

ANALYSIS OF NON-REVERSIBLE MARKOV CHAINS VIA SIMILARITY ORBIT

MICHAEL C.H. CHOI AND PIERRE PATIE

ABSTRACT. In this paper, we develop an in-depth analysis of non-reversible Markov chains on denumerable state space from a similarity orbit perspective. In particular, we study the class of Markov chains whose transition kernel is in the similarity orbit of a normal transition kernel, such as the one of birth-death chains or reversible Markov chains. We start by identifying a set of sufficient conditions for a Markov chain to belong to the similarity orbit of a birth-death one. As by-products, we obtain a spectral representation in terms of non-self-adjoint resolutions of identity in the sense of Dunford [22] 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 paper, 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.

AMS 2010 subject classifications: Primary 60J05, 60J10, 60J27; Secondary 60J20, 37A25, 37A30

Keywords and phrases: non-reversible Markov chains; normal operators; spectral operators; intertwining; similarity orbit; spectral gap; cutoff; eigentime identity; orthogonal polynomials

CONTENTS

1. Introduction	2
2. Spectral theory of the class \mathcal{S} and its convergence rate to equilibrium	5
2.1. Proof of Theorem 2.1	9
2.2. Proof of Theorem 2.2	10
2.3. Proof of Corollary 2.2	11
3. Separation cutoff	12
3.1. Proof of Theorem 3.1	12
4. L^2 -cutoff	14
4.1. Proof of Theorem 4.1	15
5. Non-asymptotic estimation error bounds for integral functionals	15
5.1. Proof of Theorem 5.1	16
6. Similarity orbit of reversible Markov chains	18
6.1. Permutation link on finite state space	20
6.2. n -dimensional permutation link on \mathbb{N}_0	21
6.3. Pure birth link on finite state space	22
6.4. $(n + 1)$ -dimensional pure birth link on \mathbb{N}_0	24
6.5. Random walk link on finite state space	25
References	26

Date: June 7, 2017.

This work was partially supported by NSF Grant DMS-1406599 and ARC IAPAS, a fund of the Communauté française de Belgique. The second author is grateful for the hospitality of the LMA at the UPPA, where part of this work was completed.

1. INTRODUCTION

The spectral theorem of normal operators is undoubtedly a powerful tool to deal with substantial and difficult issues arising in the analysis of Markov chains. The intrusion of spectral theory to the analysis of Markov chains dates back to the long line of work initiated by Ledermann and Reuter [40] and Karlin and McGregor [32] who were among the first to offer a detailed spectral analysis in the direction of reversible birth-death processes. Beyond eigenvalues expansion, the spectral theorem also appears in the study of the rate of convergence to equilibrium, mixing time, eigentime identity, separation cutoff and L^2 -cutoff, see e.g. Aldous and Fill [1], Chen and Saloff-Coste [10], Cui and Mao [17], Diaconis and Saloff-Coste [20], Levin et al. [41], Miclo [44], to name but a few. It is also central for their statistical estimations as it is demonstrated by the recent work of Altmeyer and Chorowski [2] for the integral functionals of normal Markov chains.

However, the lack of spectral theorem for non-normal operators gives major difficulties for tackling these fundamental topics in the context of general Markov chains, since the transition kernel P is 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 a reference (invariant or excessive) measure of P . Not only the non-reversibility property, or generally the non-normality of P , is a generic property from a theoretical perspective, it is also natural and becomes increasingly popular recently in various applications. For instance, non-reversible Markov chains appear in the study of queueing networks, fluid approximation and MCMC, see Fort et al. [27] and the references therein, and the very recent introduction of non-reversible Metropolis-Hastings and its variants, see e.g. Bierkens [6], Rosenthal and Rosenthal [50].

To overcome the challenge of analyzing non-self-adjoint operators, a wide variety of intriguing ideas has been elaborated to deal with specific issues. This includes, for example, the dilation concept developed by Kendall [33], reversibilizations techniques as in Fill [25], Paulin [49] or recasting to a weighted- L^∞ space Kontoyiannis and Meyn [37, 38, 39].

In this paper, we propose an alternative remedy by resorting to the algebraic concept of similarity orbit of normal Markov chains as defined in Definition 1.1 below. This identifies a class of transition kernels of Markov chains, denoted by \mathcal{S} , which is a subset of \mathcal{M} , the set of Markov transition kernels acting on a denumerable state space \mathcal{X} . We emphasize that our approach offers an unifying framework to analyze all substantial and classical topics for Markov kernels in \mathcal{S} that were enumerated above for normal Markov chains. This extends the work by the authors in [14] from skip-free Markov chains to general ones. It is also in line with the papers by Miclo [45], Patie and Savov [47] and Patie and Zhao [48] for the study of spectral theory of non-reversible Markov processes and by Chafaï and Joulin [8], Diaconis and Fill [18] and Cloez and Delplancke [16] for birth-death processes, which rely on the notion of intertwining relationships. We proceed by recalling the definition of similarity orbit as introduced in Choi and Patie [14].

Definition 1.1 (Similarity). We say that the transition kernel $P \in \mathcal{M}$ of a Markov chain X 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 $\Lambda : \ell^2(\pi_Q) \rightarrow \ell^2(\pi)$ (π_Q being a reference measure for Q) with bounded inverse such that

$$(1.1) \quad P\Lambda = \Lambda Q.$$

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 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 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 $(P_t)_{t \geq 0}$ (resp. $(Q_t)_{t \geq 0}$). It follows easily that if $L \overset{\Lambda}{\sim} G$ then $P_t \overset{\Lambda}{\sim} Q_t$ for $t \geq 0$.

The \mathcal{S} class is now defined as 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 kernels Q on \mathcal{X} , that is, $Q\hat{Q} = \hat{Q}Q$ in $\ell^2(\pi_Q)$ where $\hat{\cdot}$ denotes throughout the corresponding object for the time-reversal process.

Definition 1.2 (The \mathcal{S} class). Suppose that $Q \in \mathcal{N}$. The similarity orbit of Q (in \mathcal{M}) 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 [56], 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 [22, Section 3], see also Dunford and Schwartz [23, Page 1938, Definition 1].

Finally, we say that $X \in \mathcal{S}^M$ if $X \in \mathcal{S}$ and its time-reversal $(X, \hat{\mathbb{P}})$ is stochastically monotone, i.e. $y \mapsto \hat{\mathbb{P}}_y(X_1 \leq x)$ 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 the paper. In Section 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 [22], see also Dunford and Schwartz [23]. 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 [30] 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 [57]. Another intriguing aspect of the similarity orbit analysis is that in the continuous-time setting with $L \in \mathcal{S}(G)$ (see Remark 1.2 above), where G is the generator of a normal Markov chain, then both the heat kernel $(e^{tL})_{t \geq 0}$ and $(e^{tG})_{t \geq 0}$ share the same *eigentime* identity, offering new examples and insights to the sequence of work by Aldous and Fill [1], Cui and Mao [17] and Miclo [44]. 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 \hat{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 [14] in the context of skip-free chains. Note that the notion of stochastic monotonicity is studied

by Siegmund [53] and Clifford and Sudbury [15] and intertwining between stochastic monotone birth-death chains, which are reversible chains, has been previously investigated in detail by Diaconis and Fill [18], Huillet and Martinez [29] and Jansen and Kurt [31]. 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 [25], 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 [55]. All these first consequences of the spectral representation are stated and proved in Section 2. Relying on such spectral decomposition as well as the fastest strong stationary time result of general chains obtained by Fill [26], 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 [20] carries over to the subclass $\mathcal{GMC}^+ \subset \mathcal{GMC}$ in Section 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 [10] and further elaborated in Chen et al. [12], we illustrate that the usual L^2 -cutoff criteria for reversible chains generalizes to the class \mathcal{S} in Section 4.

Second, in Section 5, we would like to estimate integral functionals of the type

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

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 $k \in \llbracket n \rrbracket := \{1, \dots, n\}$ 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 $(X_t)_{t \geq 0}$ 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 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 $(e^{tL})_{t \geq 0}$, 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 6.1.1 and 6.3.1), $M/M/\infty$ queue (Section 6.2.1 and 6.4.1), linear birth-death process (Section 6.2.2 and 6.4.2) and quadratic birth-death process (Section 6.1.2 and 6.3.2). Finally, we study the random walk orbit in Section 6.5, with the link Λ being the random walk previously studied by Diaconis and Miclo [19] and

Zhou [58]. An interesting aspect is that the right eigenfunction of L can now be interpreted as a discrete cosine transform.

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 1.2. We write $\|\cdot\|_{op}$ to be the operator norm, i.e. $\|P\|_{op} = \sup_{\|f\|_\pi=1} \|Pf\|_\pi$, and $\llbracket a, b \rrbracket := \{a, a+1, \dots, b-1, b\}$ for any $a \leq b \in \mathbb{Z}$. We proceed by recalling that P has a time-reversal \hat{P} , that is, for $x, y \in \mathcal{X}$,

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

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 \bar{g} is the complex conjugate of g . A spectral measure (or resolution of identity) in the sense of Dunford [22, Section 3] and Dunford and Schwartz [23, 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 $C > 0$ such that $\|E_B\|_{op} \leq C$ 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 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 [57, 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

$$(2.1) \quad A \sum_{k=1}^n |c_k|^2 \leq \left\| \sum_{k=1}^n c_k f_k \right\|^2 \leq B \sum_{k=1}^n |c_k|^2.$$

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. We are now ready to state the main result of this paper in the following, and the proof can be found in Section 2.1.

Theorem 2.1. Assume that $P \in \mathcal{S}$ with $P \stackrel{\Lambda}{\sim} Q$. 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 $\mathcal{X} = \llbracket 0, \tau \rrbracket$ with distinct eigenvalues then for any $f \in \ell^2(\pi)$ and $n \in \mathbb{N}$,

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

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

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

- (b) $P \in \mathcal{S}(Q)$ with $P \stackrel{\Lambda}{\sim} Q$ if and only if $\widehat{P} \in \mathcal{S}(\widehat{Q})$ with $\widehat{Q} \stackrel{\widehat{\Lambda}}{\sim} \widehat{P}$.
- (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 $\Lambda^{-1} \geq 0$ 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 $P \stackrel{\Lambda}{\sim} Q$ if and only if P has real and distinct eigenvalues.

Remark 2.1. We will illustrate Theorem 2.1 item (d) in Section 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 6, we also investigate in-depth two non-unitary orbits that we call the pure birth orbit and random walk orbit.

As Theorem 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 [14], such that the time-reversal \hat{P} intertwines with a birth-death chain with the link kernel Λ being related to the Siegmund kernel.

Definition 2.1 (The \mathcal{GMC} class). We say that, for some $\mathfrak{r} \geq 3$, $X \in \mathcal{GMC}_{\mathfrak{r}}$ if $(X, \mathbb{P}) \in \mathcal{M}$ with $\mathcal{X} = \llbracket 0, \mathfrak{r} \rrbracket$ and for every $x \in \llbracket 0, \mathfrak{r} - 1 \rrbracket$, its time-reversal $(X, \hat{\mathbb{P}})$ satisfies

- (1) (stochastic monotonicity) $\hat{\mathbb{P}}_{x+1}(X_1 \leq x) \leq \hat{\mathbb{P}}_x(X_1 \leq x)$,
- (2) (strict stochastic monotonicity) $\hat{\mathbb{P}}_{x+1}(X_1 \leq x - 1) < \hat{\mathbb{P}}_x(X_1 \leq x - 1)$, $x \neq \mathfrak{r} - 1$, and
- (3) (strict stochastic monotonicity) $\hat{\mathbb{P}}_{x+1}(X_1 \leq x + 1) < \hat{\mathbb{P}}_x(X_1 \leq x + 1)$, $x \neq \mathfrak{r} - 1$, and
- (4) (restricted downward jump) $\hat{\mathbb{P}}_{x+1}(X_1 \leq x - k) = \hat{\mathbb{P}}_x(X_1 \leq x - k)$, $x \neq \mathfrak{r} - 1$, $k \in \llbracket 2, x \rrbracket$, and
- (5) (restricted upward jump) $\hat{\mathbb{P}}_{x+1}(X_1 \leq x + k) = \hat{\mathbb{P}}_x(X_1 \leq x + k)$, $x \neq \mathfrak{r} - 1$, $k \in \llbracket 2, \mathfrak{r} - 1 - x \rrbracket$.

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

- (6) (lazy Siegmund dual) $\hat{\mathbb{P}}_x(X_1 \leq x) - \hat{\mathbb{P}}_{x+1}(X_1 \leq x) \geq \frac{1}{2}$.

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 2.5 below.

Remark 2.2. Recall that in Choi and Patie [14], 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 2.1 are automatically satisfied since the time-reversal \hat{P} is downward skip-free.

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

Example 2.1.

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

has eigenvalues 1, 0.44, 0.30, 0.26, and satisfies (1) – (6) in Definition 2.1.

We now formally state that \mathcal{GMC} is a subclass of \mathcal{S}^M (recall its definition in Definition 1.2), and the proof can be found in Section 2.2.

Theorem 2.2. $\mathcal{GMC} \subseteq \mathcal{S}^M$.

As a first application of Theorem 2.1, we first recall the celebrated eigentime identity studied by Aldous and Fill [1], Cui and Mao [17] and Miclo [44]: 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 2.1 item (a)), we can easily see that both P and Q share the same eigentime identity:

Corollary 2.1 (Eigentime identity). *Suppose that \mathcal{X} is a finite state space and $(Q_t)_{t \geq 0}$ (resp. $(P_t)_{t \geq 0}$) has generator G (resp. L) associated with the Markov chain $(X_t)_{t \geq 0}$ (resp. $(Y_t)_{t \geq 0}$). If $L \in \mathcal{S}(G)$ with eigenvalues $(-\lambda_i)_{i \in \llbracket \mathcal{X} \rrbracket}$, then $(P_t)_{t \geq 0}$ and $(Q_t)_{t \geq 0}$ share the same eigentime identity. That is, denote $\tau_y^Q := \inf\{t \geq 0; X_t = y\}$ (resp. $\tau_y^P := \inf\{t \geq 0; Y_t = y\}$), then*

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

As a second application of the spectral decomposition stated in Theorem 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 [25], Levin et al. [41], Montenegro and Tetali [46] 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 $\text{Var}_\pi(g) = \langle g, g \rangle_\pi - \mathbb{E}_\pi^2(g)$. Finally, we recall that Fill in Fill [25, Theorem 2.1] obtained in the finite state space case the following bound valids for all $n \in \mathbb{N}_0$

$$(2.2) \quad d(n) \leq \frac{\sigma_*(P)^n}{2} \sqrt{\frac{1 - \pi_{\min}}{\pi_{\min}}},$$

where $\pi_{\min} = \min_{x \in \mathcal{X}} \pi(x)$ and $\sigma_*(P) = \sqrt{\lambda_*(P\hat{P})}$ 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 2.3.

Corollary 2.2. *Let $P \in \mathcal{S}$ on the finite state space $\mathcal{X} = \llbracket 0, \mathfrak{r} \rrbracket$ with $\mathfrak{r} < \infty$ and invariant distribution π , that is, $\pi P = \pi$.*

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

$$(2.3) \quad \lambda_*^n \leq \|P^n - \pi\|_{\ell^2(\pi) \rightarrow \ell^2(\pi)} \leq \sigma_*^n(P) \mathbf{1}_{\{n < n^*\}} + \kappa(\Lambda) \lambda_*^n \mathbf{1}_{\{n \geq n^*\}},$$

where $n^* = \lceil \frac{\ln \kappa(\Lambda)}{\ln \sigma_*(P) - \ln \lambda_*} \rceil$ and $\kappa(\Lambda) = \|\Lambda\|_{op} \|\Lambda^{-1}\|_{op} \geq 1$ 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 (2.3) is strictly better than the reversibilization rate $\sigma_*(P)$.

(2) For any $n \in \mathbb{N}_0$,

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

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

Remark 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 [10, Section 4.3]) $\|P^n - \pi\|_{\ell^2(\pi) \rightarrow \ell^2(\pi)} = \lambda_*^n$ and $\kappa(\Lambda) = 1$ which is a specific instance of item (1).

Remark 2.4. We also recall the discrete analogue of the notion of hypocoercivity introduced in Villani [55], i.e. there exists a constant $C < \infty$ and $\rho \in (0, 1)$ such that, for all $n \in \mathbb{N}$,

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

Note that, in general, these constants are not known explicitly. We observe that the upper bound in (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 [47] where a similar spectral interpretation of the hypocoercivity phenomena is given for a class of non-self-adjoint Markov semigroups, 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.

2.1. Proof of Theorem 2.1. We first show the item (a). Since \mathcal{E} is a spectral measure, it follows easily that $\{F_B = \Lambda E_B \Lambda^{-1}; B \in \mathcal{B}(\mathbb{C})\}$ 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 [30]. 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 [23]. 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

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

where we can take $A = \|\Lambda^{-1}\|^{-2}$ and $B = \|\Lambda\|^2$, so (2.1) is satisfied. It follows from Young [57, 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 . 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 $\widehat{Q} \overset{\Lambda}{\sim} \widehat{P}$. 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 [24, Section 5.8], one can always construct an ergodic Markov chain with transition matrix Q such that

$$Q = V D V^{-1}.$$

2.2. Proof of Theorem 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 $H_S(x, y) = \mathbf{1}_{\{x \leq y\}}$ and its inverse $H_S^{-1} = (H_S^{-1}(x, y))_{x, y \in \mathcal{X}}$ is $H_S^{-1}(x, y) = \mathbf{1}_{\{x=y\}} - \mathbf{1}_{\{x=y-1\}}$, see Siegmund [53]. 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 $x \in \llbracket 0, \mathfrak{r} - 2 \rrbracket$, condition 2 and 3 in \mathcal{GMC} yield, respectively, $\widetilde{p}(x, x+1) > 0$, while for $x \in \llbracket 1, \mathfrak{r} - 1 \rrbracket$, we have $\widetilde{p}(x, x-1) > 0$. Condition 4 and 5 in \mathcal{GMC} guarantee that $\widetilde{p}(x, y) = 0$ for each $x \in \llbracket 0, \mathfrak{r} - 3 \rrbracket$ and $y \in \llbracket x+2, \mathfrak{r} - 1 \rrbracket$ and for each $x \in \llbracket 2, \mathfrak{r} - 1 \rrbracket$ and $y \in \llbracket 0, x-2 \rrbracket$. That is, \widetilde{P} is a (strictly substochastic) irreducible birth-death chain when restricted to the state space $\llbracket 0, \mathfrak{r} - 1 \rrbracket$. Denote \widetilde{P}^{bd} the restriction of \widetilde{P} to $\llbracket 0, \mathfrak{r} - 1 \rrbracket$. By breaking off the last row and last column of \widetilde{P} , we can write

$$(2.4) \quad \widetilde{P} = \begin{pmatrix} \widetilde{P}^{bd} & \mathbf{v} \\ \mathbf{0} & 1 \end{pmatrix} = (H_S^{-1} \widehat{P} H_S)^T,$$

where $\mathbf{0}$ is a row vector of zero, and \mathbf{v} is a column vector storing $\widetilde{p}(x, \mathfrak{r})$ for $x \in \llbracket 0, \mathfrak{r} - 1 \rrbracket$. Considering the h -transform of \widetilde{P} with $h = H_S^T \pi > \mathbf{0}$, see e.g. Huillet and Martinez [29, 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 $Q = \tilde{P}$, which completes the proof of Theorem 2.2.

Remark 2.5. Note that condition (6) in \mathcal{GMC}^+ guarantees that Q is a lazy chain, that is $Q(x, x) \geq 1/2$ for all $x \in \mathcal{X}$, and hence the class \mathcal{GMC}^+ possesses non-negative eigenvalues.

2.3. Proof of Corollary 2.2. We first show the upper bound in item (1). Define the synthesis operator $T^* : \ell^2 \rightarrow \ell^2(\pi)$ 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 2.1. For $1 \leq i \leq \mathfrak{r}$, 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 $\|T^*\|_{op} \leq \|\Lambda\|_{op} < \infty$, since

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

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

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

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

$$(2.5) \quad \|P^n g - \pi g\|_\pi^2 = \|T^*(\alpha)\|_\pi^2 \leq \|\Lambda\|_{op}^2 \|\alpha\|_{\ell^2}^2 \leq \|\Lambda\|_{op}^2 \|\Lambda^{-1}\|_{op}^2 \lambda_*^{2n} \|g\|_\pi^2.$$

Desired upper bound follows from (2.5) and

$$\|P^n - \pi\|_{\ell^2(\pi) \rightarrow \ell^2(\pi)} \leq \lambda_*(\hat{P}P)^{n/2} = \lambda_*(P\hat{P})^{n/2},$$

see e.g. Fill [25]. The lower bound in (1) follows readily from the well-known result that the n^{th} power of the spectral radius λ_* 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 [54]. Next, using (2.5), we get

$$(2.6) \quad \text{Var}_\pi \left(\hat{P}^n g \right) \leq \kappa(\hat{\Lambda})^2 \lambda_*^{2n} \text{Var}_\pi(g) = \kappa(\Lambda)^2 \lambda_*^{2n} \text{Var}_\pi(g), \quad n \in \mathbb{N}_0,$$

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

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

where the first inequality follows from Cauchy-Schwartz inequality. The proof is completed by combining the above bound with (2.2).

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 [20] and Chen and Saloff-Coste [11] while it has recently been extended to upward skip-free chains by Mao et al. [42] and Choi and Patie [14]. Recall that in Theorem 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 [26] to first analyze the fastest strong stationary time of this class, followed by demonstrating that the proof in Diaconis and Saloff-Coste [20] carries over for this class of non-reversible chains.

We now proceed to discuss the main results of this Section, with Theorem 3.1 addressing the case of discrete time family of Markov chains and Theorem 3.2 discussing the continuized version. Recall that the notation \mathcal{GMC}^+ introduced in Definition 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 3.1 (resp. Theorem 3.2).

Theorem 3.1. *For $n \geq 1$, suppose that $P_n \in \mathcal{GMC}_{\tau_n}^+$ on the state space $\mathcal{X}_n = \llbracket 0, \tau_n \rrbracket$ that started at 0. Let $(\theta_{n,i})_{i=1}^{\tau_n}$ be the non-zero eigenvalues of $I - P_n$, and $(c_{n,i})_{i=0}^{\tau_n}$ to be the mixture weights of the n^{th} chain defined in (3.1) in Lemma 3.1. Define*

$$w_{n,i} := \sum_{j \geq i} c_{n,j}, \quad t_n := \sum_{i=1}^{\tau_n} \frac{w_{n,i}}{\theta_{n,i}}, \quad \underline{\theta}_n := \min_{1 \leq i \leq \tau_n} \theta_{n,i}, \quad \rho_n^2 := \sum_{i=1}^{\tau_n} w_{n,i}^2 \frac{1 - \theta_{n,i}}{\theta_{n,i}^2}.$$

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

Remark 3.1. For discrete-time stochastically monotone birth-death chains which start at 0, we have $w_i = 1$ for $i \in \llbracket 1, \tau_n \rrbracket$ and $c_{n,0} = 0$, and hence we recover Diaconis and Saloff-Coste [20, Theorem 5.2].

Theorem 3.2. *For $n \geq 1$, suppose that $L_n = P_n - I$ is the infinitesimal generator with $P \in \mathcal{GMC}_{\tau_n}^+$ on the state space $\mathcal{X}_n = \llbracket 0, \tau_n \rrbracket$ that started at 0. Let $(\theta_{n,i})_{i=1}^{\tau_n}$ be the non-zero eigenvalues of $-L_n$, and $(c_{n,i})_{i=0}^{\tau_n}$ to be the mixture weights defined in (3.2) in Remark 3.2. Define*

$$w_{n,i} := \sum_{j \geq i} c_{n,j}, \quad t_n := \sum_{i=1}^{\tau_n} \frac{w_{n,i}}{\theta_{n,i}}, \quad \underline{\theta}_n := \min_{1 \leq i \leq \tau_n} \theta_{n,i}, \quad \rho_n^2 := \sum_{i=1}^{\tau_n} \frac{w_{n,i}^2}{\theta_{n,i}^2}.$$

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

We will only prove Theorem 3.1 as the proof of Theorem 3.2 is very similar and thus omitted.

3.1. Proof of Theorem 3.1. Following the plan as outlined above in Section 3, we first analyze the distribution of the fastest strong stationary time of the class \mathcal{GMC}^+ in Lemma 3.1, followed by detailing the proof of Theorem 3.1.

Lemma 3.1. *Suppose that X is an ergodic Markov chain on the state space $\mathcal{X} = \llbracket 0, \tau \rrbracket$ (and $\tau \geq 3$) 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 c -mixture of convolution of geometric $\sum_{k=1}^{\tau} c_k \mathcal{G}(\lambda_1, \dots, \lambda_k)$, where $i, j, k \in \llbracket 0, \tau \rrbracket$,*

$$(3.1) \quad Q_k := \frac{(P - \lambda_1 I) \dots (P - \lambda_k I)}{(1 - \lambda_1) \dots (1 - \lambda_k)}, \quad \Lambda(i, j) := Q_i(0, j), \quad c_k := \frac{\Lambda(k, \tau) - \Lambda(k-1, \tau)}{\pi(\tau)},$$

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

Proof. Suppose that $P\Lambda = \Lambda Q$. In view of Fill [26] Theorem 5.2, it suffices to show that the $c_k \geq 0$. First, we show that $(Q - \lambda_1 I) \dots (Q - \lambda_k I)$ are non-negative matrices, where Q is the Siegmund dual of \hat{P} . We will prove this via induction on k . For $k = 1$, thanks to Micchelli and Willoughby [43, Theorem 3.2], we have $Q^{BD} - \lambda_1 I \geq 0$, which leads to

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

Suppose that

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

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

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

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 $c_k \geq 0$ if and only if $Z_k(0, \mathbf{r}) - Z_{k-1}(0, \mathbf{r}) \geq 0$ if and only if (here we make use of H_S^T)

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

which is true. \square

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 [20]. 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 $\text{Var}(T_n) = \rho_n^2$. The key to the proof is the following:

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

where we use $\theta_{n,i} \geq 0$, $\underline{\theta}_n / \theta_{n,i} \leq 1$ and $w_i \leq 1$ in the first inequality. The rest of the proof follows as that of Choi and Patie [14, Theorem 8.1], which does not require reversibility of the chain.

Remark 3.2. The corresponding result of Lemma 3.1 in the continuous-time setting is stated in the following in order to prove Theorem 3.2. Suppose that X is a continuous-time ergodic Markov chain on the state space $\mathcal{X} = \llbracket 0, \tau \rrbracket$ (and $\tau \geq 3$) 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 $\sum_{k=1}^{\tau} c_k \mathcal{E}(\theta_1, \dots, \theta_k)$, where $i, j, k \in \llbracket 0, \tau \rrbracket$,

$$(3.2) \quad Q_k := \frac{(L + \theta_1 I) \dots (L + \theta_k I)}{\theta_1 \dots \theta_k}, \quad \Lambda(i, j) := Q_i(0, j), \quad c_k := \frac{\Lambda(k, \tau) - \Lambda(k-1, \tau)}{\pi(\tau)},$$

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

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 1.1 and 1.2, $L \in \mathcal{S}(G)$, where G is a reversible generator. We denote the spectral gap $\lambda = \lambda(L)$ of L by

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

This follows and generalizes the work of Chen and Saloff-Coste [9, 10], Chen et al. [12] 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 4.1. For $n \geq 1$, let $g_n : [0, \infty) \mapsto [0, \infty]$ be a non-increasing function vanishing at infinity. Assume that

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

Then the family $\mathcal{G} = \{g_n : n \geq 1\}$ 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 $\epsilon \in (0, 1)$,

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

- (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

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

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 $t \geq 0$. 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 $\{g_n(t) = D_{n,2}(\eta_n, t) : n \geq 1\}$ has a cutoff (resp. (t_n, b_n) -cutoff) as in Definition 4.1, where $D_{n,2}(\eta_n, t)$ is the chi-squared distance of the n^{th} chain.

Our main result in Theorem 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

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

Theorem 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 $\pi_n(f_n^2) \rightarrow \infty$, 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 ,

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

- (3) For some positive constants C, ϵ ,

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

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).

4.1. Proof of Theorem 4.1. To prove Theorem 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 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 $t \geq 0$,

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

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 [30, Lemma 3.19]. \square

Lemma 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 4.1. By Lemma 4.1, we take $g_n(t) = D_{n,2}(\eta_n, t)$ in Definition 4.1, and the desired result follows from the Laplace transform cutoff criteria in Theorem 3.5 of Chen and Saloff-Coste [10].

5. NON-ASYMPTOTIC ESTIMATION ERROR BOUNDS FOR INTEGRAL FUNCTIONALS

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

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

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. [13]. 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 $k \in \llbracket n \rrbracket$ 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 operator and F_λ is the spectral measure of $-L$. For $s \geq 0$, we define the space $\mathcal{D}^s(A) = \text{Dom}(A^s) \subset L^2(\pi)$ by functional calculus on A with the seminorm $\|f\|_{\mathcal{D}^s(A)} = \|A^{s/2} f\|_\pi$.

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 5.1. Note that (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 (5.3) offers the error bound for the non-stationary Markov process such that $X_0 \sim \eta$.

Theorem 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 $T \geq 0$, $0 \leq s \leq 1$, $f \in \mathcal{D}^s(A)$, $f_0 \in \text{Dom}(A^{-1})$ with $f_0 = f - \int f d\pi$,*

$$(5.1) \quad \left\| \Gamma_T(f) - \hat{\Gamma}_{T,n}(f) \right\|_{L^2(\mathbb{P})} \leq C \sqrt{\|f\|_{\mathcal{D}^s(A)} \|f\|_\pi T \Delta_n^{1+s}},$$

$$(5.2) \quad \left\| T^{-1} \hat{\Gamma}_{T,n}(f) - \int f d\pi \right\|_{L^2(\mathbb{P})} \leq \frac{C}{\sqrt{T}} \left(\sqrt{\|f\|_{\mathcal{D}^s(A)} \|f\|_\pi \Delta_n} + \sqrt{\|A^{-1} f_0\|_\pi \|f_0\|_\pi} \right).$$

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 $T \geq 0$, $0 \leq s \leq 1$ and $f \in \mathcal{D}^s(A)$,

$$(5.3) \quad \left\| \Gamma_T(f) - \hat{\Gamma}_{T,n}(f) \right\|_{L^2(\mathbb{P})} \leq C \left\| \frac{d\eta}{d\pi} \right\|_{\infty, \pi}^{1/2} \sqrt{\|f\|_{\mathcal{D}^s(A)} \|f\|_\pi T \Delta_n^{1+s}},$$

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

5.1. Proof of Theorem 5.1. We first state a lemma (see Inoue and Trapani [30, Lemma 3.19]) which will be used repeatedly in the proof.

Lemma 5.1. *For $f \in \mathcal{D}^s(A)$,*

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

We now proceed to give the proof of Theorem 5.1. We first prove (5.1) and consider

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

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 $(k-1)\Delta_n \leq r \leq h \leq k\Delta_n$,

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

so by symmetry in r and h we have

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

where the last inequality follows from the functional calculus of L in Theorem 2.1 and for $\lambda \in \sigma(L)$,

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

From Altmeyer and Chorowski [2, Page 15], we know that $|\Phi(\lambda)| \leq 4n\Delta_n^{2+s}|\lambda|^s$ with fixed $0 \leq s \leq 1$. Now, we apply Lemma 5.1 to arrive at

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

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

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

where the last inequality follows again from the functional calculus of L in Theorem 2.1 and for $\lambda \in \sigma(L)$,

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

Using Altmeyer and Chorowski [2, (16)] there exists a universal constant $\tilde{C} < \infty$ such that $|\tilde{\Phi}(\lambda)| \leq \tilde{C}T\Delta_n^{1+s}|\lambda|^s$, and together with Lemma 5.1 yield

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

Next, we prove (5.2). By (5.1) and triangle inequality,

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

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

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

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

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

and there exists a constant \tilde{C} such that $|\bar{\Phi}(\lambda)| \leq \frac{\tilde{C}}{|\lambda|T}$. Using Lemma 5.1 gives

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

Finally, it follows from a standard change of measure argument to give (5.3).

6. SIMILARITY ORBIT OF REVERSIBLE MARKOV CHAINS

In this Section, our aim is to provide several illuminating examples for Theorem 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 $(Q_t)_{t \geq 0}$, 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 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 6.1.1 and 6.3.1)
- (2) $M/M/\infty$ queue (Section 6.2.1 and 6.4.1)
- (3) linear birth-death process (Section 6.2.2 and 6.4.2)

(4) quadratic birth-death process (Section 6.1.2 and 6.3.2)

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 [28], Khare and Mukherjee [34], Khare and Zhou [35] and Zhou [58]) 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 [45] in the notion of Markov similarity. We write $\sigma : \mathcal{X} \rightarrow \mathcal{X}$ 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 2.1.

Proposition 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 respect 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 $(Q_t)_{t \geq 0}$ (resp. $(P_t)_{t \geq 0}$) being the transition semigroup associated with G (resp. L) of the Markov chain $(X_t)_{t \geq 0}$ (resp. $(Y_t)_{t \geq 0}$). For $t \geq 0$ 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) $(P_t)_{t \geq 0}$ and $(Q_t)_{t \geq 0}$ shares the same eigentime identity. That is, denote $\tau_y^Q := \inf\{t \geq 0; X_t = y\}$ (resp. $\tau_y^P := \inf\{t \geq 0; Y_t = y\}$), 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 6.1. Note that item (2) offers us a way to generate reversible processes $(P_t)_{t \geq 0}$ from a birth-death process $(Q_t)_{t \geq 0}$ 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 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. \square

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 $\mathcal{X} = \llbracket 0, N \rrbracket$, 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. [21], Karlin and McGregor [32], Koekoek and Swarttouw [36], Sasaki [51], Schoutens [52], Zhou [58] and the references therein.

6.1.1. 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 $x \in \llbracket 0, N \rrbracket$, and the associated orthogonal polynomials are the Krawtchouk polynomials. Under the permutation link Λ_σ , Proposition 6.1 (2) and (3) now reads, for $j, x, y \in \llbracket 0, N \rrbracket$ and $t \geq 0$,

$$\begin{aligned}\pi_x &= \binom{N}{x} p^x (1 - p)^{N-x}, \\ \phi_j(x) &= {}_2F_1 \left(\begin{matrix} -j, -x \\ -N \end{matrix} \middle| p^{-1} \right), \\ Q_t(x, y) &= \pi_y \sum_{j=0}^N e^{-jt} \phi_j(x) \phi_j(y) \frac{(-1)^{-j} p^j}{j! (1 - p)^j} (-N)_j, \\ P_t(x, y) &= \pi_{\sigma^{-1}(y)} \sum_{j=0}^N e^{-jt} \phi_j(\sigma^{-1}(x)) \phi_j(\sigma^{-1}(y)) \frac{(-1)^{-j} p^j}{j! (1 - p)^j} (-N)_j, \\ \sum_{x, y \in \mathcal{X}} \mathbb{E}_x(\tau_y^P) \pi_{\sigma^{-1}(x)} \pi_{\sigma^{-1}(y)} &= \sum_{i=1}^N \frac{1}{i}.\end{aligned}$$

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.

6.1.2. 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 $a, b \geq N$. 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 6.1 (2) and (3) now reads, for $j, x, y \in \llbracket 0, N \rrbracket$ and $t \geq 0$,

$$\begin{aligned}\phi_j(x) &= {}_3F_2 \left(\begin{matrix} -j, -x, -x - a - b - 1 \\ -a, -N \end{matrix} \middle| 1 \right), \\ w_j &:= \frac{\binom{N-b-1}{N} N! (-N)_j (-a)_j (2j - a - b - 1)}{(-1)^j j! (-b)_j (j - a - b - 1)_{N+1}},\end{aligned}$$

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

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.

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.

6.2.1. $M/M/\infty$ 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 $t \geq 0$,

$$\begin{aligned}
\phi_j(x) &= {}_2F_0 \left(\begin{matrix} -j, -x \\ - \end{matrix} \middle| -\mu^{-1} \right), \\
Q_t(x, y) &= e^{-\lambda/\mu} \frac{(\lambda/\mu)^y}{y!} \sum_{j=0}^{\infty} e^{-\mu j t} \phi_j(x) \phi_j(y) \frac{(\lambda/\mu)^j}{j!}, \\
P_t(x, y) &= e^{-\lambda/\mu} \frac{(\lambda/\mu)^{\sigma^{-1}(y)}}{\sigma^{-1}(y)!} \sum_{j=0}^{\infty} e^{-\mu j t} \phi_j(\sigma^{-1}(x)) \phi_j(\sigma^{-1}(y)) \frac{(\lambda/\mu)^j}{j!}.
\end{aligned}$$

6.2.2. 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 $t \geq 0$,

$$\begin{aligned}
\phi_j(x) &= {}_2F_1 \left(\begin{matrix} -j, -\frac{x}{\mu - \lambda} \\ \beta \end{matrix} \middle| \frac{\lambda - \mu}{\lambda} \right), \\
Q_t(x, y) &= \pi_y \sum_{j=0}^{\infty} e^{-(\mu - \lambda)jt} \phi_j(x) \phi_j(y) \frac{(\beta)_j}{j!} \left(\frac{\lambda}{\mu}\right)^j,
\end{aligned}$$

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

6.3. Pure birth link on finite state space. In this Section, we specialize into the case of $\mathcal{X} = \llbracket 0, N \rrbracket$, with the link being the pure birth link as introduced by Fill [26] 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 $x \in \llbracket 0, N-1 \rrbracket$, $y \in \{x, x+1\}$, $\Lambda_{pb}(N, N) = 1$ 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 $x \leq y$, $x, y \in \llbracket 0, N \rrbracket$ 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 6.2 below. We now give a structural result in this direction, which follows easily from Theorem 2.1.

Proposition 6.2. *Suppose that $G \stackrel{\Lambda_{pb}}{\sim} L$ with G being a reversible generator with respect to π_G on $\mathcal{X} = \llbracket 0, N \rrbracket$, and eigenvalues-eigenvectors denoted by $(-\lambda_j, \phi_j)_{j=0}^N$, where ϕ_j are orthonormal in $l^2(\pi_G)$. Write $(Q_t)_{t \geq 0}$ (resp. $(P_t)_{t \geq 0}$) being the transition semigroup associated with G (resp. L). Note that $(P_t)_{t \geq 0}$ need not be Markov under Λ_{pb} . For $t \geq 0$ and $j, x, y \in \llbracket 0, N \rrbracket$, the spectral decompositions are given by*

$$\begin{aligned} Q_t(x, y) &= \sum_{j=0}^N e^{-\lambda_j t} \phi_j(x) \phi_j(y) \pi_G(y), \\ P_t(x, y) &= \sum_{j=0}^N e^{-\lambda_j t} f_j(x) f_j^*(y), \\ \|P_t - \pi_L\|_{TV} &\leq \frac{\kappa(\Lambda_{pb}) e^{-\lambda_1 t}}{2} \sqrt{\frac{1 - \pi_L^*}{\pi_L^*}}, \quad \text{where} \\ f_j(x) &:= \sum_{k=x}^N (-1)^{k-x} (2\mathbb{1}_{k \neq N} + \mathbb{1}_{k=N}) \phi_j(k), \\ f_j^*(y) &:= \phi_j(y-1) \pi_G(y-1) \left(\frac{\mathbb{1}_{y-1 \geq 0}}{2} \right) + \phi_j(y) \pi_G(y) \left(\frac{\mathbb{1}_{y \neq N}}{2} + \mathbb{1}_{y=N} \right), \\ \pi_L(x) &= \pi_G(x-1) \left(\frac{\mathbb{1}_{x-1 \geq 0}}{2} \right) + \pi_G(x) \left(\frac{\mathbb{1}_{x \neq N}}{2} + \mathbb{1}_{x=N} \right), \\ \pi_L^* &:= \min_{x \in \llbracket 0, N \rrbracket} \pi_L(x). \end{aligned}$$

Remark 6.2. If $(P_t)_{t \geq 0}$ is a Markov semigroup, then as stated in Corollary 2.1, using the result of Cui and Mao [17], 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 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

$$\begin{aligned} P_t(x, y) &= \sum_{k=x}^N (-1)^{k-x} (2\mathbb{1}_{k \neq N} + \mathbb{1}_{k=N}) \left(Q_t(k, y-1) \left(\frac{\mathbb{1}_{y-1 \geq 0}}{2} \right) + Q_t(k, y) \left(\frac{\mathbb{1}_{y \neq N}}{2} + \mathbb{1}_{y=N} \right) \right) \\ &= \sum_{j=0}^N e^{-\lambda_j t} \left(\sum_{k=x}^N (-1)^{k-x} (2\mathbb{1}_{k \neq N} + \mathbb{1}_{k=N}) \phi_j(k) \right) \\ &\quad \times \left(\phi_j(y-1) \pi_G(y-1) \left(\frac{\mathbb{1}_{y-1 \geq 0}}{2} \right) + \phi_j(y) \pi_G(y) \left(\frac{\mathbb{1}_{y \neq N}}{2} + \mathbb{1}_{y=N} \right) \right), \end{aligned}$$

where the second equality follows from substituting the spectral expansion of $(Q_t)_{t \geq 0}$. $\|P_t - \pi_L\|_{TV} \leq \frac{\kappa(\Lambda_{pb})e^{-\lambda_1 t}}{2} \sqrt{\frac{1 - \pi_L^*}{\pi_L^*}}$ follows directly from Corollary 2.2. \square

To illustrate this proposition, we again look at the Ehrenfest model and the quadratic birth-death process.

6.3.1. The Ehrenfest model and its pure birth variant. Recall that we introduce the Ehrenfest model in Section 6.1.1. Under the pure birth link Λ_{pb} , Proposition 6.2 now reads, for $j, x, y \in \llbracket 0, N \rrbracket$ and $t \geq 0$,

$$\begin{aligned} P_t(x, y) &= \sum_{j=0}^N e^{-jt} f_j(x) f_j^*(y) \frac{(-1)^{-j} p^j}{j!(1-p)^j} (-N)_j, \\ \|P_t - \pi_L\|_{TV} &\leq \frac{\kappa(\Lambda_{pb})e^{-t}}{2} \sqrt{\frac{1 - \pi_L^*}{\pi_L^*}} = O(e^{-t}), \quad \text{where} \\ \pi_G(x) &= \binom{N}{x} p^x (1-p)^{N-x}, \\ f_j(x) &= \sum_{k=x}^N (-1)^{k-x} (2\mathbb{1}_{k \neq N} + \mathbb{1}_{k=N}) {}_2F_1 \left(\begin{matrix} -j, -k \\ -N \end{matrix} \middle| p^{-1} \right), \\ f_j^*(y) &= {}_2F_1 \left(\begin{matrix} -j, -(y-1) \\ -N \end{matrix} \middle| p^{-1} \right) \pi_G(y-1) \left(\frac{\mathbb{1}_{y-1 \geq 0}}{2} \right) \\ &\quad + {}_2F_1 \left(\begin{matrix} -j, -y \\ -N \end{matrix} \middle| p^{-1} \right) \pi_G(y) \left(\frac{\mathbb{1}_{y \neq N}}{2} + \mathbb{1}_{y=N} \right), \\ \pi_L(x) &= \pi_G(x-1) \left(\frac{\mathbb{1}_{x-1 \geq 0}}{2} \right) + \pi_G(x) \left(\frac{\mathbb{1}_{x \neq N}}{2} + \mathbb{1}_{x=N} \right), \\ \pi_L^* &= \min_x \pi_L(x). \end{aligned}$$

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.

6.3.2. Quadratic birth-death process and its pure birth variant. In the second example, we study the quadratic birth-death process as introduced in Section 6.1.2. In the pure birth link Λ_{pb} , Proposition 6.2 now reads, for $j, x, y \in \llbracket 0, N \rrbracket$ and $t \geq 0$,

$$\begin{aligned}
P_t(x, y) &= \sum_{j=0}^N e^{-j(a+b+1-j)t} f_j(x) f_j^*(y) w_j, \\
\|P_t - \pi_L\|_{TV} &\leq \frac{\kappa(\Lambda_{pb}) e^{-(a+b)t}}{2} \sqrt{\binom{a+b}{N} - 1} = O(e^{-(a+b)t}), \quad \text{where} \\
\pi_G(x) &= \frac{\binom{a}{x} \binom{b}{N-x}}{\binom{a+b}{N}}, \\
w_j &= \frac{\binom{N-b-1}{N} N! (-N)_j (-a)_j (2j - a - b - 1)}{(-1)^j j! (-b)_j (j - a - b - 1)_{N+1}}, \\
f_j(x) &= \sum_{k=x}^N (-1)^{k-x} (2\mathbb{1}_{k \neq N} + \mathbb{1}_{k=N}) {}_3F_2 \left(\begin{matrix} -j, -k, -k - a - b - 1 \\ -a, -N \end{matrix} \middle| 1 \right), \\
f_j^*(y) &= {}_3F_2 \left(\begin{matrix} -j, -(y-1), -(y-1) - a - b - 1 \\ -a, -N \end{matrix} \middle| 1 \right) \pi_G(y-1) \left(\frac{\mathbb{1}_{y-1 \geq 0}}{2} \right) \\
&\quad + {}_3F_2 \left(\begin{matrix} -j, -y, -y - a - b - 1 \\ -a, -N \end{matrix} \middle| 1 \right) \pi_G(y) \left(\frac{\mathbb{1}_{y \neq N}}{2} + \mathbb{1}_{y=N} \right), \\
\pi_L(x) &= \pi_G(x-1) \left(\frac{\mathbb{1}_{x-1 \geq 0}}{2} \right) + \pi_G(x) \left(\frac{\mathbb{1}_{x \neq N}}{2} + \mathbb{1}_{x=N} \right).
\end{aligned}$$

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} .

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 $\llbracket 0, n \rrbracket$) on the state space $\mathcal{X} = \mathbb{N}_0$ generated by the $M/M/\infty$ model and the linear birth-death process.

6.4.1. $M/M/\infty$ and its $(n+1)$ -dimensional pure birth variant. The $M/M/\infty$ queueing model is first introduced in Section 6.2.1. 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 $t \geq 0$,

$$\begin{aligned}
P_t(x, y) &= \sum_{j=0}^{\infty} e^{-\mu j t} f_j(x) f_j^*(y) \frac{(\lambda/\mu)^j}{j!}, \quad \text{where} \\
\phi_j(x) &= {}_2F_0 \left(\begin{matrix} -j, -x \\ - \end{matrix} \middle| -\mu^{-1} \right), \\
\pi_G(x) &= e^{-\lambda/\mu} \frac{(\lambda/\mu)^x}{x!}, \\
f_j(x) &= \begin{cases} \sum_{k=x}^n (-1)^{k-x} (2\mathbb{1}_{k \neq n} + \mathbb{1}_{k=n}) \phi_j(k), & x \in \llbracket 0, n \rrbracket, \\ \phi_j(x), & x \geq n+1, \end{cases}
\end{aligned}$$

$$f_j^*(y) = \begin{cases} \phi_j(y-1)\pi_G(y-1) \left(\frac{\mathbb{1}_{y-1 \geq 0}}{2} \right) + \phi_j(y)\pi_G(y) \left(\frac{\mathbb{1}_{y \neq n}}{2} + \mathbb{1}_{y=n} \right), & y \in \llbracket 0, n \rrbracket, \\ \phi_j(y)\pi_G(y), & y \geq n+1. \end{cases}$$

6.4.2. Linear birth-death process and its $(n+1)$ -dimensional pure birth variant. The linear birth-death process is introduced in Section 6.2.2. With the $(n+1)$ -dimensional pure birth link Λ_{pb} , we have for $j, x, y \in \mathbb{N}_0$ and $t \geq 0$,

$$P_t(x, y) = \sum_{j=0}^{\infty} e^{-(\mu-\lambda)jt} f_j(x) f_j^*(y) \frac{(\beta)_j}{j!} \left(\frac{\lambda}{\mu} \right)^j, \quad \text{where}$$

$$\phi_j(x) = {}_2F_1 \left(-j, -\frac{x}{\mu-\lambda} \middle| \frac{\lambda-\mu}{\lambda} \right),$$

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

$$f_j(x) = \begin{cases} \sum_{k=x}^n (-1)^{k-x} (2\mathbb{1}_{k \neq n} + \mathbb{1}_{k=n}) \phi_j(k), & x \in \llbracket 0, n \rrbracket, \\ \phi_j(x), & x \geq n+1, \end{cases}$$

$$f_j^*(y) = \begin{cases} \phi_j(y-1)\pi_G(y-1) \left(\frac{\mathbb{1}_{y-1 \geq 0}}{2} \right) + \phi_j(y)\pi_G(y) \left(\frac{\mathbb{1}_{y \neq n}}{2} + \mathbb{1}_{y=n} \right), & y \in \llbracket 0, n \rrbracket, \\ \phi_j(y)\pi_G(y), & y \geq n+1. \end{cases}$$

6.5. Random walk link on finite state space. In this Section, we specialize into the case of $\mathcal{X} = \llbracket 0, N \rrbracket$, with the link being the random walk kernel previously studied by Diaconis and Miclo [19] and Zhou [58]. 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 $x \in \llbracket 1, N-1 \rrbracket, y = x \pm 1$, $\Lambda_{rw}(N, N) = 1$ and zero otherwise. That is, it is a simple random walk with holding at 0 and absorbing endpoint N . For $j = 1, 3, \dots, 2N-1$, the eigenvalue β_j , right eigenfunction ψ_j and left eigenfunction Ψ_j are given by

$$(6.1) \quad \beta_j := \cos \left(\frac{j\pi}{2N+1} \right),$$

$$(6.2) \quad \psi_j(x) := \cos \left(\frac{(2x+1)j\pi}{2(2N+1)} \right), \quad x \in \llbracket 0, N \rrbracket,$$

$$(6.3) \quad \Psi_j(x) := \begin{cases} \psi_j(x), & \text{for } x \in \llbracket 0, N-1 \rrbracket, \\ \frac{(-1)^{(j+1)/2}}{2} \cot \left(\frac{j\pi}{2(2N+1)} \right), & \text{for } x = N, \end{cases}$$

$$(6.4) \quad \Lambda_{rw} = \sum_{j \in \{0, 1, 3, \dots, 2N-1\}} \beta_j \psi_j \Psi_j^T,$$

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 6.3. Suppose that $G \stackrel{\Lambda_{rw}}{\sim} L$ with G being a reversible generator with respect to π_G on $\mathcal{X} = \llbracket 0, N \rrbracket$, and eigenvalues-eigenvectors denoted by $(-\lambda_j, \phi_j)_{j=0}^N$, where ϕ_j are orthonormal in

$l^2(\pi_G)$. Write $(Q_t)_{t \geq 0}$ (resp. $(P_t)_{t \geq 0}$) being the transition semigroup associated with G (resp. L). Note that $(P_t)_{t \geq 0}$ need not be Markov under Λ_{rw} . For $t \geq 0$, $j, x, y \in \llbracket 0, N \rrbracket$ and recall that β_j , ψ_j and Ψ_j are defined in (6.1), (6.2) and (6.3) respectively, the spectral decompositions are given by

$$\begin{aligned}
Q_t(x, y) &= \sum_{j=0}^N e^{-\lambda_j t} \phi_j(x) \phi_j(y) \pi_G(y), \\
P_t(x, y) &= \sum_{j=0}^N e^{-\lambda_j t} f_j(x) f_j^*(y), \quad \text{where} \\
f_j(x) &:= \Lambda_{rw}^{-1} \phi_j(x) = \sum_{k \in \{0, 1, 3, \dots, 2N-1\}} \beta_k^{-1} \langle \Psi_k, \phi_j \rangle \psi_k(x) \\
&= \sum_{k \in \{0, 1, 3, \dots, 2N-1\}} \beta_k^{-1} \langle \Psi_k, \phi_j \rangle \cos \left(\frac{(2x+1)k\pi}{2(2N+1)} \right), \\
f_j^*(y) &= \sum_{k \in \{0, 1, 3, \dots, 2N-1\}} \beta_k \langle \psi_k, \phi_j \rangle_{\pi_G} \Psi_k(y), \\
\pi_L(x) &= \tilde{\Phi}_0(x) = \sum_{k \in \{0, 1, 3, \dots, 2N-1\}} \beta_k \langle \psi_k, \mathbf{1} \rangle_{\pi_G} \Psi_k(x).
\end{aligned}$$

Remark 6.4. For $j \in \llbracket 0, N \rrbracket$, note that f_j is the discrete cosine transform of type VI (see Britanak et al. [7]) of the points $(\beta_k^{-1} \langle \Psi_k, \phi_j \rangle)_{k \in \{0, 1, 3, \dots, 2N-1\}}$.

REFERENCES

- [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>. 2, 3, 7, 8, 19
- [2] R. Altmeyer and J. Chorowski. Estimation error for occupation time functionals of stationary Markov processes. *arXiv:1610.05225*, 2016. 2, 4, 15, 16, 17, 18
- [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. 10
- [4] P. H. Baxendale. Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Probab.*, 15(1B):700–738, 2005. 9
- [5] A. Berman and R. J. Plemmons. Matrix group monotonicity. *Proc. Amer. Math. Soc.*, 46:355–359, 1974. 10
- [6] J. Bierkens. Non-reversible Metropolis-Hastings. *Stat. Comput.*, 26(6):1213–1228, 2016. 2
- [7] V. Britanak, P. C. Yip, and K. R. Rao. *Discrete cosine and sine transforms*. Elsevier/Academic Press, Amsterdam, 2007. 26
- [8] D. Chafaï and A. Joulin. Intertwining and commutation relations for birth-death processes. *Bernoulli*, 19(5A):1855–1879, 2013. 2
- [9] G.-Y. Chen and L. Saloff-Coste. The cutoff phenomenon for ergodic Markov processes. *Electron. J. Probab.*, 13:no. 3, 26–78, 2008. 14
- [10] G.-Y. Chen and L. Saloff-Coste. The L^2 -cutoff for reversible Markov processes. *J. Funct. Anal.*, 258(7):2246–2315, 2010. 2, 4, 9, 14, 15

- [11] G.-Y. Chen and L. Saloff-Coste. Computing cutoff times of birth and death chains. *Electron. J. Probab.*, 20:no. 76, 47, 2015. [12](#)
- [12] 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](#), [14](#)
- [13] M. Chesney, M. Jeanblanc-Picqué, and M. Yor. Brownian excursions and Parisian barrier options. *Adv. in Appl. Probab.*, 29(1):165–184, 1997. [16](#)
- [14] M. Choi and P. Patie. [Skip-free Markov chains](#). *Submitted*, 2016. [2](#), [3](#), [7](#), [12](#), [13](#)
- [15] 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](#)
- [16] B. Cloez and C. Delplancke. Intertwinings and Stein’s magic factors for birth-death processes. *arXiv preprint arXiv:1609.08390*, 2016. [2](#)
- [17] H. Cui and Y.-H. Mao. Eigentime identity for asymmetric finite Markov chains. *Front. Math. China*, 5(4):623–634, 2010. [2](#), [3](#), [7](#), [22](#)
- [18] P. Diaconis and J. A. Fill. Strong stationary times via a new form of duality. *Ann. Probab.*, 18(4):1483–1522, 1990. [2](#), [4](#)
- [19] P. Diaconis and L. Miclo. On quantitative convergence to quasi-stationarity. *Ann. Fac. Sci. Toulouse Math. (6)*, 24(4):973–1016, 2015. [4](#), [25](#)
- [20] P. Diaconis and L. Saloff-Coste. Separation cut-offs for birth and death chains. *Ann. Appl. Probab.*, 16(4):2098–2122, 2006. [2](#), [4](#), [12](#), [13](#)
- [21] P. Diaconis, K. Khare, and L. Saloff-Coste. Gibbs sampling, exponential families and orthogonal polynomials. *Statist. Sci.*, 23(2):151–178, 2008. [20](#)
- [22] N. Dunford. Spectral operators. *Pacific J. Math.*, 4:321–354, 1954. [1](#), [3](#), [5](#)
- [23] N. Dunford and J. T. Schwartz. *Linear operators. Part III: Spectral operators*. Interscience Publishers [John Wiley & Sons, Inc.], New York-London-Sydney, 1971. [3](#), [5](#), [9](#)
- [24] 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. [10](#)
- [25] 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. [2](#), [4](#), [8](#), [11](#)
- [26] 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. [4](#), [12](#), [13](#), [22](#)
- [27] G. Fort, S. Meyn, E. Moulines, and P. Priouret. The ODE method for stability of skip-free Markov chains with applications to MCMC. *Ann. Appl. Probab.*, 18(2):664–707, 2008. [2](#)
- [28] R. Griffiths. Multivariate Krawtchouk polynomials and composition birth and death processes. *Symmetry*, 8(5):Art. 33, 19, 2016. [19](#)
- [29] T. Huillet and S. Martinez. Duality and intertwining for discrete Markov kernels: relations and examples. *Adv. in Appl. Probab.*, 43(2):437–460, 2011. [4](#), [10](#)
- [30] A. Inoue and C. Trapani. Non-self-adjoint resolutions of the identity and associated operators. *Complex Anal. Oper. Theory*, 8(7):1531–1546, 2014. [3](#), [9](#), [15](#), [16](#)
- [31] S. Jansen and N. Kurt. On the notion(s) of duality for Markov processes. *Probab. Surv.*, 11:59–120, 2014. [4](#)
- [32] S. Karlin and J. McGregor. Coincidence properties of birth and death processes. *Pacific J. Math.*, 9:1109–1140, 1959. [2](#), [20](#)
- [33] 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](#)

- [34] K. Khare and N. Mukherjee. Convergence analysis of some multivariate Markov chains using stochastic monotonicity. *Ann. Appl. Probab.*, 23(2):811–833, 2013. 19
- [35] K. Khare and H. Zhou. Rates of convergence of some multivariate Markov chains with polynomial eigenfunctions. *Ann. Appl. Probab.*, 19(2):737–777, 2009. 19
- [36] R. Koekoek and R. F. Swarttouw. The Askey-scheme of hypergeometric orthogonal polynomials and its q-analogue, 1998. 20
- [37] I. Kontoyiannis and S. P. Meyn. Spectral theory and limit theorems for geometrically ergodic Markov processes. *Ann. Appl. Probab.*, 13(1):304–362, 2003. 2
- [38] I. Kontoyiannis and S. P. Meyn. Large deviations asymptotics and the spectral theory of multiplicatively regular Markov processes. *Electron. J. Probab.*, 10:no. 3, 61–123, 2005. 2
- [39] 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
- [40] 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. 2
- [41] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2009. 2, 8
- [42] 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. 12
- [43] C. A. Micchelli and R. A. Willoughby. On functions which preserve the class of Stieltjes matrices. *Linear Algebra Appl.*, 23:141–156, 1979. 13
- [44] L. Miclo. An absorbing eigentime identity. *Markov Process. Related Fields*, 21(2):249–262, 2015. 2, 3, 7
- [45] L. Miclo. On the Markovian similarity. Preprint, Mar. 2016. 2, 19
- [46] R. Montenegro and P. Tetali. Mathematical aspects of mixing times in Markov chains. *Found. Trends Theor. Comput. Sci.*, 1(3):x+121, 2006. 8
- [47] P. Patie and M. Savov. [Spectral expansion of non-self-adjoint generalized Laguerre semigroups](#). Submitted, page 162p. (current version), 2015. 2, 9
- [48] P. Patie and Y. Zhao. Spectral decomposition of fractional operators and a reflected stable semigroup. *J. Differential Equations*, 262(3):1690–1719, 2017. 2
- [49] D. Paulin. Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electron. J. Probab.*, 20:no. 79, 1–32, 2015. 2
- [50] J. S. Rosenthal and P. Rosenthal. Spectral bounds for certain two-factor non-reversible MCMC algorithms. *Electron. Commun. Probab.*, 20:no. 91, 10, 2015. 2
- [51] R. Sasaki. Exactly solvable birth and death processes. *J. Math. Phys.*, 50(10):103509, 18, 2009. 20
- [52] W. Schoutens. *Stochastic processes and orthogonal polynomials*, volume 146 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 2000. 20
- [53] D. Siegmund. The equivalence of absorbing and reflecting barrier problems for stochastically monotone Markov processes. *Ann. Probability*, 4(6):914–924, 1976. 4, 10
- [54] R. C. Thompson. Singular values, diagonal elements, and convexity. *SIAM J. Appl. Math.*, 32(1):39–63, 1977. 11
- [55] C. Villani. Hypocoercivity. *Mem. Amer. Math. Soc.*, 202(950):iv+141, 2009. 4, 9
- [56] J. Wermer. Commuting spectral measures on Hilbert space. *Pacific J. Math.*, 4:355–361, 1954. 3
- [57] R. M. Young. *An introduction to nonharmonic Fourier series*. Academic Press, Inc., San Diego, CA, first edition, 2001. 3, 5, 10
- [58] H. Zhou. *Examples of multivariate Markov chains with orthogonal polynomial eigenfunctions*. PhD thesis, Stanford University, 2008. 5, 19, 20, 25

SCHOOL OF OPERATIONS RESEARCH AND INFORMATION ENGINEERING, CORNELL UNIVERSITY, ITHACA, NY 14853.

E-mail address: cc2373@cornell.edu

SCHOOL OF OPERATIONS RESEARCH AND INFORMATION ENGINEERING, CORNELL UNIVERSITY, ITHACA, NY 14853, AND, LABORATOIRE DE MATHÉMATIQUES ET LEURS APPLICATIONS, UMR CNRS 5142, UNIVERSITÉ DE PAU ET DES PAYS DE L'ADOUR, PAU, FRANCE.

E-mail address: pp396@cornell.edu