{b]}}) = \widetilde{\mathbb{E}}_x(q^{T_b}) = \frac{\widetilde{H_q}(x) - \widetilde{H}_q^{b]}(x)}{\widetilde{H}_q(b)}.$$

2.4.1 Skip-free random walk on $\mathbb{Z}$ revisited

We now aim to illustrate how the comprehensive result stated in Theorem 2.4.1 could be used to obtain explicit representations for the pgf of the first exit time of a skip-free Markov chain when applied to some specific instances. More specifically, we consider an upward skip-free random walk $(X, \mathbb{P})$ on the state space $\mathbb{Z}$ and, for sake of conciseness, postpone to a subsequent work [20] the application of our approach to the Galton-Watson processes with immigration which are skip-free to the left Markov chains. Suppose that $X = (X_n)_{n\in\mathbb{N}_0}$ is a skip-free random walk given by $\mathbb{P}_x(X_0 = x) = 1$, $x \in \mathbb{Z}$, and $X_n = X_0 + \sum_{i=1}^{n} S_i$, where $(S_i)_{n\in\mathbb{N}_0}$ are i.i.d. random variables with common distribution F which is supported on $\{n \in \mathbb{Z};\ n \leqslant 1\}$. We write $\mathbf{F}$ for the pgf of F. Thus, writing $p_s(x) = s^x$, we have

$$Pp_s(x) = \mathbf{F}(s)p_s(x). \tag{2.4.9}$$

Since $\lim_{s\to\infty} \mathbf{F}(s) = \infty$, one also sees that the mapping $s \mapsto \mathbf{F}(s)$ is continuous, increasing and convex on $(1/h(1), \infty)$ where $h(1) \geqslant 1$ is the largest root of the equation $\mathbf{F}(s) = 1$. Note that 1 is always a root and $h(1) > 1$ when $\mathbf{F}'(1^+) < 0$ (the right-derivative at 1). Therefore, $\frac{1}{\mathbf{F}}$ is continuous, decreasing on $(0, h(1))$ and thus has a well-defined inverse $h : (0, 1) \to (h(1), \infty)$. Recall also that, in this case, the dual, with respect to the reference measure $\pi \equiv 1$, is $(X, \widehat{\mathbb{P}}) \stackrel{d}{=} (-X, \mathbb{P})$.

Proposition 2.4.1. *We take the reference point to be $\mathfrak{o} = 0$ and let $0 < q < 1$.*

1. *$H_q \in \mathcal{H}_q$ and $\widehat{H}_q \in \widehat{\mathcal{H}}_q$, where for $x \in \mathbb{Z}$,*

$$H_q(x) = h(q)^x, \quad \widehat{H}_q(x) = h(q)^{-x}. \tag{2.4.10}$$

In addition,

$$C_q = -q\frac{h'(q)}{h(q)}. \tag{2.4.11}$$

2. *For any $x \geqslant y+1$, we have $\mathbf{H}_q^{y]}(x) = h(q)^y \mathbf{H}_q^{0]}(x-y)$ where, for $s \in \mathbb{R}$ such that $\mathbf{F}(\frac{1}{s}) > \frac{1}{q}$,*

$$\sum_{x\in\mathbb{N}} p_s(x)\mathbf{H}_q^{0]}(x) = \frac{1}{C_q(q\mathbf{F}(1/s)-1)}. \tag{2.4.12}$$

3. *For any $x > 0$,*

$$\mathbb{E}_x(q^{T_{0]}}) = 1 + (1-q)C_q\sum_{y=1}^{x-1}\mathbf{H}_q^{0]}(y) + qh'(q)\frac{q-1}{h(q)-1}\mathbf{H}_q^{0]}(x). \tag{2.4.13}$$

Proof. First, we deduce easily from (2.4.9) and the definition of h, that, with $H_q(x) = h(q)^x = p_{h(q)}(x)$, $0 < q < 1$,

$$qPH_q(x) = H_q(x)$$

that is $H_q \in \mathcal{H}_q$. Since $\ln h$ is nonnegative, clearly H_q is increasing. As $(X, \widehat{\mathbb{P}}) \stackrel{d}{=} (-X, \mathbb{P})$, the first claims follow readily from Theorem 2.4.1. Next, by means of Tonelli's theorem, on the one hand for $s \in \mathbb{R}$ such that $|q\mathbf{F}(1/s)| < 1$, we have

$$\sum_{x\in\mathbb{Z}} s^{-x}G_q(0,x) = \sum_{n\geqslant 0} q^n\mathbb{E}_0(s^{-X_n}) = \sum_{n\geqslant 0}(q\mathbf{F}(1/s))^n = \frac{1}{1-q\mathbf{F}(1/s)}.$$

On the other hand, using the translation invariant property of X, we have $G_q(-x,0) = G_q(0,x)$, which leads to

$$\sum_{x\in\mathbb{Z}} s^{-x}G_q(-x,0) = \sum_{x>0} s^x G_q(x,0) + \sum_{x\geqslant 0} s^{-x}G_q(-x,0).$$

Rearranging the terms and using (2.3.4) and (2.4.10) with $\sum_{x>0} s^x H_q(x) = -\frac{sh(q)}{sh(q)-1}$ yields

$$\sum_{x>0} s^x (H_q(x) - \mathbf{H}_q^{0]}(x)) = \frac{1}{C_q(1-q\mathbf{F}(1/s))} - \frac{sh(q)}{sh(q)-1}. \tag{2.4.14}$$

As the left-hand side can be treated as $\sum_{x>0} s^x \mathbb{E}_x(q^{T_0}) \leqslant \sum_{x>0} s^x < \infty$ for $0 < s < 1$, it is analytical on the unit disc and by the principle of analytical continuation, one gets that

$$C_q = \lim_{s \to \frac{1}{h(q)}} \frac{1-1/sh(q)}{1-q\mathbf{F}(1/s)} = \frac{-qh'(q)}{h(q)},$$

which shows (2.4.11). Next, following (2.4.14) and using again $\sum_{x>0} s^x H_q(x) = -\frac{sh(q)}{sh(q)-1}$, we get

$$\sum_{x\in\mathbb{N}} s^x \mathbf{H}_q^{0]}(x) = \frac{1}{C_q\left(q\mathbf{F}(1/s)-1\right)}.$$

Then, note that by the translation invariance property of X, $G_q(x,y) = G_q(x-y,0)$ for any $x, y \in E$ which after some easy algebra yields $\mathbf{H}_q^{y]}(x) = h(q)^y \mathbf{H}_q^{0]}(x-y)$ for any $x \geqslant y$. Finally, using the first claim of Theorem 2.4.1, we have

$$\begin{aligned}
\mathbb{E}_x(q^{T_{0]}}) &= 1 + (q-1)C_q \sum_{y\in E^{0]}} \widehat{H}_q(y)\left(\mathbf{H}_q^{0]}(x) - \mathbf{H}_q^{y]}(x)\right) \\
&= 1 + (q-1)C_q \sum_{y\in E^{0]}} h(q)^{-y}\left(\mathbf{H}_q^{0]}(x) - h(q)^y\mathbf{H}_q^{0]}(x-y)\right) \\
&= 1 + (1-q)C_q \sum_{y=1}^{x-1} \mathbf{H}_q^{0]}(y) + qh'(q)\frac{q-1}{h(q)-1}\mathbf{H}_q^{0]}(x)
\end{aligned}$$

which completes the proof of the Theorem. $\square$

2.5 Classes of first passage times distributions

The aim of this part is two-fold. First, we provide a characterization of the first passage times distribution, both from above and below the starting point, in which we built upon the results of Fill [43]. This follows from the line of work by Kent and Longford [63] who characterize the class of upward and downward hitting time of birth-death Markov chains on non-negative integers. More specifically, they define and introduce a new class $\mathcal{K}(b,\tau,M)$ of discrete infinitely divisible distributions with pgf given by

$$\phi(z) = \exp\left\{b(z-1) - \tau + z(z-1)\int_{-1}^{1}(1-pz)^{-1}M(dp)\right\}$$

that contain both the upward and downward hitting time of such chain, where b,τ are non-negative parameters and M is a finite measure on $(-1,1)$ such that $\int_{-1}^{1}(1+p)^{-1}M(dp) < \infty$. In particular, using the interlacing property of the eigenvalues for birth-death process, Kent and Longford [63] show that the measure M that corresponds to the hitting times are non-negative. However, this interlacing property is lost in general when we move to the realm of skip-free chains. This motivates us to define a more general class of distribution that we call $\mathbb{G}_p$ in Definition 2.5.1 which contains both upward and downward hitting times in Theorem 2.5.1. This will be further demonstrated in Remark 2.5.2 below. Second, we derive an explicit representation of the FqE functions associated to $(X,\mathbb{P})$ and $(X,\mathbb{P}^{b]})$ in Section 2.5.2. As we will elaborate in Section 2.5.4, this allows us to investigate the infinitely divisibility of the upward hitting time for skip-free Markov chains and characterizes the associated R-function by means of the eigenvalues. In this vein we emphasize that Feller [41] studied the infinite divisibility of the hitting times of birth-death random walk and Viskov [110] extended his work to skip-free random walk using the classical Lagrange inversion formula. We focus on $\mathcal{SF}_{\mathfrak{l}}$ the subclass of $\mathcal{SF}$ that has a finite left boundary point, i.e. $\mathfrak{l} < \infty$. We shall also need the following definition where we used the notation, for $d \in \mathbb{N}$, $\mathbb{D}^d = \{\boldsymbol{\lambda} = (\lambda_k)_{k=1}^d \in \mathbb{C}^d;\ |\lambda_k| \in [0,1],\ k = 1,\ldots,d$ and $\lambda_i \neq \lambda_k, i \neq k\}$.

Definition 2.5.1. Let $p,d \in \mathbb{N}$ with $d \leqslant p$, $\boldsymbol{\lambda} = (\lambda_k) \in \mathbb{D}^d$ and $\boldsymbol{m} = (m_k) \in \mathbb{N}^d$ with $\sum_{k=1}^d m_k = p$, and for each $k \in [1,d]$, the multiplicity of λ_k is m_k, $\mathbf{c}_k = (c_{k,i})_{i=1}^{m_k} \in \mathbb{C}^{m_k}$ and we write $\mathbf{c} = (\mathbf{c}_{k,i})$. We say that a non-negative discrete random variable

$X \in \mathbb{G}_p(\boldsymbol{c};\boldsymbol{\lambda};\boldsymbol{m})$ if its probability mass function can be written, for any $n \in \mathbb{N}_0$, as

$$0 \leqslant \mathbb{P}(X=n) = \sum_{k=1}^{d}\sum_{i=1}^{m_k} c_{k,i}\binom{n+i-1}{n}\lambda_k^n(1-\lambda_k)^i \leqslant 1, \tag{2.5.1}$$

where $\sum_{k=1}^{d}\sum_{i=1}^{m_k} c_{k,i} \leqslant 1$. In particular, if $\boldsymbol{\theta} = \boldsymbol{\lambda} \in [0,1]^d$, then we write $\mathcal{G}_p(\boldsymbol{c};\boldsymbol{\theta};\boldsymbol{m}) = \mathbb{G}_p(\boldsymbol{c};\boldsymbol{\theta};\boldsymbol{m})$. Note that it can be interpreted as a (signed) mixture of geometric random variables when $m_k = 1$ and $\lambda_k \in [0,1]$ for all k.

Before stating the main result of this part, we introduce the notation $\boldsymbol{\lambda}^{[x,y]} = (\lambda_i^{[x,y]})_{i=0}^{y-x}$ for the (non-unit) eigenvalues of the transition matrix P restricted to $[x,y]$ with $x \leqslant y$ and we use $\mathcal{ID}$ to denote the class of infinitely divisible law, see Section 2.5.4 for formal definition and further discussion on related topics. In Fill [43], under the condition that the transition matrix of an upward skip-free chain has only real and non-negative eigenvalues, showed that its upward first hitting time is a convolution of geometric distributions. We also refer to Miclo [80] for similar results in the context of reversible Markov chains.

Theorem 2.5.1. *For any $\mathfrak{l} \leqslant b \leqslant x \leqslant a \leqslant \mathfrak{r}$, we have, under $\mathbb{P}_x$,*

$$T_a - (a-x) \in \mathbb{G}_a(\mathbf{c}(x,a);\boldsymbol{\lambda};\boldsymbol{m}) \cap \mathcal{ID} \quad \textit{and} \quad T_b \ (\textit{resp. } T_{b]}) \in \mathbb{G}_{\mathfrak{r}-1}(\tilde{\boldsymbol{c}}(x,b);\tilde{\boldsymbol{\lambda}};\tilde{\boldsymbol{m}}),$$

where $\boldsymbol{\lambda}$ are the distinct eigenvalues of $\boldsymbol{\lambda}^{[0,a-1]}$, $\tilde{\boldsymbol{\lambda}}$ are the distinct eigenvalues of $\boldsymbol{\lambda}^{[0,b-1]}$ and $\boldsymbol{\lambda}^{[b+1,\mathfrak{r}-1]}$, $\boldsymbol{m}$ (resp. $\tilde{\boldsymbol{m}}$) are the multiplicities of $\boldsymbol{\lambda}$ (resp. $\tilde{\boldsymbol{\lambda}}$). In particular, when $\boldsymbol{\lambda}, \tilde{\boldsymbol{\lambda}}$ are all real and non-negative, then, under $\mathbb{P}_x$,

$$T_a - (a-x) \in \mathcal{G}_a(\mathbf{c}(x,a);\boldsymbol{\lambda};\boldsymbol{m}) \cap \mathcal{ID} \quad \textit{and} \quad T_b \ (\textit{resp. } T_{b]}) \in \mathcal{G}_{\mathfrak{r}-1}(\tilde{\boldsymbol{c}}(x,b);\tilde{\boldsymbol{\lambda}};\tilde{\boldsymbol{m}}).$$

Remark 2.5.1. Note that for class $\mathcal{S}$ to be introduced in Section 2.6, the passage time distributions are in the particular cases of Theorem 2.5.1.

2.5.1 Characterizations of the class $\mathbb{G}_p$

We first provide a characterization of the class $\mathbb{G}_p$, which will simplify our proof of Theorem 2.5.1.

Proposition 2.5.1. *$X \in \mathbb{G}_p(\boldsymbol{c}; \boldsymbol{\lambda}; \boldsymbol{m})$ if and only if its probability generating function can be written as*

$$\mathbb{E}[q^X] = \sum_{k=1}^{d} \sum_{i=1}^{m_k} \frac{C_{k,i}}{(1-\lambda_k q)^i}, \tag{2.5.2}$$

where $(C_{k,i})$ depends on $\mathbf{c}, \boldsymbol{\lambda}$. Moreover, if the probability generating function of X can be written as

$$\mathbb{E}[q^X] = \prod_{j=0}^{b-1} \frac{(1-\beta_j)q}{1-\beta_j q} \prod_{j=0}^{x-1} \frac{1-\alpha_j q}{(1-\alpha_j)q}, \tag{2.5.3}$$

where $\boldsymbol{\beta} = (\beta_j) \in \mathbb{C}^b$, $\boldsymbol{\alpha} = (\alpha_j) \in \mathbb{C}^x$ with $|\lambda_j|, |\alpha_j| \in [0,1]$ for all j, $\boldsymbol{\lambda}$ are the distinct elements of $\boldsymbol{\beta}$ and $\boldsymbol{m}$ are the multiplicities, then $X \in \mathbb{G}_b(\boldsymbol{c}; \boldsymbol{\lambda}; \boldsymbol{m})$.

Remark 2.5.2. We remark that the class $\mathbb{G}_p$ is more general than the class $\mathcal{K}(b, \tau, M)$ studied by Kent and Longford [63]. Indeed, using the expression in [63, Section 10], we see that (2.5.3) can be rewritten as

$$\begin{aligned} q^{x-b}\mathbb{E}[q^X] = \exp\Bigg\{&(q-1)\left(\sum_{j=0}^{b-1} |\log(1-\beta_j)| - \sum_{j=0}^{x-1} |\log(1-\alpha_j)|\right) \\ &+ q(q-1)\int_{-1}^{1} (1-pz)^{-1} M(dp)\Bigg\}, \end{aligned}$$

where $M(dp) = \left(\sum_{j=0}^{b-1} \mathbb{1}_{(0,\beta_j)} - \sum_{j=0}^{x-1} \mathbb{1}_{(0,\alpha_j)}\right) |p|(1-p)dp$. In the particular case of birth-death chains, we can see that the measure M that corresponds to $(T_{a+1}, \mathbb{P}_a)$ is, by the reality and the interlacing property of eigenvalues, non-negative. In general, the measure M is a signed measure.

Proof. We first show that if $X \in \mathbb{G}_p(\boldsymbol{c}; \boldsymbol{\lambda}; \boldsymbol{m})$ then (2.5.2) holds. By writing $C_{k,i} =$

$c_{k,i}(1-\lambda_k)^i$, we see that

$$\mathbb{E}[q^X] = \sum_{k=1}^{d}\sum_{i=1}^{m_k} C_{k,i} \sum_{n\geqslant 0} \binom{n+i-1}{n} (q\lambda_k)^n = \sum_{k=1}^{d}\sum_{i=1}^{m_k} \frac{C_{k,i}}{(1-\lambda_k q)^i}.$$

The opposite direction can be shown by differentiating the pgf n times followed by dividing $n!$. Next, we show that pgf of the form (2.5.3) belongs to the class $\mathbb{G}_p$. By means of partial fraction, we note that (2.5.3) can be written as

$$\mathbb{E}[q^X] = \sum_{k=0}^{d}\sum_{i=1}^{m_k} \frac{C_{k,i}}{(1-\lambda_k q)^i},$$

so by (2.5.2), $X \in \mathbb{G}_b(\boldsymbol{c};\boldsymbol{\lambda};\boldsymbol{m})$. $\square$

2.5.2 FqE functions of $(X,\mathbb{P})$ and $(X,\mathbb{P}^{b]})$

In this section, we give explicit formulas on the FqE functions of $(X,\mathbb{P})$ and $(X,\mathbb{P}^{b]})$. Note that the former case follows from Fill [43], whereas the latter one is a slight generalization of the first one by allowing substochastic transition matrices.

Proposition 2.5.2. *Suppose that $(X,\mathbb{P})$ is an upward skip-free Markov chain on E with $|\mathfrak{l}|<\infty$. For $b\in E$, we consider the killed process $(X,\mathbb{P}^{b]})$ on $E^{b]}=[b+1,\mathfrak{r})$. Let $\lambda_i=\lambda_i^{[b+1,x-1]}$. If we take $b+1$ to be the reference point (i.e. $\mathfrak{o}^{b]}=b+1$), then the FqE function of $(X,\mathbb{P}^{b]})$ is*

$$H_q^{b]}(x) = \begin{cases} 1, & \text{if } x = b+1, \\ \dfrac{1}{\mathbb{P}_{b+1}(T_x<T_{b]})} \displaystyle\prod_{i=0}^{x-b-2} \frac{1-\lambda_i q}{(1-\lambda_i)q}, & \text{if } x \geqslant b+2, \end{cases}$$

where $\mathbb{P}_{b+1}(T_x<T_{b]}) = \dfrac{p(b+1,b+2)\dots p(x-1,x)}{(1-\lambda_0)\dots(1-\lambda_{x-b-2})}$.

Proof. Suppose that X is an upward skip-free Markov chain on $[-1,a]$, where a and -1 are absorbing states. Denote P to be the substochastic transition matrix on $[0,a]$, with $(\lambda_i)_{i=0}^{a-1}$ to be the (non-unit) eigenvalues, where $\lambda_i = \lambda_i^{[0,a-1]}$ for $i = 0, \dots, a-1$. We define $\widetilde{P} = (\widetilde{p}(i,j))$ for $i, j = 0, 1, \dots, a$ to be

$$\widetilde{p}(i,j) = \begin{cases} \lambda_i, & j = i, \\ 1 - \lambda_i, & j = i+1, \\ 0, & \text{otherwise}, \end{cases}$$

where $\lambda_a = 1$. Next, the so-called spectral polynomial is defined to be $Q_0 = I$ and

$$Q_k = \frac{(P - \lambda_0 I) \dots (P - \lambda_{k-1} I)}{(1 - \lambda_0) \dots (1 - \lambda_{k-1})}, \quad k = 1, \dots, a.$$

Define the link matrix $\Lambda = (\Lambda(i,j))$ to be

$$\Lambda(i,j) = Q_i(0,j), \quad i, j = 0, \dots, a.$$

Then Λ satisfies the following properties:

(i) The sum of each row of Λ is less than or equal to 1.

(ii) Λ is lower-triangular with $\Lambda(0,0) = 1$ and

$$\Lambda(k,k) = \frac{p(0,1) \dots p(k-1,k)}{(1 - \lambda_0) \dots (1 - \lambda_{k-1})} \neq 0, \quad k = 1, \dots, a,$$

and hence Λ is invertible.

(iii) $\Lambda P = \widetilde{P} \Lambda$.

The proof of ((i)), ((ii)) and ((iii)) above are identical to that of Lemma 2.1 in [43], except that the sum of each row of Λ is now less than or equal to 1, since each row of $(1 - \lambda_i)^{-1}(P - \lambda_i I)$ has that property. The next two results (2.5.4) and (2.5.5) are analogous to [43, Lemma 2.3 and Theorem 1.2], and we omit the proofs here as they are almost identical, except now we have $\mathbb{P}_0(T_a < T_{-1}) = \Lambda(a,a)$. For $t \in \mathbb{N}_0$,

$$\mathbb{P}_0(T_a \leqslant t) = \mathbb{P}_0(T_a < T_{-1}) \widetilde{P^t}(0,a), \tag{2.5.4}$$

where

$$\mathbb{P}_0(T_a < T_{-1}) = \Lambda(a,a) = \frac{p(0,1)\dots p(a-1,a)}{(1-\lambda_0)\dots(1-\lambda_{a-1})}.$$

For $q \in [0,1]$,

$$\mathbb{E}_0^{-1]}(q^{T_a}) = \mathbb{P}_0(T_a < T_{-1}) \prod_{i=0}^{a-1} \frac{(1-\lambda_i)q}{1-\lambda_i q}. \tag{2.5.5}$$

Desired result follows immediately from Theorem 2.3.1, Lemma 2.3.2 and (2.5.5). □

With the two previous results in hands, we are now ready to complete the proof of Theorem 2.5.1.

2.5.3 Proof of Theorem 2.5.1

Using the result in [105, Chapter VII Theorem 2.1], $(T_a - (a-x), \mathbb{P}_x)$ is infinitely divisible. It follows from (2.3.6) that

$$\mathbb{E}_x(q^{T_a-(a-x)}) = q^{-(a-x)} \frac{H_q(x)}{H_q(a)},$$

and by Proposition 2.5.2, the ratio $q^{-(a-x)} \dfrac{H_q(x)}{H_q(a)}$ is of the form (2.5.3). Therefore, by Proposition 2.5.1, $T_a - (a-x) \in \mathbb{G}_a(\mathbf{c}(x,a); \boldsymbol{\lambda}; \boldsymbol{m}) \cap \mathcal{ID}$. Next, using Lemma 2.3.3, we have

$$\mathbb{E}_x(q^{T_b} \mathbb{1}_{\{T_b < T_{\mathfrak{r}}\}}) = \frac{H_q(x)}{H_q(b)} - \frac{H_q(\mathfrak{r})}{H_q(b)} \frac{H_q^{b]}(x)}{H_q^{b]}(\mathfrak{r})}.$$

We substitute the FqE functions from Proposition 2.5.2 and again recognize that it is of the form (2.5.3). By Proposition 2.5.1, T_b (resp. $T_{b]}) \in \mathbb{G}_{\mathfrak{r}-1}(\tilde{\boldsymbol{c}}(x,b); \tilde{\boldsymbol{\lambda}}; \tilde{\boldsymbol{m}})$.

2.5.4 Infinite divisibility and R-functions

In this subsection, we first review several main results in the study of infinitely divisible distribution, which will be used in analyzing the upward hitting times of $(X, \mathbb{P})$. We refer interested readers to [105] for a formal reference in the literature. ϕ is the probability generating function of an infinitely divisible distribution if and only if for every $n \in \mathbb{N}$ there is a pgf ϕ_n such that for $q \in (0,1)$,

$$\phi(q) = \phi_n(q)^n.$$

We first characterize infinite divisibility via canonical sequence: ϕ is the pgf of an infinitely divisible distribution on $\mathbb{N}_0$ with $\phi(0) > 0$ and $\phi(1) = 1$ if and only if ϕ has the form

$$\phi(q) = \exp\left\{ -\sum_{k=0}^{\infty} \frac{r_k}{k+1}(1 - q^{k+1}) \right\}, \quad 0 \leqslant q \leqslant 1,$$

where $r_k \geqslant 0$ for all $k \geqslant 0$. The canonical sequence $(r_k)_{k\in\mathbb{N}_0}$ is unique. Apart from the canonical representation, an alternative way to characterize an infinitely divisible law is by means of the R-function. ϕ is the pgf of a possibly defective infinitely divisible distribution on $\mathbb{N}_0$ with $\phi(0) > 0$ and $\phi(1) = \exp\{-b\} \in (0,1], b \geqslant 0$ ($b = 0$ if the distribution is non-defective) if and only if ϕ has the form

$$\phi(q) = \exp\left\{ -b - \int_q^1 \mathrm{R}(s)\, ds \right\}, \quad 0 \leqslant q < 1, \tag{2.5.6}$$

where R is an absolutely monotone function on $[0,1)$. R is unique and is the generating function of the canonical sequence $(r_k)_{k\in\mathbb{N}_0}$ of ϕ.

Our first result in this section is a by-product of Theorem 2.3.1, where we obtain the interesting connection between absolute monotonicity and the FqE function H_q.

Corollary 2.5.1. *The mapping $q \mapsto \frac{d}{dq}\log\left(q^{-(a-x)}\frac{H_q(x)}{H_q(a)}\right)$ is absolutely monotone if and only if $x \leqslant a$.*

Proof. Assume that $x \leqslant a$. Then, from [105, Chapter VII Theorem 2.1] we have that $(T_a - (a-x), \mathbb{P}_x)$ is a positive infinitely divisible variable. Combining this fact with the identity (2.3.6) and (2.5.6), we conclude that $\frac{d}{dq}\log\left(q^{-(a-x)}\frac{H_q(x)}{H_q(a)}\right)$ is a R-function. Conversely, if $x > a$, then $\frac{d}{dq}\log\left(q^{-(x-a)}\frac{H_q(a)}{H_q(x)}\right) \geqslant 0$ is a R-function, so $\frac{d}{dq}\log\left(q^{-(a-x)}\frac{H_q(x)}{H_q(a)}\right) = -\frac{d}{dq}\log\left(q^{-(x-a)}\frac{H_q(a)}{H_q(x)}\right) < 0$ cannot be a R-function which completes the proof of the Corollary. $\square$

In the next two propositions, we proceed by offering explicit spectral formulas of R-functions and canonical sequences associated with the infinitely divisible variables $(T_a - a, \mathbb{P}_0)$ and $(T_a - (a-x), \mathbb{P}_x)$ for $0 \leqslant x \leqslant a$, building upon the results in Section 2.5.2, and thereby extending the work by Feller [41] for the case of birth-death random walk and Viskov [110] for skip-free random walk. We start by defining a few notations. For $0 \leqslant j \leqslant a-1$, let

$$\mathrm{R}_j(s) = \frac{\lambda_j}{1-\lambda_j s}.$$

Denote the number of real (resp. complex) eigenvalues by $N_r^{0\to a} = |\{j : \lambda_j \text{ is real}\}|$ (resp. $N_c^{0\to a} = |\{j : \lambda_j \text{ is complex}\}|$).

Proposition 2.5.3. *Suppose that $a \in \mathbb{N}_0$, and let $\lambda_j = \lambda_j^{[0,a-1]}$. The R-function of $(T_a - a, \mathbb{P}_0)$ is*

$$\mathrm{R}^{0\to a}(s) = \sum_{j=0}^{a-1} \mathrm{R}_j(s)$$

with canonical sequence

$$r_k^{0\to a} = \sum_{j=0}^{a-1} \lambda_j^{k+1} = \sum_{j=1}^{N_r^{0\to a}} \lambda_j^{k+1} + \sum_{j=1}^{N_c^{0\to a}} |\lambda_j|^{k+1} \cos\left((k+1)\mathrm{Arg}\lambda_j\right), \quad k \in \mathbb{N}_0,$$

and constant $b^{0\to a} = -\ln \mathbb{P}_0(T_a < \infty)$.

Proof. Using the result in [105, Chapter VII Theorem 2.1], $(T_a - a, \mathbb{P}_0)$ is infinitely divisible. By Proposition 2.5.2,

$$\mathbb{E}_0(q^{T_a - a}) = \mathbb{P}_0(T_a < \infty)\prod_{j=0}^{a-1}\frac{(1-\lambda_j)}{1-\lambda_j q} = \mathbb{P}_0(T_a < \infty)\exp\left\{-\int_q^1 \sum_{j=0}^{a-1}\mathrm{R}_j(s)\,ds\right\},$$

where the second equality follows from identity

$$\frac{1-\lambda}{1-\lambda q} = \exp\left\{ -\int_q^1 \frac{\lambda}{1-\lambda s}\, ds \right\},$$

valid for $|\lambda| < 1/q$. Since complex eigenvalues occur in conjugate pair, we can check that $\sum_{j=0}^{a-1} \mathrm{R}_j(s) \in \mathbb{R}$. □

Proposition 2.5.4. *Suppose that $0 \leqslant x \leqslant a$. Following (2.5.6), the R-function of $(T_a - (a-x), \mathbb{P}_x)$ is*

$$\mathrm{R}^{x\to a}(s) = \mathrm{R}^{0\to a}(s) - \mathrm{R}^{0\to x}(s)$$

with canonical sequence

$$r_k^{x\to a} = r_k^{0\to a} - r_k^{0\to x}, \quad k \in \mathbb{N}_0,$$

and constant $b^{x\to a} = -\ln \mathbb{P}_x(T_a < \infty)$.

Proof. Using the result in [105, Chapter VII Theorem 2.1], $(T_a - (a-x), \mathbb{P}_x)$ is infinitely divisible. Let $\beta_j = \lambda_j^{[0,x-1]}$ for $j = 0, \dots, x-1$. By Proposition 2.5.2,

$$\begin{aligned}
\mathbb{E}_x(q^{T_a-(a-x)}) &= \mathbb{P}_x(T_a < \infty) \left(\prod_{j=0}^{x-1} \frac{\frac{(1-\lambda_j)}{1-\lambda_j q}}{\frac{(1-\beta_j)}{1-\beta_j q}} \right) \left(\prod_{j=x}^{a-1} \frac{(1-\lambda_j)}{1-\lambda_j q} \right) \\
&= \mathbb{P}_x(T_a < \infty) \exp\left\{ -\int_q^1 \mathrm{R}^{0\to a}(s) - \mathrm{R}^{0\to x}(s)\, ds \right\}.
\end{aligned}$$

□

Since $r_k^{x\to a} \geqslant 0$ for $k \in \mathbb{N}_0$, we immediately obtain the following ordering of eigenvalues.

Corollary 2.5.2. *Suppose that $0 \leqslant x \leqslant a$, and let $\beta_j = \lambda_j^{[0,x-1]}$ for $j = 0, \dots, x-1$. Then*

$$\sum_{j=0}^{a-1} \lambda_j^{k+1} \geqslant \sum_{j=0}^{x-1} \beta_j^{k+1} \geqslant 0, \quad k \in \mathbb{N}_0.$$

Remark 2.5.3. We can rewrite the results in Proposition 2.5.4 and Corollary 2.5.2 in terms of trace. Define $P^{[0,a]}$ to be the restriction of P from state 0 to a. For $k \in \mathbb{N}_0$, we observe that $r_k^{0\to a} = \mathrm{Tr}((P^{[0,a]})^{k+1})$, and so Corollary 2.5.2 can be rewritten as

$$\mathrm{Tr}((P^{a]})^{k+1}) \geqslant \mathrm{Tr}((P^{x]})^{k+1}).$$

Upward hitting times

In this part, we characterize a class of discrete distribution that contains all upward hitting times within the class $\mathcal{SF}$.

Definition 2.5.2. We say that a (possibly defective) distribution lies in class $\mathcal{U}$ if it is discrete, infinitely divisible and each term of the canonical sequence can be written as

$$\frac{r_k}{k+1} = \int_{-1}^{1} x^k w(x)\, dx, k \in \mathbb{N}_0, \tag{2.5.7}$$

where the mapping w satisfies the integrability condition

$$\int_{-1}^{1} |w(x)|\, dx < \infty.$$

Proposition 2.5.5. *Suppose that* $0 \leqslant x \leqslant a$. *Then,* $(T_a - (a-x), \mathbb{P}_x)$ *belongs to the class* $\mathcal{U}$.

Proof. It suffices to show (2.5.7) for $r_k^{x\to a}$. Define

$$w^{0\to a} = \sum_{j=1}^{N_r^{0\to a}} \mathbb{1}_{(0,\lambda_j)} + \sum_{j=1}^{N_c^{0\to a}} \mathbb{1}_{(0,|\lambda_j|\cos((k+1)\mathrm{Arg}\lambda_j)^{1/(k+1)})},$$

then

$$\int_{-1}^{1} y^k w^{0\to a}(y)\, dy = \frac{\sum_{j=0}^{a-1} \lambda_j^{k+1}}{k+1} = \frac{r_k^{0\to a}}{k+1}.$$

In addition, by the triangle's inequality,

$$\int_{-1}^{1} |w^{0\to a}(y)|\, dy \leqslant \sum_{j=0}^{a-1} \lambda_j < \infty.$$

Let $w^{x\to a} = w^{0\to a} - w^{0\to x}$, we obtain

$$\int_{-1}^{1} y^k w^{x\to a}(y)\, dy = \frac{r_k^{0\to a} - r_k^{0\to x}}{k+1} = \frac{r_k^{x\to a}}{k+1},$$
$$\int_{-1}^{1} |w^{x\to a}(y)|\, dy \leqslant r_0^{0\to a} + r_0^{0\to x} < \infty,$$

which completes the proof.

$\square$

2.6 The class of skip-free Markov chains similar to birth-death chains

In this Section, we develop an original methodology to obtain the spectral decomposition in Hilbert space of the (transition operator of) Markov chains that belong to the class $\mathcal{S}_{sf}$ a subclass of $\mathcal{SF}$ which is defined in Definition 2.6.2 below. We recall that as the transition operator P of a chain in $\mathcal{SF}$ is non-self-adjoint (non-reversible) in the weighted Hilbert space

$$\ell^2(\pi) = \left\{ f : E \mapsto \mathbb{R};\ ||f||_\pi^2 = \sum_{x\in E} f^2(x)\pi(x) < \infty \right\},$$

where π is the reference measure, there is no spectral theorem available for such bounded linear operator. We already point out that the two subsequent Sections contain interesting and substantial applications of this spectral decomposition namely the study of the speed of convergence to equilibrium and the separation cutoff phenomenon. We proceed by defining an equivalence relation on the set of transition matrices in $\mathcal{SF}$.

Definition 2.6.1 (Similarity)**.** We say that the transition matrix P of a Markov chain $X \in \mathcal{SF}$ is similar to the transition matrix of a Markov chain Q on E, and we write $P \sim Q$, if there exits a bounded linear operator $\Lambda : \ell^2(\pi_Q) \to \ell^2(\pi)$ (π_Q being the reference measure for Q) with bounded inverse such that

$$P\Lambda = \Lambda Q. \tag{2.6.1}$$

When needed we may write $P \overset{\Lambda}{\sim} Q$ to specify the intertwining kernel. Note that $\sim$ is an equivalence relationship on the set of transition matrices.

With Definition 2.6.1 in mind, we are now ready to define the $\mathcal{S}_{sf}$ class.

Definition 2.6.2 (The $\mathcal{S}_{sf}$ class)**.** Suppose that $Q \in \mathcal{B}$, the set of transition matrix Q on E of an irreducible (or with at most one absorbing or entrance state) birth-death chain. The similarity orbit of Q (in $\mathcal{SF}$) is

$$\mathcal{S}(Q) = \{P \in \mathcal{SF};\ P \sim Q\},$$

and the $\mathcal{S}_{sf}$ class is the union over all possible orbits

$$\mathcal{S}_{sf} = \bigcup_{Q \in \mathcal{B}} \mathcal{S}(Q).$$

Similarly, we define $\widehat{\mathcal{S}}(Q)$ and $\widehat{\mathcal{S}}_{sf}$ by replacing P with $\widehat{P}$ above. Finally, we say that $X \in \mathcal{S}_{sf}^M$ if $X \in \mathcal{S}_{sf}$ and $(X, \widehat{\mathbb{P}})$ is stochastically monotone, i.e. $y \mapsto \widehat{\mathbb{P}}_y(X_1 \leqslant x)$ is non-increasing for every fixed x.

We remark that $\widehat{\Lambda} : \ell^2(\pi) \to \ell^2(\pi_Q)$, the adjoint of Λ, is a bounded operator with a bounded inverse as well (see e.g. [24, Proposition 2.6]). We write $\|\cdot\|_{op}$ to be the operator norm, i.e. $\|P\|_{op} = \sup_{||f||_\pi = 1} ||Pf||_\pi$.

Before stating the main result of this Section, we introduce the following class.

Definition 2.6.3 (The $\mathcal{MC}$ class)**.** We say that, for some $\mathfrak{r} \geqslant 3$, $X \in \mathcal{MC}_{\mathfrak{r}}$ if $(X, \mathbb{P}) \in \mathcal{SF}$ with $E = \{0, 1 \ldots, \mathfrak{r}\}$ and for every $x \in [0, \mathfrak{r} - 1]$, its time-reversal $(X, \widehat{\mathbb{P}})$ satisfies

1. (stochastic monotonicity) $\widehat{\mathbb{P}}_{x+1}(X_1 \leqslant x) \leqslant \widehat{\mathbb{P}}_x(X_1 \leqslant x)$,
2. (strict stochastic monotonicity) $\widehat{\mathbb{P}}_{x+1}(X_1 \leqslant x+1) < \widehat{\mathbb{P}}_x(X_1 \leqslant x+1), \quad x \neq \mathfrak{r}-1$,
3. (restricted upward jump) $\widehat{\mathbb{P}}_{x+1}(X_1 \leqslant x + k) = \widehat{\mathbb{P}}_x(X_1 \leqslant x + k), \quad x \neq \mathfrak{r} - 1, \quad k \in [2, \mathfrak{r} - 1 - x]$.

Moreover, we say $X \in \mathcal{MC}_{\mathfrak{r}}^+$ if $X \in \mathcal{MC}_{\mathfrak{r}}$ and

(4) (lazy Siegmund dual) $\widehat{\mathbb{P}}_x(X_1 \geqslant x + 1) + \widehat{\mathbb{P}}_{x+1}(X_1 = x) \leqslant \frac{1}{2}, \quad x = 0, \ldots, \mathfrak{r} - 1$.

When there is no ambiguity of the state space, we write $\mathcal{MC} = \mathcal{MC}_{\mathfrak{r}}$ (resp. $\mathcal{MC}^+ = \mathcal{MC}_{\mathfrak{r}}^+$).

We proceed by recalling that P has an π-dual or time-reversal $\widehat{P}$, that is, for $x, y \in E$,

$$\pi(x)\widehat{p}(x, y) = \pi(y)p(y, x),$$

where π is a reference measure for P. We equip the Hilbert space $\ell^2(\pi)$ with the usual inner product $\langle \cdot, \cdot \rangle_\pi$ defined by

$$\langle f, g \rangle_\pi = \sum_{x \in E} f(x)g(x)\pi(x), \quad f, g \in \ell^2(\pi).$$

We also recall that a basis (f_k) of a Hilbert space $\mathcal{H}$ is a Riesz basis if it is obtained from an orthonormal basis (e_k) under a bounded invertible operator T, that is, $Te_k = f_k$ for all k. It can be shown, see e.g. [114, Theorem 9], that the sequence (f_k) forms a Riesz basis if and only if (f_k) is complete in $\mathcal{H}$ and there exist positive constants A, B such that for arbitrary $n \in \mathbb{N}$ and scalars $c_1, \dots, c_n$, we have

$$A \sum_{k=1}^{n} |c_k|^2 \leqslant \left\| \sum_{k=1}^{n} c_k f_k \right\|^2 \leqslant B \sum_{k=1}^{n} |c_k|^2. \tag{2.6.2}$$

If (g_k) is a biorthogonal sequence to (f_k), that is, $\langle f_k, g_m \rangle_\pi = \delta_{k,m}$, $k, m \in \mathbb{N}$ and $\delta_{k,m}$ is the Kronecker symbol, then (g_k) also forms a Riesz basis.

Theorem 2.6.1. *1. Assume that $\mathfrak{r} < \infty$ and Q is the transition kernel of an irreducible birth-death process, then $P \overset{\Lambda}{\sim} Q$ if and only if P has real and distinct eigenvalues.*

2. $(X, \mathbb{P}) \in \mathcal{S}_{sf}$ with $P \overset{\Lambda}{\sim} Q$ if and only if $(X, \widehat{\mathbb{P}}) \in \widehat{\mathcal{S}}$ with $Q \overset{\widehat{\Lambda}}{\sim} \widehat{P}$, where $\widehat{\Lambda}$ is the adjoint operator of Λ.

Let us assume that $(X, \mathbb{P}) \in \mathcal{S}_{sf}$ with $P \overset{\Lambda}{\sim} Q$. Then the following holds.

(a) For any $h \in \mathcal{E}$ then $(X, \mathbb{P}^h) \in \mathcal{S}_{sf}$ with $P^h \overset{\Lambda^h}{\sim} Q$ and $\Lambda^h = D_h^{-1}\Lambda$, where D_h is a diagonal matrix of h.

(b) If P is normal, i.e. $P\widehat{P} = \widehat{P}P$, then P is self-adjoint, i.e. $P = \widehat{P}$.

(c) P is compact (resp. trace class) if and only if Q is.

(d) Assume that P is compact then for any $f \in \ell^2(\pi)$ and $n \in \mathbb{N}$,

$$P^n f = \sum_{k=0}^{\mathfrak{r}} \lambda_k^n \langle f, f_k^* \rangle_\pi f_k,$$

where the set $(f_k)_{k=0}^{\mathfrak{r}}$ are real eigenfunctions of P associated to the real eigenvalues $(\lambda_k)_{k=0}^{\mathfrak{r}}$ and form a Riesz basis of $\ell^2(\pi)$, and the set $(f_k^*)_{k=0}^{\mathfrak{r}}$ is the unique Riesz basis biorthogonal to $(f_k)_{k=0}^{\mathfrak{r}}$. In particular, for any $x, y \in E$ and $n \in \mathbb{N}$, the spectral expansion of P is given by

$$P^n(x, y) = \sum_{k=0}^{\mathfrak{r}} \lambda_k^n f_k(x) f_k^*(y) \pi(y).$$

(e) $\mathcal{MC} \subseteq \mathcal{S}_{sf}^M$.

(f) *Assume that the condition of the item (d) holds. If $\mathfrak{r} < \infty$ is absorbing and $\mathfrak{l}$ is regular, then with the same notation as above, we have*

$$\mathbb{P}_x(T_{\mathfrak{r}} = n) = \sum_{k=0}^{\mathfrak{r}} \lambda_k^n (1-\lambda_k) \langle \mathbb{1}, f_k^* \rangle_\pi f_k,$$

and assuming that $\mathfrak{l} < \infty$ is absorbing and $\mathfrak{r} < \infty$ is regular then

$$\mathbb{P}_x(T_{\mathfrak{l})} = n) = \sum_{k=0}^{\mathfrak{r}} \lambda_k^n (1-\lambda_k) \langle \mathbb{1}, f_k^* \rangle_\pi f_k.$$

Proof. First, if $P \overset{\Lambda}{\sim} Q$, then P has real and distinct eigenvalues since Q has real and distinct eigenvalues. Conversely, if P has real and distinct eigenvalues, P is diagonalizable, so there exists an invertible Λ such that

$$P = \Lambda D \Lambda^{-1}.$$

where D is the diagonal matrix storing the eigenvalues of P. Given the spectral data D, by inverse spectral theorem, see e.g. [38], one can always construct an ergodic Markov chain with transition matrix Q such that

$$Q = VDV^{-1}.$$

Next, we show item (2). If $P \overset{\Lambda}{\sim} Q$, then for $f \in \ell^2(\pi_Q)$ and $g \in \ell^2(\pi)$,

$$\langle f, \widehat{\Lambda}\widehat{P} g \rangle_{\pi_Q} = \langle P\Lambda f, g \rangle_\pi = \langle \Lambda Q f, g \rangle_\pi = \langle f, Q\widehat{\Lambda} g \rangle_{\pi_Q},$$

which shows that $Q \overset{\widehat{\Lambda}}{\sim} \widehat{P}$. The opposite direction can be shown similarly. Item (a) follows directly from

$$P\Lambda = D_h P^h D_h^{-1} \Lambda = D_h P^h \Lambda^h = \Lambda Q.$$

For the item (b), we recall, from the spectral theorem, that if two normal matrices are similar then they are unitary equivalent that is $\Lambda^{-1} = \widehat{\Lambda}$. Then, the proof of this claim is completed since we easily deduce, from the item (2), that

$$P = P\Lambda\widehat{\Lambda} = \Lambda\widehat{\Lambda}\widehat{P} = \widehat{P}. \tag{2.6.3}$$

Next, we turn to the proof of the item (c). If P is compact (resp. trace class), then $Q = \Lambda^{-1}P\Lambda$ is compact (resp. trace class) since the product of bounded and compact (resp. trace class) operator is a compact (resp. trace class) operator, see [24, Proposition 4.2] (resp. see [91, Page 218]). To show, in the item (d), that (f_k) and (f_m^*) are biorthogonal, we note that the fact that P has distinct eigenvalues yields that $\langle f_k, f_m^* \rangle_\pi = \delta_{k,m}$ for any k, m. Next, denote (g_k) to be the (orthogonal) eigenfunctions of Q, see e.g. [69]. Since $f_k = \Lambda g_k$ and Λ is bounded, (f_k) is complete as (g_k) is a basis. As Λ is bounded from above and below, for any $n \in \mathbb{N}$ and arbitrary sequence $(c_k)_{k=1}^n$, we have

$$A\sum_{k=1}^{n}|c_k|^2 \leqslant \left\|\sum_{k=1}^{n} c_k f_k\right\|_\pi^2 = \left\|\Lambda\sum_{k=1}^{n} c_k g_k\right\|_\pi^2 \leqslant B\sum_{k=1}^{n}|c_k|^2,$$

where we can take $A = \|\Lambda^{-1}\|^{-2}$ and $B = \|\Lambda\|^2$, so (2.6.2) is satisfied. It follows from [114, Theorem 9] that there exists the sequence (f_k^*) being the unique Riesz basis biorthogonal to $(f_k)_{k=0}^{\mathfrak{r}}$, and, any $f \in \ell^2(\pi)$ can be written as

$$f = \sum_{k=0}^{\mathfrak{r}} c_k f_k,$$

where $c_k = \langle f, f_k^* \rangle_\pi$. Desired result follows by applying P^n to f and using $P^n f_k = \lambda_k^n f_k$. In particular, if we take $f = \delta_y$, the Dirac mass at y, and evaluate the resulting expression at x, we obtain the spectral expansion of P. To show (e), we write $\widetilde{P}$ the so-called Siegmund dual (or H_S-dual) of $\widehat{P}$. That is, $\widetilde{P}^T = H_S^{-1}\widehat{P}H_S$ where $H_S = (H_S(x,y))_{x,y\in E}$ is defined to be $H_S(x,y) = \mathbb{1}_{\{x\leqslant y\}}$ and its inverse $H_S^{-1} = (H_S^{-1}(x,y))_{x,y\in E}$ is $H_S^{-1}(x,y) = \mathbb{1}_{\{x=y\}} - \mathbb{1}_{\{x=y-1\}}$, see [102]. Since $X \in \mathcal{MC}$, then $\widehat{P}$ is stochastically monotone, hence from [3, Proposition 4.1], we have that $\widetilde{P}$ is a sub-Markovian kernel. For $x \in [1, \mathfrak{r}-1]$, $\widetilde{p}(x, x+1) = \widehat{p}(x+1, x) > 0$ since $\widehat{P}$

is irreducible and downward skip-free, and $\widetilde{p}(x,y) = 0 \quad \forall y \geqslant x+2$. We also have $\widetilde{p}(0,1) = \widehat{p}(1,0) > 0$. For $x \in [0, \mathfrak{r}-2]$, condition (2) in $\mathcal{MC}$ gives that $\widetilde{p}(x,x-1) > 0$, while condition (3) in $\mathcal{MC}$ guarantees that $\widetilde{p}(x,y) = 0$ for each $x \in [0,\mathfrak{r}-1]$ and $y \in [0, x-2]$. That is, $\widetilde{P}$ is a (strictly substochastic) irreducible birth-death chain when restricted to the state space $[0,\mathfrak{r}-1]$. Denote $\widetilde{P}^{bd}$ the restriction of $\widetilde{P}$ to $[0,\mathfrak{r}-1]$. By breaking off the last row and last column of $\widetilde{P}$, we can write

$$\widetilde{P} = \begin{pmatrix} \widetilde{P}^{bd} & \mathbf{v} \\ \mathbf{0} & 1 \end{pmatrix} = (H_S^{-1} \widehat{P} H_S)^T, \tag{2.6.4}$$

where $\mathbf{0}$ is a row vector of zero, and $\mathbf{v}$ is a column vector storing $\widetilde{p}(x,\mathfrak{r})$ for $x \in [0,\mathfrak{r}-1]$. Considering the h-transform of $\widetilde{P}$ with $h = H_S^T \pi > \mathbf{0}$, see e.g. [51, Theorem 2], we see that $X \in \mathcal{S}_{sf}$, as we have

$$P\Lambda = \Lambda Q,$$

where $\Lambda = (H_S^T D_\pi)^{-1}$ (D_π is a diagonal matrix of π) and $Q = \widetilde{P}$, which completes the proof of (e). Finally to show item (f), after observing that, for any $n \in \mathbb{N}$ and $x \in E$,

$$\mathbb{P}_x(T_\mathfrak{r} > n) = \sum_{y \in E} \mathbb{P}_x(X_n = y, T_\mathfrak{r} > n) = P\mathbb{1}(x)$$

the first representation in (d) with $f \equiv 1$, $n = 1$ and easy algebra yields the desired result. The last claim follows by similar means. $\square$

2.7 Convergence to equilibrium

As a first application of the spectral decomposition stated in Theorem 2.6.1, we derive accurate information regarding the speed of convergence to stationarity for ergodic chains in $\mathcal{S}_{sf}$ in both the Hilbert space topology and in total variation distance. There have been a rich literature devoted to the study of convergence to equilibrium for non-reversible chains, see e.g. [42, 97] and the references therein. In these papers, to overcome the lack of a spectral theory, the authors resort to reversibilization procedures to extract bounds for the distance to stationarity. Our approach reveals a natural extension to the non-reversible case of the classical spectral gap that appears in the study of reversible chains, see e.g. [97]. To state our result we need to introduce some notation. We denote the second largest eigenvalue in modulus (SLEM) or the spectral radius of P in the Hilbert space $\ell_0^2(\pi) = \{f \in \ell^2(\pi);\ \langle \mathbb{1}, f\rangle_\pi = 0\}$, by $\lambda_* = \lambda_*(P) = \sup\{|\lambda_i|;\ \lambda_i \neq 1\}$, then the *absolute spectral gap* is $\gamma_* = 1 - \lambda_*$. For any two probability measures μ, ν on E, the total variation distance between μ and ν is given by

$$||\mu - \nu||_{TV} = \frac{1}{2} \sum_{x \in E} |\mu(x) - \nu(x)|.$$

For $n \in \mathbb{N}$, the total variation distance from stationarity of X is

$$d(n) = \max_{x \in E} ||\delta_x P^n - \pi||_{TV}.$$

For $g \in \ell^2(\pi)$, the mean of g with respect to π can be written as $\mathbb{E}_\pi(g) = \langle g, \mathbb{1}\rangle_\pi$. Similarly, the variance of g with respect to π is $\mathrm{Var}_\pi(g) = \langle g, g\rangle_\pi - \mathbb{E}_\pi^2(g)$. Finally, we recall that Fill in [42, Theorem 2.1] obtained when $\mathfrak{r} < \infty$ the following bound valid for all $n \in \mathbb{N}_0$

$$d(n) \leqslant \frac{\sigma_*(P)^n}{2} \sqrt{\frac{1 - \pi_{min}}{\pi_{min}}}, \tag{2.7.1}$$

where $\pi_{min} = \min_{x \in E} \pi(x)$ and $\sigma_*(P) = \sqrt{\lambda_*(P\widehat{P})}$ is the second largest singular value of P. We obtain the following refinement for Markov chains in the class $\mathcal{S}_{sf}$.

Theorem 2.7.1. *Let $X \in \mathcal{S}_{sf}$ and X is ergodic with stationary distribution π.*

1. *For any $n \in \mathbb{N}_0$, we have*

$$\lambda_*^n \leqslant \|P^n - \pi\|_{\ell^2(\pi) \to \ell^2(\pi)} \leqslant \sigma_*^n(P) \mathbb{1}_{\{n < n^*\}} + \kappa(\Lambda) \lambda_*^n \mathbb{1}_{\{n \geqslant n^*\}} \tag{2.7.2}$$

where $n^* = \lceil \frac{\ln \kappa(\Lambda)}{\ln \sigma_*(P) - \ln \lambda_*} \rceil$ *and* $\kappa(\Lambda) = \|\Lambda\|_{op} \, \|\Lambda^{-1}\|_{op} \geqslant 1$ *is the condition number of* Λ*.*

2. *If* $\mathfrak{r} < \infty$ *then* P *is non-reversible if and only if* $\kappa(\Lambda) > 1$*.*

3. *A sufficient condition for which* $\lambda_* < \sigma_*(P)$ *is given by* $\max_{i \in E} p(i,i) > \lambda_*$*. In such case, for* n *large enough, the convergence rate* λ_* *given in item* (1) *is strictly better than the reversibilization rate* $\sigma_*(P)$*.*

4. *Suppose now that* $\mathfrak{r} < \infty$*. Then, for any* $n \in \mathbb{N}_0$*,*

$$d(n) \leqslant \frac{\min\left(\sigma_*^n(P), \kappa(\Lambda)\lambda_*^n\right)}{2} \sqrt{\frac{1-\pi_{min}}{\pi_{min}}},$$

where $\lambda_* \leqslant \sigma_*(P)$*.*

Remark 2.7.1. It is interesting to recall that when P is reversible and compact then the sequence of eigenfunctions is orthonormal and thus an application of the Parseval identity yields $\|P^n - \pi\|_{\ell^2(\pi)\to\ell^2(\pi)} = \lambda_*^n$ and $\kappa(\Lambda) = 1$ which is a specific instance of item (1) and (2).

Remark 2.7.2. We also recall the discrete analogue of the notion of hypocoercivity introduced in [109], i.e. there exists a constant $C < \infty$ and $\rho \in (0,1)$ such that, for all $n \in \mathbb{N}$,

$$\|P^n - \pi\|_{\ell^2(\pi)\to\ell^2(\pi)} \leqslant C\rho^n.$$

Note that, in general, these constants are not known explicitly. We observe that the upper bound in (2.7.2) reveals that the ergodic chains in $\mathcal{S}_{sf}$ satisfy this hypocoercivity phenomena. More interestingly, our approach based on the similarity concept enables us to get on the one hand an explicit and on the other hand a spectral interpretation of this rate of convergence. Indeed, it can be understood as a modified spectral gap where the perturbation from the classical spectral gap is given by the condition number $\kappa(\Lambda)$ which may be interpreted as a measure of deviation from symmetry. In this vein, we mention the recent work [85] where a similar spectral interpretation of the hypocoercivity phenomena is given for a class of non-self-adjoint Markov semigroups.

Proof. We first show the upper bound in (1). Define the synthesis operator $T^* : \ell^2 \to \ell^2(\pi)$ by $\alpha = (\alpha_i) \mapsto T^*(\alpha) = \sum_{i=0}^{\mathfrak{r}} \alpha_i f_i$, where (f_i) are the eigenfunctions of P and (f_i^*) are the unique biorthogonal basis of (f_i) as in Theorem 2.6.1. For $1 \leqslant i \leqslant \mathfrak{r}$, we take $\alpha_i = \lambda_i^n \langle g, f_i^* \rangle_\pi$, and denote (q_i) to be the orthonormal eigenfunctions of Q, where $f_i = \Lambda q_i$. Note that $||T^*||_{op} \leqslant ||\Lambda||_{op} < \infty$, since

$$||T^*(\alpha)|| = \left\|\sum_{i=0}^{\mathfrak{r}} \alpha_i \Lambda q_i\right\| \leqslant \|\Lambda\|_{op} \left\|\sum_{i=0}^{\mathfrak{r}} \alpha_i q_i\right\|_{\pi_Q} \leqslant ||\Lambda||_{op}||\alpha||_{\ell^2}.$$

For $g \in \ell^2(\pi)$, we also have

$$\sum_{i=0}^{\mathfrak{r}} |\langle g, f_i^* \rangle_\pi|^2 = \sum_{i=0}^{\mathfrak{r}} |\langle g, (\Lambda^*)^{-1} q_i \rangle_\pi|^2 = \sum_{i=0}^{\mathfrak{r}} |\langle \Lambda^{-1} g, q_i \rangle_{\pi_Q}|^2 = ||\Lambda^{-1} g||_{\pi_Q}^2 \leqslant ||\Lambda^{-1}||_{op}^2 ||g||_\pi^2,$$

where the third equality follows from Parseval's identity, which leads to

$$||P^n g - \pi g||_\pi^2 = ||T^*(\alpha)||^2 \leqslant ||\Lambda||_{op}^2 ||\alpha||_{l^2}^2 \leqslant ||\Lambda||_{op}^2 ||\Lambda^{-1}||_{op}^2 \lambda_*^{2n} ||g||_\pi^2. \tag{2.7.3}$$

Desired upper bound follows from (2.7.3) and

$$||P^n - \pi||_{\ell^2(\pi)\to\ell^2(\pi)} \leqslant \lambda_*(\widehat{P}P)^{n/2} = \lambda_*(P\widehat{P})^{n/2},$$

see e.g. [97]. The lower bound in (1) follows readily from the well-known result that the n^{th} power of the spectral radius λ_*^n is less than or equal to the norm of P^n on the reduced space $\ell_0^2(\pi)$. Next, we show that P is non-self-adjoint if and only if $\kappa(\Lambda) > 1$. We recall from (2.6.3) that $PA = A\widehat{P}$ where $A = \Lambda\widehat{\Lambda}$ is a positive self-adjoint matrix. Thus, by the spectral theorem there exists $T \in \mathrm{GL}_{\mathfrak{r}}$, the general linear group of dimension $\mathfrak{r}$, such that $A = TD_A\widehat{T}$ with D_A the diagonal matrix of its eigenvalues where since Λ is defined up to a non-zero multiplicative constant we can assume, without loss of generality, that its largest eigenvalue $\lambda_1(A)$=1. Next, we recall from [49, p. 382] that

$$\kappa(\Lambda) = \frac{\sigma_1(\Lambda)}{\sigma_{\mathfrak{r}}(\Lambda)} = \frac{\sqrt{\lambda_1(A)}}{\sqrt{\lambda_{\mathfrak{r}}(A)}}, \tag{2.7.4}$$

where $\sigma_1(\Lambda)$ (resp. $\sigma_{\mathfrak{r}}(\Lambda)$) is the largest (resp. smallest) singular value of Λ and $\lambda_{\mathfrak{r}}(A)$ is the smallest eigenvalue which is positive as $\Lambda \in \mathrm{GL}_{\mathfrak{r}}$ and hence $A \in \mathrm{GL}_{\mathfrak{r}}$. Thus, $\kappa(\Lambda) = 1$ implies that $D_A = I_{\mathfrak{r}}$ where $I_{\mathfrak{r}}$ is the identity matrix, that is $\Lambda\widehat{\Lambda} = A = I_{\mathfrak{r}}$ and, from (2.6.3), we deduce that P is self-adjoint. Conversely, if P is self-adjoint then by means of the same argument used for the proof of Theorem 2.6.1(b), we have that Λ is unitary and hence $\Lambda\widehat{\Lambda} = I_{\mathfrak{r}}$, that is $\kappa(\Lambda) = 1$, which completes the proof of this statement. The claim in (3) is a straightforward consequence of the Sing-Thompson theorem [107]. Next, using (2.7.3), we get

$$\mathrm{Var}_\pi\left(\widehat{P}^n g\right) \leqslant \kappa(\widehat{\Lambda})^2\lambda_*^{2n}\mathrm{Var}_\pi(g) = \kappa(\Lambda)^2\lambda_*^{2n}\mathrm{Var}_\pi(g), \quad n \in \mathbb{N}_0, \tag{2.7.5}$$

where we used the obvious identity $\kappa(\Lambda) = \kappa(\widehat{\Lambda})$ in the equality. This leads to

$$\begin{aligned}||\delta_x P^n - \pi||_{TV}^2 &= \frac{1}{4}\mathbb{E}_\pi^2\left|\frac{\delta_x P^n}{\pi} - 1\right| \leqslant \frac{1}{4}\mathrm{Var}_\pi\left(\frac{\delta_x P^n}{\pi}\right) = \frac{1}{4}\mathrm{Var}_\pi\left(\widehat{P}^n\frac{\delta_x}{\pi}\right)\\ &\leqslant \frac{1}{4}\kappa(\Lambda)^2\lambda_*^{2n}\mathrm{Var}_\pi\left(\frac{\delta_x}{\pi}\right) = \frac{1}{4}\kappa(\Lambda)^2\lambda_*^{2n}\frac{1-\pi(x)}{\pi(x)} \leqslant \frac{1}{4}\kappa(\Lambda)^2\lambda_*^{2n}\frac{1-\pi_{min}}{\pi_{min}},\end{aligned}$$

where the first inequality follows from Cauchy-Schwartz inequality. The proof is completed by combining the above bound with (2.7.1). □

A numerical example

To illustrate the previous result, we consider the following example in the $\mathcal{MC}$ class where the dual transition matrix is given by

$$\widehat{P} = \begin{pmatrix} 0.275 & 0.7 & 0.005 & 0.02 \\ 0.17 & 0.8 & 0.01 & 0.02 \\ 0 & 0.94 & 0.02 & 0.04 \\ 0 & 0 & 0.95 & 0.05 \end{pmatrix}. \tag{2.7.6}$$

Table 2.1 shows the rate of convergence of P towards $\pi = (0.18, 0.77, 0.03, 0.02)$ with $\widehat{P}$ given in (2.7.6). We observe that for $n = 1, 2$, the reversibilization bound $\lambda_*(P\widehat{P})^{n/2}$

is smaller while for $n \geqslant 3$, our upper bound in Theorem 2.7.1 $\kappa(\Lambda)\lambda_*^n$ is smaller. Also, we point out that since $\max_i p(i,i) = 0.8 > \lambda_* = 0.17$, the Sing-Thompson condition in item (3) of Theorem 2.7.1 holds.

n	$\|\|P^n - \pi\|\|_{\ell^2(\pi)\to\ell^2(\pi)}$	$\kappa(\Lambda)\lambda_*^n$	$\lambda_*(P\widehat{P})^{n/2}$
1	0.81	8.64	0.81
2	0.08	1.42	0.65
3	0.02	0.24	0.52
4	0.003	0.04	0.42
5	0.0005	0.006	0.34

Table 2.1: $\ell^2(\pi)$-rate of convergence of P

2.8 Separation cutoff

As another illustration of the similarity concept, we aim at generalizing, to the non-reversible chains in the class $\mathcal{S}_{sf}^M$, the separation cutoff criteria established by Diaconis and Saloff-Coste in [31] and Chen and Saloff-Coste in [13] for reversible birth-death chains. We also offer an alternative necessary and sufficient condition as obtained by Mao et al. [76] recently for continuous-time upward skip-free chain with stochastic monotone time-reversal. To this end, we recall the definition of separation distance of Markov chains, which is used as a standard measure for convergence to equilibrium. For $n \in \mathbb{N}$, the maximum separation distance $s(n)$ is defined by

$$s(n) = \max_{x,y\in E}\left[1 - \frac{P^n(x,y)}{\pi(y)}\right] = \max_{x\in E}\,\mathrm{sep}(P^n(x,\cdot),\pi) = \max_{x\in E} s_x(n).$$

Note that separation distance is not a metric. One of its nice feature is its connection to strong stationary times that we now describe. We say that a strong stationary time T for a Markov chain X with stationary distribution π is a randomized stopping time T, possibly depending on the initial starting position x, if, for all $x, y \in E$,

$$\mathbb{P}_x(T = n, X_T = y) = \mathbb{P}_x(T = n)\pi(y).$$

The fastest strong stationary time is a strong stationary time such that for all $n \in \mathbb{N}$, $s_x(n) = \mathbb{P}(T > n)$. We now provide a description of the cutoff phenomenon for Markov chains. Recall that the separation mixing times are defined, for any $x \in E$ and $\epsilon > 0$, as

$$T^s(x, \epsilon) = \min\{n \geqslant 0;\ \mathrm{sep}(P^n(x, \cdot), \pi) \leqslant \epsilon\}$$

and

$$T^s(\epsilon) = \min\{n \geqslant 0;\ s(n) \leqslant \epsilon\}.$$

A family, indexed by $n \in \mathbb{N}$, of ergodic chains $X^{(n)}$ defined on $E_{\mathfrak{r}_n} = \{0, \ldots, \mathfrak{r}_n\}$ with transition matrix P_n, stationary distribution π_n and separation mixing times $\mathrm{T}_n(\epsilon) = T_n^s(\epsilon)$ or $T_n^s(x, \epsilon)$, for some $x \in E$, is said to present a separation cutoff if there is a positive sequence (t_n) such that for all $\epsilon \in (0, 1)$,

$$\lim_{n \to \infty} \frac{\mathrm{T_n}(\epsilon)}{t_n} = 1.$$

The family has a (t_n, b_n) separation cutoff if the sequences (t_n) and (b_n) are positive, $b_n/t_n \to 0$ and for all $\epsilon \in (0, 1)$,

$$\limsup_{n \to \infty} \frac{|\mathrm{T}_n(\epsilon) - t_n|}{b_n} < \infty.$$

Let us now write $\mathrm{R}_n = (I - P_n)^{-1}_{\ell_0^2}$ for the centered resolvent, that is, the resolvent operator restricted to $l_0^2(\pi)$. The main result of this section is the following.

Theorem 2.8.1. *Suppose that, for each $n \geqslant 1$, $X^{(n)} \in \mathcal{S}_{\mathfrak{r}_n}^M$ and let $(\theta_{n,i})_{i=1}^{\mathfrak{r}_n}$ be the non-zero eigenvalues of $I - P_n$. Define*

$$\underline{\theta}_n = \min_{1 \leqslant i \leqslant \mathfrak{r}_n} \theta_{n,i}, \quad \rho_n^2 = \sum_{i=1}^{\mathfrak{r}_n} \frac{1 - \theta_{n,i}}{\theta_{n,i}^2}.$$

Then the family of chains $(X^{(n)})$ with transition kernel (P_n), all started from 0, has a separation cutoff if and only if $\mathrm{Tr}(\mathrm{R}_n)\underline{\theta}_n \to \infty$ if and only if $T^s_n(0,\epsilon)\underline{\theta}_n \to \infty$. In this case there is a $(\mathrm{Tr}(\mathrm{R}_n), \max\{\rho_n, 1\})$ separation cutoff.

We begin the proof of Theorem 2.8.1 by giving an important lemma that gives a lower bound on the mixing time in terms of the eigenvalues of $I - P$. The corresponding result for reversible and ergodic Markov chain can be found in [74, Theorem 12.4].

Lemma 2.8.1. *For an ergodic chain $X \in \mathcal{S}_{sf}$ and $\epsilon \in (0,1)$, denote by $(\theta_i)_{i=0}^{\mathfrak{r}}$ the eigenvalues of $I - P$ arranged in ascending order and $\underline{\theta} = \min_{i\neq 0}\theta_i$. We have*

$$T^s(\epsilon) \geqslant \left(\underline{\theta}^{-1} - 1\right)\log\left(\frac{1}{2\epsilon}\right).$$

Proof. Take f to be an eigenfunction of P associated to an arbitrary eigenvalue $\lambda \neq 1$, where $\lambda \in (\lambda_i)_{i=0}^{\mathfrak{r}}$ which is the set of (real) eigenvalues of P arranged in descending order. Note that f is orthogonal to $\mathbb{1}$, since

$$\langle Pf, \mathbb{1}\rangle_\pi = \lambda\langle f, \mathbb{1}\rangle_\pi = \langle f, \widehat{P}\mathbb{1}\rangle_\pi = \langle f, \mathbb{1}\rangle_\pi,$$

where the last equality follows from stochasticity of $\widehat{P}$. By writing $||f||_\infty = \max_{x\in E}|f(x)| = f(x^*)$, we have

$$\begin{aligned}|\lambda^n f(x)| = |P^n f(x)| &= \left|\sum_{y\in E} P^n(x,y)f(y) - \pi(y)f(y)\right| \\ &\leqslant 2||f||_\infty ||P^n(x,\cdot) - \pi||_{\mathrm{TV}} \leqslant 2||f||_\infty s(n),\end{aligned}$$

where the first inequality follows from the definition of total variation distance $||P^n(x,\cdot) - \pi||_{\mathrm{TV}}$, and the second inequality comes from the result that total variation distance is less than or equal to separation distance, see e.g. [74, Lemma 6.13]. Taking $n = T^s(\epsilon)$ and $x = x^*$, the above yields $|\lambda|^{T^s(\epsilon)} \leqslant 2\epsilon$, which leads to

$$\frac{1-|\lambda|}{|\lambda|}T^s(\epsilon) \geqslant -\log(|\lambda|)T^s(\epsilon) \geqslant -\log(2\epsilon).$$

The proof is completed by specializing to $|\lambda| = \max\{|\lambda_1|, |\lambda_{\mathfrak{r}}|\}$. $\square$

The next lemma gives a necessary condition for separation cutoff in terms of the spectral information.

Lemma 2.8.2 (Necessary condition for separation cutoff)**.** *Suppose that, for each $n \geqslant 1$, $X^{(n)} \in \mathcal{SF}_{\mathfrak{r}_n}$ and let $(\theta_{n,i})_{i=1}^{\mathfrak{r}_n}$ be the eigenvalues of $I - P_n$. For the n^{th} chain in the family, define $\underline{\theta}_n = \min_{1\leqslant i\leqslant \mathfrak{r}_n} \theta_{n,i}$ and $T_n^s(\epsilon)$ to be the separation mixing time. If the family of chain with transition kernel (P_n) exhibits a separation cutoff, then $T_n^s(\epsilon)\underline{\theta}_n \to \infty$.*

Proof. Let $T_n^s = T_n^s(0.25)$, that is, choosing $\epsilon = 0.25$. If $T_n^s(\epsilon)\underline{\theta}_n$ is bounded above in n by $c > 0$, then by Lemma 2.8.1 we have

$$\frac{T_n^s(\epsilon)}{T_n^s} \geqslant \frac{\underline{\theta}_n^{-1} - 1}{T_n^s} \log\left(\frac{1}{2\epsilon}\right) \geqslant c \log\left(\frac{1}{2\epsilon}\right),$$

so $\dfrac{T_n^s(\epsilon)}{T_n^s} \to \infty$ as $\epsilon \to 0$, which implies that there is no separation cutoff. $\square$

We are now ready to complete the proof of Theorem 2.8.1. Denote by P_n^k the distribution of the n^{th} chain at time k, and by π_n the stationary measure of the n^{th} chain. We also write $t_n = \mathrm{Tr}(\mathrm{R}_n)$. It is known, see e.g. [43, Theorem 1.4], that

$$\mathrm{sep}(P_n^k, \pi_n) = \mathbb{P}(T_n > k),$$

where T_n is the fastest strong stationary time of the n^{th} chain, which is equal in distribution to a $\mathfrak{r}_n$-fold convolution of geometric random variables each with success probability $\theta_{n,i}$ for $i = 1, \ldots, \mathfrak{r}_n$, mean t_n and variance ρ_n^2. The key to establish the proof is the following.

$$\rho_n^2 = \underline{\theta}_n^{-2} \sum_{i=1}^{\mathfrak{r}_n} \frac{(1-\theta_{n,i})\,\underline{\theta}_n^2}{\theta_{n,i}^2} \leqslant \underline{\theta}_n^{-2} \sum_{i=1}^{\mathfrak{r}_n} \frac{\underline{\theta}_n}{\theta_{n,i}} = \underline{\theta}_n^{-1} t_n, \tag{2.8.1}$$

where we use the facts that $\theta_{n,i} \geqslant 0$ and $\underline{\theta}_n/\theta_{n,i} \leqslant 1$ in the inequality. Assume that $t_n\underline{\theta}_n \to \infty$, which together with (2.8.1) yields $\rho_n/t_n \to 0$. The rest of the argument are similar to the ones developed in the proof of [31, Theorem 5.1]. For sake of completeness we now provide its main ingredients. First, by means of Chebyshev's inequality, we have

$$t_n - (\epsilon^{-1} - 1)^{1/2}\rho_n \leqslant T_n^s(0, \epsilon) \leqslant t_n + (\epsilon^{-1} - 1)^{1/2}\rho_n. \tag{2.8.2}$$

This shows that the family of chain exhibits a separation cutoff if we divide (2.8.2) by t_n and take $n \to \infty$. On the other hand, separation cutoff implies $T_n^s(0, \epsilon)\underline{\theta}_n \to \infty$ by Lemma 2.8.2, so it remains to show $T_n^s(0, \epsilon)\underline{\theta}_n \to \infty$ implies $t_n\underline{\theta}_n \to \infty$. Using $t_n\underline{\theta}_n \geqslant 1$, (2.8.1) yields $\rho_n \leqslant t_n$, and together with the upper bound of (2.8.2) leads to

$$T_n^s(0, \epsilon) \leqslant t_n + (\epsilon^{-1} - 1)^{1/2}\rho_n \leqslant t_n(1 + (\epsilon^{-1} - 1)^{1/2}),$$

so $t_n\underline{\theta}_n \to \infty$ holds if and only if $T_n^s(0, \epsilon)\underline{\theta}_n \to \infty$. It follows from [13, Remark 1.1] that there is a $(t_n, \max\{\rho_n, 1\})$ separation cutoff. Precisely, (2.8.2) gives

$$|T_n^s(0, \epsilon) - t_n| \leqslant (\epsilon^{-1} - 1)^{1/2}\rho_n + 1,$$

and a $(t_n, \max\{\rho_n, 1\})$ cutoff is observed by noting that $\theta_{n,i} \leqslant 2$, $t_n \geqslant n/2$ and $\rho_n/t_n \to 0$.

2.8.1 ℓ^p-cutoff

We proceed by investigating the ℓ^p-cutoff for fixed $p \in (1, \infty)$ for the class $\mathcal{S}$. Recall that Chen and Saloff-Coste [11, Theorem 4.2, 4.3] have shown that for a family of *normal* ergodic transition kernel P_n, the max-ℓ^p cutoff is equivalent to the *spectral gap times mixing time* going to infinity. We can extend their result to the case of the non-normal chains in $\mathcal{S}$ as follows.

Theorem 2.8.2 (Max-ℓ^p cutoff)**.** *Suppose that, for each $n \geqslant 1$, $X^{(n)} \in \mathcal{S}_{\mathfrak{r}_n}$ with compact transition kernel $P_n \overset{\Lambda_n}{\sim} Q_n$ and stationary measure π_n, and let $\lambda_{n,*}$ be the second largest eigenvalue in modulus of P_n. Assume that*

$$\sup_{n \geqslant 1} ||\Lambda_n||_{op} \, ||\Lambda_n^{-1}||_{op} < \infty.$$

Fix $p \in (1, \infty)$ and $\epsilon > 0$. Consider the max-ℓ^p distance to stationarity

$$f_n(t) = \sup_{x \in E_{\mathfrak{r}_n}} \left\| \frac{p_n^t(x, \cdot)}{\pi_n} - 1 \right\|_{\ell^p(\pi)}$$

and define

$$t_n = \inf\{t > 0; \; f_n(t) \leqslant \epsilon\}, \quad \theta_{n,*} = -\log \lambda_{n,*} \quad \textit{and } \mathcal{F} = \{f_n; \; n = 1, 2, \ldots\}.$$

Assume that each n, $f_n(t) \to 0$ as $t \to \infty$ and $t_n \to \infty$. Then the family $\mathcal{F}$ has a max-ℓ^p cutoff if and only if $t_n \theta_{n,*} \to \infty$. In this case there is a $(t_n, \max\{1, \theta_{n,*}^{-1}\})$ cutoff.

The proof in [11, Theorem $4.2, 4.3$] works nicely as long as we have Lemma 2.8.3 below, which gives a two-sided control on the $\ell^p(\pi)$ norm of $P^n - \pi$. The following lemma is then the key to the proof.

Lemma 2.8.3. *Suppose that $X \in \mathcal{S}$ on E with transition kernel $P \overset{\Lambda}{\sim} Q$. Fix $p \in (1, \infty)$. Then, for any $n \in \mathbb{N}$, we have*

$$2^{-1+\theta_p} \lambda_*(P)^{n\theta_p} \leqslant \|P^n - \pi\|_{\ell^p(\pi) \to \ell^p(\pi)} \leqslant 2^{|1-2/p|} (\kappa(\Lambda) \lambda_*(P)^n)^{1-|1-2/p|}, \tag{2.8.3}$$

where $\theta_p \in [1/2, 1]$ and $\kappa(\Lambda) = \|\Lambda\|_{op} \, ||\Lambda^{-1}||_{op}$.

Proof. By the Riesz-Thorin interpolation theorem, see e.g. [11, equation 3.4], we have

$$\|P^n - \pi\|_{\ell^p(\pi) \to \ell^p(\pi)} \leqslant 2^{|1-2/p|} \, \|P^n - \pi\|_{\ell^2(\pi) \to \ell^2(\pi)}^{1-|1-2/p|},$$

which when combined with Theorem 2.7.1 gives the upper bound of (2.8.3). Next, to show the lower bound in (2.8.3), we use another version of the Riesz-Thorin interpolation theorem, see e.g. [11, Lemma 4.1], to get

$$\|P^n - \pi\|_{\ell^p(\pi)\to\ell^p(\pi)} \geqslant 2^{-1+\theta_p} \|P^n - \pi\|^{\theta_p}_{\ell^2(\pi)\to\ell^2(\pi)} \geqslant 2^{-1+\theta_p} \lambda_*(P)^{n\theta_p},$$

where we use Theorem 2.7.1 in the second inequality. This completes the proof. $\square$

CHAPTER 3

ANALYSIS OF NON-REVERSIBLE MARKOV CHAINS VIA SIMILARITY ORBIT

3.1 Introduction

In this Chapter, we study Markov chains on a denumerable state space $\mathcal{X}$ with transition kernel P and time-reversal $\widehat{P}$ by the concept of similarity orbit as defined in Definition 3.1.1 below. We write $\mathcal{M}$ to be the set of Markov transition kernel on $\mathcal{X}$. The spectral theorem of normal operators turns out to be a powerful tool to deal with substantial and difficult issues arising in the study of normal Markov chains. This includes, for example, the rate of convergence and L^2-cutoff (see e.g. Chen and Saloff-Coste [12]) as well as in estimating the integral functionals for normal Markov semigroups (see e.g. Altmeyer and Chorowski [2]). However, removing the assumption on normality gives major difficulty for a spectral analysis of general Markov chains because of a lack of spectral theorem for non-normal operators, since P is possibly a non-normal linear operator in the weighted Hilbert space

$$\ell^2(\pi) = \left\{ f : \mathcal{X} \mapsto \mathbb{C};\ ||f||_\pi^2 = \sum_{x \in \mathcal{X}} |f(x)|^2 \pi(x) < \infty \right\},$$

with π being an invariant measure of P. The intrusion of spectral theory to the analysis of Markov chains first dates back to the long line of work initiated by Ledermann and Reuter [69] and Karlin and McGregor [58] who were among the first to offer a detailed spectral analysis in the direction of birth-death processes. To overcome the challenge of analyzing non-self-adjoint operators, there are a wide variety of ideas proposed in the literature, such as dilation Kendall [62], reversiblizations Fill [42] or recasting to a weighted-L^∞ space Kontoyiannis and Meyn [67]. We shall elaborate an approach by resorting to the concept of similarity to give a spectral analysis of a class of Markov chain that we call $\mathcal{S} \subset \mathcal{M}$ to be introduced in Definition 3.1.2 below, along the sequence of work by Choi and Patie [20], Miclo [83], Patie and Savov [85] and Patie and Zhao [88] for the study of spectral theory of non-reversible Markov processes via intertwining and Chafaï and Joulin [10] and Cloez and Delplancke [23] for intertwining of birth-death processes. We first recall the definition of similarity as introduced in Choi and Patie [20].

Definition 3.1.1 (Similarity)**.** We say that the transition kernel P of a Markov chain $X \in \mathcal{SF}$ is similar to the transition kernel of a Markov chain Q on $\mathcal{X}$, and we write $P \sim Q$, if there exists a bounded linear operator $\Lambda : \ell^2(\pi_Q) \to \ell^2(\pi)$ (π_Q being an invariant measure for Q) with bounded inverse such that

$$P\Lambda = \Lambda Q. \tag{3.1.1}$$

When needed we may write $P \overset{\Lambda}{\sim} Q$ to specify the intertwining or the link kernel Λ. Note that $\sim$ is an equivalence relationship on the set of transition kernels $\mathcal{M}$.

Remark 3.1.1. In the discrete-time setting, for $n \in \mathbb{N}$, if $P \overset{\Lambda}{\sim} Q$, then $P^n \overset{\Lambda}{\sim} Q^n$.

Remark 3.1.2. Note that this definition carries over when we study similarity on the level of infinitesimal generator in the continuous-time setting. For example, we write $L \overset{\Lambda}{\sim} G$ if L (resp. G) is the infinitesimal generator associated with the continuous-time Markov semigroup $(P_t)_{t\geqslant 0}$ (resp. $(Q_t)_{t\geqslant 0}$). It follows easily that if $L \overset{\Lambda}{\sim} G$ then $P_t \overset{\Lambda}{\sim} Q_t$ for $t \geqslant 0$.

The $\mathcal{S}$ class is simply the similarity orbit in $\mathcal{M}$ of all possible Markov chains with normal transition kernel on $\mathcal{X}$. Note that reversible Markov kernels are normal operators in $\ell^2(\pi)$. From now on, we write $\mathcal{N}$ to be the set of normal transition kernel Q on $\mathcal{X}$, that is, $Q\widehat{Q} = \widehat{Q}Q$ in $\ell^2(\pi_Q)$ with $\widehat{Q}$ being the time-reversal of Q.

Definition 3.1.2 (The $\mathcal{S}$ class)**.** Suppose that $Q \in \mathcal{N}$. The similarity orbit of Q (in $\mathcal{SF}$)

is

$$\mathcal{S}(Q) = \{P \in \mathcal{M};\ P \sim Q\},$$

and the $\mathcal{S}$ class is the union over all possible orbits

$$\mathcal{S} = \bigcup_{Q \in \mathcal{N}} \mathcal{S}(Q).$$

We note that according to Wermer [111], the class $\mathcal{S}$ is also characterized as the class of Markov chain whose transition kernel is a spectral scalar-type operator in the sense of Dunford [36, Section 3], see also Dunford and Schwartz [37, Page 1938, Definition 1].

Finally, we say that $X \in \mathcal{S}^M$ if $X \in \mathcal{S}$ and its time-reversal $(X, \widehat{\mathbb{P}})$ is stochastically monotone, i.e. $y \mapsto \widehat{\mathbb{P}}_y(X_1 \leqslant x)$ is non-increasing in $\mathcal{X}$ for every fixed $x \in \mathcal{X}$.

We now summarize the major contributions of this work in the analysis of general Markov chains which also serve as an outline of this Chapter. In Section 3.2, we begin by showing how the concept of similarity orbit is natural for developing the spectral decomposition of non-reversible Markov operators in the class $\mathcal{S}$. Indeed, each of its element admits a spectral representation with respect to non-self-adjoint resolution of identity as introduced by Dunford [36], see also Dunford and Schwartz [37]. We also remark on the growing interest for non-self-adjoint operators with real spectrum that arise in the study of pseudo-hermitian quantum mechanics, see e.g. Inoue and Trapani [55] and the references therein. As by-product, one can develop a functional calculus for this class as for normal operators. Moreover, we obtain, under mild conditions, an eigenvalues expansion expressed in terms of Riesz basis, a notion that generalizes orthogonal basis and was introduced in non-harmonic analysis, see Young [114]. Another intriguing aspect of the similarity orbit analysis is that in the continuous-time setting with $L \in \mathcal{S}(G)$ (see Remark 3.1.2 above), where G is the generator of a normal Markov chain, then both the heat kernel $(e^{tL})_{t \geqslant 0}$ and $(e^{tG})_{t \geqslant 0}$ share the same *eigentime* identity, offering new examples and insights to the sequence of work by Aldous and Fill [1], Cui and Mao [25] and Miclo [82]. In view of the tractability and the fascinating properties that the class $\mathcal{S}$ possesses, it will be very interesting to characterize this class in terms of the one-step transition probabilities of $P \in \mathcal{S}$. Although fundamental, this issue seems to be very challenging. However, we manage to identify a set of sufficient conditions that defines what we call the generalized monotonicity condition class $\mathcal{GMC} \subset \mathcal{S}$, such that the time-reversal $\widehat{P}$ intertwines with a birth-death chain. This $\mathcal{GMC}$ class rests on the assumption of stochastic monotonicity in which Λ is the so-called Siegmund kernel. This readily generalizes the $\mathcal{MC}$ class introduced by Choi and Patie [20] in the context of skip-free chains. Note that the notion of stochastic monotonicity is studied by Siegmund [102] and Clifford and Sudbury [22] and intertwining between stochastic monotone birth-death chains, which are reversible chains, has been previously investigated in detail by Diaconis and Fill [28], Huillet and Martinez [51] and Jansen and Kurt [56]. Added to the above, we obtain a two-phase refinement for the convergence rate of the Markov kernels in the class $\mathcal{S}$ measured in the Hilbert space topology or in total variation distance: recall that in the normal case the rate of convergence in the Hilbert space topology is given by *exactly* the second largest eigenvalue in modulus; for class $\mathcal{S}$ however, in small time we adapt the singular value upper bound of Fill [42], while for large time, the decay rate is the second largest eigenvalue in modulus modulo a constant which is the condition number of the link kernel Λ. This offers an original spectral explanation of the hypocoercivity phenomenon that has been observed and studied intensively in the PDE literature, see for instance Villani [109]. All these first consequences of the spectral representation are stated and proved in Section 3.2. Relying on such spectral decomposition as well as the fastest strong stationary time result of general chains obtained by Fill [43], we study the separation cutoff phenomenon and demonstrate that the famous "spectral gap times mixing time" conjecture as well as the proof in Diaconis and Saloff-Coste [31] carries over to the subclass $\mathcal{GMC}^+ \subset \mathcal{GMC}$ in Section 3.3. Next, building upon the concept of the non-self-adjoint spectral measure and the Laplace transform cutoff criteria proposed in Chen and Saloff-Coste [12] and further elaborated in Chen et al. [15], we illustrate that the usual L^2-cutoff criteria for reversible chains generalizes to the class $\mathcal{S}$ in Section 3.4.

Second, in Section 3.5, we would like to estimate integral functionals of the type

$$\Gamma_T(f) = \int_0^T f(X_t)\, dt, \quad T \geqslant 0,$$

where T is a fixed time and f is a function such that the integral $\Gamma_T(f)$ is well-defined, by the Riemann-sum estimator given by, for $n \in \mathbb{N}$,

$$\hat{\Gamma}_{T,n}(f) = \sum_{k=1}^{n} f(X_{(k-1)\Delta_n})\Delta_n,$$

where we observe $(X_t)_{t\in[0,T]}$ at discrete epochs $t = (k-1)\Delta_n$ with $k \in [\![n]\!] := \{1, \ldots, n\}$ and $\Delta_n = T/n$. This work is motivated by the recent work of Altmeyer and Chorowski [2], in which they studied the same problem with the outstanding assumption that the infinitesimal generator of the Markov process $(X_t)_{t\geqslant 0}$ is a normal operator to yield interesting results on the estimator error bound by spectral theory. We demonstrate that a number of their results can be readily generalized to the class $\mathcal{S}$ on the infinitesimal generator level.

Finally, in Section 3.6, we examine three particular similarity orbits of reversible Markov chains, that we call the permutation orbit, the pure birth orbit and the random walk orbit respectively. More precisely, suppose that we start with a reversible generator G such that $G \overset{\Lambda}{\sim} L$, where L is the generator of a contraction yet possibly non-Markovian semigroup $(e^{tL})_{t\geqslant 0}$, we would like to investigate various properties of L with Λ being either a permutation, pure birth or random walk kernel. This idea is powerful enough to allow us to generate completely new Markov or contraction kernel from known ones in which we have precise control and exact expressions on the stationary distribution, eigenfunctions and the speed of convergence. In particular, we perform an in-depth study on the permutation and pure birth variants of four models and their associated orthogonal polynomials, namely the Ehrenfest model (Section 3.6.1 and 3.6.3), $M/M/\infty$ queue (Section 3.6.2 and 3.6.4), linear birth-death process (Section 3.6.2 and 3.6.4) and quadratic birth-death process (Section 3.6.1 and 3.6.3). Finally, we study the random walk orbit in Section 3.6.5, with the link Λ being the random walk previously studied by Diaconis and Miclo [29] and Zhou [116]. An interesting aspect is that the right eigenfunction of L can now be interpreted as a discrete cosine transform.

3.2 Spectral theory of the class $\mathcal{S}$ and its convergence rate to equilibrium

In this Section, we develop an original methodology to obtain the spectral decomposition in Hilbert space of the transition operator of Markov chains that belong to the class $\mathcal{S}$, a subclass of $\mathcal{M}$ which is defined in Definition 3.1.2. We write $\|\cdot\|_{op}$ to be the operator norm, i.e. $\|P\|_{op} = \sup_{||f||_\pi=1} ||Pf||_\pi$, and $[\![a, b]\!] := \{a, a+1, \ldots, b-1, b\}$ for any $a \leqslant b \in \mathbb{Z}$. We proceed by recalling that P has a time-reversal $\widehat{P}$, that is, for $x, y \in \mathcal{X}$,

$$\pi(x)\widehat{P}(x, y) = \pi(y)P(y, x),$$

where π is a reference measure for P. We equip the Hilbert space $\ell^2(\pi)$ with the usual inner product $\langle \cdot, \cdot \rangle_\pi$ defined by

$$\langle f, g\rangle_\pi = \sum_{x\in\mathcal{X}} f(x)\overline{g(x)}\pi(x), \quad f, g \in \ell^2(\pi),$$

where $\overline{g}$ is the complex conjugate of g. A spectral measure (or resolution of identity) in the sense of Dunford [36, Section 3] and Dunford and Schwartz [37, Page 1929

Definition 1] of a Hilbert space $\mathcal{H}$ on $\mathbb{C}$ is a family of bounded operators $\mathcal{E} = \{E_B; B \in \mathcal{B}(\mathbb{C})\}$, where $\mathcal{B}(\mathbb{C})$ is the Borel algebra on $\mathbb{C}$, satisfying the following:

1. $E_\emptyset = 0, E_\mathbb{C} = I$.
2. For all $A, B \in \mathcal{B}(\mathbb{C})$,
$$E_{A\cup B} = E_A E_B,$$
while for disjoint A, B,
$$E_{A\cap B} = E_A + E_B.$$
3. There exists a constant M such that $\|E_B\|_{op} \leqslant M$ for all $B \in \mathcal{B}(\mathbb{C})$.

For normal operator $Q \in \mathcal{N}$, its resolution of identity $\mathcal{E}$ is self-adjoint and hence $\mathcal{E}$ is a self-adjoint orthogonal projection. We also denote E_B^* to be the adjoint of E_B. Recall that by the spectral theorem for normal operators the spectral resolution of Q is

$$Q = \int_{\sigma(Q)} \lambda \, dE_\lambda,$$

where $\sigma(Q)$ is the spectrum of Q. More generally, for $M \in \mathcal{M}$, we write $\sigma(M)$ (resp. $\sigma_c(M)$, $\sigma_p(M)$, $\sigma_r(M)$) to be the spectrum (resp. continuous spectrum, point spectrum, residual spectrum) of M. We proceed to recall the notion of Riesz basis, which will be useful when we derive the spectral decomposition for compact $P \in \mathcal{S}$ in our main result Theorem 3.2.1 below. A basis (f_k) of a Hilbert space $\mathcal{H}$ is a Riesz basis if it is obtained from an orthonormal basis (e_k) under a bounded invertible operator T, that is, $Te_k = f_k$ for all k. It can be shown, see e.g. Young [114, Theorem 9], that the sequence (f_k) forms a Riesz basis if and only if (f_k) is complete in $\mathcal{H}$ and there exist positive constants A, B such that for arbitrary $n \in \mathbb{N}$ and scalars $c_1, \ldots, c_n$, we have

$$A \sum_{k=1}^n |c_k|^2 \leqslant \left\| \sum_{k=1}^n c_k f_k \right\|^2 \leqslant B \sum_{k=1}^n |c_k|^2. \tag{3.2.1}$$

If (g_k) is a biorthogonal sequence to (f_k), that is, $\langle f_k, g_m\rangle_\pi = \delta_{k,m}$, $k, m \in \mathbb{N}$ and $\delta_{k,m}$ is the Kronecker symbol, then (g_k) also forms a Riesz basis. We are now ready to state the main result of this Chapter in the following, and the proof can be found in Section 3.2.1.

Theorem 3.2.1. *Assume that $P \in \mathcal{S}$ with $P \overset{\Lambda}{\sim} Q$. Then the following holds.*

(a) Denote the self-adjoint spectral measure of Q by $\mathcal{E} = \{E_B;\ B \in \mathcal{B}(\mathbb{C})\}$, then $\{F_B := \Lambda E_B \Lambda^{-1};\ B \in \mathcal{B}(\mathbb{C})\}$ defines a spectral measure and P is a spectral scalar-type operator with spectral resolution given by

$$P = \int_{\sigma(P)} \lambda \, dF_\lambda,$$
$$\widehat{P} = \int_{\sigma(\widehat{P})} \lambda \, dF^*_\lambda.$$

Note that

$$\sigma(P) = \sigma(Q), \sigma(P) = \overline{\sigma(\widehat{P})}, \sigma_c(P) = \sigma_c(Q), \sigma_p(P) = \sigma_p(Q), \sigma_r(P) = \sigma_r(Q),$$

and the multiplicity of each eigenvalue in $\sigma_p(P)$ is the same as that of $\sigma_p(Q)$. For analytic and single valued function f on $\sigma(P)$, we have

$$f(P) = \int_{\sigma(P)} f(\lambda) \, dF_\lambda.$$

In particular, if P is compact on $\mathcal{X} = [\![0, \mathfrak{r}]\!]$ with distinct eigenvalues then for any $f \in \ell^2(\pi)$ and $n \in \mathbb{N}$,

$$P^n f = \sum_{k=0}^{\mathfrak{r}} \lambda_k^n \langle f, f_k^* \rangle_\pi f_k,$$

where the set $(f_k)_{k=0}^{\mathfrak{r}}$ are eigenfunctions of P associated to the eigenvalues $(\lambda_k)_{k=0}^{\mathfrak{r}}$ and form a Riesz basis of $\ell^2(\pi)$, and the set $(f_k^*)_{k=0}^{\mathfrak{r}}$ is the unique Riesz basis biorthogonal to $(f_k)_{k=0}^{\mathfrak{r}}$. For any $x, y \in \mathcal{X}$ and $n \in \mathbb{N}$, the spectral expansion of P is given by

$$P^n(x, y) = \sum_{k=0}^{\mathfrak{r}} \lambda_k^n f_k(x) f_k^*(y) \pi(y).$$

(b) $P \in \mathcal{S}(Q)$ *with* $P \overset{\Lambda}{\sim} Q$ *if and only if* $\widehat{P} \in \mathcal{S}(\widehat{Q})$ *with* $\widehat{Q} \overset{\widehat{\Lambda}}{\sim} \widehat{P}$.

(c) *Suppose that* Λ *is an unitary operator, that is,* $\Lambda^{-1} = \widehat{\Lambda}$*, where* $\widehat{\Lambda}$ *is the adjoint operator of* Λ*. Then* P *is a normal (resp. self-adjoint) operator in* $\ell^2(\pi)$ *if and only if* Q *is a normal (resp. self-adjoint) operator in* $\ell^2(\pi_Q)$.

(d) *(Lattice isomorphism) Suppose that* $\mathcal{X}$ *is a finite state space.* Λ *is an invertible Markov kernel on* $\mathcal{X}$ *with* $\Lambda^{-1} \geqslant 0$ *if and only if* $\Lambda \in \mathcal{P}$*, the set of permutation kernels. We recall that* $\Lambda \in \mathcal{P}$ *if* $\Lambda = \Lambda_\sigma := (\mathbb{1}_{y=\sigma(x)})_{x,y\in\mathcal{X}}$ *with* $\sigma : \mathcal{X} \mapsto \mathcal{X}$ *being a permutation of the state space, and note that* Λ_σ *is an unitary Markov kernel. Moreover, for any* $Q \in \mathcal{M}$*, the permutation orbit of* Q*,* $\mathcal{S}_{\mathcal{P}}(Q) = \{P \in \overline{\mathcal{M}}; P\Lambda = \Lambda Q, \Lambda \in \mathcal{P}\} \subset \mathcal{M}$*, where* $\overline{\mathcal{M}}$ *is the set of squared matrices on* $\mathcal{X}$.

(e) *Suppose that* $\mathcal{X}$ *is a finite state space and* Q *is the transition kernel of an irreducible birth-death process, then* $P \overset{\Lambda}{\sim} Q$ *if and only if* P *has real and distinct eigenvalues.*

Remark 3.2.1. We will illustrate Theorem 3.2.1 item (d) in Section 3.6, in which we look into the permutation orbit $\mathcal{S}_{\mathcal{P}}$ of four classical birth-death processes. Also, as suggested by item (c), we can generate new non-normal examples via non-unitary link from known normal Markov chains such as birth-death processes, so in the remaining of Section 3.6, we also investigate in-depth two non-unitary orbits that we call the pure birth orbit and random walk orbit.

As Theorem 3.2.1 suggests, the class $\mathcal{S}$ is highly tractable and enjoys a number of attractive properties. It will therefore be very interesting to characterize this class in terms of the one-step transition probabilities of P, which is a fundamental yet challenging issue. However, we manage to identify a set of sufficient conditions that we call the generalized monotonicity condition class $\mathcal{GMC} \subset \mathcal{S}$, generalizing the $\mathcal{MC}$ class for skip-free chains as introduced in Choi and Patie [20], such that the time-reversal $\widehat{P}$

intertwines with a birth-death chain with the link kernel Λ being related to the Siegmund kernel.

Definition 3.2.1 (The $\mathcal{GMC}$ class)**.** We say that, for some $\mathfrak{r} \geqslant 3$, $X \in \mathcal{GMC}_{\mathfrak{r}}$ if $(X, \mathbb{P}) \in \mathcal{SF}$ with $\mathcal{X} = [\![0, \mathfrak{r}]\!]$ and for every $x \in [\![0, \mathfrak{r}-1]\!]$, its time-reversal $(X, \widehat{\mathbb{P}})$ satisfies

1. (stochastic monotonicity) $\widehat{\mathbb{P}}_{x+1}(X_1 \leqslant x) \leqslant \widehat{\mathbb{P}}_x(X_1 \leqslant x)$,
2. (strict stochastic monotonicity) $\widehat{\mathbb{P}}_{x+1}(X_1 \leqslant x-1) < \widehat{\mathbb{P}}_x(X_1 \leqslant x-1), \quad x \neq \mathfrak{r}-1, \quad$ and
3. (strict stochastic monotonicity) $\widehat{\mathbb{P}}_{x+1}(X_1 \leqslant x+1) < \widehat{\mathbb{P}}_x(X_1 \leqslant x+1), \quad x \neq \mathfrak{r}-1, \quad$ and
4. (restricted downward jump) $\widehat{\mathbb{P}}_{x+1}(X_1 \leqslant x-k) = \widehat{\mathbb{P}}_x(X_1 \leqslant x-k), \quad x \neq \mathfrak{r}-1, \quad k \in [\![2, x]\!], \quad$ and
5. (restricted upward jump) $\widehat{\mathbb{P}}_{x+1}(X_1 \leqslant x+k) = \widehat{\mathbb{P}}_x(X_1 \leqslant x+k), \quad x \neq \mathfrak{r}-1, \quad k \in [\![2, \mathfrak{r}-1-x]\!]$.

Moreover, we say $X \in \mathcal{GMC}_{\mathfrak{r}}^+$ if $X \in \mathcal{GMC}_{\mathfrak{r}}$ and for every $x \in [\![0, \mathfrak{r}-1]\!]$,

6. (lazy Siegmund dual) $\widehat{\mathbb{P}}_x(X_1 \leqslant x) - \widehat{\mathbb{P}}_{x+1}(X_1 \leqslant x) \geqslant \dfrac{1}{2}$.

When there is no ambiguity of the state space, we write $\mathcal{GMC} = \mathcal{GMC}_{\mathfrak{r}}$ (resp. $\mathcal{GMC}^+ = \mathcal{GMC}_{\mathfrak{r}}^+$). Note that the upper-script of the plus sign in $\mathcal{GMC}^+$ means that this class has non-negative eigenvalues, see Remark 3.2.5 below.

Remark 3.2.2. Recall that in Choi and Patie [20], if $P \in \mathcal{MC}$, that is, P is upward skip-free and satisfies $(1), (3), (5)$, then it is clear that $\mathcal{MC} \subset \mathcal{GMC}$, as item (2) and (4) in Definition 3.2.1 are automatically satisfied since the time-reversal $\widehat{P}$ is downward skip-free.

We now give an example that illustrates the $\mathcal{GMC}$ class.

Example 3.2.1.

$$\widehat{P} = \begin{pmatrix} 0.5 & 0.3 & 0.1 & 0.1 \\ 0.2 & 0.5 & 0.2 & 0.1 \\ 0.2 & 0.1 & 0.5 & 0.2 \\ 0.2 & 0.05 & 0.25 & 0.5 \end{pmatrix}$$

has eigenvalues $1, 0.44, 0.30, 0.26$, and satisfies $(1) - (6)$ in Definition 3.2.1.

We now formally state that $\mathcal{GMC}$ is a subclass of $\mathcal{S}^M$ (recall its definition in Definition 3.1.2), and the proof can be found in Section 3.2.2.

Theorem 3.2.2. $\mathcal{GMC} \subseteq \mathcal{S}^M$.

As a first application of Theorem 3.2.1, we first recall the celebrated eigentime identity studied by Aldous and Fill [1], Cui and Mao [25] and Miclo [82]: suppose that we sample two points x and y randomly from the stationary distribution of the chain and calculate the expected hitting time from x to y, the expected value of this procedure is the sum of the inverse of the non-zero (and negative of the) eigenvalues of the generator. Since similarity preserves the eigenvalues (see Theorem 3.2.1 item (a)), we can easily see that both P and Q share the same eigentime identity:

Corollary 3.2.1 (Eigentime identity)**.** *Suppose that $\mathcal{X}$ is a finite state space and $(Q_t)_{t\geqslant 0}$ (resp. $(P_t)_{t\geqslant 0}$) has generator G (resp. L) associated with the Markov chain $(X_t)_{t\geqslant 0}$ (resp. $(Y_t)_{t\geqslant 0}$). If $L \in \mathcal{S}(G)$ with eigenvalues $(-\lambda_i)_{i\in[\![|\mathcal{X}|]\!]}$, then $(P_t)_{t\geqslant 0}$ and $(Q_t)_{t\geqslant 0}$ share the same eigentime identity. That is, denote $\tau_y^Q := \inf\{t \geqslant 0; X_t = y\}$ (resp. $\tau_y^P := \inf\{t \geqslant 0; Y_t = y\}$), then*

$$\sum_{x,y\in\mathcal{X}} \mathbb{E}_x(\tau_y^Q)\pi_Q(x)\pi_Q(y) = \sum_{x,y\in\mathcal{X}} \mathbb{E}_x(\tau_y^P)\pi(x)\pi(y) = \sum_{i=1,\lambda_i\neq 0}^{|\mathcal{X}|} \frac{1}{\lambda_i}.$$

As a second application of the spectral decomposition stated in Theorem 3.2.1, we derive accurate information regarding the speed of convergence to stationarity for ergodic chains in $\mathcal{S}$ in both the Hilbert space topology and in total variation distance. There have been a rich literature devoted to the study of convergence to equilibrium for non-reversible chains by means of reversibilizations, see e.g. Aldous and Fill [1], Fill [42], Levin et al. [74], Montenegro and Tetali [84] and the references therein. Our approach reveals a natural extension to the non-reversible case of the classical spectral gap that appears in the study of reversible chains. To state our result we now fix some notations. We denote the second largest eigenvalue in modulus (SLEM) or the spectral radius of P in the Hilbert space $\ell_0^2(\pi) = \{f \in \ell^2(\pi);\ \ \langle f, \mathbf{1}\rangle_\pi = 0\}$, by $\lambda_* = \lambda_*(P) = \sup\{|\lambda_i|;\ \lambda_i \in \sigma(P), \lambda_i \neq 1\}$, then the *absolute spectral gap* is $\gamma_* = 1 - \lambda_*$. For any two probability measures μ, ν on $\mathcal{X}$, the total variation distance between μ and ν is given by

$$||\mu - \nu||_{TV} = \frac{1}{2}\sum_{x \in \mathcal{X}} |\mu(x) - \nu(x)|.$$

For $n \in \mathbb{N}$, the total variation distance from stationarity of X is

$$d(n) = \max_{x \in \mathcal{X}} ||\delta_x P^n - \pi||_{TV}.$$

For $g \in \ell^2(\pi)$, the mean of g with respect to π can be written as $\mathbb{E}_\pi(g) = \langle g, \mathbf{1}\rangle_\pi$. Similarly, the variance of g with respect to π is $\mathrm{Var}_\pi(g) = \langle g, g\rangle_\pi - \mathbb{E}_\pi^2(g)$. Finally, we recall that Fill in Fill [42, Theorem 2.1] obtained in the finite state space case the following bound valids for all $n \in \mathbb{N}_0$

$$d(n) \leqslant \frac{\sigma_*(P)^n}{2}\sqrt{\frac{1 - \pi_{min}}{\pi_{min}}}, \tag{3.2.2}$$

where $\pi_{min} = \min_{x \in \mathcal{X}} \pi(x)$ and $\sigma_*(P) = \sqrt{\lambda_*(P\widehat{P})}$ is the second largest singular value of P. We obtain the following refinement for Markov chains in the class $\mathcal{S}$. The proof is deferred to Section 3.2.3.

Corollary 3.2.2. *Let $P \in \mathcal{S}$ on the finite state space $\mathcal{X} = [\![0, \mathfrak{r}]\!]$ with $\mathfrak{r} < \infty$ and invariant distribution π, that is, $\pi P = \pi$.*

1. *For any $n \in \mathbb{N}_0$, we have*

$$\lambda_*^n \leqslant \|P^n - \pi\|_{\ell^2(\pi) \to \ell^2(\pi)} \leqslant \sigma_*^n(P) \mathbb{1}_{\{n < n^*\}} + \kappa(\Lambda) \lambda_*^n \mathbb{1}_{\{n \geqslant n^*\}}, \tag{3.2.3}$$

   where $n^* = \lceil \frac{\ln \kappa(\Lambda)}{\ln \sigma_*(P) - \ln \lambda_*} \rceil$ and $\kappa(\Lambda) = \|\Lambda\|_{op} \|\Lambda^{-1}\|_{op} \geqslant 1$ is the condition number of Λ. A sufficient condition for which $\lambda_* < \sigma_*(P)$ is given by $\max_{i \in \mathcal{X}} P(i,i) > \lambda_*$. In such case, for n large enough, the convergence rate λ_* given (3.2.3) is strictly better than the reversibilization rate $\sigma_*(P)$.

2. *For any $n \in \mathbb{N}_0$,*

$$d(n) \leqslant \frac{\min\left(\sigma_*^n(P), \kappa(\Lambda)\lambda_*^n\right)}{2} \sqrt{\frac{1 - \pi_{min}}{\pi_{min}}},$$

   where $\lambda_* \leqslant \sigma_*(P)$.

Remark 3.2.3. Recall that when P is reversible and compact then the sequence of eigenfunctions is orthonormal and thus an application of the Parseval identity yields the well-known result (see e.g. Chen and Saloff-Coste [12, Section 4.3]) $\|P^n - \pi\|_{\ell^2(\pi) \to \ell^2(\pi)} = \lambda_*^n$ and $\kappa(\Lambda) = 1$ which is a specific instance of item (1).

Remark 3.2.4. We also recall the discrete analogue of the notion of hypocoercivity introduced in Villani [109], i.e. there exists a constant $C < \infty$ and $\rho \in (0,1)$ such that, for all $n \in \mathbb{N}$,

$$\|P^n - \pi\|_{\ell^2(\pi) \to \ell^2(\pi)} \leqslant C\rho^n.$$

Note that, in general, these constants are not known explicitly. We observe that the upper bound in (3.2.3) reveals that the ergodic chains in $\mathcal{S}$ satisfy this hypocoercivity phenomena. More interestingly, our approach based on the similarity concept enables us to get on the one hand an explicit and on the other hand a spectral interpretation of this rate of convergence. Indeed, it can be understood as a modified spectral gap where the perturbation from the classical spectral gap is given by the condition number $\kappa(\Lambda)$ which can be interpreted as a measure of deviation from symmetry. In this vein, we mention the recent work Patie and Savov [85] where a similar spectral interpretation of the hypocoercivity phenomena is given for a class of non-self-adjoint Markov semi-groups, and the related work of Baxendale [4] which also gives computable C and ρ by renewal theory in the notion of geometric ergodicity measured in the L^∞_V norm.

3.2.1 Proof of Theorem 3.2.1

We first show the item (a). Since $\mathcal{E}$ is a spectral measure, it follows easily that $\{F_B = \Lambda E_B \Lambda^{-1};\ B \in \mathcal{B}(\mathcal{C})\}$ is a spectral measure. The fact that the spectrum coincides and

$$\sigma(P) = \sigma(Q), \sigma(P) = \overline{\sigma(\widehat{P})}, \sigma_c(P) = \sigma_c(Q), \sigma_p(P) = \sigma_p(Q), \sigma_r(P) = \sigma_r(Q),$$

follows from Proposition 2.4 in Inoue and Trapani [55]. Define $\overline{P} := \int_{\sigma(P)} \lambda\, dF_\lambda$. We have

$$\overline{P} = \int_{\sigma(P)} \lambda\, d(\Lambda^{-1} E_\lambda \Lambda) = \Lambda^{-1} \left(\int_{\sigma(Q)} \lambda\, dE_\lambda \right) \Lambda = \Lambda^{-1} Q \Lambda = P,$$

so the desired spectral resolution of P follows, thus it is a spectral scalar-type operator. The spectral resolution of $\widehat{P}$ follows from that of P. The functional calculus of P follows immediately from that of spectral scalar-type operator, see e.g. Theorem 1 in Chapter $XV.5$, Page 1941 of Dunford and Schwartz [37]. We proceed to handle the case when P is compact. To see that (f_k) and (f_m^*) are biorthogonal, we note that the fact that P has distinct eigenvalues yields that $\langle f_k, f_m^* \rangle_\pi = \delta_{k,m}$ for any k, m. Next, denote (g_k) to be the (orthogonal) eigenfunctions of the normal transition kernel Q. Since $f_k = \Lambda g_k$ and Λ is bounded, (f_k) is complete as (g_k) is a basis. As Λ is bounded from above and below, for any $n \in \mathbb{N}$ and arbitrary sequence $(c_k)_{k=1}^n$, we have

$$A\sum_{k=1}^{n}|c_k|^2 \leqslant \left\|\sum_{k=1}^{n}c_kf_k\right\|_{\pi}^2 = \left\|\Lambda\sum_{k=1}^{n}c_kg_k\right\|_{\pi}^2 \leqslant B\sum_{k=1}^{n}|c_k|^2,$$

where we can take $A = \|\Lambda^{-1}\|^{-2}$ and $B = \|\Lambda\|^2$, so (3.2.1) is satisfied. It follows from Young [114, Theorem 9] that there exists the sequence (f_k^*) being the unique Riesz basis biorthogonal to $(f_k)_{k=0}^{\mathfrak{r}}$, and, any $f \in \ell^2(\pi)$ can be written as

$$f = \sum_{k=0}^{\mathfrak{r}} c_kf_k,$$

where $c_k = \langle f, f_k^*\rangle_\pi$. Desired result follows by applying P^n to f and using $P^nf_k = \lambda_k^nf_k$. In particular, if we take $f = \delta_y$, the Dirac mass at y, and evaluate the resulting expression at x, we obtain the spectral expansion of P. Next, we show item (b). If $P \overset{\Lambda}{\sim} Q$, then for $f \in \ell^2(\pi_Q)$ and $g \in \ell^2(\pi)$,

$$\langle f, \widehat{\Lambda}\widehat{P}g\rangle_{\pi_Q} = \langle P\Lambda f, g\rangle_\pi = \langle \Lambda\widehat{Q}f, g\rangle_\pi = \langle f, \widehat{Q}\widehat{\Lambda}g\rangle_{\pi_Q},$$

which shows that $\widehat{Q} \overset{\widehat{\Lambda}}{\sim} \widehat{P}$. The opposite direction can be shown similarly. For item (c). Since Λ is unitary, the spectral measures of P and Q are related by $F_B = \Lambda E_B\widehat{\Lambda}$, so F_B is self-adjoint if and only if E_B is self-adjoint, which implies that P is normal if and only if Q is normal. If Q is self-adjoint, then item (b) yields $P \overset{\Lambda}{\sim} Q$ if and only if $Q \overset{\Lambda^{-1}}{\sim} \widehat{P}$, which implies that $\widehat{P} = P$ in $\ell^2(\pi)$. The opposite direction can be shown similarly. Next, we show item (d). If Λ is a permutation link, then it is trivial to see that Λ is an invertible Markov kernel. For the opposite direction, it is known (see e.g. Berman and Plemmons [5, Section 3]) that $\Lambda = D\Lambda_\sigma$, where D is a diagonal matrix. We then have $\mathbf{1} = \Lambda\mathbf{1} = D\Lambda_\sigma\mathbf{1} = D\mathbf{1}$, which gives $D = I$, and hence $\Lambda = \Lambda_\sigma$. Let now $Q \in \mathcal{M}$ and $P \in \mathcal{S}_{\mathcal{P}}(Q)$, then since $P = \Lambda Q\Lambda^{-1}$ with $\Lambda, \Lambda^{-1} \in \mathcal{P}$, we deduce readily that $P \in \mathcal{M}$. Finally, to show item (e), if $P \overset{\Lambda}{\sim} Q$, then P has real and distinct eigenvalues since Q has real and distinct eigenvalues. Conversely, if P has real and distinct eigenvalues, P is diagonalizable, so there exists an invertible Λ such that

$$P = \Lambda D\Lambda^{-1}.$$

where D is the diagonal matrix storing the eigenvalues of P. Given the spectral data D, by inverse spectral theorem, see e.g. Dym and McKean [38, Section 5.8], one can always construct an ergodic Markov chain with transition matrix Q such that

$$Q = VDV^{-1}.$$

3.2.2 Proof of Theorem 3.2.2

We write $\widetilde{P}$ the so-called Siegmund dual (or H_S-dual) of $\widehat{P}$. That is, $\widetilde{P}^T = H_S^{-1}\widehat{P}H_S$ where $H_S = (H_S(x,y))_{x,y\in\mathcal{X}}$ is defined to be $H_S(x,y) = \mathbb{1}_{\{x\leqslant y\}}$ and its inverse $H_S^{-1} = (H_S^{-1}(x,y))_{x,y\in\mathcal{X}}$ is $H_S^{-1}(x,y) = \mathbb{1}_{\{x=y\}} - \mathbb{1}_{\{x=y-1\}}$, see Siegmund [102]. Since $X \in \mathcal{GMC}$, then $\widehat{P}$ is stochastically monotone, hence from Asmussen [3, Proposition 4.1], we have that $\widetilde{P}$ is a sub-Markovian kernel. For $x \in [\![0, \mathfrak{r}-2]\!]$, condition 2 and 3 in $\mathcal{GMC}$ yield, respectively, $\widetilde{p}(x, x+1) > 0$, while for $x \in [\![1, \mathfrak{r}-1]\!]$, we have $\widetilde{p}(x, x-1) > 0$. Condition 4 and 5 in $\mathcal{GMC}$ guarantee that $\widetilde{p}(x,y) = 0$ for each $x \in [\![0, \mathfrak{r}-3]\!]$ and $y \in [\![x+2, \mathfrak{r}-1]\!]$ and for each $x \in [\![2, \mathfrak{r}-1]\!]$ and $y \in [\![0, x-2]\!]$. That is, $\widetilde{P}$ is a (strictly substochastic) irreducible birth-death chain when restricted to the state space $[\![0, \mathfrak{r}-1]\!]$. Denote $\widetilde{P}^{bd}$ the restriction of $\widetilde{P}$ to $[\![0, \mathfrak{r}-1]\!]$. By breaking off the last row and last column of $\widetilde{P}$, we can write

$$\widetilde{P} = \begin{pmatrix} \widetilde{P}^{bd} & \mathbf{v} \\ \mathbf{0} & 1 \end{pmatrix} = (H_S^{-1}\widehat{P}H_S)^T, \tag{3.2.4}$$

where $\mathbf{0}$ is a row vector of zero, and $\mathbf{v}$ is a column vector storing $\widetilde{p}(x, \mathfrak{r})$ for $x \in [\![0, \mathfrak{r}-1]\!]$. Considering the h-transform of $\widetilde{P}$ with $h = H_S^T\pi > \mathbf{0}$, see e.g. Huillet and Martinez [51, Theorem 2], we see that $X \in \mathcal{S}$, as we have

$$P\Lambda = \Lambda Q,$$

where $\Lambda = (H_S^T D_\pi)^{-1}$ (D_π is the diagonal matrix of π) and $Q = \widetilde{P}$, which completes the proof of Theorem 3.2.2.

Remark 3.2.5. Note that condition (6) in $\mathcal{GMC}^+$ guarantees that Q is a lazy chain, that is $Q(x,x) \geqslant 1/2$ for all $x \in \mathcal{X}$, and hence the class $\mathcal{GMC}^+$ possesses non-negative eigenvalues.

3.2.3 Proof of Corollary 3.2.2

We first show the upper bound in item (1). Define the synthesis operator $T^* : \ell^2 \to \ell^2(\pi)$ by $\alpha = (\alpha_i) \mapsto T^*(\alpha) = \sum_{i=0}^{\mathfrak{r}} \alpha_i f_i$, where (f_i) are the eigenfunctions of P and (f_i^*) are the unique biorthogonal basis of (f_i) as in Theorem 3.2.1. For $1 \leqslant i \leqslant \mathfrak{r}$, we take $\alpha_i = \lambda_i^n \langle g, f_i^* \rangle_\pi$, and denote (q_i) to be the orthonormal eigenfunctions of Q, where $f_i = \Lambda q_i$. Note that $||T^*||_{op} \leqslant ||\Lambda||_{op} < \infty$, since

$$||T^*(\alpha)|| = \left\| \sum_{i=0}^{\mathfrak{r}} \alpha_i \Lambda q_i \right\| \leqslant \|\Lambda\|_{op} \left\| \sum_{i=0}^{\mathfrak{r}} \alpha_i q_i \right\|_{\pi_Q} \leqslant ||\Lambda||_{op} ||\alpha||_{\ell^2}.$$

For $g \in \ell^2(\pi)$, we also have

$$\sum_{i=0}^{\mathfrak{r}} |\langle g, f_i^* \rangle_\pi|^2 = \sum_{i=0}^{\mathfrak{r}} |\langle g, (\Lambda^*)^{-1} q_i \rangle_\pi|^2 = \sum_{i=0}^{\mathfrak{r}} |\langle \Lambda^{-1} g, q_i \rangle_{\pi_Q}|^2 = ||\Lambda^{-1} g||_{\pi_Q}^2 \leqslant ||\Lambda^{-1}||_{op}^2 ||g||_\pi^2,$$

where the third equality follows from Parseval's identity, which leads to

$$||P^n g - \pi g||_\pi^2 = ||T^*(\alpha)||^2 \leqslant ||\Lambda||_{op}^2 ||\alpha||_{l^2}^2 \leqslant ||\Lambda||_{op}^2 ||\Lambda^{-1}||_{op}^2 \lambda_*^{2n} ||g||_\pi^2. \qquad (3.2.5)$$

Desired upper bound follows from (3.2.5) and

$$||P^n - \pi||_{\ell^2(\pi) \to \ell^2(\pi)} \leqslant \lambda_*(\widehat{P} P)^{n/2} = \lambda_*(P \widehat{P})^{n/2},$$

see e.g. Fill [42]. The lower bound in (1) follows readily from the well-known result that the n^{th} power of the spectral radius λ_*^n is less than or equal to the norm of P^n on the reduced space $\ell_0^2(\pi)$. For the sufficient condition in item (1), that is, $\max_{i\in\mathcal{X}} P(i,i) > \lambda_*$ implies $\lambda_* < \sigma_*(P)$, it is a straightforward consequence of the Sing-Thompson Theorem, see Thompson [107]. Next, using (3.2.5), we get

$$\mathrm{Var}_\pi\left(\widehat{P}^n g\right) \leqslant \kappa(\widehat{\Lambda})^2\lambda_*^{2n}\mathrm{Var}_\pi(g) = \kappa(\Lambda)^2\lambda_*^{2n}\mathrm{Var}_\pi(g), \quad n \in \mathbb{N}_0, \tag{3.2.6}$$

where we used the obvious identity $\kappa(\Lambda) = \kappa(\widehat{\Lambda})$ in the equality. This leads to

$$\begin{aligned}||\delta_x P^n - \pi||_{TV}^2 &= \frac{1}{4}\mathbb{E}_\pi^2\left|\frac{\delta_x P^n}{\pi} - 1\right| \leqslant \frac{1}{4}\mathrm{Var}_\pi\left(\frac{\delta_x P^n}{\pi}\right) = \frac{1}{4}\mathrm{Var}_\pi\left(\widehat{P}^n\frac{\delta_x}{\pi}\right)\\ &\leqslant \frac{1}{4}\kappa(\Lambda)^2\lambda_*^{2n}\mathrm{Var}_\pi\left(\frac{\delta_x}{\pi}\right) = \frac{1}{4}\kappa(\Lambda)^2\lambda_*^{2n}\frac{1-\pi(x)}{\pi(x)} \leqslant \frac{1}{4}\kappa(\Lambda)^2\lambda_*^{2n}\frac{1-\pi_{min}}{\pi_{min}},\end{aligned}$$

where the first inequality follows from Cauchy-Schwartz inequality. The proof is completed by combining the above bound with (3.2.2).

3.3 Separation cutoff

In this Section, we investigate the separation cutoff phenomenon for the $\mathcal{GMC}$ class. For birth-death chains, they have been studied in Diaconis and Saloff-Coste [31] and Chen and Saloff-Coste [13] while it has recently been extended to upward skip-free chains by Mao et al. [76] and Choi and Patie [20]. Recall that in Theorem 3.2.2 we have shown that $\mathcal{GMC} \subset \mathcal{S}^M$. In order to establish the famous "spectral gap times mixing time" criteria for this class, we will build upon the result of Fill [43] to first analyze the fastest strong stationary time of this class, followed by demonstrating that the proof in Diaconis and Saloff-Coste [31] carries over for this class of non-reversible chains.

We now proceed to discuss the main results of this Section, with Theorem 3.3.1 addressing the case of discrete time family of Markov chains and Theorem 3.3.2 discussing the continuized version. Recall that the notation $\mathcal{GMC}^+$ introduced in Definition 3.2.1

represents the generalized monotonicity class with *non-negative* eigenvalues. This is an important subclass since the eigenvalues of the transition kernel (resp. negative of the generator) are the parameters in the geometric distribution (resp. exponential distribution) of the fastest strong stationary time in Theorem 3.3.1 (resp. Theorem 3.3.2).

Theorem 3.3.1. *For $n \geqslant 1$, suppose that $P_n \in \mathcal{GMC}^+_{\mathfrak{r}_n}$ on the state space $\mathcal{X}_n = [\![0, \mathfrak{r}_n]\!]$ that started at 0. Let $(\theta_{n,i})_{i=1}^{\mathfrak{r}_n}$ be the eigenvalues of $I - P_n$, and $(c_{n,i})_{i=0}^{\mathfrak{r}_n}$ to be the mixture weights of the n^{th} chain defined in (3.3.1) in Lemma 3.3.1. Define*

$$w_{n,i} := \sum_{j \geqslant i}^{\mathfrak{r}_n} c_{n,j}\,, \quad t_n := \sum_{i=1}^{\mathfrak{r}_n} \frac{w_{n,i}}{\theta_{n,i}}\,, \quad \underline{\theta}_n := \min_{1 \leqslant i \leqslant \mathfrak{r}_n} \theta_{n,i}\,, \quad \rho_n^2 := \sum_{i=1}^{\mathfrak{r}_n} w_{n,i}^2 \frac{1-\theta_{n,i}}{\theta_{n,i}^2}\,.$$

Then this family has a separation cutoff if and only if $t_n \underline{\theta}_n \to \infty$. Furthermore, if $t_n \underline{\theta}_n \to \infty$, then there is a $(t_n, \max\{\rho_n, 1\})$ separation cutoff.

Remark 3.3.1. For discrete-time stochastically monotone birth-death chains which start at 0, we have $w_i = 1$ for $i \in [\![1, \mathfrak{r}_n]\!]$ and $c_{n,0} = 0$, and hence we recover Diaconis and Saloff-Coste [31, Theorem 5.2].

Theorem 3.3.2. *For $n \geqslant 1$, suppose that $L_n = P_n - I$ is the infinitesimal generator with $P \in \mathcal{GMC}^+_{\mathfrak{r}_n}$ on the state space $\mathcal{X}_n = [\![0, \mathfrak{r}_n]\!]$ that started at 0. Let $(\theta_{n,i})_{i=1}^{\mathfrak{r}_n}$ be the eigenvalues of $-L_n$, and $(c_{n,i})_{i=0}^{\mathfrak{r}_n}$ to be the mixture weights defined in (3.3.2) in Remark 3.3.2. Define*

$$w_{n,i} := \sum_{j \geqslant i}^{\mathfrak{r}_n} c_{n,j}\,, \quad t_n := \sum_{i=1}^{\mathfrak{r}_n} \frac{w_{n,i}}{\theta_{n,i}}\,, \quad \underline{\theta}_n := \min_{1 \leqslant i \leqslant \mathfrak{r}_n} \theta_{n,i}\,, \quad \rho_n^2 := \sum_{i=1}^{\mathfrak{r}_n} \frac{w_{n,i}^2}{\theta_{n,i}^2}\,.$$

Then this family has a separation cutoff if and only if $t_n \underline{\theta}_n \to \infty$. Furthermore, if $t_n \underline{\theta}_n \to \infty$, then there is a (t_n, ρ_n) separation cutoff.

We will only prove Theorem 3.3.1 as the proof of Theorem 3.3.2 is very similar and thus omitted.

3.3.1 Proof of Theorem 3.3.1

Following the plan as outlined above in Section 3.3, we first analyze the distribution of the fastest strong stationary time of the class $\mathcal{GMC}^+$ in Lemma 3.3.1, followed by detailing the proof of Theorem 3.3.1.

Lemma 3.3.1. *Suppose that X is an ergodic Markov chain on the state space $\mathcal{X} = [\![0, \mathfrak{r}]\!]$ (and $\mathfrak{r} \geqslant 3$) with transition matrix P and stationary distribution π which starts at 0. If $P \in \mathcal{GMC}^+$, then the fastest strong stationary time is distributed as the $\mathbf{c}$-mixture of convolution of geometric $\sum_{k=1}^{\mathfrak{r}} c_k \mathcal{G}(\lambda_1, \ldots, \lambda_k)$, where $i, j, k \in [\![0, \mathfrak{r}]\!]$,*

$$Q_k := \frac{(P - \lambda_1 I) \ldots (P - \lambda_k I)}{(1 - \lambda_1) \ldots (1 - \lambda_k)}, \quad \Lambda(i, j) := Q_i(0, j), \quad c_k := \frac{\Lambda(k, \mathfrak{r}) - \Lambda(k-1, \mathfrak{r})}{\pi(\mathfrak{r})}, \tag{3.3.1}$$

and $\{\lambda_k\}_{k=1}^{\mathfrak{r}}$ are the non-unit eigenvalues of P.

Proof. Suppose that $P\Lambda = \Lambda Q$. In view of Fill [43] Theorem 5.2, it suffices to show that the $c_k \geqslant 0$. First, we show that $(Q - \lambda_1 I) \ldots (Q - \lambda_k I)$ are non-negative matrices, where Q is the Siegmund dual of $\widehat{P}$. We will prove this via induction on k. For $k = 1$, thanks to Micchelli and Willoughby [79, Theorem 3.2], we have $Q^{BD} - \lambda_1 I \geqslant \mathbf{0}$, which leads to

$$Q - \lambda_1 I = \begin{pmatrix} Q^{BD} - \lambda_1 I & \mathbf{h} \\ \mathbf{0}^T & 1 - \lambda_1 \end{pmatrix} \geqslant \mathbf{0}.$$

Suppose that

$$\prod_{i=1}^{k} (Q - \lambda_i I) = \begin{pmatrix} \prod_{i=1}^{k} (Q^{BD} - \lambda_i I) & \mathbf{n} \\ \mathbf{0}^T & \prod_{i=1}^{k} (1 - \lambda_i) \end{pmatrix} \geqslant \mathbf{0},$$

where $\mathbf{n} \geqslant \mathbf{0}$ is a non-negative vector. Therefore,

$$\prod_{i=1}^{k+1} (Q - \lambda_i I) = \begin{pmatrix} \prod_{i=1}^{k+1} (Q^{BD} - \lambda_i I) & \prod_{i=1}^{k} (Q^{BD} - \lambda_i I)\mathbf{h} + (1 - \lambda_{k+1})\mathbf{n} \\ \mathbf{0}^T & \prod_{i=1}^{k+1} (1 - \lambda_i) \end{pmatrix} \geqslant \mathbf{0},$$

which completes the induction. Define

$$Z_k := H_S^T \prod_{i=1}^{k} \frac{Q - \lambda_i I}{1 - \lambda_i} (H_S^T)^{-1} .$$

Note that $P = D_\pi^{-1} H_S^T Q (H_S^T)^{-1} D_\pi$, so $c_k \geqslant 0$ if and only if $Z_k(0, \mathfrak{r}) - Z_{k-1}(0, \mathfrak{r}) \geqslant 0$ if and only if (here we make use of H_S^T)

$$\left(\prod_{i=1}^{k} \frac{Q - \lambda_i I}{1 - \lambda_i} \right) (0, \mathfrak{r}) - \left(\prod_{i=1}^{k-1} \frac{Q - \lambda_i I}{1 - \lambda_i} \right) (0, \mathfrak{r}) = \left(\prod_{i=1}^{k} (Q^{BD} - \lambda_i I) \mathbf{h} \right) (0) \geqslant 0 ,$$

which is true. □

When we have a handle on the fastest strong stationary time, we can then analyze the separation cutoff phenomenon, and the rest of the proof follow the Chebyshev inequality framework introduced by Diaconis and Saloff-Coste [31]. More precisely, denote P_n^k to be the distribution of the n^{th} chain at time k, π_n to be the stationary measure and T_n to be the fastest strong stationary time of the n^{th} chain. We note that $\mathbb{E}(T_n) = t_n$ and $\mathrm{Var}(T_n) = \rho_n^2$. The key to the proof is the following:

$$\rho_n^2 = \underline{\theta}_n^{-2} \sum_{i=1}^{\mathfrak{r}_n} w_{n,i}^2 \frac{(1 - \theta_{n,i}) \, \underline{\theta}_n^2}{\theta_{n,i}^2} \leqslant \underline{\theta}_n^{-2} \sum_{i=1}^{\mathfrak{r}_n} w_{n,i} \frac{\underline{\theta}_n}{\theta_{n,i}} = \underline{\theta}_n^{-1} t_n ,$$

where we use $\theta_{n,i} \geqslant 0$, $\underline{\theta}_n / \theta_{n,i} \leqslant 1$ and $w_i \leqslant 1$ in the first inequality. The rest of the proof follows as that of Choi and Patie [20, Theorem 8.1], which does not require reversibility of the chain.

Remark 3.3.2. The corresponding result of Lemma 3.3.1 in the continuous-time setting is stated in the following in order to prove Theorem 3.3.2. Suppose that X is a continuous-time ergodic Markov chain on the state space $\mathcal{X} = [\![0, \mathfrak{r}]\!]$ (and $\mathfrak{r} \geqslant 3$) with generator $L = P - I$ and stationary distribution π which starts at 0. If $P \in \mathcal{GMC}^+$, then the fastest strong stationary time is distributed as the $\mathbf{c}$-mixture of convolution of exponential $\sum_{k=1}^{\mathfrak{r}} c_k \mathcal{E}(\theta_1, \ldots, \theta_k)$, where $i, j, k \in [\![0, \mathfrak{r}]\!]$,

$$Q_k := \frac{(L+\theta_1 I)\ldots(L+\theta_k I)}{\theta_1 \ldots \theta_k}, \quad \Lambda(i,j) := Q_i(0,j), \quad c_k := \frac{\Lambda(k,\mathfrak{r}) - \Lambda(k-1,\mathfrak{r})}{\pi(\mathfrak{r})}, \tag{3.3.2}$$

and $\{\theta_k\}_{k=1}^{\mathfrak{r}}$ are the non-zero eigenvalues of $-L$.

3.4 L^2-cutoff

The aim of this Section is to investigate the spectral criterion for the existence of L^2-cutoff for the class of Markov chains in a continuous-time setting with generator L and similarity on the generator level. That is, in the notation of Definition 3.1.1 and 3.1.2, $L \in \mathcal{S}(G)$, where G is a reversible generator. We denote the spectral gap $\lambda = \lambda(L)$ of L by

$$\lambda = \inf\{\langle -Lf, f\rangle_\pi;\ f \in \mathrm{Dom}(L), \text{real valued}, \mathbb{E}_\pi(f) = 0, \mathbb{E}_\pi(f^2) = 1\}.$$

This follows and generalizes the work of Chen and Saloff-Coste [11, 12], Chen et al. [15] who studied the L^2-cutoff phenomena in the context of normal Markov processes. We proceed to provide a quick review on the notion of L^2-cutoff.

Definition 3.4.1. For $n \geqslant 1$, let $g_n : [0,\infty) \mapsto [0,\infty]$ be a non-increasing function vanishing at infinity. Assume that

$$M = \limsup_{n\to\infty} g_n(0) > 0.$$

Then the family $\mathcal{G} = \{g_n : n \geqslant 1\}$ is said to have

1. A *cutoff* if there exists a sequence of positive numbers t_n, known as the cutoff time, such that for $\epsilon \in (0,1)$,

$$\lim_{n\to\infty} g_n((1+\epsilon)t_n) = 0, \quad \lim_{n\to\infty} g_n((1-\epsilon)t_n) = M.$$

2. A (t_n, b_n)-*cutoff* if $t_n > 0$, $b_n > 0$, where b_n is known as the cutoff window, $b_n = o(t_n)$ and

$$\lim_{c\to\infty} \limsup_{n\to\infty} g_n(t_n + cb_n) = 0, \qquad \lim_{c\to-\infty} \liminf_{n\to\infty} g_n(t_n - cb_n) = M.$$

If $\eta P_t \ll \pi$ with density $f(t, \eta, \cdot)$, then the chi-squared distance is given by

$$D_2(\eta, t)^2 = \int_{\mathcal{X}} |f(t, \eta, x) - 1|^2 \, \pi(dx).$$

Suppose that we have a family of measurable spaces $(\mathcal{X}_n, \mathcal{B}_n)_{n\in\mathbb{N}}$. For $n \in \mathbb{N}$, we denote $p_n(t, \eta_n, \cdot)$ defined on $(\mathcal{X}_n, \mathcal{B}_n)$ to be the transition function with initial probability law $\eta_n \ll \pi_n$ and $t \geqslant 0$. We denote f_n to be the L^2-density of η_n with respect to π_n. The family $\{p_n(t, \eta_n, \cdot) : t \in [0, \infty)\}$ has an L^2-cutoff (resp. (t_n, b_n) L^2-cutoff) if $\{g_n(t) = D_{n,2}(\eta_n, t) : n \geqslant 1\}$ has a cutoff (resp. (t_n, b_n)-cutoff) as in Definition 3.4.1, where $D_{n,2}(\eta_n, t)$ is the chi-squared distance of the n^{th} chain.

Our main result in Theorem 3.4.1 gives the spectral criterion for L^2-cutoff to the family of process with $L_n \in \mathcal{S}(G_n)$, where G_n is a reversible generator. We denote the (non-self-adjoint) spectral measure of L_n of the n^{th} chain by $F_{n,B}$ for $B \in \mathcal{B}(\mathbb{C})$, and $H_{n,B} = F_{n,B}F_{n,B}^*$. We use the following notation: for $\delta, C > 0$ and $B \in \mathcal{B}(\mathbb{C})$, we set

$$\begin{aligned}
V_n(B) &= \langle H_{n,B} f_n, f_n \rangle_{\pi_n}, \\
t_n(\delta) &= \inf\{t : D_{n,2}(\eta_n, t) \leqslant \delta\}, \\
\lambda_n(C) &= \inf\{\lambda : V_n([\lambda_n, \lambda]) > C\}, \\
\tau_n(C) &= \sup\left\{ \frac{\log(1 + V_n([\lambda_n, \lambda]))}{2\lambda} : \lambda \geqslant \lambda_n(C) \right\}, \\
\gamma_n &= \lambda_n(C)^{-1}, \\
b_n &= \lambda_n(C)^{-1} \log(\lambda_n(C)\tau_n(C)).
\end{aligned}$$

Theorem 3.4.1. *Suppose that $L_n \in \mathcal{S}(G_n)$ for each member in the family $\{p_n(t, \eta_n, \cdot)$:*

$t \in [0,\infty)\}$, where G_n is a reversible generator. If $\pi_n(f_n^2) \to \infty$, then the following are equivalent.

1. *$\{p_n(t,\eta_n,\cdot) : t \in [0,\infty)\}$ has an L^2-cutoff.*

2. *For some positive constants C, δ, ϵ,*

$$\lim_{n\to\infty} t_n(\delta)\lambda_n(C) = \infty, \quad \lim_{n\to\infty} \int_{[\lambda_n,\lambda_n(C)]} e^{-\epsilon\gamma t_n(\delta)} dV_n(\gamma) = 0.$$

3. *For some positive constants C, ϵ,*

$$\lim_{n\to\infty} \tau_n(C)\lambda_n(C) = \infty, \quad \lim_{n\to\infty} \int_{[\lambda_n,\lambda_n(C)]} e^{-\epsilon\gamma \tau_n(C)} dV_n(\gamma) = 0.$$

If (2) (resp. (3)) holds, then $\{p_n(t,\eta_n,\cdot) : t \in [0,\infty)\}$ has a $(t_n(\delta),\gamma_n)$ L^2-cutoff (resp. $(\tau_n(C), b_n)$ L^2-cutoff).

3.4.1 Proof of Theorem 3.4.1

To prove Theorem 3.4.1, it relies on the following lemma that relates the chi-squared distance to the spectral decomposition of the infinitesimal generator $-L$, which allows us to connect with the Laplace transform of the spectral measure $H_B = F_B F_B^*$.

Lemma 3.4.1. *Let X be a Markov process with $X_0 \sim \eta$, generator $L \in \mathcal{S}(G)$, where G is a reversible generator, such that $\eta \ll \pi$ with $L^2(\pi)$-density f and spectral gap $\lambda > 0$. Denote $\{F_B : B \in \mathcal{B}(\mathbb{R})\}$ to be the non-self-adjoint spectral measure for $-L$, and we define, for $B \in \mathcal{B}(\mathbb{R})$,*

$$H_B = F_B F_B^*.$$

Then, for $t \geqslant 0$,

$$D_2(\eta,t)^2 = \int_{[\lambda,\infty)} e^{-2\gamma t}\, d\langle H_\gamma f, f\rangle_\pi \,.$$

Proof. By the definition of chi-square distance D_2 and $\pi(f) = 1$, we have

$$D_2(\eta, t)^2 = \left\| \widehat{P}_t f - \pi(f) \right\|_\pi^2 = \left\| \widehat{P}_t f \right\|_{L_0^2(\pi)}^2 = \int_{[\lambda,\infty)} e^{-2\gamma t} \, d\langle H_\gamma f, f\rangle_\pi,$$

where the last equality follows from Inoue and Trapani [55, Lemma 3.19]. □

Lemma 3.4.1 reveals that the problem of L^2-cutoff reduces to the cutoff phenomenon of the Laplace transform. We proceed to complete the proof of Theorem 3.4.1. By Lemma 3.4.1, we take $g_n(t) = D_{n,2}(\eta_n, t)$ in Definition 3.4.1, and the desired result follows from the Laplace transform cutoff criteria in Theorem 3.5 of Chen and Saloff-Coste [12].

3.5 Non-asymptotic estimation error bounds for integral functionals

In this Section, we would like to estimate integral functionals of the type

$$\Gamma_T(f) = \int_0^T f(X_t) \, dt, \quad T \geqslant 0,$$

where T is a fixed time and f is a function such that the integral $\Gamma_T(f)$ is well-defined. This follows the line of work of Altmeyer and Chorowski [2], who studied the same problem with the assumption that the infinitesimal generator of the Markov process is a normal operator. This type of integral functionals appear in a number of applications. For instance, if we take $f = \mathbb{1}_B$, the indicator function of the Borel set B, then $\Gamma_T(f)$ is the occupation time of the process in B. As another example, it is not hard to see that such functional appears in the study of path-dependent derivatives in mathematical finance, see e.g. Chesney et al. [19]. In practice however, one often only have access to a sample path of the Markov process at discrete time point. A natural estimator for $\Gamma_T(f)$, known as the Riemann-sum estimator, is given by

$$\hat{\Gamma}_{T,n}(f) = \sum_{k=1}^{n} f(X_{(k-1)\Delta_n})\Delta_n,$$

where we observe $(X_t)_{t\in[0,T]}$ at discrete epochs $t = (k-1)\Delta_n$ with $k \in [\![n]\!]$ and $\Delta_n = T/n$, with the idea that we approximate $\Gamma_T(f)$ by its Riemann-sum.

For a stationary Markov process and $f \in L^2(\pi)$, both $\Gamma_T(f)$ and $\hat{\Gamma}_{T,n}(f)$ are π-a.s. defined everywhere in $L^2(\mathbb{P})$. If $L \in \mathcal{S}(G)$, we identify by Riesz theorem a linear self-adjoint operator A such that for $f, g \in L^2(\pi)$,

$$\langle Af, g\rangle_\pi = \int_{\sigma(L)} |\lambda|^2 \, d\langle H^*_\lambda f, g\rangle_\pi,$$

where we recall $H^*_\lambda = F^*_\lambda F_\lambda$ is a self-adjoint spectral measure and F_λ is the spectral measure of $-L$. For $s \geqslant 0$, we define the space $\mathcal{D}^s(A) = \mathrm{Dom}(A^s) \subset L^2(\pi)$ by functional calculus on A with the seminorm $\|f\|_{\mathcal{D}^s(A)} = \left\|A^{s/2} f\right\|_\pi$.

The main results are the following error bounds, in which the proof is similar as that of Altmeyer and Chorowski [2, Theorem 2.2, Corollary 2.3, Theorem 2.4] and is deferred to Section 3.5.1. Note that (3.5.2) gives the error bound on the space average of X with the finite-time and finite-sample estimator $T^{-1}\hat{\Gamma}_{T,n}(f)$, while (3.5.3) offers the error bound for the non-stationary Markov process such that $X_0 \sim \eta$.

Theorem 3.5.1. *Let X be a Markov process with $X_0 \sim \pi$ and generator $L \in \mathcal{S}(G)$. There exists a constant C such that for all $T \geqslant 0$, $0 \leqslant s \leqslant 1$, $f \in \mathcal{D}^s(A)$, $f_0 \in \mathrm{Dom}(A^{-1})$ with $f_0 = f - \int f \, d\pi$,*

$$\left\|\Gamma_T(f) - \hat{\Gamma}_{T,n}(f)\right\|_{L^2(\mathbb{P})} \leqslant C\sqrt{\|f\|_{\mathcal{D}^s(A)} \|f\|_\pi T\Delta_n^{1+s}}, \tag{3.5.1}$$

$$\left\|T^{-1}\hat{\Gamma}_{T,n}(f) - \int f \, d\pi\right\|_{L^2(\mathbb{P})} \leqslant \frac{C}{\sqrt{T}}\left(\sqrt{\|f\|_{\mathcal{D}^s(A)} \|f\|_\pi \Delta_n} + \sqrt{\|A^{-1}f_0\|_\pi \|f_0\|_\pi}\right). \tag{3.5.2}$$

If $X_0 \sim \eta$ such that $\eta \ll \pi$ with density $d\eta/d\pi$, then there exists a constant C such that for all $T \geqslant 0$, $0 \leqslant s \leqslant 1$ and $f \in \mathcal{D}^s(A)$,

$$\left\| \Gamma_T(f) - \hat{\Gamma}_{T,n}(f) \right\|_{L^2(\mathbb{P})} \leqslant C \left\| \frac{d\eta}{d\pi} \right\|_{\infty,\pi}^{1/2} \sqrt{\|f\|_{\mathcal{D}^s(A)} \|f\|_\pi T\Delta_n^{1+s}}, \tag{3.5.3}$$

where $\|\cdot\|_{\infty,\pi}$ is the sup-norm in $L^\infty(\pi)$.

3.5.1 Proof of Theorem 3.5.1

We first state a lemma (see Inoue and Trapani [55, Lemma 3.19]) which will be used repeatedly in the proof.

Lemma 3.5.1. *For $f \in \mathcal{D}^s(A)$,*

$$\left| \int_{\sigma(L)} \lambda \, d\langle F_\lambda f, f\rangle_\pi \right| \leqslant \left(\int_{\sigma(L)} |\lambda|^{2s} \, d\langle H_\lambda^* f, f\rangle_\pi \right) \|f\|_\pi = \|f\|_{\mathcal{D}^s(A)} \|f\|_\pi \, .$$

We now proceed to give the proof of Theorem 3.5.1. We first prove (3.5.1) and consider

$$\begin{aligned}
\left\| \Gamma_T(f) - \hat{\Gamma}_{T,n}(f) \right\|_{L^2(\mathbb{P})}^2 &= \mathbb{E}\left[\left(\sum_{k=1}^n \int_{(k-1)\Delta_n}^{k\Delta_n} \left(f(X_r) - f(X_{(k-1)\Delta_n}) \right) dr \right)^2 \right] \\
&= \sum_{k,l=1}^n \int_{(k-1)\Delta_n}^{k\Delta_n} \int_{(l-1)\Delta_n}^{l\Delta_n} \mathbb{E}\Big[\left(f(X_r) - f(X_{(k-1)\Delta_n}) \right) \\
&\qquad \left(f(X_h) - f(X_{(l-1)\Delta_n}) \right) \Big] \, dr dh \, ,
\end{aligned}$$

then we proceed to bound the diagonal ($k = l$) and off-diagonal ($k \neq l$) terms. For the diagonal terms, by stationarity we have for $(k-1)\Delta_n \leqslant r \leqslant h \leqslant k\Delta_n$,

$$\begin{aligned}
&\mathbb{E}\left[\left(f(X_r) - f(X_{(k-1)\Delta_n}) \right) \left(f(X_h) - f(X_{(k-1)\Delta_n}) \right) \right] \\
&\quad = \langle (P_{h-r} - I)f + (I - P_{h-(k-1)\Delta_n})f + (I - P_{r-(k-1)\Delta_n})f, f\rangle_\pi,
\end{aligned}$$

so by symmetry in r and h we have

$$\sum_{k=1}^{n} \int_{(k-1)\Delta_n}^{k\Delta_n} \int_{(k-1)\Delta_n}^{k\Delta_n} \mathbb{E}\left[\left(f(X_r) - f(X_{(k-1)\Delta_n})\right)\left(f(X_h) - f(X_{(k-1)\Delta_n})\right)\right] drdh$$
$$= 2n\left\langle \left(\int_0^{\Delta_n} \int_0^h (P_{h-r} - I)\, drdh + \Delta_n \int_0^{\Delta_n} (I - P_h)\, dh\right) f, f\right\rangle_\pi$$
$$= \langle \Phi(L) f, f\rangle_\pi$$
$$= \int_{\sigma(L)} \Phi(\lambda)\, d\langle F_\lambda f, f\rangle_\pi,$$

where the last inequality follows from the functional calculus of L in Theorem 3.2.1 and for $\lambda \in \sigma(L)$,

$$\Phi(\lambda) = 2n\left(\int_0^{\Delta_n} \int_0^h (e^{\lambda(h-r)} - 1)\, drdh + \Delta_n \int_0^{\Delta_n} (1 - e^{\lambda h})\, dh\right).$$

From Altmeyer and Chorowski [2, Page 15], we know that $|\Phi(\lambda)| \leqslant 4n\Delta_n^{2+s}|\lambda|^s$ with fixed $0 \leqslant s \leqslant 1$. Now, we apply Lemma 3.5.1 to arrive at

$$\left|\int_{\sigma(L)} \Phi(\lambda)\, d\langle F_\lambda f, f\rangle_\pi\right| \leqslant 4T\Delta_n^{1+s} \left\|f\right\|_\pi \int_{\sigma(L)} |\lambda|^{2s}\, d\langle H_\lambda^* f, f\rangle_\pi = 4T\Delta_n^{1+s} \left\|f\right\|_\pi \left\|f\right\|_{\mathcal{D}^s(A)}.$$

Next, we bound the off-diagonal terms, in which

$$2\sum_{k>l}^{n} \int_{(k-1)\Delta_n}^{k\Delta_n} \int_{(l-1)\Delta_n}^{l\Delta_n} \mathbb{E}\left[\left(f(X_r) - f(X_{(k-1)\Delta_n})\right)\left(f(X_h) - f(X_{(l-1)\Delta_n})\right)\right] drdh$$
$$= 2\left\langle \left(\int_0^{\Delta_n} \int_0^{\Delta_n} \left(\sum_{k>l=1}^{n} P_{(k-l)\Delta_n - r}\right)(P_h - I)(I - P_r)\, drdh\right) f, f\right\rangle_\pi$$
$$= \langle \tilde{\Phi}(L) f, f\rangle_\pi$$
$$= \int_{\sigma(L)} \tilde{\Phi}(\lambda)\, d\langle F_\lambda f, f\rangle_\pi,$$

where the last inequality follows again from the functional calculus of L in Theorem 3.2.1 and for $\lambda \in \sigma(L)$,

$$\tilde{\Phi}(\lambda) = 2\left(\int_0^{\Delta_n} \int_0^{\Delta_n} \left(\sum_{k>l=1}^{n} e^{\lambda((k-l)\Delta_n - r)}\right)(e^{\lambda h} - 1)(1 - e^{\lambda r})\, drdh\right).$$

Using Altmeyer and Chorowski [2, (16)] there exists a universal constant $\tilde{C} < \infty$ such that $|\tilde{\Phi}(\lambda)| \leqslant \tilde{C} T \Delta_n^{1+s} |\lambda|^s$, and together with Lemma 3.5.1 yield

$$\left| \int_{\sigma(L)} \tilde{\Phi}(\lambda)\, d\langle F_\lambda f, f\rangle_\pi \right| \leqslant \tilde{C} T \Delta_n^{1+s} \left\| f \right\|_\pi \int_{\sigma(L)} |\lambda|^{2s}\, d\langle H_\lambda^* f, f\rangle_\pi = \tilde{C} T \Delta_n^{1+s} \left\| f \right\|_\pi \left\| f \right\|_{\mathcal{D}^s(A)} .$$

Next, we prove (3.5.2). By (3.5.1) and triangle inequality,

$$\begin{aligned} \left\| T^{-1}\hat{\Gamma}_{T,n}(f) - \int f\, d\pi \right\|_{L^2(\mathbb{P})} &\leqslant T^{-1} \left\| \hat{\Gamma}_{T,n}(f) - \Gamma_T(f) \right\|_{L^2(\mathbb{P})} + \left\| T^{-1}\Gamma_T(f) - \int f\, d\pi \right\|_{L^2(\mathbb{P})} \\ &\leqslant \frac{C}{\sqrt{T}} \sqrt{\left\| f \right\|_{\mathcal{D}^s(A)} \left\| f \right\|_\pi \Delta_n} + \left\| T^{-1}\Gamma_T(f_0) \right\|_{L^2(\mathbb{P})} . \end{aligned}$$

We proceed to bound $\| T^{-1}\Gamma_T(f_0) \|_{L^2(\mathbb{P})}$, in which

$$\begin{aligned} \left\| T^{-1}\Gamma_T(f_0) \right\|_{L^2(\mathbb{P})}^2 &= 2T^{-2} \int_0^T \int_0^h \langle P_{h-r} f_0, f_0\rangle_\pi\, dr dh \\ &= \int_{\sigma(L)} \overline{\Phi}(\lambda)\, d\langle F_\lambda f_0, f_0\rangle_\pi , \end{aligned}$$

where $\overline{\Phi}$ is defined by, for $\lambda \in \sigma(L)$,

$$\overline{\Phi}(\lambda) = 2T^{-2} \int_0^T \int_0^h e^{\lambda(h-r)}\, dr dh = 2\frac{(\lambda T)^{-1}(e^{\lambda T} - 1) - 1}{\lambda T},$$

and there exists a constant $\tilde{C}$ such that $|\overline{\Phi}(\lambda)| \leqslant \dfrac{\tilde{C}}{|\lambda| T}$. Using Lemma 3.5.1 gives

$$\begin{aligned} \left\| T^{-1}\Gamma_T(f_0) \right\|_{L^2(\mathbb{P})}^2 &\leqslant \frac{\tilde{C}}{T} \left| \int_{\sigma(L)} |\lambda|^{-1}\, d\langle F_\lambda f_0, f_0\rangle_\pi \right| \\ &\leqslant \frac{\tilde{C}}{T} \left(\int_{\sigma(L)} |\lambda|^{-2}\, d\langle H_\lambda^* f_0, f_0\rangle_\pi \right) \left\| f_0 \right\|_\pi = \frac{\tilde{C}}{T} \left\| A^{-1} f_0 \right\|_\pi \left\| f_0 \right\|_\pi . \end{aligned}$$

Finally, it follows from a standard change of measure argument to give (3.5.3).

3.6 Similarity orbit of reversible Markov chains

In this Section, our aim is to provide several illuminating examples for Theorem 3.2.1 and we will work in the continuous-time setting. More precisely, suppose that we start with a reversible generator G with transition semigroup $(Q_t)_{t\geqslant 0}$, we would like to characterize the family of Markov chains with generator L associated with G under the similarity transformation $G\Lambda = \Lambda L$ with Λ being a bounded invertible Markov link. This idea allows us to generate Markov or contraction kernel from known ones in which the spectral decomposition, stationary distribution and eigenfunctions are linked by Λ. In addition, the so-called eigentime identity is preserved under intertwining as the spectrum is invariant under such transformation as stated in Theorem 3.2.1. We will illustrate this approach by studying the permutation link, pure birth link and random walk link in particular. As examples, we consider four classical models in birth-death processes and investigate their permutation and pure birth orbits, namely:

1. the Ehrenfest model (Section 3.6.1 and 3.6.3)
2. $M/M/\infty$ queue (Section 3.6.2 and 3.6.4)
3. linear birth-death process (Section 3.6.2 and 3.6.4)
4. quadratic birth-death process (Section 3.6.1 and 3.6.3)

While we consider the above univariate examples in subsequent Sections, nonetheless we can still handle the orbits of multivariate reversible Markov chains (e.g. Griffiths [46], Khare and Mukherjee [64], Khare and Zhou [65] and Zhou [116]) by considering the link kernel to be the tensor product from univariate link and analyze the corresponding tensorized orbits. Note that the permutation link has been studied by Miclo [83] in the notion of Markov similarity. We write $\sigma : \mathcal{X} \to \mathcal{X}$ to be a permutation of the state space $\mathcal{X}$, and its associated link is denoted by $\Lambda_\sigma := (\mathbb{1}_{y=\sigma(x)})_{x,y\in\mathcal{X}}$. We first give a few structural results under the permutation link, which show that permutation of state space can be effectively casted into the similarity orbit framework. Note that the following proposition is a particular case of Theorem 3.2.1.

Proposition 3.6.1. *Suppose that $G \stackrel{\Lambda_\sigma}{\sim} L$ and $\mathcal{X}$ is a finite state space.*

1. *G is reversible with respect to $\pi_G = (\pi_G(x))_{x\in\mathcal{X}}$ if and only if L is reversible with respect to $\pi_L := \pi_G\Lambda_\sigma = (\pi_G(\sigma^{-1}(x)))_{x\in\mathcal{X}}$.*

2. *Suppose that $G \stackrel{\Lambda_\sigma}{\sim} L$ with G being a reversible generator with respective to π_G, and eigenvalues-eigenvectors denoted by $(-\lambda_j, \phi_j)_{j=1}^{|\mathcal{X}|}$, where ϕ_j are orthonormal in $l^2(\pi_G)$. Write $(Q_t)_{t\geqslant 0}$ (resp. $(P_t)_{t\geqslant 0}$) being the transition semigroup associated with G (resp. L) of the Markov chain $(X_t)_{t\geqslant 0}$ (resp. $(Y_t)_{t\geqslant 0}$). For $t \geqslant 0$ and $x, y \in \mathcal{X}$, the spectral decompositions are given by*
$$Q_t(x,y) = \pi_G(y)\sum_{j=1}^{|\mathcal{X}|} e^{-\lambda_j t}\phi_j(x)\phi_j(y),$$
$$P_t(x,y) = \pi_G(\sigma^{-1}(y))\sum_{j=1}^{|\mathcal{X}|} e^{-\lambda_j t}\phi_j(\sigma^{-1}(x))\phi_j(\sigma^{-1}(y)).$$

3. *(Eigentime identity) $(P_t)_{t\geqslant 0}$ and $(Q_t)_{t\geqslant 0}$ shares the same eigentime identity. That is, denote $\tau_y^Q := \inf\{t \geqslant 0; X_t = y\}$ (resp. $\tau_y^P := \inf\{t \geqslant 0; Y_t = y\}$), then*
$$\sum_{x,y\in\mathcal{X}} \mathbb{E}_x(\tau_y^Q)\pi_G(x)\pi_G(y) = \sum_{x,y\in\mathcal{X}} \mathbb{E}_x(\tau_y^P)\pi_L(x)\pi_L(y) = \sum_{i=1,\lambda_i\neq 0}^{|\mathcal{X}|} \frac{1}{\lambda_i}.$$

Remark 3.6.1. Note that item (2) offers us a way to generate reversible processes $(P_t)_{t\geqslant 0}$ from a birth-death process $(Q_t)_{t\geqslant 0}$ via the permutation link Λ_σ, where the birth-death structure is possibly destroyed while maintaining reversibility in a different Hilbert space. This will be illustrated in subsequent Sections when we look into various birth-death models.

Proof. First, we show (1). Since Λ_σ is an unitary operator, (1) has already been proved in Theorem 3.2.1 (c). We offer another proof here via checking directly the detailed balance condition. Note that for $x, y \in \mathcal{X}$,
$$L(x,y) = G(\sigma^{-1}(x), \sigma^{-1}(y)),$$

so the detailed balance for L is given by

$$\pi_L(x)L(x,y) = \pi_G(\sigma^{-1}(x))G(\sigma^{-1}(x), \sigma^{-1}(y)).$$

Therefore, G is reversible with respect to π_G if and only if L is reversible with respect to π_L. Next, to show (2), the spectral decomposition of Q_t follows from the spectral theorem of normal operator, while that of P_t follows from the relationship $P_t(x,y) = Q_t(\sigma^{-1}(x), \sigma^{-1}(y))$. Finally, to show (3), we simply need to invoke the eigentime identity Aldous and Fill [1, Proposition 3.13] for reversible Markov chains and the fact that the eigenvalues remain the same under permutation. □

3.6.1 Permutation link on finite state space

In this Section, we provide two examples using the permutation link Λ_σ on a finite state space $\mathcal{X} = [\![0, N]\!]$, generated by birth-death processes with birth and death rates to be λ_x and μ_x respectively. We assume that $\mu_0 = \lambda_N = 0$. We also write $(a)_n$ to be the Pochhammer's symbol and ${}_pF_q$ to be the generalized hypergeometric series. For further details on various birth-death models and their connections with orthogonal polynomials, we refer interested readers to Diaconis et al. [34], Karlin and McGregor [58], Koekoek and Swarttouw [66], Sasaki [99], Schoutens [100], Zhou [116] and the references therein.

The Ehrenfest model and its permuted variant

In this example, we study the Ehrenfest model. That is, it is a birth-death process with $\lambda_x = p(N-x), \mu_x = (1-p)x$, where $0 < p < 1$. The stationary distribution is the binomial distribution with probability mass function $\binom{N}{x} p^x (1-p)^{N-x}$ for $x \in [\![0, N]\!]$, and the associated orthogonal polynomials are the Krawtchouk polynomials. Under the permutation link Λ_σ, Proposition 3.6.1 (2) and (3) now reads, for $j, x, y \in [\![0, N]\!]$ and $t \geqslant 0$,

$$\pi_x = \binom{N}{x} p^x (1-p)^{N-x},$$

$$\phi_j(x) = {}_2F_1\left(\begin{matrix} -j, -x \\ -N \end{matrix} \middle| p^{-1}\right),$$

$$Q_t(x, y) = \pi_y \sum_{j=0}^{N} e^{-jt} \phi_j(x)\phi_j(y) \frac{(-1)^{-j} p^j}{j!(1-p)^j} (-N)_j,$$

$$P_t(x, y) = \pi_{\sigma^{-1}(y)} \sum_{j=0}^{N} e^{-jt} \phi_j(\sigma^{-1}(x))\phi_j(\sigma^{-1}(y)) \frac{(-1)^{-j} p^j}{j!(1-p)^j} (-N)_j,$$

$$\sum_{x,y \in \mathcal{X}} \mathbb{E}_x(\tau_y^P) \pi_{\sigma^{-1}(x)} \pi_{\sigma^{-1}(y)} = \sum_{i=1}^{N} \frac{1}{i}.$$

We can see that the permuted Ehrenfest model has a permuted binomial distribution as stationary distribution, and the corresponding eigenvectors are the permuted Krawtchouk polynomials.

Quadratic birth-death process and its permuted variant

In the second example, we investigate the so-called quadratic model with $\lambda_x = (N - x)(a - x)$, $\mu_x = x(b - (N - x))$ and parameters $a, b \geqslant N$. The invariant distribution is given by the hypergeometric distribution with probability mass function

$$\pi_x := \frac{\binom{a}{x}\binom{b}{N-x}}{\binom{a+b}{N}},$$

and the associated orthogonal polynomials are the dual Hahn polynomials. With the permutation link Λ_σ, Proposition 3.6.1 (2) and (3) now reads, for $j, x, y \in [\![0, N]\!]$ and $t \geqslant 0$,

$$\phi_j(x) = {}_3F_2\left(\begin{matrix} -j, -x, -x-a-b-1 \\ -a, -N \end{matrix} \middle| 1\right),$$

$$
\begin{aligned}
w_j &:= \frac{\binom{N-b-1}{N} N!(-N)_j(-a)_j(2j-a-b-1)}{(-1)^j j!(-b)_j(j-a-b-1)_{N+1}}, \\
Q_t(x,y) &= \pi_y \sum_{j=0}^{N} e^{-j(a+b+1-j)t}\phi_j(x)\phi_j(y)w_j, \\
P_t(x,y) &= \pi_{\sigma^{-1}(y)} \sum_{j=0}^{N} e^{-j(a+b+1-j)t}\phi_j(\sigma^{-1}(x))\phi_j(\sigma^{-1}(y))w_j, \\
\sum_{x,y\in\mathcal{X}} \mathbb{E}_x(\tau_y^P)\pi_{\sigma^{-1}(x)}\pi_{\sigma^{-1}(y)} &= \sum_{i=1}^{N} \frac{1}{i(a+b+1-i)}.
\end{aligned}
$$

We again observe that the permuted quadratic birth-death process is not necessarily a birth-death process under the permutation link, and its stationary distribution and eigenvectors are respectively the permuted hypergeometric distribution and permuted dual Hahn polynomials.

3.6.2 n-dimensional permutation link on $\mathbb{N}_0$

In this Section, we provide two instances using a n-dimensional permutation link Λ_σ (i.e. a permutation that only permutes n elements) on the state space $\mathcal{X} = \mathbb{N}_0$ generated by birth-death processes, namely a $M/M/\infty$ model and the linear birth-death process.

$M/M/\infty$ and its n-dimensional permuted variant

In this example, we look at the $M/M/\infty$ queueing model with $\lambda_x = \lambda, \mu_x = x\mu$, where $\lambda, \mu > 0$ are the arrival and service rate respectively. The stationary distribution is the Poisson distribution with mean λ/μ and the associated orthogonal polynomials are the Charlier polynomials. With the n-dimensional permutation link Λ_σ, we have for $j, x, y \in \mathbb{N}_0$ and $t \geqslant 0$,

$$
\phi_j(x) = {}_2F_0\left(\begin{matrix} -j, -x \\ - \end{matrix} \,\middle|\, -\mu^{-1}\right),
$$

$$Q_t(x,y) = e^{-\lambda/\mu}\frac{(\lambda/\mu)^y}{y!}\sum_{j=0}^{\infty} e^{-\mu j t}\phi_j(x)\phi_j(y)\frac{(\lambda/\mu)^j}{j!},$$
$$P_t(x,y) = e^{-\lambda/\mu}\frac{(\lambda/\mu)^{\sigma^{-1}(y)}}{\sigma^{-1}(y)!}\sum_{j=0}^{\infty} e^{-\mu j t}\phi_j(\sigma^{-1}(x))\phi_j(\sigma^{-1}(y))\frac{(\lambda/\mu)^j}{j!}.$$

Linear birth-death process and its n-dimensional permuted variant

In this example, we study the linear birth-death process with $\lambda_x = (x+\beta)\lambda$, $\mu_x = x\mu$, where $\beta, \lambda, \mu > 0$ are the parameters with $\lambda < \mu$. The stationary distribution is given by the negative binomial distribution with probability mass function

$$\pi_x := \left(1 - \frac{\lambda}{\mu}\right)^{\beta} \frac{(\beta)_x}{x!}\left(\frac{\lambda}{\mu}\right)^{x},$$

and the associated orthogonal polynomials are the Meixner polynomials. With the n-dimensional permutation link Λ_σ, we have for $j, x, y \in \mathbb{N}_0$ and $t \geqslant 0$,

$$\phi_j(x) = {}_2F_1\left(\left.\begin{matrix} -j, -\dfrac{x}{\mu-\lambda} \\ \beta \end{matrix}\,\right|\, \frac{\lambda-\mu}{\lambda}\right),$$
$$Q_t(x,y) = \pi_y \sum_{j=0}^{\infty} e^{-(\mu-\lambda)jt}\phi_j(x)\phi_j(y)\frac{(\beta)_j}{j!}\left(\frac{\lambda}{\mu}\right)^{j},$$
$$P_t(x,y) = \pi_{\sigma^{-1}(y)} \sum_{j=0}^{\infty} e^{-(\mu-\lambda)jt}\phi_j(\sigma^{-1}(x))\phi_j(\sigma^{-1}(y))\frac{(\beta)_j}{j!}\left(\frac{\lambda}{\mu}\right)^{j}.$$

3.6.3 Pure birth link on finite state space

In this Section, we specialize into the case of $\mathcal{X} = [\![0, N]\!]$, with the link being the pure birth link as introduced by Fill [43] to study the distribution of hitting time and fastest strong stationary time. The particular pure birth link Λ_{pb} that we study is of the form $\Lambda_{pb}(x,y) = 1/2$ for $x \in [\![0, N-1]\!]$, $y \in \{x, x+1\}$, $\Lambda_{pb}(N,N) = 1$ and zero otherwise. It can be shown that the inverse is given by $\Lambda_{pb}^{-1}(x,y) = (-1)^{y-x}(2\mathbb{1}_{y\neq N} + \mathbb{1}_{y=N})$ for

$x \leqslant y$, $x, y \in [\![0, N]\!]$ and zero otherwise. A special feature in the pure birth orbit is that the heat kernel $P_t := e^{tL}$ of L need not be Markovian, yet it still converges to π_L exponentially fast as illustrated in Proposition 3.6.2 below. We now give a structural result in this direction, which follows easily from Theorem 3.2.1.

Proposition 3.6.2. *Suppose that $G \overset{\Lambda_{pb}}{\sim} L$ with G being a reversible generator with respective to π_G on $\mathcal{X} = [\![0, N]\!]$, and eigenvalues-eigenvectors denoted by $(-\lambda_j, \phi_j)_{j=0}^N$, where ϕ_j are orthonormal in $l^2(\pi_G)$. Write $(Q_t)_{t\geqslant 0}$ (resp. $(P_t)_{t\geqslant 0}$) being the transition semigroup associated with G (resp. L). Note that $(P_t)_{t\geqslant 0}$ need not be Markov under Λ_{pb}. For $t \geqslant 0$ and $j, x, y \in [\![0, N]\!]$, the spectral decompositions are given by*

$$Q_t(x,y) = \sum_{j=0}^{N} e^{-\lambda_j t}\phi_j(x)\phi_j(y)\pi_G(y),$$

$$P_t(x,y) = \sum_{j=0}^{N} e^{-\lambda_j t} f_j(x) f_j^*(y),$$

$$||P_t - \pi_L||_{TV} \leqslant \frac{\kappa(\Lambda_{pb})e^{-\lambda_1 t}}{2}\sqrt{\frac{1-\pi_L^*}{\pi_L^*}}, \quad \textit{where}$$

$$f_j(x) := \sum_{k=x}^{N} (-1)^{k-x}(2\mathbb{1}_{k\neq N} + \mathbb{1}_{k=N})\phi_j(k),$$

$$f_j^*(y) := \phi_j(y-1)\pi_G(y-1)\left(\frac{\mathbb{1}_{y-1\geqslant 0}}{2}\right) + \phi_j(y)\pi_G(y)\left(\frac{\mathbb{1}_{y\neq N}}{2} + \mathbb{1}_{y=N}\right),$$

$$\pi_L(x) = \pi_G(x-1)\left(\frac{\mathbb{1}_{x-1\geqslant 0}}{2}\right) + \pi_G(x)\left(\frac{\mathbb{1}_{x\neq N}}{2} + \mathbb{1}_{x=N}\right),$$

$$\pi_L^* := \min_{x\in[\![0,N]\!]} \pi_L(x).$$

Remark 3.6.2. If $(P_t)_{t\geqslant 0}$ is a Markov semigroup, then as stated in Corollary 3.2.1, using the result of Cui and Mao [25], we again reach at the eigentime identity:

$$\sum_{x,y\in\mathcal{X}} \mathbb{E}_x(\tau_y^Q)\pi_G(x)\pi_G(y) = \sum_{x,y\in\mathcal{X}} \mathbb{E}_x(\tau_y^P)\pi_L(x)\pi_L(y) = \sum_{i=0,\lambda_i\neq 0}^{N} \frac{1}{\lambda_i}.$$

Remark 3.6.3. We can see that π_L is the distribution at time 1 of the Markov chain with transition matrix Λ_{pb} under the initial law π_G.

Proof. Upon expanding $P_t = \Lambda_{pb}^{-1} Q_t \Lambda_{pb}$, we get

$$
\begin{aligned}
P_t(x,y) &= \sum_{k=x}^{N} (-1)^{k-x}(2\mathbb{1}_{k\neq N} + \mathbb{1}_{k=N}) \left(Q_t(k,y-1)\left(\frac{\mathbb{1}_{y-1\geqslant 0}}{2}\right) + Q_t(k,y)\left(\frac{\mathbb{1}_{y\neq N}}{2} + \mathbb{1}_{y=N}\right)\right) \\
&= \sum_{j=0}^{N} e^{-\lambda_j t} \left(\sum_{k=x}^{N} (-1)^{k-x}(2\mathbb{1}_{k\neq N} + \mathbb{1}_{k=N})\phi_j(k)\right) \\
&\quad \times \left(\phi_j(y-1)\pi_G(y-1)\left(\frac{\mathbb{1}_{y-1\geqslant 0}}{2}\right) + \phi_j(y)\pi_G(y)\left(\frac{\mathbb{1}_{y\neq N}}{2} + \mathbb{1}_{y=N}\right)\right),
\end{aligned}
$$

where the second equality follows from substituting the spectral expansion of $(Q_t)_{t\geqslant 0}$. $||P_t - \pi_L||_{TV} \leqslant \dfrac{\kappa(\Lambda_{pb})e^{-\lambda_1 t}}{2}\sqrt{\dfrac{1-\pi_L^*}{\pi_L^*}}$ follows directly from Corollary 3.2.2. □

To illustrate this proposition, we again look at the Ehrenfest model and the quadratic birth-death process.

The Ehrenfest model and its pure birth variant

Recall that we introduce the Ehrenfest model in Section 3.6.1. Under the pure birth link Λ_{pb}, Proposition 3.6.2 now reads, for $j, x, y \in [\![0, N]\!]$ and $t \geqslant 0$,

$$
\begin{aligned}
P_t(x,y) &= \sum_{j=0}^{N} e^{-jt} f_j(x) f_j^*(y) \frac{(-1)^{-j}p^j}{j!(1-p)^j}(-N)_j, \\
||P_t - \pi_L||_{TV} &\leqslant \frac{\kappa(\Lambda_{pb})e^{-t}}{2}\sqrt{\frac{1-\pi_L^*}{\pi_L^*}} = O(e^{-t}), \quad \text{where} \\
\pi_G(x) &= \binom{N}{x} p^x (1-p)^{N-x}, \\
f_j(x) &= \sum_{k=x}^{N} (-1)^{k-x}(2\mathbb{1}_{k\neq N} + \mathbb{1}_{k=N})\, {}_2F_1\left(\left.\begin{matrix}-j, -k \\ -N\end{matrix}\right| p^{-1}\right), \\
f_j^*(y) &= {}_2F_1\left(\left.\begin{matrix}-j, -(y-1) \\ -N\end{matrix}\right| p^{-1}\right)\pi_G(y-1)\left(\frac{\mathbb{1}_{y-1\geqslant 0}}{2}\right) \\
&\quad + {}_2F_1\left(\left.\begin{matrix}-j, -y \\ -N\end{matrix}\right| p^{-1}\right)\pi_G(y)\left(\frac{\mathbb{1}_{y\neq N}}{2} + \mathbb{1}_{y=N}\right),
\end{aligned}
$$

$$\pi_L(x) = \pi_G(x-1)\left(\frac{\mathbb{1}_{x-1\geqslant 0}}{2}\right) + \pi_G(x)\left(\frac{\mathbb{1}_{x\neq N}}{2} + \mathbb{1}_{x=N}\right),$$
$$\pi_L^* = \min_x \pi_L(x).$$

We can see that π_L is the stationary distribution if we sample from a binomial distribution with parameters N and p to initialize the pure birth process with kernel Λ_{pb}. Another remark is that P_t is not necessarily reversible or Markovian, yet it converges to π_L exponentially fast.

Quadratic birth-death process and its pure birth variant

In the second example, we study the quadratic birth-death process as introduced in Section 3.6.1. In the pure birth link Λ_{pb}, Proposition 3.6.2 now reads, for $j, x, y \in [\![0, N]\!]$ and $t \geqslant 0$,

$$P_t(x,y) = \sum_{j=0}^{N} e^{-j(a+b+1-j)t} f_j(x) f_j^*(y) w_j,$$
$$||P_t - \pi_L||_{TV} \leqslant \frac{\kappa(\Lambda_{pb})e^{-(a+b)t}}{2}\sqrt{\binom{a+b}{N} - 1} = O(e^{-(a+b)t}), \quad \text{where}$$
$$\pi_G(x) = \frac{\binom{a}{x}\binom{b}{N-x}}{\binom{a+b}{N}},$$
$$w_j = \frac{\binom{N-b-1}{N} N!(-N)_j(-a)_j(2j-a-b-1)}{(-1)^j j!(-b)_j(j-a-b-1)_{N+1}},$$
$$f_j(x) = \sum_{k=x}^{N}(-1)^{k-x}(2\mathbb{1}_{k\neq N} + \mathbb{1}_{k=N})\,{}_3F_2\left(\left.\begin{matrix} -j, -k, -k-a-b-1 \\ -a, -N \end{matrix}\right| 1\right),$$
$$f_j^*(y) = {}_3F_2\left(\left.\begin{matrix} -j, -(y-1), -(y-1)-a-b-1 \\ -a, -N \end{matrix}\right| 1\right)\pi_G(y-1)\left(\frac{\mathbb{1}_{y-1\geqslant 0}}{2}\right)$$
$$+ {}_3F_2\left(\left.\begin{matrix} -j, -y, -y-a-b-1 \\ -a, -N \end{matrix}\right| 1\right)\pi_G(y)\left(\frac{\mathbb{1}_{y\neq N}}{2} + \mathbb{1}_{y=N}\right),$$
$$\pi_L(x) = \pi_G(x-1)\left(\frac{\mathbb{1}_{x-1\geqslant 0}}{2}\right) + \pi_G(x)\left(\frac{\mathbb{1}_{x\neq N}}{2} + \mathbb{1}_{x=N}\right).$$

π_L is the stationary distribution if we sample from a hypergeometric distribution π_G to initialize the pure birth process with kernel Λ_{pb}.

3.6.4 $(n+1)$-dimensional pure birth link on $\mathbb{N}_0$

In this Section, we detail two instances using a $(n+1)$-dimensional pure birth link Λ_{pb} (i.e. a pure birth kernel on $[\![0,n]\!]$) on the state space $\mathcal{X} = \mathbb{N}_0$ generated by the $M/M/\infty$ model and the linear birth-death process.

$M/M/\infty$ and its $(n+1)$-dimensional pure birth variant

The $M/M/\infty$ queueing model is first introduced in Section 3.6.2. With the $(n+1)$-dimensional pure birth link Λ_{pb}, the pure birth variant of $M/M/\infty$ is given by, for $j, x, y \in \mathbb{N}_0$ and $t \geqslant 0$,

$$
\begin{aligned}
P_t(x,y) &= \sum_{j=0}^{\infty} e^{-\mu j t} f_j(x) f_j^*(y) \frac{(\lambda/\mu)^j}{j!}, \quad \text{where} \\
\phi_j(x) &= {}_2F_0\left(\left. \begin{matrix} -j, -x \\ - \end{matrix} \right| -\mu^{-1} \right), \\
\pi_G(x) &= e^{-\lambda/\mu} \frac{(\lambda/\mu)^x}{x!}, \\
f_j(x) &= \begin{cases} \sum_{k=x}^{n} (-1)^{k-x} (2\mathbb{1}_{k \neq n} + \mathbb{1}_{k=n}) \phi_j(k), & x \in [\![0,n]\!], \\ \phi_j(x), \quad x \geqslant n+1, \end{cases} \\
f_j^*(y) &= \begin{cases} \phi_j(y-1)\pi_G(y-1)\left(\frac{\mathbb{1}_{y-1 \geqslant 0}}{2}\right) + \phi_j(y)\pi_G(y)\left(\frac{\mathbb{1}_{y \neq n}}{2} + \mathbb{1}_{y=n}\right), & y \in [\![0,n]\!], \\ \phi_j(y)\pi_G(y), \quad y \geqslant n+1. \end{cases}
\end{aligned}
$$

Linear birth-death process and its $(n+1)$-dimensional pure birth variant

The linear birth-death process is introduced in Section 3.6.2. With the $(n+1)$-dimensional pure birth link Λ_{pb}, we have for $j, x, y \in \mathbb{N}_0$ and $t \geqslant 0$,

$$P_t(x,y) = \sum_{j=0}^{\infty} e^{-(\mu-\lambda)jt} f_j(x) f_j^*(y) \frac{(\beta)_j}{j!} \left(\frac{\lambda}{\mu}\right)^j, \quad \text{where}$$

$$\phi_j(x) = {}_2F_1\left(\begin{matrix} -j, -\frac{x}{\mu-\lambda} \\ \beta \end{matrix} \middle| \frac{\lambda-\mu}{\lambda}\right),$$

$$\pi_G(x) = \left(1 - \frac{\lambda}{\mu}\right)^{\beta} \frac{(\beta)_x}{x!} \left(\frac{\lambda}{\mu}\right)^x,$$

$$f_j(x) = \begin{cases} \sum_{k=x}^{n} (-1)^{k-x} (2\mathbb{1}_{k\neq n} + \mathbb{1}_{k=n}) \phi_j(k), & x \in [\![0,n]\!], \\ \phi_j(x), & x \geqslant n+1, \end{cases}$$

$$f_j^*(y) = \begin{cases} \phi_j(y-1)\pi_G(y-1)\left(\frac{\mathbb{1}_{y-1\geqslant 0}}{2}\right) + \phi_j(y)\pi_G(y)\left(\frac{\mathbb{1}_{y\neq n}}{2} + \mathbb{1}_{y=n}\right), & y \in [\![0,n]\!], \\ \phi_j(y)\pi_G(y), & y \geqslant n+1. \end{cases}$$

3.6.5 Random walk link on finite state space

In this Section, we specialize into the case of $\mathcal{X} = [\![0,N]\!]$, with the link being the random walk kernel previously studied by Diaconis and Miclo [29] and Zhou [116]. The particular random walk link Λ_{rw} that we study is of the form $\Lambda_{rw}(0,0) = \Lambda_{rw}(0,1) = \Lambda_{rw}(x,y) = 1/2$ for $x \in [\![1,N-1]\!]$, $y = x \pm 1$, $\Lambda_{rw}(N,N) = 1$ and zero otherwise. That is, it is a simple random walk with holding at 0 and absorbing endpoint N. For $j = 1, 3, \ldots, 2N-1$, the eigenvalue β_j, right eigenfunction ψ_j and left eigenfunction Ψ_j are given by

$$\beta_j := \cos\left(\frac{j\pi}{2N+1}\right), \tag{3.6.1}$$

$$\psi_j(x) := \cos\left(\frac{(2x+1)j\pi}{2(2N+1)}\right), \quad x \in [\![0,N]\!], \tag{3.6.2}$$

$$\Psi_j(x) := \begin{cases} \psi_j(x), & \text{for } x \in [\![0, N-1]\!], \\ \dfrac{(-1)^{(j+1)/2}}{2} \cot\left(\dfrac{j\pi}{2(2N+1)}\right), & \text{for } x = N, \end{cases} \tag{3.6.3}$$

$$\Lambda_{rw} = \sum_{j \in \{0,1,3,\ldots,2N-1\}} \beta_j \psi_j \Psi_j^T, \tag{3.6.4}$$

where for $j = 0$, $\beta_0 := 1$, $\psi_0 := \mathbf{1}$, the vector of 1s, and $\Psi_0 := \delta_N$, the Dirac mass at N. An interesting feature in this random walk orbit is that the right eigenfunction can be interpreted as a special discrete cosine transform using ψ_j.

Proposition 3.6.3. *Suppose that $G \stackrel{\Lambda_{rw}}{\sim} L$ with G being a reversible generator with respective to π_G on $\mathcal{X} = [\![0, N]\!]$, and eigenvalues-eigenvectors denoted by $(-\lambda_j, \phi_j)_{j=0}^N$, where ϕ_j are orthonormal in $l^2(\pi_G)$. Write $(Q_t)_{t \geqslant 0}$ (resp. $(P_t)_{t \geqslant 0}$) being the transition semigroup associated with G (resp. L). Note that $(P_t)_{t \geqslant 0}$ need not be Markov under Λ_{rw}. For $t \geqslant 0$, $j, x, y \in [\![0, N]\!]$ and recall that β_j, ψ_j and Ψ_j are defined in (3.6.1), (3.6.2) and (3.6.3) respectively, the spectral decompositions are given by*

$$\begin{aligned} Q_t(x, y) &= \sum_{j=0}^{N} e^{-\lambda_j t} \phi_j(x) \phi_j(y) \pi_G(y), \\ P_t(x, y) &= \sum_{j=0}^{N} e^{-\lambda_j t} f_j(x) f_j^*(y), \quad \text{where} \\ f_j(x) := \Lambda_{rw}^{-1} \phi_j(x) &= \sum_{k \in \{0,1,3,\ldots,2N-1\}} \beta_k^{-1} \langle \Psi_k, \phi_j \rangle \psi_k(x) \\ &= \sum_{k \in \{0,1,3,\ldots,2N-1\}} \beta_k^{-1} \langle \Psi_k, \phi_j \rangle \cos\left(\frac{(2x+1)k\pi}{2(2N+1)}\right), \\ f_j^*(y) &= \sum_{k \in \{0,1,3,\ldots,2N-1\}} \beta_k \langle \psi_k, \phi_j \rangle_{\pi_G} \Psi_k(y), \\ \pi_L(x) = \widetilde{\Phi}_0(x) &= \sum_{k \in \{0,1,3,\ldots,2N-1\}} \beta_k \langle \psi_k, \mathbf{1} \rangle_{\pi_G} \Psi_k(x). \end{aligned}$$

Remark 3.6.4. For $j \in [\![0, N]\!]$, note that f_j is the discrete cosine transform of type VI (see Britanak et al. [8]) of the points $(\beta_k^{-1} \langle \Psi_k, \phi_j \rangle)_{k \in \{0,1,3,\ldots,2N-1\}}$.

CHAPTER 4

METROPOLIS-HASTINGS REVERSIBLIZATIONS OF NON-REVERSIBLE MARKOV CHAINS

4.1 Introduction

In this Chapter, we consider a Markov chain with transition kernel P and stationary distribution π with its time-reversal P^* on a general state space $\mathcal{X}$. To tackle non-reversibility, the path taken in this chapter is in the spirit of the reversiblization approach by Fill [42] and [89] as explained in Chapter 1 and stems on an additional reversiblization procedure. More specifically, we use and develop further the celebrated Metropolis-Hastings (MH) algorithm to provide an original in-depth analysis of non-reversible chains. The aim is to investigate Metropolis-Hastings (MH) reversiblizations, and how it helps to analyze non-reversible chains. The MH algorithm, developed by Metropolis et al. [77] and Hastings [48], is a Markov Chain Monte Carlo method that is of fundamental importance in statistics and other applications, see e.g. Roberts and Rosenthal [94] and the references therein. The idea is to construct from a proposal kernel a reversible chain which converges to a desired distribution. Much of the literature focuses on the speed of convergence of *specific* algorithms, where the proposal kernel (e.g. a random walk proposal or an Ornstein-Uhlenbeck proposal) are often by itself reversible and the target density is in general *not* the proposal stationary measure. For example, Roberts and Tweedie [95] investigates the random walk MH with exponential family target density. Hairer et al. [47] compares the theoretical performance of random walk MH and pCN algorithm with specific target density by establishing their Wasserstein spectral gap.

The notion of MH reversiblizations to study non-reversible chains is not entirely new. To the best of our knowledge, this term is first formally introduced by Aldous and Fill [1], although they did not provide a detailed analysis. Our contributions can be summarized as follows:

1. We start by studying two types of MH reversiblizations. The first MH kernel is the classical Metropolis chain of P, and we identify a new self-adjoint yet possibly non-Markovian operator that we call the second MH kernel. It captures the opposite transition effect of the first kernel, and thus it can be interpreted as the *dual* in a broad sense. We show that the linear operator $P + P^*$ can be written as the sum of the two MH kernels, which allows us to state a version of Weyl's inequality for the spectral gap of P and its additive reversiblization in the finite state space case. We prove that our bound is sharp by investigating in details the asymmetric simple random walk on the n-cycle. We also give a spectral-type expansion of P expressed in terms of the spectral measures of the two MH kernels, in which we called a *pseudospectral* expansion, in terms of the spectral measures of the two MH kernels.

2. We proceed by defining a *pseudo*-spectral gap, that we call the MH-spectral gap, based on the spectrum of the two MH kernels, along the line of work by Paulin [89]. We show that the existence of a MH-spectral gap implies that P is geometrically ergodic. We carry out some numerical examples that reveal that our MH-spectral gap is, for non-reversible chains, a better estimate than the existing bounds found in the literature. Variance bounds are also proved in terms of the proposed gap. Finally, we revisit the notion of metastability and the Cheeger inequality, to offer a variant of these celebrated inequalities by means of comparison of the non-reversible chain and the two MH kernels.

The rest of this Chapter is organized as follows. We fix the notation and give a review of the theory of general state space Markov chain as well as the MH algorithm in Section 4.2. We begin Section 4.3 by formally defining the two MH kernels and state some elementary results, followed by comparing P and the two MH kernels using the Peksun ordering, and we end this section by stating Weyl's inequality for the spectral gap of P. Section 4.4 describes the pseudospectral expansion of P. The MH-spectral gap is defined in Section 4.5, and we give a number of results that relate geometric ergodicity, mixing time and the MH-spectral gap. Finally, we state the results about variance bounds in terms of the MH-spectral gap in Section 4.6, and discuss metastability and the Cheeger inequality bounds in Section 4.7.

4.2 Preliminaries

In this section, we review several fundamental notions for Markov chain on a general state space.

Let $X = (X_n)_{n\in\mathbb{N}_0}$ be a time-homogeneous Markov chain on a measurable state space $(\mathcal{X}, \mathcal{F})$, and as usual we write P to be the Markov transition kernel which describes the one-step transition. Recall that for $P : \mathcal{X} \times \mathcal{F} \to [0, 1]$ to be a Markov transition kernel, for each fixed $A \in \mathcal{F}$, the mapping $x \mapsto P(x, A)$ is $\mathcal{F}$-measurable and for each fixed $x \in \mathcal{X}$, the function $A \mapsto P(x, A)$ is a probability measure on $\mathcal{X}$. Given a function $f : \mathcal{F} \to \mathbb{C}$ and a signed measure μ on $(\mathcal{X}, \mathcal{F})$, P acts on f from the left and μ from the right by

$$Pf(x) := \int_{\mathcal{X}} f(y)P(x, dy), \quad \mu P(A) := \int_{\mathcal{X}} P(x, A)\mu(dx), \quad x \in \mathcal{X}, A \in \mathcal{F},$$

whenever the integrals exist.

We say that π is a stationary distribution of X if π is a probability measure on $(\mathcal{X}, \mathcal{F})$

and

$$\int_{\mathcal{X}} P(x, A)\, \pi(dx) = \pi(A), \quad A \in \mathcal{F}.$$

A closely related notion is *reversibility*. We say that X is reversible if there is a probability measure π on $(\mathcal{X}, \mathcal{F})$ such that the *detailed balance* relation is satisfied:

$$\pi(dx)P(x, dy) = \pi(dy)P(y, dx).$$

Note that detailed balance means the two probability measures are identical on the product space $(\mathcal{X} \times \mathcal{X}, \mathcal{F} \times \mathcal{F})$. It is known that if π is a reversible probability measure, then π is a stationary distribution, yet the converse is not true. Let $L^2(\pi)$ be the Hilbert space of complex valued measurable functions on $\mathcal{X}$ that are squared-integrable with respect to π, endowed with the inner product $\langle f, g\rangle_\pi := \int f g^*\, d\pi$ and the norm $||f||_\pi := \langle f, f\rangle_\pi^{1/2}$, where * is the complex conjugate operation. P can then be viewed as a linear operator on $L^2(\pi)$, in which we still denote the operator by P. The operator norm of P on $L^2(\pi)$ is

$$||P||_{L^2 \to L^2} = \sup_{\substack{f \in L^2(\pi) \\ ||f||_\pi = 1}} ||Pf||_\pi.$$

Let P^* be the adjoint or time-reversal of P on $L^2(\pi)$, and it can be checked that

$$\pi(dx)P^*(x, dy) = \pi(dy)P(y, dx).$$

This shows that reversibility is equivalent to self-adjointness of P. Write $\sigma(P) = \sigma(P|L^2)$ to be the spectrum of P on $L^2(\pi)$, i.e.

$$\sigma(P|L^2) = \{\lambda \in \mathbb{C}\backslash 0 : (\lambda I - P) \text{ does not have a bounded inverse}\}.$$

If P is self-adjoint, then $\sigma(P) \subseteq [-1, 1]$. In addition, the spectral theorem for self-adjoint operators gives

$$P = \int_{\sigma(P)} \lambda\, \mathcal{E}(d\lambda), \tag{4.2.1}$$

where $\mathcal{E}$ is the spectral measure associated with P. When considering the spectral gap of P, it is often convenient for us to restrict to the space $L^2_0(\pi) = \{f \in L^2(\pi) : \mathbb{E}_\pi f = 0\}$. We formally define the meaning of a L^2-spectral gap.

Definition 4.2.1 (L^2-spectral gap)**.** Suppose that P is a Markov kernel with stationary measure π. If

$$\beta = ||P||_{L^2_0 \to L^2_0} < 1,$$

then the (absolute) L^2-spectral gap is $\gamma^* = \gamma^*(P) = 1 - \beta$.

Let

$$\lambda = \lambda(P) := \inf\{\alpha : \alpha \in \sigma(P|L^2_0)\}, \quad \Lambda = \Lambda(P) := \sup\{\alpha : \alpha \in \sigma(P|L^2_0)\}.$$

If P is reversible with respect to π, then it is known that (see, e.g. Rudolf [96])

$$\lambda = \inf_{\substack{f \in L^2_0(\pi) \\ ||f||_\pi = 1}} \langle Pf, f\rangle_\pi, \quad \Lambda = \sup_{\substack{f \in L^2_0(\pi) \\ ||f||_\pi = 1}} \langle Pf, f\rangle_\pi, \tag{4.2.2}$$

and when P is self-adjoint, we have

$$||P||_{L^2_0 \to L^2_0} = \sup_{\substack{f \in L^2_0(\pi) \\ ||f||_\pi = 1}} |\langle Pf, f\rangle_\pi|.$$

This allows us to deduce that

$$\beta = \max\{|\lambda|, \Lambda\}. \tag{4.2.3}$$

We also define the (right) spectral gap for a Markov kernel.

Definition 4.2.2 (spectral gap)**.** Suppose that P is a Markov kernel with stationary measure π. The (right) spectral gap is defined to be

$$\gamma = \gamma(P) := 1 - \sup\{\mathrm{Re}(\alpha) : \alpha \in \sigma(P|L^2_0)\}.$$

If P is reversible, then $\gamma = 1 - \Lambda(P)$. It can be shown that $\gamma(P) = \gamma((P + P^*)/2)$, so for a general P,

$$\gamma(P) = 1 - \Lambda((P + P^*)/2).$$

Remark 4.2.1. We recall that in Fill [42], additive reversiblization $(P + P^*)/2$ and multiplicative reversiblization PP^* are proposed to study mixing for non-reversible chains. In the discrete-time setting, the upper bound involves $\gamma(PP^*)$, while for continuous-time Markov chains, the upper bound depends on $\gamma((P + P^*)/2)$.

Remark 4.2.2. In Paulin [89], a *pseudo*-spectral gap based on the spectral gap of $P^{*k}P^k$ for $k \geqslant 1$ is introduced. Precisely, we define

$$\gamma^{ps} = \gamma^{ps}(P) := \max_{k \geqslant 1}\{\gamma(P^{*k}P^k)/k\}.$$

4.2.1 The Metropolis-Hastings kernel

Let π be a probability measure on $(\mathcal{X}, \mathcal{F})$ that is absolutely continuous with density π (with a slight abuse of notation, the density is still denoted by π) with respect to a reference measure μ on $\mathcal{X}$, that is, $\pi(dx) = \pi(x)\mu(dx)$. Denote Q to be any Markov kernel on $\mathcal{X}$, where $Q(x, \cdot)$ is absolutely continuous with density q with respect to μ. π is the so-called *target* distribution, while Q is commonly known as the *proposal* kernel. Define the *acceptance* probabilities $\alpha(x, y)$ by

$$\alpha(x, y) = \begin{cases} \min\left(\dfrac{\pi(y)q(y, x)}{\pi(x)q(x, y)}, 1\right) & \text{if } \pi(x)q(x, y) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Let $p(x, y) := \alpha(x, y)q(x, y)$, and define the *reject* probabilities $r : \mathcal{X} \to [0, 1]$ via $r(x) := 1 - \int p(x, y)\mu(dy)$. The Metropolis-Hastings kernel P is given by

$$P(x, dy) = p(x, y)\mu(dy) + r(x)\delta_x(dy),$$

where δ_x is the point mass at x.

The MH kernel allows for the following algorithmic interpretation. First, we choose X_0 and given the current state X_n, we generate the *proposal* Y_{n+1} by $Q(X_n, \cdot)$. With probability $\alpha(X_n, Y_{n+1})$, we accept the proposal and set $X_{n+1} = Y_{n+1}$. Otherwise, we reject Y_{n+1} and set $X_{n+1} = X_n$. Finally, we set $n = n + 1$ and the above procedure is repeated.

Markov chain Monte Carlo (MCMC) methods, such as the classical Metropolis-Hastings algorithm, involve constructing a Markov chain which converges to a desired stationary distribution π that one would like to sample from. It differs from Monte Carlo methods in the sense that π is often difficult to simulate directly, and is particularly useful in situations where we only know π up to normalization constants. As described in Roberts and Rosenthal [94], we can see that the choice of the proposal kernel Q has an significant impact on the performance of MH algorithm. Common choice of Q includes the symmetric MH ($q(x, y) = q(y, x)$), random walk MH ($q(x, y) = q(y - x)$) and independence MH ($q(x, y) = q(y)$).

4.3 Metropolis-Hastings reversiblizations

From now on, unless otherwise stated, we assume X is a ϕ-irreducible Markov chain, which may not be reversible, with transition kernel P and stationary distribution π. We also assume that both $P(x, \cdot)$ and π share a common dominating reference measure μ on $(\mathcal{X}, \mathcal{F})$, with density denoted by $p(x, \cdot)$ and $\pi(\cdot)$, respectively. Furthermore, we assume that the set $\{x : \pi(x) = 0\}$ is a μ-null set.

Given a Markov chain X with transition kernel P and stationary distribution π, we can obtain a *MH-reversiblized* chain by taking the proposal kernel to be P. The resulting process is what we called *the first MH chain*.

Definition 4.3.1 (The first MH kernel)**.** The first MH chain, with transition kernel denoted by $M_1 := M_1(P)$, is the MH kernel with proposal kernel P and target distribution π. That is, let

$$\alpha_1(x,y) = \begin{cases} \min\left(\dfrac{\pi(y)p(y,x)}{\pi(x)p(x,y)}, 1\right) & \text{if } \pi(x)p(x,y) > 0, \\ 0 & \text{otherwise,} \end{cases}$$

$$m_1(x,y) = \alpha_1(x,y)p(x,y) = \min\{p^*(x,y), p(x,y)\},$$
$$r_1(x) = \int_{y \neq x} (1 - \alpha_1(x,y))p(x,y)\,\mu(dy),$$

then M_1 is given by

$$M_1(x,dy) = m_1(x,y)\mu(dy) + r_1(x)\delta_x(dy).$$

By taking a closer look at m_1, we can see that the first MH chain *weakens* the transition to $A_x := \{y \in \mathcal{X} : \alpha_1(x,y) < 1\}$, and *follows* the same transition as the original chain X for $A_x^c = \{y : \alpha_1(x,y) = 1\}$. This motivates us to develop what we call *the second MH kernel* $M_2 := M_2(P)$ with density m_2, which captures the *opposite* transition of M_1. Precisely, we would like to have

$$m_2(x,y) = \begin{cases} p(x,y) & \text{if } y \in A_x, \\ p^*(x,y) & \text{if } y \in A_x^c \backslash \{x\}. \end{cases} = \max\{p^*(x,y), p(x,y)\}.$$

As a result, we obtain the following:

Definition 4.3.2 (The second MH kernel)**.** The second MH kernel M_2 and density m_2 are given by

$$m_2(x,y) = \max\{p^*(x,y), p(x,y)\},$$
$$M_2(x,dy) = m_2(x,y)\mu(dy) - r_1(x)\delta_x(dy).$$

Note that M_2 in general may not be a Markov kernel, since there is no guarantee that $M_2(x, \{x\}) = P(x, \{x\}) - r_1(x) \geqslant 0$. For instance, if P is the transition kernel of a finite Markov chain with $P(x, x) = 0$ for all $x \in \mathcal{X}$, then $M_2(x, x) = -r_1(x) \leqslant 0$. In the following we collect a few elementary properties of M_1 and M_2.

Lemma 4.3.1. *Suppose that P is a Markov kernel with stationary measure π, with M_1 and M_2 being the first and second MH kernel of P respectively. Then the following holds.*

(i) $P + P^* = M_1 + M_2$. *In particular,* $M_2(x, \mathcal{X}) = 1$, *for* $x \in \mathcal{X}$.

(ii) M_1 *and* M_2 *are reversible with respect to* π.

(iii) $M_1 = M_2 = P$ *if and only if* P *is reversible with respect to* π.

(iv) $M_i(P) = M_i(P^*)$ *for* $i = 1, 2$.

Proof. (i) This can easily be seen from Definition 4.3.1 and 4.3.2 together with

$$p(x, y) + p^*(x, y) = \min\{p^*(x, y), p(x, y)\} + \max\{p^*(x, y), p(x, y)\}.$$

(ii) It is well known that M_1 is reversible (and hence invariant) w.r.t. π, see e.g. Roberts and Rosenthal [94]. To see that M_2 is reversible w.r.t. π, we use (i). That is,

$$M_2^* = (P + P^* - M_1)^* = P^* + P - M_1 = M_2.$$

(iii) If $M_1 = M_2 = P$, then (i) gives $P^* = M_1 + M_2 - P = P$. Conversely, when P is reversible w.r.t. π, then $\alpha(x, y) = 1$ $\mu \times \mu$ *a.e.*, and hence $M_1 = P$ by Definition 4.3.1. It follows again from (i) that $M_2 = P + P^* - M_1 = P$.

(iv) Using the fact that $P^{**} = P$ and Definition 4.3.1, $M_1(P) = M_1(P^*)$. Next, (i) gives

$$M_2(P) = P + P^* - M_1(P) = P^* + P - M_1(P^*) = M_2(P^*).$$

□

Remark 4.3.1. As remarked earlier, although M_2 is not a Markov kernel in general, it is reversible w.r.t. π, and satisfies $\pi M_2 = \pi(P + P^* - M_1) = \pi$.

Remark 4.3.2. If one would like to define M_2 as a Markov kernel, we can divide m_2 by 2 and put the remaining probability mass back to x to obtain:

$$
\begin{aligned}
m_2(x, y) &= \frac{1}{2} \max\{p^*(x, y), p(x, y)\}, \\
r_2(x) &= 1 - \int m_2(x, y)\, \mu(dy), \\
M_2(x, dy) &= m_2(x, y)\mu(dy) + r_2(x)\delta_x(dy).
\end{aligned}
$$

In this definition, we provide an algorithmic interpretation for M_2. We first fix X_0. Given X_n, we flip an unbiased coin to decide which kernel (P or P^*) to follow. If the resulting kernel is P (resp. P^*), we generate the *proposal* Y_{n+1} by $P(X_n, \cdot)$ (resp. Y^*_{n+1} by $P^*(X_n, \cdot)$). If $Y_{n+1} \in A_{X_n}$ (resp. $Y^*_{n+1} \in A^c_{X_n}$), then the proposal is accepted and we set $X_{n+1} = Y_{n+1}$ (resp. $X_{n+1} = Y^*_{n+1}$). Otherwise, the proposal is rejected and $X_{n+1} = X_n$. Finally, we set $n = n + 1$ and the above procedure is repeated.

4.3.1 Peskun Ordering

We aim to investigate some further relationships and properties of the spectra of P, M_1 and M_2 via the so-called Peskun ordering, which was first introduced by Peskun [90] as a partial ordering for Markov transition kernels on finite state space, and was further generalized by Tierney [108] to general state space.

Definition 4.3.3 (Peskun Ordering)**.** Suppose that P_1, P_2 are transition kernels with invariant distribution π. P_1 dominates P_2 off the diagonal, written as $P_1 \succeq P_2$, if for π-almost all x $P_1(x, A) \geqslant P_2(x, A)$ for all $A \in \mathcal{F}$ with $x \notin A$.

Note that we are not restricting to Markov transition kernels in Definition 4.3.3, since M_2 in general may not be a Markov kernel. Even in this setting, we can still demonstrate that the results obtained by Tierney [108] holds in the following lemma:

Lemma 4.3.2. *Suppose that P is a Markov kernel with stationary measure π, with M_1 and M_2 being the first and second MH kernel of P respectively. We have the following:*

(i) $M_2 \succeq P \succeq M_1$.

(ii) *$P - M_2$, $M_1 - P$ and $M_1 - M_2$ are positive operators.*

Proof. (i) For $x \in \mathcal{X}$ and $A \in \mathcal{F}$ with $x \notin A$,

$$\begin{aligned} M_2(x, A) &= \int_A \max\{p(x,y), p^*(x,y)\}\mu(dy) \geqslant P(x, A) \\ &\geqslant \int_A \min\{p(x,y), p^*(x,y)\}\mu(dy) = M_1(x, A). \end{aligned}$$

(ii) We modify the proof of Lemma 3 in [108] to cater for the case where M_2 may not be a Markov kernel. Let $H(dx, dy) = \pi(dx)(\delta_x(dy) - P(x, dy) + M_2(x, dy))$. Lemma 4.3.1 yields $H(A) \geqslant 0$, $H(\mathcal{X} \times \mathcal{X}) = 1$, $H(\mathcal{X} \times B) = H(B \times \mathcal{X}) = \pi(B)$ for $A \in \mathcal{F} \otimes \mathcal{F}$ and $B \in \mathcal{F}$. The rest of the proof are the same as the proof of Lemma 3 in [108].

□

Corollary 4.3.1. *Suppose that P is a Markov kernel with stationary measure π, with M_1 and M_2 being the first and second MH kernel of P respectively. Using the notation defined in (4.2.2), we obtain:*

$$\lambda(M_2) \leqslant \inf_{\substack{f \in L_0^2(\pi) \\ ||f||_\pi = 1}} \langle Pf, f \rangle_\pi \leqslant \lambda(M_1), \tag{4.3.1}$$

$$\Lambda(M_2) \leqslant \sup_{\substack{f \in L_0^2(\pi) \\ ||f||_\pi = 1}} \langle Pf, f \rangle_\pi \leqslant \Lambda(M_1). \tag{4.3.2}$$

Proof. Lemma 4.3.2 leads to

$$\langle M_2 f, f\rangle_\pi \leqslant \langle Pf, f\rangle_\pi \leqslant \langle M_1 f, f\rangle_\pi.$$

Desired result follows by taking infimum or supremum over $\{f \in L_0^2(\pi) : ||f||_\pi = 1\}$, (4.2.2) and Lemma 4.3.1(i). $\square$

Inspired by Corollary 4.3.1 and (4.2.3), we will introduce a *pseudo*-spectral gap (that we will call the MH-spectral gap) based on $\lambda(M_2)$ and $\Lambda(M_1)$ in Section 4.5. We will obtain a number of new bounds based on this gap.

4.3.2 Weyl's inequality for additive reversiblization

In this section, we introduce Weyl's inequality for the additive reversiblization for finite Markov chains, which allows us to give upper and lower bound on the eigenvalues of $(P + P^*)/2$, in terms of the eigenvalues of $M_1(P)$ and $M_2(P)$. Assume that P is a stochastic matrix on a finite state space $\mathcal{X}$ with stationary distribution π, with eigenvalues-eigenvectors denoted by $(\lambda_j(P), \phi_j(P))_{j=1}^{|\mathcal{X}|}$. If P is a self-adjoint matrix, we arrange its eigenvalues in non-increasing order by $\lambda_1(P) \geqslant \ldots \geqslant \lambda_n(P)$, where $n := |\mathcal{X}|$.

Theorem 4.3.1 (Weyl's inequality for additive reversiblization)**.** *Assume that P is a $n \times n$ stochastic matrix with stationary distribution π, with M_1, M_2 to be the first and second MH-kernel.*

(i) *For integers i, j, k such that $1 \leqslant i, j, k \leqslant n$ and $i + 1 = j + k$,*

$$\lambda_i(P + P^*) \leqslant \lambda_j(M_1) + \lambda_k(M_2).$$

Equality holds if and only if there exists a vector f with $||f||_{l^2(\pi)} = 1$ such that $M_1 f = \lambda_j f, M_2 f = \lambda_k f$ and $(P + P^*)f = \lambda_i f$.

(ii) *For integers i, l, m such that $1 \leqslant i, l, m \leqslant n$ and $i + n = l + m$,*

$$\lambda_i(P + P^*) \geqslant \lambda_l(M_1) + \lambda_m(M_2).$$

Equality holds if and only if there exists a vector f with $||f||_{l^2(\pi)} = 1$ such that $M_1 f = \lambda_l f, M_2 f = \lambda_m f$ and $(P + P^*)f = \lambda_i f$.

Proof. Thanks to Lemma 4.3.1(i), $P + P^* = M_1 + M_2$, where both M_1 and M_2 are self-adjoint matrices in $l^2(\pi)$. Desired results follow directly from Weyl's inequality, see e.g. Theorem 4.3.1 in Horn and Johnson [49]. $\square$

Since $\gamma(P) = \gamma((P + P^*)/2)$, we can obtain bounds on the spectral gap of P in terms of the eigenvalues of M_1, M_2.

Corollary 4.3.2. *With the assumptions of Theorem 4.3.1, we have*

$$1 - \frac{1}{2}U \leqslant \gamma(P) \leqslant 1 - \frac{1}{2}L,$$

where $L := \max_{l+m=2+n}\{\lambda_l(M_1) + \lambda_m(M_2)\}$ and $U := \min_{j+k=3}\{\lambda_j(M_1) + \lambda_k(M_2)\}$.

Proof. We take $i = 2$ in Theorem 4.3.1 to obtain

$$\frac{1}{2}L \leqslant \lambda_2((P + P^*)/2) \leqslant \frac{1}{2}U.$$

Desired result follows by using $\gamma(P) = \gamma((P + P^*)/2) = 1 - \lambda_2((P + P^*)/2)$. $\square$

4.3.3 Examples: asymmetric random walk and birth-death processes with vortices

In this section, we first show that the bounds in Corollary 4.3.2 are sharp by studying the asymmetric simple random walk on n-cycle and on discrete torus. We then proceed to give spectral gap bounds for birth-death processes with vortices.

Example 4.3.1 (Asymmetric simple random walk on the n-cycle)**.** We first recall the asymmetric simple random walk on the n-cycle. We take $\mathcal{X} = \{0, 1, \ldots, n-1\}$ and the transition matrix to be $P(j,k) = p$ for $k = j+1 \mod n$, $P(j,k) = q = 1-p$ for $k = j-1 \mod n$ and 0 otherwise. Its stationary distribution is given by $\pi(i) = 1/n$ for all $i \in \mathcal{X}$, and its time-reversal has transition matrix given by $P^* = P^T$, the transpose of P. In the particular case when $p = q = 1/2$, we recover the symmetric random walk with eigenvalues $(\cos(2\pi j/n))_{j=0}^{n-1}$, which have been studied in Diaconis and Stroock [32], Fill [42], Levin et al. [74].

We denote $l := \min\{p, q\}$ and $r := \max\{p, q\}$. Then M_1 and M_2 are given by, for $j \in \mathcal{X}$,

$$M_1(j,k) = l, \quad \text{for } k = j \pm 1 \mod n, \quad M_1(j,j) = 1-2l,$$
$$M_2(j,k) = r, \quad \text{for } k = j \pm 1 \mod n, \quad M_2(j,j) = 1-2r.$$

Note that M_2 is not a Markov kernel unless $r = p = q = 1/2$. For $p \neq 1/2$, we can interpret M_2 as $M_2 = G + I$, where $G := M_2 - I$ is the Markov generator on $\mathcal{X}$. Using the notation of Section 4.3.2 and observe that the additive reversiblization is the simple symmetric random walk, the unordered eigenvalues of $(P+P^*)/2$, M_1 and M_2 (see Example $3.1, 3.2$ in Fill [42]) are, for $i \in \{1, \ldots, n\}$,

$$\lambda_i((P+P^*)/2) = \cos(2\pi(i-1)/n),$$

$$\lambda_i(M_1) = 1 - 2l(1 - \cos(2\pi(i-1)/n)),$$

$$\lambda_i(M_2) = 1 - 2r(1 - \cos(2\pi(i-1)/n)),$$

so Corollary 4.3.2 now reads $L = 2\cos(2\pi/n)$, $U = 2 - 2r(1 - \cos(2\pi/n))$ and

$$r(1 - \cos(2\pi/n)) = 1 - \frac{1}{2}U \leqslant \gamma(P) = 1 - \cos(2\pi/n) = 1 - \frac{1}{2}L,$$

that is, the upper bound is exactly attained and the lower bound is sharp in n.

Example 4.3.2 (Asymmetric simple random walk on discrete torus)**.** This example investigates the asymmetric simple random walk on discrete torus $\mathbb{Z}_n^d = (\mathbb{Z}\backslash n\mathbb{Z})^d$, in which we build a product chain via the asymmetric kernel on n-cycle that we studied in Example 4.3.1 and we also adapt the notations therein. That is, we choose one of the d coordinates at random and it will move according to the kernel $P(j,k) = p$ for $k = j+1 \mod n$, $P(j,k) = q = 1-p$ for $k = j-1 \mod n$ and 0 otherwise. Denote the transition kernel (resp. first Metropolis kernel, second Metropolis kernel) on $\mathbb{Z}_n^d$ by $\widetilde{P}$ (resp. $\widetilde{M_1}$, $\widetilde{M_2}$), then we have

$$\widetilde{P} = \frac{1}{d}\sum_{i=1}^{d} \underbrace{I \otimes \cdots \otimes I}_{i-1} \otimes P \otimes \underbrace{I \otimes \cdots \otimes I}_{d-i},$$

$$\widetilde{M_1} = \frac{1}{d}\sum_{i=1}^{d} \underbrace{I \otimes \cdots \otimes I}_{i-1} \otimes M_1 \otimes \underbrace{I \otimes \cdots \otimes I}_{d-i},$$

$$\widetilde{M_2} = \frac{1}{d}\sum_{i=1}^{d} \underbrace{I \otimes \cdots \otimes I}_{i-1} \otimes M_2 \otimes \underbrace{I \otimes \cdots \otimes I}_{d-i}.$$

Note that the stationary distribution is the uniform distribution on $\mathbb{Z}_n^d$. The unordered eigenvalues of $(\widetilde{P} + \widetilde{P}^*)/2$, $\widetilde{M_1}$ and $\widetilde{M_2}$ are, for $i \in \{1, \ldots, n\}$,

$$\lambda_i((\widetilde{P} + \widetilde{P}^*)/2) = (d-1)/d + \cos(2\pi(i-1)/n)/d,$$

$$\lambda_i(\widetilde{M_1}) = 1 - 2l(1 - \cos(2\pi(i-1)/n))/d,$$

$$\lambda_i(\widetilde{M_2}) = 1 - 2r(1 - \cos(2\pi(i-1)/n))/d.$$

and Corollary 4.3.2 now reads $L = 2 - 2/d + 2\cos(2\pi/n)/d$, $U = 2 - 2r(1 - \cos(2\pi/n))/d$ and

$$r(1-\cos(2\pi/n))/d = 1 - \frac{1}{2}U \leqslant \gamma(P) = \frac{1-\cos(2\pi/n)}{d} = 1 - \frac{1}{2}L,$$

that is, the upper bound is exactly attained and the lower bound is sharp in n.

Example 4.3.3 (Inserting vortices to birth-death processes)**.** Giving two-sided precise spectral gap bounds for non-reversible Markov chains is well-known to be a difficult task. For the spectral gap estimates of birth-death processes, we refer interested readers to Chen [17]. We aim at using this example to show how we can give such type of estimates by means of MH reversibilization. This example is inspired by Bierkens [6], Sun et al. [106], which offers an interesting way to artificially create non-reversible Markov chains from reversible ones via perturbation or inserting vortices. It is perhaps more suitable to work in the setting of continuous-time Markov chains. We write G^{BD} to be the infinitesimal generator of a birth-death process with birth rate b_i and death rate d_i for $i \in \mathcal{X} = \mathbb{N}_0$ with stationary distribution $\pi(i)$. Next, we denote V to be the n-dimensional cyclic vortices given by $V(i,i) = -1/\pi(i)$ and $V(i,j) = 1/\pi(i)$ for $j = (i+1) \mod n$ for $i \in \{0,\ldots,n\}$. By Corollary 1 in Sun et al. [106], $G := G^{BD}+V$ is the generator of a non-reversible Markov chain on $\mathcal{X}$ with stationary distribution π.

To analyze the left spectral gap $\gamma(G)$ of $-G$, the construction of M_1 and M_2 applies essentially in verbatim to G as in Section 4.3. In particular, $G^* = G^{BD} + V^*$, $M_1(G) = G^{BD}$ and $M_2(G) = G^{BD} + V + V^*$, so by Corollary 4.3.1, we have

$$\gamma(G^{BD}) \leqslant \gamma(G) \leqslant \gamma(G^{BD}) + \gamma(V+V^*) \leqslant \gamma(G^{BD}) + \frac{2}{\min_{i\in\{0,\ldots,n\}}\pi(i)}(1-\cos(2\pi/n)),$$

where we further upper bound the left spectral gap of $V + V^*$ by the symmetric random walk on n-cycle with birth and death rate $1/\min_{i\in\{0,\ldots,n\}}\pi(i)$. We can then specialize into various well-known examples of birth-death processes, in which we summarize the results below:

Process	spectral gap bounds
Ehrenfest with vortices	$1 \leqslant \gamma(G) \leqslant 1+2(1-\cos(2\pi/n)) \max\{p^{-n},(1-p)^{-n}\}$
$M/M/1$ with vortices	$(\sqrt{\mu}-\sqrt{\lambda})^2 \leqslant \gamma(G) \leqslant (\sqrt{\mu}-\sqrt{\lambda})^2+2(1-\lambda/\mu)^{-1}(\mu/\lambda)^{n-1}(1-\cos(2\pi/n))$
$M/M/\infty$ with vortices	$1 \leqslant \gamma(G) \leqslant 1+2(1-\cos(2\pi/n))e^{\lambda} \max_{i\in\{0,\ldots,n\}} i!\lambda^{-i}$
GWI with vortices	$1-\lambda \leqslant \gamma(G) \leqslant 1-\lambda+2(1-\cos(2\pi/n)) \max_{i\in\{0,\ldots,n\}} \dfrac{\Gamma(r)i!}{\Gamma(r+i)}(1-\lambda)^{-r}\lambda^{-i}$

Table 4.1: Spectral gap bounds for various birth-death processes with n-dimensional cyclic vortices

For the Ehrenfest model with cyclic vortices, it is constructed from a birth-death process with $b_i = p(n-i)$, $d_i = (1-p)i$ with $0 < p < 1$ on $\mathcal{X} = \{0, \ldots, n\}$ and π being the binomial distribution with parameters n and p. For $M/M/1$ with vortices, it is constructed from a birth-death process with $b_i = \lambda$, $d_i = \mu$ with $\mu > \lambda$ and $\pi(i) = (1-\lambda/\mu)(\lambda/\mu)^{i-1}$. For $M/M/\infty$ with vortices, it is constructed from a birth-death process with $b_i = \lambda$, $d_i = i$ and π being the Poisson distribution with mean λ. For the Galton-Watson process with immigration (GWI) and vortices, we have $b_i = \lambda(r+i)$, $d_i = i$ and π being the negative binomial distribution with parameters λ and r.

4.4 Pseudospectral expansion

As a consequence of Lemma 4.3.1(ii), M_1 and M_2 are self-adjoint operators on $L^2(\pi)$, which help us to obtain a *pseudospectral* expansion of P in terms of the spectral measures of M_1 and M_2.

Theorem 4.4.1. *Denote P to be a Markov kernel with stationary distribution π, and M_i to be the MH kernel (defined in def. 4.3.1 and def. 4.3.2) with spectral measure $\mathcal{E}_i$ for $i = 1, 2$. For $x \in \mathcal{X}$, $B \in \mathcal{F}$ with $x \notin B$, we have*

$$P(x, B) = \int_{\sigma(M_1)} \lambda\, \delta_x \mathcal{E}_1(d\lambda)(B \cap A_x^c) + \int_{\sigma(M_2)} \lambda\, \delta_x \mathcal{E}_2(d\lambda)(B \cap A_x), \qquad (4.4.1)$$

$$P(x,\{x\}) = \frac{1}{2}\left(\int_{\sigma(M_1)} \lambda\, \delta_x \mathcal{E}_1(d\lambda)(\{x\}) + \int_{\sigma(M_2)} \lambda\, \delta_x \mathcal{E}_2(d\lambda)(\{x\})\right), \tag{4.4.2}$$

$$P^*(x,B) = \int_{\sigma(M_1)} \lambda\, \delta_x \mathcal{E}_1(d\lambda)(B\cap A_x) + \int_{\sigma(M_2)} \lambda\, \delta_x \mathcal{E}_2(d\lambda)(B\cap A_x^c), \tag{4.4.3}$$

where we recall that $A_x := \{y \in E : \alpha_1(x,y) < 1\}$.

Proof. We first show (4.4.1). By Lemma 4.3.1 and (4.2.1), for $i = 1,2$,

$$M_i = \int_{\sigma(M_i)} \lambda\, \mathcal{E}_i(d\lambda). \tag{4.4.4}$$

Therefore, we can deduce that

$$\begin{aligned}
P(x,B) &= P(x, B\cap A_x^c) + P(x, B\cap A_x) \\
&= \int_{B\cap A_x^c} P(x,dy) + \int_{B\cap A_x} P(x,dy) \\
&= \int_{B\cap A_x^c} M_1(x,dy) + \int_{B\cap A_x} M_2(x,dy) \quad (\mathbb{1}_B(x) = 0) \\
&= \delta_x M_1(B\cap A_x^c) + \delta_x M_2(B\cap A_x) \\
&= \int_{\sigma(M_1)} \lambda\, \delta_x \mathcal{E}_1(d\lambda)(B\cap A_x^c) + \int_{\sigma(M_2)} \lambda\, \delta_x \mathcal{E}_2(d\lambda)(B\cap A_x). \quad \text{(By (4.4.4))}
\end{aligned}$$

Next, in view of Lemma 4.3.1, we have

$$\begin{aligned}
P(x,\{x\}) &= \frac{1}{2}\left(M_1(x,\{x\}) + M_2(x,\{x\})\right) \\
&= \frac{1}{2}\left(\int_{\sigma(M_1)} \lambda\, \delta_x \mathcal{E}_1(d\lambda)(\{x\}) + \int_{\sigma(M_2)} \lambda\, \delta_x \mathcal{E}_2(d\lambda)(\{x\})\right),
\end{aligned}$$

which gives (4.4.2). Finally, to show (4.4.3), we follow a very similar proof of (4.4.1) that leads to

$$\begin{aligned}
P^*(x,B) &= P^*(x, B\cap A_x) + P^*(x, B\cap A_x^c) \\
&= \delta_x M_1(B\cap A_x) + \delta_x M_2(B\cap A_x^c) \\
&= \int_{\sigma(M_1)} \lambda\, \delta_x \mathcal{E}_1(d\lambda)(B\cap A_x) + \int_{\sigma(M_2)} \lambda\, \delta_x \mathcal{E}_2(d\lambda)(B\cap A_x^c).
\end{aligned}$$

□

Remark 4.4.1. When P is reversible, Lemma 4.3.1 yields $P = M_1 = M_2$, so (4.4.1) and (4.4.2) reduces to

$$P(x, B) = \int_{\sigma(P)} \lambda\, \delta_x \mathcal{E}_1(d\lambda)(B),$$
$$P(x, \{x\}) = \int_{\sigma(P)} \lambda\, \delta_x \mathcal{E}_1(d\lambda)(\{x\}),$$

which are expected expressions (since we can invoke the spectral theorem directly on P).

Remark 4.4.2. An alternative expression for $P(x, \{x\})$ is the following: Using (4.4.1) (with B replaced by $\mathcal{X}\backslash\{x\}$), we observe that

$$\begin{aligned}
P(x, \{x\}) &= 1 - P(x, \mathcal{X}\backslash\{x\}) \\
&= 1 - \int_{\sigma(M_1)} \lambda\, \delta_x \mathcal{E}_1(d\lambda)(\mathcal{X}\backslash\{x\} \cap A_x^c) - \int_{\sigma(M_2)} \lambda\, \delta_x \mathcal{E}_2(d\lambda)(\mathcal{X}\backslash\{x\} \cap A_x) \\
&= \int_{\sigma(M_1)} \lambda\, \delta_x \mathcal{E}_1(d\lambda)(\{x\} \cup A_x) - \int_{\sigma(M_2)} \lambda\, \delta_x \mathcal{E}_2(d\lambda)(A_x).
\end{aligned}$$

To compute the pseudospectral expansion of the n-step transition kernel P^n, we can make use of the Chapman-Kolmogorov equation together with (4.4.1) and (4.4.2). Equivalently, we can replace P by P^n in Theorem 4.4.1, which leads to:

Corollary 4.4.1. *Denote P to be a Markov kernel with stationary measure π, so that P^n is the n-step transition kernel for $n \in \mathbb{N}$. Let $M_i(P^n)$ to be the MH kernel (defined in def. 4.3.1 and def. 4.3.2) with spectral measure $\mathcal{E}_i(P^n)$ for $i = 1, 2$. For $x \in \mathcal{X}$, $B \in \mathcal{F}$ with $x \notin B$, we have*

$$P^n(x, B) = \int_{\sigma(M_1(P^n))} \lambda\, \delta_x \mathcal{E}_1(P^n)(d\lambda)(B \cap A_x^c) + \int_{\sigma(M_2(P^n))} \lambda\, \delta_x \mathcal{E}_2(P^n)(d\lambda)(B \cap A_x), \tag{4.4.5}$$

$$P^n(x, \{x\}) = \frac{1}{2} \left(\int_{\sigma(M_1(P^n))} \lambda\, \delta_x \mathcal{E}_1(P^n)(d\lambda)(\{x\}) + \int_{\sigma(M_2(P^n))} \lambda\, \delta_x \mathcal{E}_2(P^n)(d\lambda)(\{x\}) \right), \tag{4.4.6}$$

$$P^{*n}(x, B) = \int_{\sigma(M_1(P^n))} \lambda\, \delta_x \mathcal{E}_1(P^n)(d\lambda)(B \cap A_x) + \int_{\sigma(M_2(P^n))} \lambda\, \delta_x \mathcal{E}_2(P^n)(d\lambda)(B \cap A_x^c), \tag{4.4.7}$$

where $A_x := \{y : \alpha_1(P^n)(x, y) < 1\}$.

Next, we specialize into the case of finite Markov chains, as more explicit results can be obtained.

Corollary 4.4.2. *Suppose that* P *is a stochastic matrix on a finite state space* $\mathcal{X}$ *with stationary distribution* π. *Let* $M_i(P^n)$ *be the MH kernel with eigenvalues-eigenvectors denoted by* $(\lambda_j^i, \phi_j^i)_{j=1}^{|\mathcal{X}|}$ *for* $i = 1, 2$ *(note that the dependence of* (λ_j^i, ϕ_j^i) *on* P^n *is suppressed). For* $x \in \mathcal{X}$ *and* $f \in l^2(\pi)$, *we have*

$$P^n(x, y) = \begin{cases} \sum_{j=1}^{|\mathcal{X}|} \lambda_j^2 \phi_j^2(x) \phi_j^2(y) \pi(y) & \text{if } y \in A_x, \\ \sum_{j=1}^{|\mathcal{X}|} \lambda_j^1 \phi_j^1(x) \phi_j^1(y) \pi(y) & \text{if } y \in A_x^c \backslash \{x\}, \\ \dfrac{1}{2}(\sum_{j=1}^{|\mathcal{X}|} \lambda_j^1 \phi_j^1(x) \phi_j^1(x) \pi(x) + \sum_{j=1}^{|\mathcal{X}|} \lambda_j^2 \phi_j^2(x) \phi_j^2(x) \pi(x)) & \text{if } y = x. \end{cases} \tag{4.4.8}$$

$$P^n f(x) = \sum_{j=1}^{|\mathcal{X}|} \lambda_j^1 \phi_j^1(x) \langle f \mathbb{1}_{A_x^c \backslash \{x\}}, \phi_j^1 \rangle_\pi + \sum_{j=1}^{|\mathcal{X}|} \lambda_j^2 \phi_j^2(x) \langle f \mathbb{1}_{A_x}, \phi_j^2 \rangle_\pi + P^n(x, x) f(x), \tag{4.4.9}$$

where $A_x := \{y : \alpha_1(P^n)(x, y) < 1\}$.

Proof. The proof of (4.4.8) follows from (4.4.5) and (4.4.6). To see (4.4.9), we decompose $P^n f(x)$ into

$$\begin{aligned} P^n f(x) &= P^n f \mathbb{1}_{A_x^c \backslash \{x\}}(x) + P^n f \mathbb{1}_{A_x}(x) + P^n f \mathbb{1}_{\{x\}}(x) \\ &= M_1(P^n) f \mathbb{1}_{A_x^c \backslash \{x\}}(x) + M_2(P^n) f \mathbb{1}_{A_x}(x) + P^n f \mathbb{1}_{\{x\}}(x). \end{aligned} \tag{4.4.10}$$

Now, since $(\phi_j^i)_{j=1}^{|\mathcal{X}|}$ is a basis on $l^2(\pi)$ for $i = 1, 2$, we have

$$M_1(P^n)f\mathbb{1}_{A_x^c\backslash\{x\}}(x) = \sum_{j=1}^{|\mathcal{X}|} \lambda_j^1 \phi_j^1(x)\langle f\mathbb{1}_{A_x^c\backslash\{x\}}, \phi_j^1\rangle_\pi, \tag{4.4.11}$$

$$M_2(P^n)f\mathbb{1}_{A_x}(x) = \sum_{j=1}^{|\mathcal{X}|} \lambda_j^2 \phi_j^2(x)\langle f\mathbb{1}_{A_x}, \phi_j^2\rangle_\pi. \tag{4.4.12}$$

Desired result follows by collecting (4.4.10), (4.4.11) and (4.4.12). □

4.5 Geometric ergodicity, mixing time and MH-spectral gap

We will measure the speed of convergence to stationarity by the total variation distance, which is defined to be:

Definition 4.5.1 (Total variation distance)**.** The total variation distance between two probability measures μ and ν is given by

$$||\mu - \nu||_{TV} := \sup_A |\mu(A) - \nu(A)| = \frac{1}{2} \sup_{|f| \leqslant 1} |\mu(f) - \nu(f)|.$$

where $|f| := \sup_{x\in\mathcal{X}} |f(x)|$.

We refer the readers to Levin et al. [74], Meyn and Tweedie [78] and Roberts and Rosenthal [94] for further properties of the total variation distance. We can now define various notions of ergodicity of a transition kernel P.

Definition 4.5.2 (Geometric ergodic, π-a.e. geometrically ergodic, uniformly ergodic, mixing time)**.** Suppose that P is a transition kernel with stationary measure π. P is geometrically ergodic if for each probability measure μ, there exists $C_\mu < \infty$ and $\rho \in [0, 1)$ such that

$$||\mu P^n - \pi||_{TV} \leqslant C_\mu \rho^n, \quad n \in \mathbb{N}. \tag{4.5.1}$$

If (4.5.1) holds with $\mu = \delta_x$ for π-a.e. x, then P is called π-a.e. geometrically ergodic. P is uniformly ergodic if there exists $C < \infty$ and $\rho \in [0, 1)$ such that

$$d(n) := \sup_{x \in \mathcal{X}} ||P^n(x, \cdot) - \pi||_{TV} \leqslant C\rho^n, \quad n \in \mathbb{N}. \tag{4.5.2}$$

The mixing time $t_{mix}(\epsilon)$ is defined to be

$$t_{mix}(\epsilon) := \min\{n : d(n) \leqslant \epsilon\}.$$

For Markov kernels that are reversible w.r.t. π, we have the following characterization of geometric ergodicity in terms of the L^2-spectral gap.

Theorem 4.5.1. *Suppose that P is reversible with respect to π. The following statements are equivalent:*

(i) *P is geometrically ergodic.*

(ii) *P admits a L^2-spectral gap, i.e. $\gamma = 1 - \beta = 1 - \max\{|\lambda|, \Lambda\} > 0$.*

The proof can be found in Roberts and Rosenthal [93].

4.5.1 Main results

Following from the result in Corollary 4.3.1 and (4.2.3), we can define a *pseudo*-spectral gap by taking $1 - \max\{|\lambda(M_2)|, \Lambda(M_1)\}$. However, this gap may not be informative as $|\lambda(M_2)|$ maybe greater than or equal to 1 since M_2 is not a Markov kernel in general. To define a meaningful gap, we should consider M_2 with $|\lambda(M_2)| < 1$. This leads us to the following definition:

Definition 4.5.3 (MH-spectral gap)**.** Suppose that P is a Markov kernel with stationary measure π. Let

$$\mathcal{C} := \{n \in \mathbb{N} : |\lambda(M_2(P^n))| < 1, \Lambda(M_1(P^n)) < 1\}, \quad \mathcal{C}^c := \mathbb{N}\backslash\mathcal{C},$$
$$\beta^{MH} := \sup_{n\in\mathcal{C}}\{|\lambda(M_2(P^n))|^{1/n}, \Lambda(M_1(P^n))^{1/n}\}.$$

The MH-spectral gap $\gamma^{MH} = \gamma^{MH}(P)$ is given by

$$\gamma^{MH} := 1 - \beta^{MH}.$$

In this definition, we insert the idea of “burn-in" in MCMC to define a MH-spectral gap. Precisely, we discard the spectral gaps in $\mathcal{C}^c$, and only consider the gaps in $\mathcal{C}$.

Note that for reversible P, the L^2-spectral gap is equal to the MH-spectral gap. If P is geometrically ergodic, Lemma 4.3.1(iii) and Theorem 4.5.1 lead to $\mathcal{C} = \mathbb{N}$ and

$$\begin{aligned}
\gamma^{MH} &= 1 - \beta^{MH} \\
&= 1 - \sup_{n\in\mathbb{N}}\{|\lambda(P^n)|^{1/n}, \Lambda(P^n)^{1/n}\} \\
&= 1 - \sup_{n\in\mathbb{N}}\{|\lambda(P)|, \Lambda(P)\} \\
&= 1 - \max\{|\lambda(P)|, \Lambda(P)\} = 1 - \beta = \gamma^*.
\end{aligned}$$

If P is not geometrically ergodic, then $\beta^{MH} = \beta = 1$, so $\gamma^{MH} = \gamma^* = 0$.

As a first result, by means of Weyl's inequality, we can show that $M_2 = M_2(P)$ is a contraction whenever P is a finite-state lazy and ergodic Markov kernel.

Theorem 4.5.2. *If P is a finite-state lazy and ergodic Markov kernel, then* $|\lambda(M_2(P))| \leqslant 1$.

Proof. By Weyl's inequality introduced in Theorem 4.3.1, we have

$$\lambda_n(P + P^*) \leqslant \lambda_1(M_1) + \lambda_n(M_2) = 1 + \lambda_n(M_2).$$

Note that laziness of P implies the laziness of $(P + P^*)/2$, which implies $0 \leqslant \lambda_n(P + P^*) \leqslant 1 + \lambda_n(M_2)$. $\square$

Next, we present the main results in this section. Theorem 4.5.3 shows that a MH-spectral gap leads to geometric ergodicity.

Theorem 4.5.3. *Suppose that P is the Markov kernel of a ϕ-irreducible and aperiodic Markov chain with stationary measure π on a countably generated state space $\mathcal{X}$. If $|\mathcal{C}^c| < \infty$ and P admits a MH-spectral gap, i.e. $\gamma^{MH} = 1 - \beta^{MH} > 0$, then P and P^* are geometrically ergodic (and π-a.e. geometrically ergodic).*

Next, we demonstrate a partial converse to Theorem 4.5.3.

Theorem 4.5.4. *[Partial converse of Theorem 4.5.3] Suppose that P is a Markov kernel with stationary measure π. If P and P^* are uniformly ergodic, then there exists a $N \in \mathbb{N}$ such that for all $n \geqslant N$, $M_i(P^n)$ are uniformly ergodic for $i = 1, 2$.*

Recall that in the reversible case Theorem 4.5.1 gives the existence of L^2-spectral gap is equivalent to geometric ergodicity. While we hope for a result that characterizes geometric ergodicity in the non-reversible case by means of the MH-spectral gap, we only manage to show that under a stronger assumption of uniform ergodicity of both P and P^*, $M_i(P^n)$ is uniformly ergodic for sufficiently large n. This implies the existence of L^2-spectral gap of $M_i(P^n)$ for sufficiently large n, yet it is unclear whether β^{MH} is less than 1 (since we are taking supremum in the definition of β^{MH}).

Next, we present a result that gives a mixing time upper bound in terms of the MH-spectral gap.

Corollary 4.5.1. *For a finite Markov chain with transition matrix P that is irreducible,* *if* $|\mathcal{C}^c| < \infty$ *and* P *admits a MH-spectral gap, then*

$$t_{mix}(\epsilon) \leqslant \frac{\log\left(\frac{1}{\epsilon\pi_{min}}\right)}{\gamma^{MH}},$$

where $\pi_{min} := \min_x \pi(x)$.

Corollary 4.5.1 can be compared with the result in the reversible case (Theorem 12.3 in Levin et al. [74]), which shows that

$$t_{mix}(\epsilon) \leqslant \frac{\log\left(\frac{1}{\epsilon\pi_{min}}\right)}{\gamma^{*}}.$$

4.5.2 Proofs of Theorem 4.5.3, Theorem 4.5.4 and Corollary 4.5.1

First, we start with the following result that allows us to control the total variation distance of P^n and P^{*n} to π by means of that of $M_1(P^n)$ and $M_2(P^n)$ and vice versa. The bounds are by no means tight, yet they will serve their purpose to demonstrate geometric ergodicity in the proof of Theorem 4.5.3 and Theorem 4.5.4.

Lemma 4.5.1. *Suppose that* P *is a Markov kernel with stationary measure* π*, and* M_i *to be the MH kernel for* $i = 1, 2$*. For* $x \in \mathcal{X}$*,* $n \in \mathbb{N}$*,*

$$\begin{aligned}
||P^n(x,\cdot) - \pi||_{TV} &\leqslant \frac{3}{2}||M_1(P^n)(x,\cdot) - \pi||_{TV} + \frac{3}{2}||M_2(P^n)(x,\cdot) - \pi||_{TV}, \\
||P^{*n}(x,\cdot) - \pi||_{TV} &\leqslant \frac{3}{2}||M_1(P^n)(x,\cdot) - \pi||_{TV} + \frac{3}{2}||M_2(P^n)(x,\cdot) - \pi||_{TV}, \\
||M_1(P^n)(x,\cdot) - \pi||_{TV} &\leqslant 2||P^n(x,\cdot) - \pi||_{TV} + 2||P^{*n}(x,\cdot) - \pi||_{TV}, \\
||M_2(P^n)(x,\cdot) - \pi||_{TV} &\leqslant 3||P^n(x,\cdot) - \pi||_{TV} + 3||P^{*n}(x,\cdot) - \pi||_{TV}.
\end{aligned}$$

Proof. We use the same idea as in the proof of Theorem 4.4.1. For any $x \in E$ let $A_{x,n} := \{y : \alpha_1(P^n)(x,y) < 1\}$. We have, for all $n \in \mathbb{N}$,

$$\begin{aligned}
||P^n(x,\cdot) - \pi||_{TV} &= \sup_A |P^n(x,A) - \pi(A)| \\
&\leqslant \sup_A |P^n(x, A \cap A_{x,n}) - \pi(A \cap A_{x,n})| \\
&\quad + \sup_A |P^n(x, A \cap A_{x,n}^c \backslash \{x\}) - \pi(A \cap A_{x,n}^c \backslash \{x\})| \\
&\quad + |P^n(x,\{x\}) - \pi(\{x\})| \\
&\leqslant ||M_2(P^n)(x,\cdot) - \pi||_{TV} + ||M_1(P^n)(x,\cdot) - \pi||_{TV} \\
&\quad + \frac{1}{2}|M_1(P^n)(x,\{x\}) - \pi(\{x\})| \\
&\quad + \frac{1}{2}|M_2(P^n)(x,\{x\}) - \pi(\{x\})| \\
&\leqslant \frac{3}{2}||M_1(P^n)(x,\cdot) - \pi||_{TV} + \frac{3}{2}||M_2(P^n)(x,\cdot) - \pi||_{TV}.
\end{aligned}$$

To show the inequality for $||P^{*n}(x,\cdot) - \pi||_{TV}$, we replace P by P^* above and observe that $M_i(P^n) = M_i(P^{*n})$ for $i = 1, 2$ by Lemma 4.3.1(iv). Next, we observe that

$$\begin{aligned}
||M_1(P^n)(x,\cdot) - \pi||_{TV} &\leqslant \sup_{|f| \leqslant 1} |M_1(P^n)(f \mathbb{1}_{A_{x,n}})(x) - \pi(f \mathbb{1}_{A_{x,n}})| \\
&\quad + \sup_{|f| \leqslant 1} |M_1(P^n)(f \mathbb{1}_{A_{x,n}^c \backslash \{x\}})(x) - \pi(f \mathbb{1}_{A_{x,n}^c \backslash \{x\}})| \\
&\quad + \sup_{|f| \leqslant 1} |M_1(P^n)(x,\{x\}) - \pi(\{x\})||f(x)| \\
&\leqslant ||P^{*n}(x,\cdot) - \pi||_{TV} + ||P^n(x,\cdot) - \pi||_{TV} \\
&\quad + \sup_{|f| \leqslant 1} |P^n(x, A_{x,n}) - \pi(A_{x,n})||f(x)| \\
&\quad + \sup_{|f| \leqslant 1} |P^{*n}(x, A_{x,n}^c \backslash \{x\}) - \pi(A_{x,n}^c \backslash \{x\})||f(x)| \\
&= ||P^{*n}(x,\cdot) - \pi||_{TV} + ||P^n(x,\cdot) - \pi||_{TV} \\
&\quad + \sup_{|f| \leqslant 1} |P^n(f(x) \mathbb{1}_{A_{x,n}})(x) - \pi(f(x) \mathbb{1}_{A_{x,n}})| \\
&\quad + \sup_{|f| \leqslant 1} |P^{*n}(f(x) \mathbb{1}_{A_{x,n}^c \backslash \{x\}})(x) - \pi(f(x) \mathbb{1}_{A_{x,n}^c \backslash \{x\}})|
\end{aligned}$$

$$\leqslant 2||P^n(x,\cdot) - \pi||_{TV} + 2||P^{*n}(x,\cdot) - \pi||_{TV}.$$

Finally, using the inequality above together with Lemma 4.3.1(i) and the triangle inequality yields

$$\begin{aligned}||M_2(P^n)(x,\cdot) - \pi||_{TV} &\leqslant ||P^n(x,\cdot) - \pi||_{TV} + ||P^{*n}(x,\cdot) - \pi||_{TV} + ||M_1(P^n)(x,\cdot) - \pi||_{TV} \\ &\leqslant 3||P^n(x,\cdot) - \pi||_{TV} + 3||P^{*n}(x,\cdot) - \pi||_{TV}.\end{aligned}$$

□

Proof of Theorem 4.5.3. Fix $n \in \mathcal{C}$. Since $\beta^{MH} < 1$, both $M_1(P^n)$ and $M_2(P^n)$ admit L^2-spectral gap, that is, $\gamma(M_i(P^n)) > 0$ for $i = 1, 2$. Theorem 2 in Roberts and Rosenthal [93] gives that $M_1(P^n)$ and $M_2(P^n)$ are geometrically ergodic (even though $M_2(P^n)$ may not be a Markov kernel, the proof there will work through as long as $M_2(P^n)$ admits a L^2-spectral gap). By Lemma 4.5.1, we have

$$\begin{aligned}||P^n(x,\cdot) - \pi||_{TV} &\leqslant \frac{3}{2}||M_1(P^n)(x,\cdot) - \pi||_{TV} + \frac{3}{2}||M_2(P^n)(x,\cdot) - \pi||_{TV} \\ &\leqslant \frac{3}{2}C_x^1\beta(M_1(P^n)) + \frac{3}{2}C_x^2\beta(M_2(P^n)) \\ &\leqslant \frac{3}{2}(C_x^1 + C_x^2)(\beta^{MH})^n = \frac{3}{2}(C_x^1 + C_x^2)(1 - \gamma^{MH})^n,\end{aligned}$$

where C_x^i are the constants of geometric ergodicity for $M_i(P^n)$ for $i = 1, 2$ as in Definition 4.5.2, and the third inequality follows from Corollary 4.3.1. For $n \in \mathcal{C}^c$, we can bound it by a similar way. Precisely, let $\beta^{max} = \max_{n\in\mathcal{C}^c}\{\beta(M_2(P^n))\} = \max_{n\in\mathcal{C}^c}\{|\lambda(M_2(P^n))|\}$. Using again Lemma 4.5.1 leads to

$$\begin{aligned}||P^n(x,\cdot) - \pi||_{TV} &\leqslant \frac{3}{2}||M_1(P^n)(x,\cdot) - \pi||_{TV} + \frac{3}{2}||M_2(P^n)(x,\cdot) - \pi||_{TV} \\ &\leqslant \frac{3}{2}(C_x^1 + C_x^2)\beta^{max} \\ &\leqslant \frac{3}{2}(C_x^1 + C_x^2)\frac{\beta^{max}}{(\beta^{MH})^{|\mathcal{C}^c|}}(\beta^{MH})^n.\end{aligned}$$

We have shown that P is π-a.e. geometrically ergodic, and we can extend it to geometric ergodicity by adapting the argument in the last paragraph of page 9 in Roberts and Rosenthal [93] i.e. the direction from Proposition 1 to Theorem 2. (This is the place where we use the assumption of ϕ-irreducibility and aperiodicity on a countably generated state space.) The proof of geometric ergodicity of P^* is the same as above (by replacing P by P^*) and is omitted. $\square$

Proof of Theorem 4.5.4. Since P and P^* are uniformly ergodic, Proposition 7 in Roberts and Rosenthal [94] gives

$$\sup_{x\in\mathcal{X}} ||P^n(x,\cdot)-\pi||_{TV} < \frac{1}{12},$$
$$\sup_{x\in\mathcal{X}} ||P^{*n}(x,\cdot)-\pi||_{TV} < \frac{1}{12},$$

for all sufficiently large n. Therefore, for all sufficiently large n, Lemma 4.5.1 yields

$$\sup_{x\in\mathcal{X}} ||M_1(P^n)(x,\cdot)-\pi||_{TV} \leqslant 2\sup_{x\in\mathcal{X}} ||P^n(x,\cdot)-\pi||_{TV} + 2\sup_{x\in\mathcal{X}} ||P^{*n}(x,\cdot)-\pi||_{TV} < \frac{1}{3},$$
$$\sup_{x\in\mathcal{X}} ||M_2(P^n)(x,\cdot)-\pi||_{TV} \leqslant 3\sup_{x\in\mathcal{X}} ||P^n(x,\cdot)-\pi||_{TV} + 3\sup_{x\in\mathcal{X}} ||P^{*n}(x,\cdot)-\pi||_{TV} < \frac{1}{2}.$$

Desired result follows from Proposition 7 in Roberts and Rosenthal [94]. $\square$

Proof of Corollary 4.5.1. We follow a similar line of reasoning than in the proof of Theorem 12.3 in Levin et al. [74]. For any $x, y \in \mathcal{X}$, if $y \in A_{x,n}$, we have

$$\left|\frac{P^n(x,y)}{\pi(y)} - 1\right| = \left|\frac{M_2(P^n)(x,y)}{\pi(y)} - 1\right| \leqslant \frac{e^{-n\gamma^{MH}}}{\pi_{min}},$$

where the inequality follows from Theorem 12.3 in Levin et al. [74]. Similarly,

$$\left|\frac{P^n(x,y)}{\pi(y)} - 1\right| = \left|\frac{M_1(P^n)(x,y)}{\pi(y)} - 1\right| \leqslant \frac{e^{-n\gamma^{MH}}}{\pi_{min}}, \quad \text{if } y \in A_{x,n}^c\backslash\{x\},$$
$$\left|\frac{P^n(x,y)}{\pi(y)} - 1\right| \leqslant \frac{1}{2}\left|\frac{M_1(P^n)(x,y)}{\pi(y)} - 1\right| + \frac{1}{2}\left|\frac{M_2(P^n)(x,y)}{\pi(y)} - 1\right| \leqslant \frac{e^{-n\gamma^{MH}}}{\pi_{min}}, \quad \text{if } y = x.$$

Lemma 6.13 in Levin et al. [74] gives $||P^n(x,\cdot)-\pi||_{TV} \leqslant \frac{e^{-n\gamma^{MH}}}{\pi_{min}}$, and desired result follows from the definition of t_{mix}. $\square$

4.5.3 Examples

We illustrate the usefulness of the MH-spectral gap using three examples. In the first two cases, both the additive reversiblization and multiplicative reversiblization fail to give insights on the total variation distance from stationarity, however the pseudo-spectral gap and MH-spectral gap can still provide informative bounds.

Example 4.5.1 (non-reversible walk on a triangle)**.** The first example is taken from Montenegro and Tetali [84] Example 5.2. We consider a Markov chain on the triangle $\{0,1,2\}$ with transition probability given by $P(0,1) = P(1,2) = 1$ and $P(2,0) = P(2,1) = 0.5$. The stationary distribution is $\pi(0) = \pi(1) = 0.25$, $\pi(2) = 0.5$. The chain is non-reversible (for example, $P(1,0) = 0$, yet $P^*(1,0) = 1$), with eigenvalues $1, \frac{-1\pm i\sqrt{7}}{4}$. The additive reversiblization bound does not work here, since the chain is not strongly aperiodic. For multiplicative reversiblization, it has been noted in Montenegro and Tetali [84] that $\gamma(PP^*) = 0$, and the conductance bound does not work in this example as well.

The classical bounds fail since the chain, if started at state 0, requires two steps before its total variation distance decreases. Therefore, γ^{ps} and γ^{MH} are expected to give meaningful upper bounds in this case, since by definition they are catered to such situations. Indeed, we have $\gamma^{ps} = \gamma(P^{*3}P^3)/3 = 0.25$, while $\gamma^{MH} = 1 - \Lambda(M_1(P^6))^{1/6} = 0.151$. Comparing the results in Proposition 3.4 in Paulin [89] with Corollary 4.5.1, we give a tighter upper bound in the total variation distance from stationarity, since the convergence rate is bounded by $||P^n(x,\cdot) - \pi||_{TV} \leqslant O((1-\gamma^{MH})^n) = O((0.849)^n) \leqslant O((\sqrt{1-\gamma^{ps}})^n) = O(\sqrt{0.75}^n)$.

Example 4.5.2 (non-reversible Markov chain sampler)**.** The second example is taken from Montenegro and Tetali [84] Example 5.3 and Diaconis et al. [33]. Consider a Markov chain on $\mathcal{X} = \mathbb{Z}/2m\mathbb{Z}$ labeled by $\{-(m-1), \ldots, 0, 1, \ldots, m\}$, with transitions

$P(i, i+1) = 1 - \frac{1}{m}, P(i, -i) = \frac{1}{m}$. The chain is doubly stochastic with stationary distribution being the uniform distribution on the state space. It is shown in Theorem 1 of Diaconis et al. [33] that $t_{mix}(\epsilon) = \Theta(m \log(1/\epsilon))$, and in Montenegro and Tetali [84] that existing upper bounds cannot provide useful information.

We now fix $m = 3$ and demonstrate that γ^{ps} and γ^{MH} give meaningful bounds. By computation, we have $\gamma^{ps} = \gamma(P^{*3}P^3)/3 = 0.315$, and $\gamma^{MH} = 1 - |\lambda(M_2(P^4))|^{1/4} = 0.270$. Similar to Example 4.5.1, the upper bound provided by Corollary 4.5.1 outperforms that in Paulin [89], since $||P^n(x, \cdot) - \pi||_{TV} \leqslant O((1-\gamma^{MH})^n) = O((0.730)^n) \leqslant O((\sqrt{1-\gamma^{ps}})^n) = O(\sqrt{0.685}^n)$.

Example 4.5.3 (Winning streak)**.** The third example is the so-called winning streak Markov chain. It has been studied in Example 4.1.5 and Section 5.3.5 in Levin et al. [74]. Consider a Markov chain on $\mathcal{X} = \{0, \ldots, m\}$ with transitions $P(i, 0) = P(i, i+1) = P(m, m) = 1/2$. One remarkable property of such a chain is that its time-reversal, P^*, attains *exactly* the stationary distribution in m steps.

By a coupling argument, $d(n) \leqslant \frac{1}{2^n}$ for all m. Yet, for P^*, its mixing time is of order m. Let's fix $m = 4$ for now and look at various spectral gaps. We have $\gamma(PP^*) = \gamma^{ps} = 0.5$ (this holds for any $m \geqslant 2$ numerically) and $\gamma^{MH} = 0.138$, so both the multiplicative reversiblization and pseudo-spectral gap give a correct order of convergence rate. The performance of γ^{MH} is poor in this example, due to the fact that P^* has a much slower mixing time when compared to P.

4.6 Variance bounds

In this section, we prove variance bounds for Markov chains in terms of the MH-spectral gap. The readers should compare Theorem 4.6.1 with Lemma 12.20 in Levin et al. [74]

and Theorem 3.5,3.7 in Paulin [89].

Theorem 4.6.1. *Let $(X_n)_{n\geqslant 0}$ be a Markov chain with Markov kernel P, stationary measure π and MH-spectral gap γ^{MH}. Suppose that $f \in L^2(\pi)$, and define the variance and asymptotic variance to be respectively*

$$V_f := \mathrm{Var}_\pi(f),$$
$$\sigma_{as}^2 := \lim_{n\to\infty} \frac{1}{n}\mathrm{Var}_\pi\left(\sum_{i=1}^n f(X_i)\right).$$

The variance bounds are given by

$$\mathrm{Var}_\pi\left(\sum_{i=1}^n f(X_i)\right) \leqslant nV_f\left(|\mathcal{C}^c| + \frac{2}{\gamma^{MH}}\right), \tag{4.6.1}$$

$$\left|\mathrm{Var}_\pi\left(\sum_{i=1}^n f(X_i)\right) - n\sigma_{as}^2\right| \leqslant 4V_f\left(1 + |\mathcal{C}^c| + \frac{4(\beta^{MH})^{|\mathcal{C}^c|+1}}{\gamma^{MH}}\right)^2. \tag{4.6.2}$$

More generally, if $f_i \in L^2(\pi)$ for $i = 1,\ldots,n$, then

$$\mathrm{Var}_\pi\left(\sum_{i=1}^n f_i(X_i)\right) \leqslant \sum_{i=1}^n \mathrm{Var}_\pi(f_i(X_i))\left(|\mathcal{C}^c| + \frac{2}{\gamma^{MH}}\right). \tag{4.6.3}$$

Before we give the proof of Theorem 4.6.1, we state a lemma that bounds the operator norm of P by that of M_1 and M_2 in $L_0^2(\pi)$.

Lemma 4.6.1. *Suppose that P is a Markov kernel with stationary measure π. Then*

$$||P||_{L_0^2(\pi)\to L_0^2(\pi)} \leqslant ||M_1||_{L_0^2(\pi)\to L_0^2(\pi)} + ||M_2||_{L_0^2(\pi)\to L_0^2(\pi)} + |\lambda(M_1(P^2))|^{1/2} + |\lambda(M_2(P^2))|^{1/2}.$$

Proof. By Lemma 4.3.1(i), we have $||(P+P^*)f||_\pi^2 = ||(M_1+M_2)f||_\pi^2$ where $f \in L_0^2(\pi)$. Rearranging the terms give

$$\begin{aligned}\langle Pf, Pf\rangle_\pi &= \langle M_1 f, M_1 f\rangle_\pi + \langle M_2 f, M_2 f\rangle_\pi + \langle M_1 f, M_2 f\rangle_\pi + \langle M_2 f, M_1 f\rangle_\pi \\ &\quad - \langle P^* f, P^* f\rangle_\pi - \langle f, (P^*)^2 f\rangle_\pi - \langle f, P^2 f\rangle_\pi \\ &\leqslant \langle M_1 f, M_1 f\rangle_\pi + \langle M_2 f, M_2 f\rangle_\pi + \langle M_1 f, M_2 f\rangle_\pi + \langle M_2 f, M_1 f\rangle_\pi\end{aligned}$$

$$- \langle f, M_1(P^2)f\rangle_\pi - \langle f, M_2(P^2)f\rangle_\pi,$$

where we used that $P^2 + (P^*)^2 = M_1(P^2) + M_2(P^2)$ by Lemma 4.3.1(i) and $\langle P^* f, P^* f\rangle_\pi \geqslant 0$ in the inequality. Therefore, we have

$$||Pf||_\pi \leqslant (||M_1||_{L_0^2(\pi)\to L_0^2(\pi)} + ||M_2||_{L_0^2(\pi)\to L_0^2(\pi)})||f||_\pi + |\lambda(M_1(P^2))|^{1/2} + |\lambda(M_2(P^2))|^{1/2}.$$

Result follows by taking supremum over all f with $||f||_\pi \leqslant 1$ and $\mathbb{E}_\pi f = 0$. $\square$

Proof of Theorem 4.6.1. Assume without loss of generality that $\mathbb{E}_\pi(f) = 0$ and $\mathbb{E}_\pi(f_i) = 0$ for $i = 1, \ldots, n$. We first show (4.6.1). Since $X_0 \sim \pi$ and by Lemma 4.3.2,

$$\mathbb{E}_\pi(f(X_i)f(X_j)) = \langle f, P^{|j-i|}f\rangle_\pi \leqslant \langle f, M_1(P^{|j-i|})f\rangle_\pi = \langle f, (M_1(P^{|j-i|}) - \pi)f\rangle_\pi.$$

Summing up j from 1 to n leads to

$$\begin{aligned}
\mathbb{E}_\pi\left(f(X_i)\sum_{j=1}^n f(X_j)\right) &\leqslant \sum_{j=1}^n \langle f, (M_1(P^{|j-i|}) - \pi)f\rangle_\pi \\
&\leqslant \mathbb{E}_\pi(f^2)\sum_{j=1}^n ||M_1(P^{|j-i|})||_{L_0^2(\pi)\to L_0^2(\pi)} \\
&\leqslant V_f\left(|\mathcal{C}^c| + \sum_{j=1}^n (\beta^{MH})^{|j-i|}\right) \\
&\leqslant V_f\left(|\mathcal{C}^c| + \frac{2}{\gamma^{MH}}\right).
\end{aligned}$$

(4.6.1) follows when we sum i from 1 to n. Next, to show (4.6.3), we observe that

$$\begin{aligned}
\mathbb{E}_\pi(f_i(X_i)f_j(X_j)) &= \langle f_i, P^{|j-i|}f_j\rangle_\pi \\
&\leqslant \langle f_i, (M_1(P^{|j-i|}) - \pi)f_j\rangle_\pi \\
&\leqslant ||f_i||_\pi ||f_j||_\pi ||M_1(P^{|j-i|})||_{L_0^2(\pi)\to L_0^2(\pi)} \\
&\leqslant \frac{1}{2}(\mathbb{E}_\pi f_i^2 + \mathbb{E}_\pi f_j^2)||M_1(P^{|j-i|})||_{L_0^2(\pi)\to L_0^2(\pi)}.
\end{aligned}$$

(4.6.3) follows when we sum i, j from 1 to n, and $\sum_{j=1}^{n} ||M_1(P^{|j-i|})||_{L_0^2(\pi)\to L_0^2(\pi)} \leqslant |\mathcal{C}^c| + \frac{2}{\gamma^{MH}}$. Finally, we will show (4.6.2). Following from the proof of Theorem 3.5 and 3.7 in Paulin [89], using the definition of σ_{as}^2, we have

$$\left|\mathrm{Var}_\pi\left(\sum_{i=1}^{n} f(X_i)\right) - n\sigma_{as}^2\right| = |\langle f, 2(I-(P-\pi)^{n-1})(I-(P-\pi))^{-2} f\rangle_\pi|.$$

Note that $||I-(P-\pi)^{n-1}||_{L^2(\pi)\to L^2(\pi)} \leqslant 2$, and by Lemma 4.6.1,

$$\begin{aligned}
||(I-(P-\pi))^{-1}||_{L^2(\pi)\to L^2(\pi)} &\leqslant \sum_{k=0}^{\infty} ||(P-\pi)^k||_{L^2(\pi)\to L^2(\pi)} \\
&\leqslant 1 + |\mathcal{C}^c| + \sum_{k=|\mathcal{C}^c|+1}^{\infty} ||P^k||_{L_0^2(\pi)\to L_0^2(\pi)} \\
&\leqslant 1 + |\mathcal{C}^c| + 4 \sum_{k=|\mathcal{C}^c|+1}^{\infty} (\beta^{MH})^k \\
&= 1 + |\mathcal{C}^c| + \frac{4(\beta^{MH})^{|\mathcal{C}^c|+1}}{\gamma^{MH}}.
\end{aligned}$$

□

4.7 Metastability, conductance and Cheeger's inequality

In this section, we aim at analyzing metastability, conductance, Cheeger's inequality and their relationships with the two MH kernels. We begin by briefly recalling these concepts.

Definition 4.7.1 (Metastability of a set)**.** Let $A, B \in \mathcal{F}$ be measurable subsets of $\mathcal{X}$. Denote by

$$Q(A,B) := \frac{1}{\pi(A)} \int_A p(x,B)\,\pi(dx) = \frac{\langle P\mathbb{1}_A, \mathbb{1}_B\rangle_\pi}{\langle \mathbb{1}_A, \mathbb{1}_A\rangle_\pi},$$

if $\pi(A) > 0$ and 0 otherwise. A is said to be *metastable* (resp. *invariant*) if

$$Q(A,A) \approx 1\,(\text{resp. } Q(A,A) = 1)\,.$$

Remark 4.7.1. For a non-reversible chain with transition kernel P, since $\langle P\mathbb{1}_A, \mathbb{1}_B\rangle_\pi = \langle \mathbb{1}_A, P^*\mathbb{1}_B\rangle_\pi$, $Q(A,B)$ of P equals to $Q(A,B)$ of the reversible chain $(P+P^*)/2$.

Remark 4.7.2. Denote A^c to be the complement of $A \subset \mathcal{X}$, then $Q(A,A^c)$ is also known as the conductance of the set A. See Definition 4.7.4 below.

Note that metastability means "almost invariant", in the sense that $Q(A,A)$ is close to 1. In reality, we are more interested in measuring the metastability of an arbitrary partition of the state space $\mathcal{X}$, in which we state in the following:

Definition 4.7.2 (Metastability of a partition)**.** Suppose that $\mathcal{D} = \{A_1, \ldots, A_n\}$ is a partition of $\mathcal{X}$. The metastability of $\mathcal{D}$ is denoted by

$$m(\mathcal{D}) := \sum_{i=1}^{n} Q(A_i, A_i)\,.$$

$\mathcal{D}$ is said to be metastable if $m(\mathcal{D}) \approx n$.

The next definition measures the "leakage" of a set A at time t, which is first introduced by Davies [26], Singleton [103].

Definition 4.7.3 (Leakage)**.** The leakage of a set $A \in \mathcal{F}$ at time t is denoted by

$$l(A,t) := \frac{||\mathbb{1}_A - P^t\mathbb{1}_A||_{L^1(\pi)}}{2\pi(A)(1-\pi(A))}\,.$$

This can be rewritten as

$$l(A,t) = \int_{\mathcal{X}\setminus A} \left(P^t \frac{\mathbb{1}_A}{\pi(A)}\right)(x)\frac{\pi(dx)}{1-\pi(A)}\,,$$

measuring the probability of $\mathbb{1}_A/\pi(A)$ being outside A at time t.

Remark 4.7.3. Another related measure of bottleneckness is conductance. See Definition 4.7.4 below.

Next, we introduce a key assumption (see e.g. Davies [26], Huisinga and Schmidt [52], Singleton [103]) that will be used in subsequent sections:

Assumption 4.7.1. Suppose that $Q : L^2(\pi) \to L^2(\pi)$ is a self-adjoint Markov kernel with n *dominant* eigenvalues denoted by

$$1 = \lambda_1 > \lambda_2 \geqslant \ldots \geqslant \lambda_n \,.$$

In addition, the spectrum $\sigma(Q)$ of Q is contained in

$$\sigma(Q) \subset [a, b] \cup \{\lambda_n, \ldots, \lambda_1\}\,,$$

where $-1 < a \leqslant b < \lambda_n$.

If Q is a finite Markov chain, or if Q is geometrically ergodic, or if Q is V-uniformly ergodic, then it can be shown that Q satisfies Assumption 4.7.1, see e.g. Huisinga and Schmidt [52], Schütte and Sarich [101]. Under Assumption 4.7.1 with $n = 2$, if the eigenvalue λ_2 is “close" to 1, then this is known as “almost degeneracy", which allows us to partition $\mathcal{X}$ into two metastable regions. This has been the subject of investigation in Davies [26], Singleton [103].

Next, we provide a quick review on the notion of conductance and Cheeger's inequality, which are first introduced to the Markov chain literature in [32].

Definition 4.7.4 (Conductance)**.** The conductance of the set A is

$$\Phi(A) := Q(A, A^c)\,,$$

where $Q(A, A^c)$ is defined in Definition 4.7.1. The conductance of the chain is defined to be

$$\Phi_*(2) := \min_{A \neq \emptyset, \mathcal{X}} \max\{\Phi(A), \Phi(A^c)\} = \min_{A: 0 < \pi(A) \leqslant 1/2} \Phi(A)\,.$$

For $k \in \mathbb{N}$, let $\mathcal{D}_k = \{A_1, \ldots, A_k\}$ be the set of k-uples of disjoint and π-non-negligible subsets of $\mathcal{X}$. Then the k-way expansion is

$$\Phi_*(k) := \min_{(A_1,\ldots,A_k)\in\mathcal{D}_k} \max_{i\in[\![k]\!]} \Phi(A_i)\,.$$

Next, we recall the Cheeger's inequality and its higher-order variants, which provide a two-sided bound in the spectral gap in terms of the k-way expansion, see e.g. [70] and [81, Proposition 5] .

Theorem 4.7.1 (Higher-order Cheeger's inequality)**.** *Suppose that P is the transition kernel of a discrete-time reversible finite Markov chain with eigenvalues $1 = \lambda_1 \geqslant \ldots \geqslant \lambda_n$. For $k \in [\![n]\!]$,*

$$\frac{1-\lambda_k}{2} \leqslant \Phi_*(k) \leqslant O(k^4)\sqrt{1-\lambda_k}\,.$$

4.7.1 Main results

In this section, we demonstrate that, by means of comparison (i.e. Peskun's ordering as in Lemma 4.3.2), that existing results on metastability, leakage and Cheeger's inequality can be readily extended to non-reversible case. We first give spectral bounds on the metastability of partition, in terms of spectral objects associated with the first and second MH kernel M_1 and M_2.

Theorem 4.7.2 (Metastability)**.** *Suppose that P is the Markov kernel of a non-reversible Markov chain on $\mathcal{X}$, with the first and second MH kernel denoted by M_1 and M_2 respectively. In addition, for $i = 1, 2$, M_i satisfies Assumption 4.7.1 with dominant eigenvalues-eigenvectors denoted by $(\lambda_j^i, \phi_j^i)_{j=1}^n$. For any partition $\mathcal{D} = \{A_1, \ldots, A_n\}$, the metastability of $\mathcal{D}$ is bounded by*

$$1 + \sum_{j=2}^{n} \rho_j \lambda_j^2 + c \leqslant m(\mathcal{D}) \leqslant 1 + \sum_{j=2}^{n} \lambda_j^1\,, \tag{4.7.1}$$

where Λ is the orthogonal projection from $L^2(\pi) \rightarrow \mathrm{Span}\{\mathbb{1}_{A_1}, \ldots, \mathbb{1}_{A_n}\}$, $\rho_j := ||\Lambda\phi_j^2||_\pi^2 \in [0,1]$ for $j = 2, \ldots, n$, and

$$c := a\left(\sum_{j=2}^{n} 1 - \rho_j\right),$$

with a being defined in Assumption 4.7.1 for M_2.

Remark 4.7.4. This result should be compared with [52, Theorem 2] and [101, Theorem 5.8], in which we retrieve the corresponding results since $P = M_1 = M_2$ and hence $\lambda_j = \lambda_j^1 = \lambda_j^2$ in the reversible case.

The next theorem gives spectral bounds on leakage:

Theorem 4.7.3 (Leakage)**.** *Suppose that P is the Markov kernel of a non-reversible Markov chain on $\mathcal{X}$, with the first and second MH kernel denoted by M_1 and M_2 respectively. In addition, for $i = 1, 2$ and $t \in \mathbb{N}$, $M_i(P^t)$ satisfies Assumption 4.7.1 with $n = 2$ and dominant eigenvalues-eigenvectors denoted by $(\lambda_j^i(P^t), \phi_j^i(P^t))_{j=1}^2$. If $\{A, B\}$ is a partition of $\mathcal{X}$, then for all $t \in \mathbb{N}$,*

$$1 - \lambda_2^1(P^t) \leqslant l(A,t) \leqslant 1 - \gamma_A^2(M_2(P^t))\lambda_2^2(P^t),$$

where $\gamma_A(M_i(P^t)) = \langle \psi_A, \phi_2^i(P^t)\rangle_\pi$ for $i = 1, 2$ and

$$\psi_A = \sqrt{\frac{\pi(B)}{\pi(A)}}\mathbb{1}_A - \sqrt{\frac{\pi(A)}{\pi(B)}}\mathbb{1}_B.$$

Remark 4.7.5. This result should be compared with [103, Theorem 5], in which we retrieve the corresponding upper bound in the reversible case since $P = M_1 = M_2$ and hence $\lambda_j = \lambda_j^1 = \lambda_j^2$.

Finally, we give a version of Cheeger's inequality in bounding the k-way expansion, in terms of the eigenvalues of the two Metropolis kernels. This result should be compared with Theorem 4.7.1.

Theorem 4.7.4 (Cheeger's inequality)**.** *Suppose that P is the Markov kernel of a non-reversible Markov chain on a finite state space $\mathcal{X}$, with the first and second MH kernel denoted by M_i and eigenvalues $(\lambda_j^i)_{j=1}^n$ for $i = 1, 2$. For $k \in [\![n]\!]$,*

$$\frac{1-\lambda_k^1}{2} \leqslant \Phi_*(k) \leqslant O(k^4)\sqrt{1-\lambda_k^2}\,. \tag{4.7.2}$$

4.7.2 Proofs

Proof of Theorem 4.7.2. The key step is the Peskun ordering between P, M_1, M_2, which yields, for any $f \in L^2(\pi)$, the following inequalities:

$$\langle M_2 f, f\rangle_\pi \leqslant \langle P f, f\rangle_\pi \leqslant \langle M_1 f, f\rangle_\pi\,. \tag{4.7.3}$$

This allows us to link $m(\mathcal{D})$ to the eigenvalues of M_1 and M_2. More precisely, we first show the upper bound of (4.7.1). Making use of the definition, we have

$$m(\mathcal{D}) = \sum_{i=1}^n \frac{\langle P\mathbb{1}_{A_i}, \mathbb{1}_{A_i}\rangle_\pi}{\langle \mathbb{1}_{A_i}, \mathbb{1}_{A_i}\rangle_\pi} \leqslant \sum_{i=1}^n \frac{\langle M_1\mathbb{1}_{A_i}, \mathbb{1}_{A_i}\rangle_\pi}{\langle \mathbb{1}_{A_i}, \mathbb{1}_{A_i}\rangle_\pi} \leqslant 1 + \sum_{j=2}^n \lambda_j^1\,,$$

where the first inequality follows from (4.7.3) with $f = \mathbb{1}_{A_i}$, and we use [52, Theorem 2] in the second inequality since M_1 is a self-adjoint Markov kernel satisfying Assumption 4.7.1. Next, to show the lower bound, using (4.7.3) again, we arrive at

$$m(\mathcal{D}) \geqslant \sum_{i=1}^n \langle M_2 \chi_{A_i}, \chi_{A_i}\rangle_\pi\,,$$

where $\chi_{A_i} := \frac{\mathbb{1}_{A_i}}{\sqrt{\langle \mathbb{1}_{A_i}, \mathbb{1}_{A_i}\rangle_\pi}}$. The remaining part of the proof follows a similar argument as in Huisinga and Schmidt [52]. Denote the orthogonal projection by $\Pi : L^2(\pi) \to \mathrm{Span}\{\phi_1^2, \ldots, \phi_n^2\}$ and its orthogonal complement by $\Pi^\perp = I - \Pi$. Note that

$$\begin{aligned}\sum_{i=1}^n \langle M_2 \chi_{A_i}, \chi_{A_i}\rangle_\pi &= \sum_{i=1}^n \langle (M_2 - aI)(\Pi + \Pi^\perp)\chi_{A_i}, \chi_{A_i}\rangle_\pi + a\sum_{i=1}^n \langle \chi_{A_i}, \chi_{A_i}\rangle_\pi \\ &= \sum_{i=1}^n \langle (M_2 - aI)\Pi\chi_{A_i}, \Pi\chi_{A_i}\rangle_\pi + \sum_{i=1}^n \langle (M_2 - aI)\Pi^\perp\chi_{A_i}, \Pi^\perp\chi_{A_i}\rangle_\pi + an\end{aligned}$$

$$
\begin{aligned}
&\geqslant \sum_{i=1}^{n} \langle (M_2 - aI)\Pi\chi_{A_i}, \Pi\chi_{A_i}\rangle_\pi + an \\
&= \sum_{i=1}^{n} \left\langle \sum_{k=1}^{n} (\lambda_k^2 - a)\langle \chi_{A_i}, \phi_k^2\rangle_\pi \phi_k^2, \sum_{k=1}^{n} \langle \chi_{A_i}, \phi_k^2\rangle_\pi \phi_k^2 \right\rangle_\pi + an \\
&= \sum_{i=1}^{n} \sum_{k=1}^{n} (\lambda_k^2 - a)\langle \chi_{A_i}, \phi_k^2\rangle_\pi^2 + an \\
&= \sum_{k=1}^{n} (\lambda_k^2 - a)||\Lambda\phi_k^2||_\pi^2 + an = 1 + \sum_{j=2}^{n} \rho_j \lambda_j^2 + c\,,
\end{aligned}
$$

where the inequality follows from the fact that $(M_2 - aI)$ is self-adjoint and positive semi-definite, the fourth equality comes from the fact that the set $\{\phi_1^2, \ldots, \phi_n^2\}$ is orthonormal, the fifth equality makes use of Parseval's identity, and we use $\lambda_1^2 = 1, \phi_1^2 = \mathbb{1}, ||\Lambda\phi_1^2||_\pi^2 = 1$ in the last equality. Desired result follows. $\square$

Proof of Theorem 4.7.3. Similar to the proof of Theorem 4.7.2, the crux again lies at the appropriate use of (4.7.3). First, by [103, Lemma 4] and (4.7.3), we have

$$
\langle (I - M_1(P^t))\psi_A, \psi_A\rangle_\pi \leqslant \langle (I - P^t)\psi_A, \psi_A\rangle_\pi = l(A,t) \leqslant \langle (I - M_2(P^t))\psi_A, \psi_A\rangle_\pi\,,
$$

so it suffices to show that

$$
\langle (I - M_2(P^t))\psi_A, \psi_A\rangle_\pi \leqslant 1 - \gamma_A^2(M_2(P^t))\lambda_2^2(P^t)\,, \tag{4.7.4}
$$

$$
\langle (I - M_1(P^t))\psi_A, \psi_A\rangle_\pi \geqslant 1 - \lambda_2^1(P^t)\,. \tag{4.7.5}
$$

The rest of the proof is similar to that of [103, Theorem 5]. For $i = 1, 2$ and $t \in \mathbb{N}$, denote by $\Pi_i(P^t)$ to be the orthogonal projection $\Pi_i(P^t) : L^2(\pi) \to \mathrm{Span}\{\phi_1^i(P^t), \phi_2^i(P^t)\}$ and its orthogonal complement by $\Pi_i^\perp(P^t) = I - \Pi_i(P^t)$. Since ψ_A is orthogonal to $\phi_1^i(P^t) = \mathbb{1}$, we have

$$
\Pi_i(P^t)\psi_A = \langle \psi_A, \phi_2^i(P^t)\rangle_\pi \phi_2^i(P^t) = \gamma_A(M_i(P^t))\phi_2^i(P^t)\,. \tag{4.7.6}
$$

We proceed to show (4.7.4). Note that

$$
\langle (I - M_2(P^t))\psi_A, \psi_A\rangle_\pi = \langle (I - M_2(P^t))(\Pi_2(P^t) + \Pi_2^\perp(P^t))\psi_A, \psi_A\rangle_\pi
$$

$$
\begin{aligned}
&= \gamma_A^2(M_2(P^t))(1-\lambda_2^2(P^t)) + \langle (I - M_2(P^t))\Pi_2^{\perp}(P^t)\psi_A, \psi_A\rangle_\pi \\
&\leqslant \gamma_A^2(M_2(P^t))(1-\lambda_2^2(P^t)) + \langle \Pi_2^{\perp}(P^t)\psi_A, \psi_A\rangle_\pi \\
&= \gamma_A^2(M_2(P^t))(1-\lambda_2^2(P^t)) + \langle (I - \Pi_2(P^t))\psi_A, \psi_A\rangle_\pi \\
&= 1 - \gamma_A^2(M_2(P^t))\lambda_2^2(P^t)\,,
\end{aligned}
$$

where the second equality follows from (4.7.6) and the inequality follows from Cauchy-Schwartz inequality. Finally, we show (4.7.5). Using the Rayleigh quotient lower bound on the self-adjoint kernel $I - M_1(P^t)$ yields

$$
\langle (I - M_1(P^t))\psi_A, \psi_A\rangle_\pi \geqslant 1 - \lambda_2^1(P^t)\,.
$$

□

Proof of Theorem 4.7.4. We first show the upper bound of (4.7.2). We have

$$
\Phi(A) = 1 - Q(A, A) \leqslant 1 - \frac{\langle M_2 \mathbb{1}_A, \mathbb{1}_A\rangle_\pi}{\langle \mathbb{1}_A, \mathbb{1}_A\rangle_\pi} = \Phi(A)(M_2) \leqslant O(k^4)\sqrt{1-\lambda_k^2}\,,
$$

where we apply the Peskun ordering (4.7.3) in the first inequality and the second inequality comes from the Cheeger's inequality for reversible chain if M_2 is Markov. In the general case however, we can write $M_2 = G + I$, where G is the Markov generator of M_2, and apply the corresponding version of Cheeger's inequality for G instead (see e.g. [81, Theorem 2]), so desired upper bound follows from the min-max characterization of the k-way expansion. Next, for the lower bound of (4.7.2), we again use the Peskun ordering (4.7.3) to get

$$
\Phi(A) = 1 - Q(A, A) \geqslant 1 - \frac{\langle M_1 \mathbb{1}_A, \mathbb{1}_{A^c}\rangle_\pi}{\langle \mathbb{1}_A, \mathbb{1}_A\rangle_\pi} = \Phi(A)(M_1) \geqslant \frac{1-\lambda_k^1}{2}\,,
$$

where the second inequality follows from the Cheeger's inequality for reversible chain.

□

4.7.3 Examples

In this section, we present an example of asymmetric random walk on n-cycle and a numerical example of upward skip-free Markov chain to investigate the sharpness of the spectral bounds presented in Theorem 4.7.2 and 4.7.3.

Example 4.7.1 (Asymmetric random walk on n-cycle)**.** Recall that we have studied about the asymmetric random walk on n-cycle in Example 4.3.1, in which we adapt the notations in that example. In particular, we have $l := \min\{p, q\}$ and $r := \max\{p, q\}$. For any partition $\mathcal{D} = \{A_1, A_2, \ldots, A_j\}$ with $0 \leqslant j - 1 < n/2$, the upper bound in Theorem 4.7.2 now gives

$$\begin{aligned}1 + \sum_{k=1}^{j-1} 1 - 2l(1 - \cos(2\pi k/n)) &= j(1 - 2l) + 2l \sum_{k=0}^{j-1} \cos(2\pi k/n) \\ &= j(1 - 2l) + 2l \left(\frac{\sin(\pi j/n) \cos(\pi(j-1)/n)}{\sin(\pi/n)} \right).\end{aligned}$$

On the other hand, we have for $j \geqslant 2$,

$$\begin{aligned}\rho_j = ||\Lambda \phi_j^2||_\pi^2 = \sum_{k=1}^{j} \sum_{x \in A_k} \pi(x) \left(\frac{\langle \phi_j^2, \mathbb{1}_{A_k} \rangle_\pi}{\pi(A_k)} \right)^2 &= \sum_{k=1}^{j} \sum_{x \in A_k} \frac{n}{|A_k|^2} \langle \phi_j^2, \mathbb{1}_{A_k} \rangle_\pi^2 \\ &= \sum_{k=1}^{j} \frac{1}{n|A_k|} \left(\sum_{x \in A_k} \cos(2\pi(j-1)x/n) \right)^2,\end{aligned}$$

and $a = 1 - 2r + 2r \min_{i \geqslant 2} \cos(2\pi(i-1)/n)$, so the lower bound of 4.7.2 is readily computable.

Next, we now look at the leakage in Theorem 4.7.3 with the partition $\mathcal{D} = \{A, B\}$. The lower bound becomes

$$1 - \lambda_2^1(P^t) = 2l(1 - \cos(2\pi/n)),$$

while we note that

$$\langle \psi_A, \phi_2^2(P) \rangle_\pi = \frac{1}{n} \left(\sqrt{\frac{|B|}{|A|}} \sum_{x \in A} \cos(2\pi x/n) - \sqrt{\frac{|A|}{|B|}} \sum_{x \in B} \cos(2\pi x/n) \right),$$

so the upper bound in Theorem 4.7.3 now reads

$$1-\frac{1}{n^2}\left(\sqrt{\frac{|B|}{|A|}}\sum_{x\in A}\cos(2\pi x/n)-\sqrt{\frac{|A|}{|B|}}\sum_{x\in B}\cos(2\pi x/n)\right)^2(1-2r(1-\cos(2\pi/n)))\ .$$

Example 4.7.2 (Upward skip-free)**.** We consider an upward skip-free chain on $\{1,2,3,4\}$ with transition kernel given by

$$P=\begin{pmatrix}0.5&0.5&0&0\\0.2&0.6&0.2&0\\0.1&0.3&0.5&0.1\\0.1&0.2&0.4&0.3\end{pmatrix}$$

with eigenvalues $(1,0.52,0.25,0.13)$. The two Metropolis kernels are

$$M_1=\begin{pmatrix}0.6&0.4&0&0\\0.2&0.66&0.14&0\\0&0.3&0.64&0.06\\0&0&0.4&0.6\end{pmatrix},\quad M_2=\begin{pmatrix}0.40&0.5&0.09&0.01\\0.25&0.54&0.2&0.01\\0.1&0.44&0.36&0.1\\0.1&0.2&0.7&0\end{pmatrix},$$

with eigenvalues $\lambda^1=(1,0.74,0.50,0.28)$ and $\lambda^2=(1,0.37,0.08,-0.16)$ respectively. In Theorem 4.7.2, we have an upper bound $1+\lambda^1_2=1.74$ and lower bound $1+\rho_2\lambda^2_2+c=1+0.09\times 0.37+(-0.16)\times(1-0.09)=0.89$.

First, we consider the partition $\mathcal{D}=\{\{1,2\},\{3,4\}\}$ with $A_1=\{1,2\}$ and $A_2=\{3,4\}$. We see that

$$m(\mathcal{D})=p(A_1,A_1)+p(A_2,A_2)=0.87+0.61=1.48\,,$$

which is closer to the upper bound 1.74. If we instead consider the partition $\mathcal{D}=\{\{1,2,3\},\{4\}\}$ with $A_1=\{1,2,3\}$ and $A_2=\{4\}$, then

$$m(\mathcal{D})=p(A_1,A_1)+p(A_2,A_2)=0.98+0.3=1.28\,,$$

which is closer to the lower bound of 0.89.

CHAPTER 5

ESTIMATION OF THE LOG-SOBOLEV CONSTANT AND EIGENSPACE OF REVERSIBLE MARKOV CHAIN VIA A SINGLE SAMPLE PATH

5.1 Introduction

In this Chapter we consider the problem of estimating the eigenspace and log-Sobolev constant from a single sample path of a reversible Markov chain. That is, given a sample path of size n $(X_1, \ldots, X_n)$, how can we estimate the eigenspace and the log-Sobolev constant with high statistical guarantee? A similar problem has caught the attention of researchers recently, in which Garren and Smith [44], Hsu et al. [50], Levin and Peres [73] considered the same problem of estimating the spectral gap and stationary distribution, and Kamath and Verdú [57] investigated the estimation of entropy rate in the same setting.

The rest of this Chapter is organized as follows. In Section 5.2 we fix our notations and provide a quick summary of the results of Hsu et al. [50], while in Section 5.3 we state our interval estimator for the log-Sobolev constant (and also the modified log-Sobolev constant and Cheeger constant) and provide related discussions. In Section 5.4 we give an eigenspace estimator with statistical guarantee.

5.2 Hsu et al. [50]'s estimators for spectral gap and stationary distribution

Let $(\mathcal{X}, P, \pi)$ be an ergodic and reversible Markov chain on a finite state space $\mathcal{X}$ with transition kernel P and stationary distribution π. Denote $\pi_* := \min_{i \in \mathcal{X}} \pi(i)$, and write

$[\![a, b]\!] := \{a, a+1, \ldots, b-1, b\}$ for any $a \leqslant b \in \mathbb{Z}$ and $[\![b]\!] := [\![1, b]\!]$ for $1 \leqslant b \in \mathbb{Z}$. Recall that the Dirichlet form of P is defined by, for any real-valued functions f, g,

$$\langle f, g \rangle_\pi := \sum_{x \in \mathcal{X}} f(x) g(x) \pi(x), \tag{5.2.1}$$

$$\mathcal{E}_P(f, g) := \mathcal{E}(f, g) = \mathbb{E}_\pi \left((I - P) f g \right) = \langle (I - P) f, g \rangle_\pi . \tag{5.2.2}$$

By writing $\mathrm{Var}_\pi(f) = \mathbb{E}_\pi(f^2) - \mathbb{E}^2_\pi(f)$, the classical variational characterization of the spectral gap λ is given by

$$\lambda(P) := \lambda = \inf_f \left\{ \frac{\mathcal{E}(f, f)}{\mathrm{Var}_\pi(f)} \, ; \mathrm{Var}_\pi(f) \neq 0 \right\}. \tag{5.2.3}$$

The log-Sobolev constant is defined by replacing the variance $\mathrm{Var}_\pi(f)$ in the definition of spectral gap by the relative entropy $\mathrm{Ent}_\pi(f^2) = \mathbb{E}_\pi(f^2 \log f^2) - \mathbb{E}_\pi(f^2 \log \mathbb{E}_\pi(f^2))$:

$$\alpha(P) := \alpha = \inf_f \left\{ \frac{\mathcal{E}(f, f)}{\mathrm{Ent}_\pi(f^2)} \, ; \mathrm{Ent}_\pi(f^2) \neq 0 \right\}. \tag{5.2.4}$$

We now recall two main results in Hsu et al. [50]. The first one is [50, Algorithm 1], which gives empirical bounds on the deviation between the empirical spectral gap and empirical stationary distribution with their true counterparts. This also forms the basis of offering an interval estimator of the log-Sobolev constant in Section 5.3. Assume that we are given a sample path $(X_1, X_2, \ldots, X_n)$ from a reversible Markov chain and a confidence parameter $\delta \in (0, 1)$.

1. We first define the smoothed empirical Markov chain $\widehat{P} = (\widehat{P}_{i,j})_{i,j \in \mathcal{X}}$.

$$N_i := |\{t \in [\![n-1]\!] : X_t = i\}|, \quad i \in \mathcal{X},$$

$$N_{i,j} := |\{t \in [\![n-1]\!] : (X_t, X_{t+1}) = (i, j)\}|, \quad i, j \in \mathcal{X},$$

$$\widehat{P}_{i,j} := \frac{N_{i,j} + 1/|\mathcal{X}|}{N_i}, \quad i, j \in \mathcal{X}.$$

2. Define $\widehat{A}^\#$ to be the group inverse of $\widehat{A} := I - \widehat{P}$, i.e. $\widehat{A}^\#$ is the unique matrix satisfying $\widehat{A}\widehat{A}^\#\widehat{A} = \widehat{A}$, $\widehat{A}^\#\widehat{A}\widehat{A}^\# = \widehat{A}^\#$ and $\widehat{A}^\# A = A\widehat{A}^\#$.

3. Define $\widehat{\pi} = (\widehat{\pi}_i)_{i\in\mathcal{X}}$ to be the unique stationary distribution of $\widehat{P}$.

4. Compute the empirical eigenvalues $1 = \widehat{\gamma}_1 \geqslant \widehat{\gamma}_2 \cdots \geqslant \widehat{\gamma}_{|\mathcal{X}|}$ of $\mathrm{Sym}(\widehat{L}) := (\widehat{L} + \widehat{L}^T)/2$ with $\widehat{L} := D_{\hat{\pi}}^{1/2}\widehat{P}D_{\hat{\pi}}^{-1/2}$, where $D_{\hat{\pi}}$ is a diagonal matrix with $\widehat{\pi}$ on its diagonal.

5. The spectral gap estimator is

$$\widehat{\gamma} := 1 - \max\{\widehat{\gamma}_2, \widehat{\gamma}_{|\mathcal{X}|}\}.$$

6. Empirical bounds for $|\widehat{P}_{i,j} - P_{i,j}|$ for $i,j \in \mathcal{X}$: $c := 1.1, \tau_{n,\delta} := \inf\{t \geqslant 0; 2|\mathcal{X}|^2(1 + \lceil \log_c(2n/t)\rceil_+)e^{-t} \leqslant \delta\}$, and

$$\widehat{B}_{i,j} := \left(\sqrt{\frac{c\tau_{n,\delta}}{2N_i}} + \sqrt{\frac{c\tau_{n,\delta}}{2N_i} + \sqrt{\frac{2c\widehat{P}_{i,j}(1-\widehat{P}_{i,j})\tau_{n,\delta}}{N_i}} + \frac{(4/3)\tau_{n,\delta} + |\widehat{P}_{i,j} - 1/|\mathcal{X}||}{N_i}}\right)^2.$$

7. $$\widehat{\kappa} := \frac{1}{2}\max\left\{\widehat{A}^{\#}_{j,j} - \min\{\widehat{A}^{\#}_{i,j} : i \in \mathcal{X}\} : j \in \mathcal{X}\right\}.$$

8. Empirical bounds for $\max_{i\in\mathcal{X}} |\widehat{\pi}_i - \pi_i|$ and $\max \cup_{i\in\mathcal{X}}\{|\sqrt{\pi_i/\widehat{\pi}_i} - 1|, |\sqrt{\widehat{\pi}_i/\pi_i} - 1|\}$:

$$\widehat{b} := \widehat{\kappa}\max\{\widehat{B}_{i,j} : i,j \in \mathcal{X}\}, \quad \widehat{\rho} := \frac{1}{2}\bigcup_{i\in\mathcal{X}}\left\{\frac{\widehat{b}}{\widehat{\pi}_i}, \frac{\widehat{b}}{(\widehat{\pi}_i - \widehat{b})_+}\right\}. \tag{5.2.5}$$

9. Empirical bounds for $|\widehat{\gamma} - \gamma|$:

$$\widehat{w} := 2\widehat{\rho} + \widehat{\rho}^2 + (1 + 2\widehat{\rho} + \widehat{\rho}^2)\left(\sum_{i,j\in\mathcal{X}} \frac{\widehat{\pi}_i}{\widehat{\pi}_j}\widehat{B}^2_{i,j}\right)^{1/2}. \tag{5.2.6}$$

The second result that we recall is [50, Theorem 4]:

Theorem 5.2.1 (Hsu et al. [50])**.** *Suppose we are given a sample path of a reversible ergodic Markov chain on a finite state space $\mathcal{X}$ and confidence parameter $\delta \in (0,1)$, with $\gamma > 0$ being the spectral gap, π being the unique stationary distribution and π_* being the minimum of π on $\mathcal{X}$. With probability at least $1-\delta$,*

$$\pi_i \in \left[\widehat{\pi}_i - \widehat{b}, \widehat{\pi}_i + \widehat{b}\right], \textit{ for } i \in \mathcal{X}, \quad \gamma \in [\widehat{\gamma} - \widehat{w}, \widehat{\gamma} + \widehat{w}].$$

The width of the intervals almost surely satisfy, as $n \to \infty$,

$$\widehat{b} = O\left(\max_{i,j\in\mathcal{X}} \frac{|\mathcal{X}|}{\gamma}\sqrt{\frac{P_{i,j}\log\log n}{\pi_i n}}\right), \quad \widehat{w} = O\left(\frac{|\mathcal{X}|}{\pi_*\gamma}\sqrt{\frac{\log\log n}{\pi_* n}}\right).$$

5.3 Estimation of the log-Sobolev constant

The log-Sobolev constant is an important notion that appears in hypercontractivity, concentration of measure as well as the mixing time bounds of Markov chains. Yet, estimating or calculating the log-Sobolev constant is a notoriously hard problem. Much effort has been spent to derive either the exact expression (see e.g. Chen and Sheu [14], Chen et al. [16], Diaconis and Saloff-Coste [30], Saloff-Coste [97]) or asymptotic behaviour (e.g. Lee and Yau [71]) of such a constant in some particular models. In this section, we add a statistical flavor to this problem in the spirit of Hsu et al. [50] and construct an explicit interval estimator that traps the true value of the log-Sobolev constant with high probability given a single sample path $(X_1, X_2, \ldots, X_n)$ from a reversible Markov chain and a confidence parameter $\delta \in (0,1)$. The proof builds upon the result of Hsu et al. [50] and the universal upper and lower bounds of the log-Sobolev constant.

In the main result below, we will utilize the following well-known universal bounds (see e.g. [97, Corollary 2.2.10]) on the log-Sobolev constant:

$$\frac{(1-2\pi_*)\gamma}{\log\left(\frac{1}{\pi_*}-1\right)} \leqslant \alpha \leqslant \frac{\gamma}{2}. \tag{5.3.1}$$

The interval estimator that we propose in Theorem 5.3.1 below is simply the empirical estimator of the inequality (5.3.1), i.e. by replacing the right hand side of (5.3.1) with the empirical upper confidence bound

$$\widehat{\gamma}_{ub}/2,$$

where $\widehat{\gamma}_{ub} := \widehat{\gamma} + \widehat{w}$ with $\widehat{w}$ defined in (5.2.6), and the left hand side of (5.3.1) by the plug-in estimator

$$\frac{(1-2\widehat{\pi}_{*,lb})\widehat{\gamma}_{lb}}{\log\left(\frac{1}{\widehat{\pi}_{*,lb}}-1\right)},$$

where $\widehat{\gamma}_{lb} := (\widehat{\gamma} - \widehat{w})_+$, $\widehat{\pi}_{*,lb} := (\min_{i\in\mathcal{X}} \widehat{\pi}_i - \widehat{b})_+$ with $\widehat{b}$ defined in (5.2.5). Note that both $\widehat{w}$ and $\widehat{b}$ depend on n and δ.

Our main result offers the following statistical guarantee of the aforementioned interval estimator:

Theorem 5.3.1. *Suppose we are given a sample path from a reversible ergodic Markov chain* $(X_1, X_2, \ldots, X_n)$ *and a desired level of accuracy* $\delta \in (0,1)$. *With probability at least* $1-3\delta$*, we have*

$$\alpha \in \left[\frac{(1-2\widehat{\pi}_{*,lb})\widehat{\gamma}_{lb}}{\log\left(\frac{1}{\widehat{\pi}_{*,lb}}-1\right)}, \frac{\widehat{\gamma}_{ub}}{2}\right]. \tag{5.3.2}$$

Remark 5.3.1. The proposed estimator (5.3.2) converges asymptotically to (5.3.1) as $\widehat{\gamma}_{ub} \overset{a.s.}{\to} \gamma$ and $\widehat{\pi}_{*,lb} \overset{a.s.}{\to} \pi_*$. Our proposed bound is therefore sharp asymptotically in the sense that there are known cases that attain exactly the upper and lower bound respectively. For example, it is shown in Diaconis and Saloff-Coste [30] that the lower bound is attained for Markov chains with $P(x,y) = \pi(y)$ for $x \in \mathcal{X}$, where π is a given positive probability distribution on $\mathcal{X}$, whereas the upper bound is attained in a number of cases, such as a symmetric Markov chain on the discrete cube, see e.g. Chen et al. [16]. Without *a priori* information on the structure of the Markov chain, we doubt whether it is possible to propose a tighter interval estimator or even a point estimator of α.

Remark 5.3.2. We can propose another (possibly looser) plug-in type estimator of the lower bound such that with probability at least $1-3\delta$, we have

$$\alpha \in \left[\frac{\widehat{\gamma}_{lb}}{2+\log\left(\frac{1}{\widehat{\pi}_{*,lb}}\right)}, \frac{\widehat{\gamma}_{ub}}{2}\right].$$

The proof is similar to that of Theorem 5.3.1, and the idea stems from Diaconis and Saloff-Coste [30] that $\alpha \geqslant \dfrac{\gamma}{2+\log\left(\frac{1}{\pi_*}\right)}$. While it perhaps is a loose lower bound, its convergence rate can be calculated explicitly to be

$$\begin{aligned}
\left|\frac{\widehat{\gamma}_{lb}}{2+\log\left(\frac{1}{\widehat{\pi}_{*,lb}}\right)} - \frac{\gamma}{2+\log\left(\frac{1}{\pi_*}\right)}\right| &\leqslant \left|\widehat{\gamma}_{lb}\left(2+\log\left(1/\pi_*\right)\right) - \gamma\left(2+\log\left(1/\widehat{\pi}_{*,lb}\right)\right)\right| \\
&\leqslant 2\left|\widehat{\gamma}_{lb}-\gamma\right| + \log(1/\pi_*)\left|\widehat{\gamma}_{lb}-\gamma\right| + \gamma\log\left(\frac{\widehat{\pi}_{*,lb}}{\pi_*}\right) \\
&\leqslant O\left(\frac{|\mathcal{X}|\log(1/\pi_*)}{\pi_*\gamma}\sqrt{\frac{\log\log n}{\pi_* n}}\right) + \gamma\left(\frac{\widehat{\pi}_{*,lb}}{\pi_*}-1\right) \\
&\leqslant O\Bigg(\max\left\{\frac{|\mathcal{X}|\log(1/\pi_*)}{\pi_*\gamma}\sqrt{\frac{\log\log n}{\pi_* n}},\right. \\
&\qquad \left.\max_{i,j\in\mathcal{X}}|\mathcal{X}|\sqrt{\frac{P_{i,j}\log\log n}{\pi_i n}}\right\}\Bigg),
\end{aligned}$$

where we use the convergence rate results of $\widehat{b}$ and $\widehat{w}$ in Theorem 5.2.1.

Remark 5.3.3. We can apply the same idea of Theorem 5.3.1 to estimate other quantities of interest, such as the modified log-Sobolev constant α_0 proposed in Bobkov and Tetali [7], Caputo et al. [9], Goel [45], where α_0 is defined via

$$\alpha_0(P) := \alpha_0 = \inf\left\{\frac{\mathcal{E}(e^f, f)}{2\mathrm{Ent}_\pi(e^f)}\,;\mathrm{Ent}_\pi(e^f)\neq 0\right\}.$$

Making use of Bobkov and Tetali [7, Proposition 3.6] we see that

$$2\alpha \leqslant \alpha_0 \leqslant \gamma,$$

so similar to Theorem 5.3.1, with probability at least $1-3\delta$, α_0 is trapped in

$$\alpha_0 \in \left[\frac{2(1-2\widehat{\pi}_{*,lb})\widehat{\gamma}_{lb}}{\log\left(\frac{1}{\widehat{\pi}_{*,lb}}-1\right)}, \widehat{\gamma}_{ub}\right]. \tag{5.3.3}$$

In this spirit, we can estimate the Cheeger constant Φ_* (see e.g. Lawler and Sokal [68]) defined by

$$\Phi_* := \min_{S:\pi(S)\leqslant\frac{1}{2}} \frac{\sum_{x\in S, y\in S^c}\pi(x)P(x,y)}{\pi(S)}.$$

Define the right spectral gap by $\gamma_r := 1 - \lambda_2$. To estimate γ_r, with probability at least $1 - \delta$, we see that γ_r is trapped in

$$\gamma_r \in [\widehat{\gamma}_{r,lb}, \widehat{\gamma}_{r,ub}],$$

where $\widehat{\gamma}_{r,lb} := 1 - \min\{1, \widehat{\lambda}_2 + \widehat{w}\}$, $\widehat{\gamma}_{r,ub} := 1 - \max\{-1, \widehat{\lambda}_2 - \widehat{w}\}$, and using the Cheeger inequality for Markov chains [74, Theorem 13.14]

$$\frac{\gamma_r}{2} \leqslant \Phi_* \leqslant \sqrt{2\gamma_r},$$

we reach the conclusion that with probability at least $1 - 2\delta$,

$$\Phi_* \in \left[\frac{\widehat{\gamma}_{r,lb}}{2}, \sqrt{2\widehat{\gamma}_{r,ub}}\right]. \tag{5.3.4}$$

Proof of Theorem 5.3.1. By union bound, it suffices for us to show that

$$\mathbb{P}\left(\alpha \notin \left[\frac{(1-2\widehat{\pi}_{*,lb})\widehat{\gamma}_{lb}}{\log\left(\frac{1}{\widehat{\pi}_{*,lb}} - 1\right)}, \frac{\widehat{\gamma}_{ub}}{2}\right]\right) \leqslant \mathbb{P}\left(\alpha < \frac{(1-2\widehat{\pi}_{*,lb})\widehat{\gamma}_{lb}}{\log\left(\frac{1}{\widehat{\pi}_{*,lb}} - 1\right)}\right) + \mathbb{P}\left(\alpha > \frac{\widehat{\gamma}_{ub}}{2}\right) \leqslant 3\delta.$$

First, using the upper bound in (5.3.1), we see that

$$\mathbb{P}\left(\alpha > \frac{\widehat{\gamma}_{ub}}{2}\right) \leqslant \mathbb{P}\left(\frac{\gamma}{2} > \frac{\widehat{\gamma}_{ub}}{2}\right) \leqslant \mathbb{P}\left(|\gamma - \widehat{\gamma}_{ub}| > 0\right) \leqslant \delta,$$

where the last inequality follows from Theorem 5.2.1. Next, we first consider two axillary functions $f, g : [0, 0.5] \to \mathbb{R}$ with

$$f(x) := \frac{1-2x}{\log\left(\frac{1}{x} - 1\right)}, \quad f'(x) = \frac{\frac{1}{x-1} + \frac{1}{x} - 2\log\left(\frac{1}{x} - 1\right)}{\log^2\left(\frac{1}{x} - 1\right)} = \frac{g(x)}{\log^2\left(\frac{1}{x} - 1\right)}, \tag{5.3.5}$$

$$g(x) := \frac{1}{x-1} + \frac{1}{x} - 2\log\left(\frac{1}{x} - 1\right), \quad g'(x) = \frac{-(2x-1)^2}{x^2(x-1)^2} < 0, \tag{5.3.6}$$

where $f'(x) > 0$ if and only if $g(x) > 0$ if and only if $x \in [0, 0.5)$, so $f(x) < f(y)$ for $x < y, x, y \in [0, 0.5]$. Now, we apply the lower bound in (5.3.1) to get

$$\mathbb{P}\left(\alpha < \frac{(1-2\widehat{\pi}_{*,lb})\widehat{\gamma}_{lb}}{\log\left(\frac{1}{\widehat{\pi}_{*,lb}} - 1\right)}\right) \leqslant \mathbb{P}\left(\frac{(1-2\pi_*)\gamma}{\log\left(\frac{1}{\pi_*} - 1\right)} < \frac{(1-2\widehat{\pi}_{*,lb})\widehat{\gamma}_{lb}}{\log\left(\frac{1}{\widehat{\pi}_{*,lb}} - 1\right)}\right)$$

$$= \mathbb{P}\left(f(\pi_*)\gamma < f(\widehat{\pi}_{*,lb})\widehat{\gamma}_{lb}, \pi_* \geqslant \widehat{\pi}_{*,lb}\right)$$

$$+ \mathbb{P}\left(f(\pi_*)\gamma < f(\widehat{\pi}_{*,lb})\widehat{\gamma}_{lb}, \pi_* < \widehat{\pi}_{*,lb}\right)$$

$$\leqslant \mathbb{P}(\gamma < \widehat{\gamma}_{lb}) + \mathbb{P}(\pi_* < \widehat{\pi}_{*,lb}) \leqslant 2\delta,$$

where we use $f(\pi_*) > f(\widehat{\pi}_{*,lb})$ on the event $\{\pi_* \geqslant \widehat{\pi}_{*,lb}\}$ in the second inequality, and the last inequality again follows from Theorem 5.2.1. □

5.4 Estimation of eigenspace

In this section, we would like to estimate the subspace spanned by the eigenvectors of P. A major motivation of this topic stems from the mixing time estimation idea from Hsu et al. [50], in which they propose to estimate the mixing time

$$t_{\mathrm{mix}} := \min\left\{n \in \mathbb{N};\ \sup_q \max_{A \subset \mathcal{X}} |\mathbb{P}_q(X_n \in A) - \pi(A)| \leqslant 1/4\right\}$$

via the inequality as in [74, Theorem 12.3 and 12.4]

$$\left(\frac{1}{\gamma} - 1\right)\ln 2 \leqslant t_{\mathrm{mix}} \leqslant \frac{1}{\gamma}\ln\left(\frac{4}{\pi_*}\right).$$

A sharper lower bound, known as the Wilson's method (see e.g. [74, Theorem 13.5] and Saloff-Coste [98]), requires knowledge on the eigenvector of P. More precisely, suppose that λ is an eigenvalue of P with $1/2 < \lambda < 1$ and v is an associated eigenvector. Under the assumption that for all $x \in \mathcal{X}$, $\mathbb{E}_x(|v(X_1) - v(x)|^2) \leqslant R$ for some $R > 0$, then for $x \in \mathcal{X}$,

$$\frac{1}{2\log(1/\lambda)}\left(\log\left(\frac{(1-\lambda)v(x)^2}{2R}\right) + \log 3\right) \leqslant t_{\mathrm{mix}}.$$

From this relationship it justifies the need to precisely estimate the eigenvectors of P in order to give a tighter lower bound on mixing time. A natural estimator of such subspace is the subspace spanned by the eigenvectors of $\mathrm{Sym}(\widehat{L})$, where $\widehat{L}$ is defined in (4) of Section 5.2. To measure the differences of the angle between these two subspaces, we recall the notion of principal angles: if $V, \widehat{V}$ have d orthonormal columns, then the vector of d principal angles is $(\cos^{-1}\sigma_1, \ldots, \cos^{-1}\sigma_d)^T$, where $\sigma_1 \geqslant \ldots \geqslant \sigma_d$ are the singular values of $\widehat{V}^T V$. Denote a $d \times d$ diagonal matrix $\Theta(\widehat{V}, V)$ with diagonal entries given by $(\cos^{-1}\sigma_i)_{i=1}^d$, and $\sin\Theta(\widehat{V}, V)$ to be the matrix defined elementwise. We also write $\|\cdot\|_F$ to be the Frobenius norm of a matrix. By means of a variant of Davis-Kahan theorem Davis and Kahan [27] introduced recently by Yu et al. [115], we offer the following statistical guarantee in estimating the subspace spanned by the eigenvectors of P in Theorem 5.4.1. However, one possible drawback of this approach is that we need to impose a population eigengap condition (i.e. $\Delta > 0$ in Theorem 5.4.1), which are satisfied, for example, for the class of birth-death chains (see e.g. Levin et al. [74]).

Theorem 5.4.1. *Suppose we are given a sample path from a reversible ergodic Markov chain $(X_1, X_2, \ldots, X_n)$ and a desired level of accuracy $\delta \in (0,1)$. Assume that P has eigenvalues-eigenvectors pair as $(\gamma_i, v_i)_{i=1}^{|\mathcal{X}|}$ with $1 = \gamma_1 > \gamma_2 \geqslant \ldots \geqslant \gamma_{|\mathcal{X}|}$ and $\mathrm{Sym}(\widehat{L})$ has eigenvalues-eigenvectors pair as $(\widehat{\gamma}_i, \widehat{v}_i)_{i=1}^{|\mathcal{X}|}$ with $1 = \widehat{\gamma}_1 \geqslant \widehat{\gamma}_2 \geqslant \ldots \geqslant \widehat{\gamma}_{|\mathcal{X}|}$, where $\widehat{L}$ is defined in (4) of Section 5.2. Fix $1 \leqslant r \leqslant s \leqslant |\mathcal{X}|$, $d := s - r + 1$ and assume that the eigengap condition holds, i.e. $\Delta := \min(\gamma_{r-1} - \gamma_r, \gamma_s - \gamma_{s+1}) > 0$. Let $V := (v_r, v_{r+1}, \ldots, v_s)$ and $\widehat{V} := (\widehat{v}_r, \widehat{v}_{r+1}, \ldots, \widehat{v}_s)$. With probability at least $1 - \delta$, we have*

$$\left\|\sin\Theta(\widehat{V}, V)\right\|_F \leqslant \frac{2\sqrt{d}\widehat{w}}{\Delta} = O\left(\frac{|\mathcal{X}|}{\Delta\pi_*\gamma}\sqrt{\frac{d\log\log n}{\pi_* n}}\right), \tag{5.4.1}$$

where $\widehat{w}$ is defined in (5.2.6).

Remark 5.4.1. Note that Theorem 5.4.1 also hold under the operator norm since the operator norm is less than or equal to the Frobenius norm.

Proof. Define $L := D_\pi^{1/2} P D_\pi^{-1/2}$, where D_π is a diagonal matrix with entries given by

π. Note that L is a symmetric matrix since P is reversible. Then, by [115, Theorem 2], we see that

$$\left\|\sin\Theta(\widehat{V},V)\right\|_F \leqslant \frac{2\sqrt{d}\left\|\mathrm{Sym}(\widehat{L})-L\right\|_{op}}{\Delta} \leqslant \frac{2\sqrt{d}\left\|\widehat{L}-L\right\|_{op}}{\Delta} \leqslant \frac{2\sqrt{d}\widehat{w}}{\Delta},$$

where $\left\|\mathrm{Sym}(\widehat{L})-L\right\|_{op}$ is the operator norm of the random matrix $\mathrm{Sym}(\widehat{L})-L$, the second inequality follows from triangle inequality and symmetry of L, and the last inequality stems from [50, Lemma 7] with probability at least $1-\delta$. $\square$

BIBLIOGRAPHY

[1] D. Aldous and J. A. Fill. Reversible Markov Chains and Random Walks on Graphs, 2002. Unfinished monograph, recompiled 2014, available at http://www.stat.berkeley.edu/~aldous/RWG/book.html. 1, 5, 70, 77, 78, 98, 109

[2] R. Altmeyer and J. Chorowski. Estimation error for occupation time functionals of stationary Markov processes. *arXiv:1610.05225*, 2016. 4, 67, 71, 91, 92, 94, 95

[3] S. Asmussen. *Applied probability and queues*, volume 51 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 2003. Stochastic Modelling and Applied Probability. 54, 82

[4] P. H. Baxendale. Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Probab.*, 15(1B):700–738, 2005. 80

[5] A. Berman and R. J. Plemmons. Matrix group monotonicity. *Proc. Amer. Math. Soc.*, 46:355–359, 1974. 81

[6] J. Bierkens. Non-reversible Metropolis-Hastings. *Stat. Comput.*, 26(6):1213–1228, 2016. 2, 123

[7] S. G. Bobkov and P. Tetali. Modified logarithmic Sobolev inequalities in discrete settings. *J. Theoret. Probab.*, 19(2):289–336, 2006. 155

[8] V. Britanak, P. C. Yip, and K. R. Rao. *Discrete cosine and sine transforms*. Elsevier/Academic Press, Amsterdam, 2007. 107

[9] P. Caputo, P. Dai Pra, and G. Posta. Convex entropy decay via the Bochner-Bakry-Emery approach. *Ann. Inst. Henri Poincaré Probab. Stat.*, 45(3):734–753, 2009. 155

[10] D. Chafaï and A. Joulin. Intertwining and commutation relations for birth-death processes. *Bernoulli*, 19(5A):1855–1879, 2013. 68

[11] G.-Y. Chen and L. Saloff-Coste. The cutoff phenomenon for ergodic Markov processes. *Electron. J. Probab.*, 13:no. 3, 26–78, 2008. 64, 65, 66, 88

[12] G.-Y. Chen and L. Saloff-Coste. The L^2-cutoff for reversible Markov processes. *J. Funct. Anal.*, 258(7):2246–2315, 2010. 4, 67, 71, 79, 88, 91

[13] G.-Y. Chen and L. Saloff-Coste. Computing cutoff times of birth and death chains. *Electron. J. Probab.*, 20:no. 76, 47, 2015. 4, 60, 64, 84

[14] G.-Y. Chen and Y.-C. Sheu. On the log-Sobolev constant for the simple random walk on the n-cycle: the even cases. *J. Funct. Anal.*, 202(2):473–485, 2003. 153

[15] G.-Y. Chen, J.-M. Hsu, and Y.-C. Sheu. The L^2-cutoffs for reversible Markov chains. *To appear in Ann. Appl. Probab.*. 4, 71, 88

[16] G.-Y. Chen, W. W. Liu, and L. Saloff-Coste. The logarithmic Sobolev constant of some finite Markov chains. *Ann. Fac. Sci. Toulouse Math. (6)*, 17(2):239–290, 2008. 153, 154

[17] M. Chen. Estimation of spectral gap for Markov chains. *Acta Math. Sinica (N.S.)*, 12(4):337–360, 1996. 123

[18] T.-L. Chen and C.-R. Hwang. Accelerating reversible Markov chains. *Statist. Probab. Lett.*, 83(9):1956–1962, 2013. 1

[19] M. Chesney, M. Jeanblanc-Picqué, and M. Yor. Brownian excursions and Parisian barrier options. *Adv. in Appl. Probab.*, 29(1):165–184, 1997. 91

[20] M. Choi and P. Patie. Skip-free Markov chains. *Submitted*, 2016. 9, 32, 36, 68, 70, 75, 76, 84, 87

[21] K. L. Chung and J. B. Walsh. *Markov processes, Brownian motion, and time symmetry*, volume 249 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, New York, second edition, 2005. 20

[22] P. Clifford and A. Sudbury. A sample path proof of the duality for stochastically monotone Markov processes. *Ann. Probab.*, 13(2):558–565, 1985. 4, 70

[23] B. Cloez and C. Delplancke. Intertwinings and Stein's magic factors for birth-death processes. *arXiv preprint arXiv:1609.08390*, 2016. 68

[24] J. B. Conway. *A course in functional analysis*, volume 96 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1990. 50, 54

[25] H. Cui and Y.-H. Mao. Eigentime identity for asymmetric finite Markov chains. *Front. Math. China*, 5(4):623–634, 2010. 5, 70, 77, 102

[26] E. B. Davies. Metastable states of symmetric Markov semigroups. II. *J. London Math. Soc. (2)*, 26(3):541–556, 1982. 141, 142

[27] C. Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.*, 7:1–46, 1970. 158

[28] P. Diaconis and J. A. Fill. Strong stationary times via a new form of duality. *Ann. Probab.*, 18(4):1483–1522, 1990. 70

[29] P. Diaconis and L. Miclo. On quantitative convergence to quasi-stationarity. *Ann. Fac. Sci. Toulouse Math. (6)*, 24(4):973–1016, 2015. 72, 106

[30] P. Diaconis and L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.*, 6(3):695–750, 1996. 6, 153, 154, 155

[31] P. Diaconis and L. Saloff-Coste. Separation cut-offs for birth and death chains. *Ann. Appl. Probab.*, 16(4):2098–2122, 2006. 4, 11, 60, 64, 71, 84, 85, 87

[32] P. Diaconis and D. Stroock. Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.*, 1(1):36–61, 1991. 121, 142

[33] P. Diaconis, S. Holmes, and R. M. Neal. Analysis of a nonreversible Markov chain sampler. *Ann. Appl. Probab.*, 10(3):726–752, 2000. 2, 136, 137

[34] P. Diaconis, K. Khare, and L. Saloff-Coste. Gibbs sampling, exponential families and orthogonal polynomials. *Statist. Sci.*, 23(2):151–178, 2008. 1, 98

[35] A. B. Duncan, T. Lelièvre, and G. A. Pavliotis. Variance reduction using nonreversible Langevin samplers. *J. Stat. Phys.*, 163(3):457–491, 2016. 2

[36] N. Dunford. Spectral operators. *Pacific J. Math.*, 4:321–354, 1954. 69, 72

[37] N. Dunford and J. T. Schwartz. *Linear operators. Part III: Spectral operators*. Interscience Publishers [John Wiley & Sons, Inc.], New York-London-Sydney, 1971. 69, 72, 80

[38] H. Dym and H. P. McKean. *Gaussian processes, function theory, and the inverse spectral problem*. Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1976. Probability and Mathematical Statistics, Vol. 31. 14, 15, 53, 82

[39] E. B. Dynkin. The boundary theory of Markov processes (discrete case). *Uspehi Mat. Nauk*, 24(2 (146)):3–42, 1969. 8, 16, 19

[40] W. Feller. Diffusion processes in one dimension. *Trans. Amer. Math. Soc.*, 77: 1–31, 1954. 14

[41] W. Feller. Infinitely divisible distributions and Bessel functions associated with random walks. *SIAM J. Appl. Math.*, 14:864–875, 1966. 12, 39, 46

[42] J. A. Fill. Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process. *Ann. Appl. Probab.*, 1(1):62–87, 1991. 4, 2, 5, 11, 56, 67, 70, 78, 83, 108, 113, 121

[43] 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. 9, 10, 39, 40, 42, 43, 63, 70, 84, 86, 101

[44] S. T. Garren and R. L. Smith. Estimating the second largest eigenvalue of a Markov transition matrix. *Bernoulli*, 6(2):215–242, 2000. 150

[45] S. Goel. Modified logarithmic Sobolev inequalities for some models of random walk. *Stochastic Process. Appl.*, 114(1):51–79, 2004. 155

[46] R. Griffiths. Multivariate Krawtchouk polynomials and composition birth and death processes. *Symmetry*, 8(5):Art. 33, 19, 2016. 1, 96

[47] M. Hairer, A. M. Stuart, and S. J. Vollmer. Spectral gaps for a Metropolis-Hastings algorithm in infinite dimensions. *Ann. Appl. Probab.*, 24(6):2455–2490, 2014. 108

[48] W. K. Hastings. Monte carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970. 5, 108

[49] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, second edition, 2013. 58, 120

[50] D. Hsu, A. Kontorovich, and C. Szepesvári. Mixing time estimation in reversible Markov chains from a single sample path. In *Proceedings of the 28th Interna-*

tional Conference on Neural Information Processing Systems, NIPS'15, pages 1459–1467, 2015. 4, ix, 6, 150, 151, 152, 153, 157, 159

[51] T. Huillet and S. Martinez. Duality and intertwining for discrete Markov kernels: relations and examples. *Adv. in Appl. Probab.*, 43(2):437–460, 2011. 55, 70, 82

[52] W. Huisinga and B. Schmidt. Metastability and dominant eigenvalues of transfer operators. In *New algorithms for macromolecular simulation*, volume 49 of *Lect. Notes Comput. Sci. Eng.*, pages 167–182. Springer, Berlin, 2006. 142, 144, 145

[53] C.-R. Hwang, S.-Y. Hwang-Ma, and S. J. Sheu. Accelerating Gaussian diffusions. *Ann. Appl. Probab.*, 3(3):897–913, 1993. 1

[54] C.-R. Hwang, S.-Y. Hwang-Ma, and S.-J. Sheu. Accelerating diffusions. *Ann. Appl. Probab.*, 15(2):1433–1444, 2005. 1

[55] A. Inoue and C. Trapani. Non-self-adjoint resolutions of the identity and associated operators. *Complex Anal. Oper. Theory*, 8(7):1531–1546, 2014. 69, 80, 91, 93

[56] S. Jansen and N. Kurt. On the notion(s) of duality for Markov processes. *Probab. Surv.*, 11:59–120, 2014. 70

[57] S. Kamath and S. Verdú. Estimation of entropy rate and Rényi entropy rate for Markov chains. In *Information Theory (ISIT), 2016 IEEE International Symposium on*, pages 685–689. IEEE, 2016. 150

[58] S. Karlin and J. McGregor. Coincidence properties of birth and death processes. *Pacific J. Math.*, 9:1109–1140, 1959. 9, 67, 98

[59] S. Karlin and J. L. McGregor. The differential equations of birth-and-death pro-

cesses, and the Stieltjes moment problem. *Trans. Amer. Math. Soc.*, 85:489–546, 1957. 7, 15

[60] J. Keilson. Log-concavity and log-convexity in passage time densities of diffusion and birth-death processes. *J. Appl. Probability*, 8:391–398, 1971. 9

[61] J. G. Kemeny, J. L. Snell, and A. W. Knapp. *Denumerable Markov chains*. Springer-Verlag, New York-Heidelberg-Berlin, second edition, 1976. With a chapter on Markov random fields, by David Griffeath, Graduate Texts in Mathematics, No. 40. 13, 16, 18, 33

[62] D. G. Kendall. Unitary dilations of one-parameter semigroups of Markov transition operators, and the corresponding integral representations for Markov processes with a countable infinity of states. *Proc. London Math. Soc. (3)*, 9:417–431, 1959. 2, 67

[63] J. T. Kent and N. T. Longford. An eigenvalue decomposition for first hitting times in random walks. *Z. Wahrsch. Verw. Gebiete*, 63(1):71–84, 1983. 12, 39, 41

[64] K. Khare and N. Mukherjee. Convergence analysis of some multivariate Markov chains using stochastic monotonicity. *Ann. Appl. Probab.*, 23(2):811–833, 2013. 96

[65] K. Khare and H. Zhou. Rates of convergence of some multivariate Markov chains with polynomial eigenfunctions. *Ann. Appl. Probab.*, 19(2):737–777, 2009. 1, 96

[66] R. Koekoek and R. F. Swarttouw. The Askey-scheme of hypergeometric orthogonal polynomials and its q-analogue, 1998. 98

[67] I. Kontoyiannis and S. P. Meyn. Geometric ergodicity and the spectral gap of non-reversible Markov chains. *Probab. Theory Related Fields*, 154(1-2):327–339, 2012. 2, 67

[68] G. F. Lawler and A. D. Sokal. Bounds on the L^2 spectrum for Markov chains and Markov processes: a generalization of Cheeger's inequality. *Trans. Amer. Math. Soc.*, 309(2):557–580, 1988. 155

[69] W. Ledermann and G. E. H. Reuter. Spectral theory for the differential equations of simple birth and death processes. *Philos. Trans. Roy. Soc. London. Ser. A.*, 246:321–369, 1954. 7, 54, 67

[70] J. R. Lee, S. Oveis Gharan, and L. Trevisan. Multi-way spectral partitioning and higher-order Cheeger inequalities. In *STOC'12—Proceedings of the 2012 ACM Symposium on Theory of Computing*, pages 1117–1130. ACM, New York, 2012. 143

[71] T.-Y. Lee and H.-T. Yau. Logarithmic Sobolev inequality for some models of random walks. *Ann. Probab.*, 26(4):1855–1873, 1998. 153

[72] C. Lefèvre, P. Patie, and S. Vakeroudis. Fluctuation theory of one-sided Ornstein-Uhlenbeck type processes. *Working paper*, 2016. 9, 32

[73] D. A. Levin and Y. Peres. Estimating the Spectral Gap of a Reversible Markov Chain from a Short Trajectory. *arXiv preprint arXiv:1612.05330*, 2016. 150

[74] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2009. 1, 62, 78, 121, 128, 132, 135, 137, 156, 157, 158

[75] P. Lorek. *Speed of convergence to stationarity for stochastically monotone Markov chains*. PhD thesis, University of Wroclaw, 2007. 1

[76] Y. H. Mao, C. Zhang, and Y. H. Zhang. Separation cutoff for upward skip-free chains. *J. Appl. Probab.*, 53(1):299–306, 03 2016. 4, 11, 60, 84

[77] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092, 1953. 5, 108

[78] S. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Cambridge University Press, Cambridge, second edition, 2009. With a prologue by Peter W. Glynn. 1, 128

[79] C. A. Micchelli and R. A. Willoughby. On functions which preserve the class of Stieltjes matrices. *Linear Algebra Appl.*, 23:141–156, 1979. 86

[80] L. Miclo. On absorption times and Dirichlet eigenvalues. *ESAIM Probab. Stat.*, 14:117–150, 2010. 40

[81] L. Miclo. On hyperboundedness and spectrum of Markov operators. *Invent. Math.*, 200(1):311–343, 2015. 1, 143, 147

[82] L. Miclo. An absorbing eigentime identity. *Markov Process. Related Fields*, 21 (2):249–262, 2015. 5, 10, 70, 77

[83] L. Miclo. On the Markovian similarity. Preprint, Mar. 2016. 4, 68, 96

[84] R. Montenegro and P. Tetali. Mathematical aspects of mixing times in Markov chains. *Found. Trends Theor. Comput. Sci.*, 1(3):x+121, 2006. 1, 78, 136, 137

[85] P. Patie and M. Savov. Spectral expansion of non-self-adjoint generalized Laguerre semigroups. *Submitted*, page 162p. (current version), 2015. 3, 4, 58, 68, 80

[86] P. Patie and C. Van Weverberg. Exit times of continuous state branching processes with immigration. *Working paper*, 2016. 9, 32

[87] P. Patie and V. Vigon. On completely assymetric markov processes. *Working paper*, 2016. 9, 32

[88] P. Patie and Y. Zhao. Spectral decomposition of fractional operators and a reflected stable semigroup. *J. Differential Equations*, 262(3):1690–1719, 2017. 68

[89] D. Paulin. Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electron. J. Probab.*, 20:no. 79, 1–32, 2015. 4, 3, 5, 108, 109, 113, 136, 137, 138, 140

[90] P. H. Peskun. Optimum Monte-Carlo sampling using Markov chains. *Biometrika*, 60:607–612, 1973. 117

[91] M. Reed and B. Simon. *Methods of modern mathematical physics. I*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York, second edition, 1980. Functional analysis. 54

[92] G. E. H. Reuter. Denumerable Markov processes. II. *J. London Math. Soc.*, 34: 81–91, 1959. 31

[93] G. O. Roberts and J. S. Rosenthal. Geometric ergodicity and hybrid Markov chains. *Electron. Comm. Probab.*, 2:no. 2, 13–25 (electronic), 1997. 1, 11, 129, 134, 135

[94] G. O. Roberts and J. S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probab. Surv.*, 1:20–71, 2004. 1, 5, 108, 114, 116, 128, 135

[95] G. O. Roberts and R. L. Tweedie. Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, 83 (1):95–110, 1996. 108

[96] D. Rudolf. Explicit error bounds for Markov chain Monte Carlo. Dissertationes Math. 485 (2012), 93 pp, 2012. 112

[97] L. Saloff-Coste. Lectures on finite Markov chains. In *Lectures on probability theory and statistics (Saint-Flour, 1996)*, volume 1665 of *Lecture Notes in Math.*, pages 301–413. Springer, Berlin, 1997. 1, 11, 56, 58, 153

[98] L. Saloff-Coste. Total variation lower bounds for finite Markov chains: Wilson's lemma. In *Random walks and geometry*, pages 515–532. Walter de Gruyter, Berlin, 2004. 6, 157

[99] R. Sasaki. Exactly solvable birth and death processes. *J. Math. Phys.*, 50(10): 103509, 18, 2009. 98

[100] W. Schoutens. *Stochastic processes and orthogonal polynomials*, volume 146 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 2000. 98

[101] C. Schütte and M. Sarich. *Metastability and Markov state models in molecular dynamics*, volume 24 of *Courant Lecture Notes in Mathematics*. Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 2013. Modeling, analysis, algorithmic approaches. 142, 144

[102] D. Siegmund. The equivalence of absorbing and reflecting barrier problems for stochastically monotone Markov processes. *Ann. Probability*, 4(6):914–924, 1976. 4, 54, 70, 82

[103] G. Singleton. Asymptotically exact estimates for metastable Markov semigroups. *Quart. J. Math. Oxford Ser. (2)*, 35(139):321–329, 1984. 141, 142, 144, 146

[104] F. Spitzer. A combinatorial lemma and its application to probability theory. *Trans. Amer. Math. Soc.*, 82:323–339, 1956. 9

[105] F. W. Steutel and K. van Harn. *Infinite divisibility of probability distributions on the real line*, volume 259 of *Monographs and Textbooks in Pure and Applied Mathematics*. Marcel Dekker, Inc., New York, 2004. 44, 45, 46, 47

[106] Y. Sun, F. Gomez, and J. Schmidhuber. Improving the asymptotic performance of Markov chain monte-carlo by inserting vortices. In *NIPS*, pages 2235–2243, USA, 2010. 2, 123

[107] R. C. Thompson. Singular values, diagonal elements, and convexity. *SIAM J. Appl. Math.*, 32(1):39–63, 1977. 59, 84

[108] L. Tierney. A note on Metropolis-Hastings kernels for general state spaces. *Ann. Appl. Probab.*, 8(1):1–9, 1998. 117, 118

[109] C. Villani. Hypocoercivity. *Mem. Amer. Math. Soc.*, 202(950):iv+141, 2009. 11, 57, 70, 79

[110] O. V. Viskov. A random walk with an upper-continuous component, and the Lagrange inversion formula. *Teor. Veroyatnost. i Primenen.*, 45(1):166–175, 2000. 12, 39, 46

[111] J. Wermer. Commuting spectral measures on Hilbert space. *Pacific J. Math.*, 4: 355–361, 1954. 69

[112] E. L. Wilmer. *Exact rates of convergence for some simple non-reversible Markov chains*. PhD thesis, Harvard University, 1999. 1

[113] W. Woess. *Denumerable Markov chains*. EMS Textbooks in Mathematics. European Mathematical Society (EMS), Zürich, 2009. 16

[114] R. M. Young. *An introduction to nonharmonic Fourier series*. Academic Press, Inc., San Diego, CA, first edition, 2001. 51, 54, 69, 73, 81

[115] Y. Yu, T. Wang, and R. J. Samworth. A useful variant of the Davis-Kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2015. 158, 159

[116] H. Zhou. *Examples of multivariate Markov chains with orthogonal polynomial eigenfunctions*. PhD thesis, Stanford University, 2008. 72, 96, 98, 106