CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Analysis of non-reversible Markov chains via similarity orbits

Michael C. H. Choi[1,*] and Pierre Patie[2]

[1]Institute for Data and Decision Analytics, The Chinese University of Hong Kong, Shenzhen, Guangdong, 518172, PR China and Shenzhen Institute of Artificial Intelligence and Robotics for Society and [2]School of Operations Research and Information Engineering, Cornell University, Ithaca, NY 14853, USA and Laboratoire de Mathématiques et leurs Applications, UMR CNRS 5142, Université de Pau et des Pays de l'Adour, Pau, France
*Corresponding author. Email: michaelchoi@cuhk.edu.cn

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 that 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 chain. As by-products, we obtain a spectral representation in terms of non-self-adjoint resolutions of identity in the sense of Dunford [21] 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 a particular similarity orbit of reversible Markov kernels, which we call the pure birth orbit, and analyse various possibly non-reversible variants of classical birth–death processes in this orbit.

1. Introduction

The spectral theorem of normal operators is undoubtedly a powerful tool for dealing with substantial and difficult issues arising in the analysis of Markov chains. The intrusion of spectral theory into 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 eigenvalue expansions, 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.* [1, 10, 20, 41, 45]), to name but a few. It is also central to their statistical estimations, as is demonstrated by the recent work of [2] on the integral functionals of normal Markov chains.

However, the lack of a spectral theorem for non-normal operators causes major difficulties in 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

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

CrossMark

where π is a reference (invariant or excessive) measure of P and $\pi(x) > 0$ for all $x \in \mathcal{X}$. Not only is the non-reversibility property, or generally the non-normality of P, a generic property from a theoretical perspective, it is also natural, and it has recently become increasingly popular in various applications. For instance, non-reversible Markov chains appear in the study of queueing networks and fluid approximation [26], hyperplane rearrangement [53] and the very recent introduction of non-reversible Metropolis–Hastings and its variants; see *e.g.* [6, 54].

To overcome the challenge of analysing 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], reversibilization techniques as in [24, 51] or recasting to a weighted-L^{∞} space [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 countable state space $\mathcal{X}$. We emphasize that our approach offers a unifying framework to analyse 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 [13] from skip-free Markov chains to general ones. It is also in line with the papers by Miclo [47], Patie and Savov [49] and Patie and Zhao [50] for the study of spectral theory of non-reversible Markov processes and by Diaconis and Fill [17], Chafaï and Joulin [7] and Cloez and Delplancke [15] for birth–death processes, which rely on the notion of intertwining relationships. We proceed by recalling the definition of similarity orbit as introduced in [13].

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 Q of a Markov chain on $\mathcal{X}$, and we write $P \sim Q$ if there exists a bounded linear operator $\Lambda \colon \ell^2(\pi_Q) \to \ell^2(\pi)$ (π_Q being a reference measure for Q) with bounded inverse such that

$$P\Lambda = \Lambda Q. \tag{1.1}$$

We also write $\widehat{\Lambda}$ to denote the adjoint operator of Λ. 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 1.2. 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 1.3. Note that this definition carries over when we study similarity on the level of infinitesimal generators in the continuous-time setting. For example, we write $L \overset{\Lambda}{\sim} G$ if L (resp. G) is the infinitesimal generator associated with the continuous-time Markov semigroup $(P_t)_{t \geqslant 0}$ (resp. $(Q_t)_{t \geqslant 0}$). It follows easily that if $L \overset{\Lambda}{\sim} G$ then $P_t \overset{\Lambda}{\sim} Q_t$ for $t \geqslant 0$.

Remark 1.4. We compare our definition of similarity with other notions of intertwining in the literature. In [47], both the link kernel Λ and Q are assumed to be Markovian, while in [7, 15], both P and Q are assumed to be birth–death processes. In [17], the authors construct the strong stationary duality theory for general Markov chains. However, in Definition 1.1 we do not require Λ to be a Markov operator, and P, Q are general Markov operators instead of birth–death processes.

The $\mathcal{S}$ class is now defined as the similarity orbit in $\mathcal{M}$ consisting of all Markov transition kernels that are similar to a normal transition kernel on $\mathcal{X}$. Note that reversible Markov kernels are normal operators in $\ell^2(\pi)$. From now on we write $\mathcal{N}$ for the set of normal transition kernels Q on $\mathcal{X}$, that is, $Q\widehat{Q} = \widehat{Q}Q$ in $\ell^2(\pi_Q)$, where $\widehat{}$ denotes throughout the corresponding object for the time-reversal process.

Definition 1.5 (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, that is,

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

We point out that according to [60], the class $\mathcal{S}$ is also characterized as the class of Markov chains whose transition kernel is a spectral scalar-type operator in the sense of [21, Section 3]; see also [22, p. 1938, Definition 1]. As we will see in Section 2, this characterization will be crucial to proving many of our later results. Note that we could also study a wider class of transition kernels $\mathcal{S}'$ in which Q is not necessarily a Markov operator. However, we intend to focus our investigation on the class $\mathcal{S}$ in this paper, as it is the appropriate setting to extend various substantial results that have been developed for reversible chains.

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 elements admits a spectral representation with respect to non-self-adjoint resolution of identity as introduced by Dunford [21]; see also [22]. We also remark on the growing interest in non-self-adjoint operators with real spectrum that arise in the study of pseudo-Hermitian quantum mechanics; see *e.g.* [30] and the references therein. As a by-product, one can develop a functional calculus for this class as for normal operators. Moreover, we obtain, under mild conditions, an eigenvalue expansion expressed in terms of a Riesz basis, a notion that generalizes an orthogonal basis and was introduced in non-harmonic analysis; see [61]. Another intriguing aspect of the similarity orbit analysis is that in the continuous-time setting with $L \in \mathcal{S}(G)$ (see Remark 1.3 above), where G is the generator of a normal Markov chain, then both the heat kernel $(e^{tL})_{t \geqslant 0}$ and $(e^{tG})_{t \geqslant 0}$ share the same *eigentime* identity, offering new examples and insights into the sequence of work by Aldous and Fill [1], Cui and Mao [16] and Miclo [45]. 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, for small times we adapt the singular value upper bound of [24], while for large times the decay rate is the second largest eigenvalue in modulus, up to 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 [59]. All these first consequences of the spectral representation are stated and proved in Section 2. 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 define what we call the generalized monotonicity condition class $\mathcal{GMC}$, such that the time-reversal $\widehat{P}$ intertwines with a birth–death chain in Section 3. 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 [13] in the context of skip-free chains. Note that the notion of stochastic monotonicity is studied by Siegmund [57] and Clifford and Sudbury [14] and intertwining between stochastic monotone birth–death chains, which are reversible chains, has been previously investigated in detail by Diaconis and Fill [17], Huillet and Martinez [29] and Jansen and Kurt [31]. Relying on the spectral decomposition as well as the fastest strong stationary time result of general chains obtained by Fill [25], we study the separation cutoff phenomenon and demonstrate that the famous 'spectral gap times mixing time' conjecture as well as the proof in

[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 criterion proposed in [10] and further elaborated in [8], we illustrate that the usual L^2-cutoff criterion 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

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

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

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

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

Finally, in Section 6, we examine a particular similarity orbit of reversible Markov chains, which we call the pure birth orbit. More precisely, suppose that we start with a reversible generator G such that $G \overset{\Lambda}{\sim} L$, where L is the generator of a contraction yet possibly non-Markovian semigroup $(e^{tL})_{t \geqslant 0}$; we would like to investigate various properties of L with Λ being a pure birth kernel. This idea is powerful enough to allow us to generate completely new Markov or contraction kernels 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 of the pure birth variants of a constant-rate birth–death model.

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 the 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 1.5. We write $\|\cdot\|_{\mathrm{op}}$ for the operator norm, *i.e.* $\|P\|_{\mathrm{op}} = \sup_{\|f\|_\pi = 1} \|Pf\|_\pi$, and $[\![a, b]\!] := \{a, a+1, \ldots, b-1, b\}$ for any $a \leqslant b \in \mathbb{Z}$. We proceed by recalling that P has a time-reversal $\widehat{P}$, that is, for $x, y \in \mathcal{X}$,

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

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

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

where $\overline{g}$ is the complex conjugate of g. A spectral measure (or resolution of identity) in the sense of [21, Section 3] and [22, p. 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 $A, B \in \mathcal{B}(\mathbb{C})$,

$$E_{A \cap B} = E_A E_B,$$

while for disjoint A, B,

$$E_{A \cup B} = E_A + E_B.$$

(3) There exists a constant $C > 0$ such that $\|E_B\|_{\mathrm{op}} \leqslant C$ 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 denote the spectrum (resp. continuous spectrum, point spectrum, residual spectrum) of M. We proceed to recall the notion of a 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.* [61, Theorem 9]) that the sequence (f_k) forms a Riesz basis if and only if (f_k) is complete in $\mathcal{H}$ and there exist positive constants A, B such that, for arbitrary $n \in \mathbb{N}$ and scalars $c_1, \dots, c_n$, we have

$$A \sum_{k=1}^{n} |c_k|^2 \leqslant \left\| \sum_{k=1}^{n} c_k f_k \right\|^2 \leqslant B \sum_{k=1}^{n} |c_k|^2. \tag{2.1}$$

If (g_k) is a biorthogonal sequence to (f_k), that is, $\langle f_k, g_m \rangle_\pi = \delta_{k,m}$, $k, m \in \mathbb{N}$ and $\delta_{k,m}$ is the Kronecker symbol, then (g_k) also forms a Riesz basis. We are now ready to state the main result of this paper, and the proof can be found in Section 2.1.

Theorem 2.1. *Assume that $P \in \mathcal{S}$ with $P \overset{\Lambda}{\sim} Q \in \mathcal{N}$. Then the following holds.*

(a) *Let $\mathcal{E} = \{E_B; B \in \mathcal{B}(\mathbb{C})\}$ denote the self-adjoint spectral measure of Q. 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

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

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 $\mathcal{X}$ with distinct eigenvalues, then for any $f \in \ell^2(\pi)$ and $n \in \mathbb{N}$,

$$P^n f = \sum_{k \in \mathcal{X}} \lambda_k^n \langle f, f_k^* \rangle_\pi f_k,$$

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

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

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

(c) *Suppose that Λ is a unitary operator, that is, $\Lambda^{-1} = \widehat{\Lambda}$. 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 Λ^{-1} having non-negative entries 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 a unitary Markov kernel. Moreover, for any $Q \in \mathcal{M}$, the permutation orbit $\mathcal{S}_\mathcal{P}(Q)$ of Q is given by*

$$\mathcal{S}_\mathcal{P}(Q) = \{P \in \overline{\mathcal{M}}; P\Lambda = \Lambda Q, \Lambda \in \mathcal{P}\} \subset \mathcal{M},$$

where $\overline{\mathcal{M}}$ is the set of square matrices on $\mathcal{X}$.

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

Remark 2.2. As suggested by item (c), we can generate new non-normal examples via non-unitary links from known normal Markov chains such as birth–death processes. In Section 6 we investigate a particular non-unitary orbit that we call the pure birth orbit.

Remark 2.3. The result in Theorem 2.1(d) has also been obtained by Miclo using a different proof; see Lemma 12 in [46].

Remark 2.4. The key to Theorem 2.1(e) lies on the simplicity of the spectrum of Q. In the context of non-negative Jacobi matrices, the inverse eigenvalue problem has been studied by Friedland and Melkman [27, Theorem 4].

Remark 2.5. Theorem 2.1 can be generalized easily to the continuous-time setting; see also Remark 1.3. Indeed, suppose that $L \in \mathcal{S}(G)$, where G is a normal generator with spectral measure $\mathcal{E} = \{E_B; B \in \mathcal{B}(\mathbb{C})\}$. Then, for $t \geqslant 0$,

$$P_t = \int_{\sigma(L)} e^{t\lambda} \, dF_\lambda,$$

where $\{F_B := \Lambda E_B \Lambda^{-1}; B \in \mathcal{B}(\mathbb{C})\}$.

As a first 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 has been a rich literature devoted to the study of convergence to equilibrium for non-reversible chains by means of reversibilizations; see *e.g.* [1, 24, 41, 48] and the references therein. Our approach reveals a natural extension to the non-reversible case of the classical spectral gap that appears in the study of reversible chains. To state our result we now fix some notations. We denote the second largest eigenvalue in modulus (SLEM) or the spectral radius of P in the Hilbert space

$$\ell_0^2(\pi) = \{f \in \ell^2(\pi); \langle f, \mathbf{1} \rangle_\pi = 0\},$$

by

$$\lambda_* = \lambda_*(P) = \sup\{|\lambda_i|; \lambda_i \in \sigma(P), \lambda_i \neq 1\}.$$

Then the *absolute spectral gap* is $\gamma_* = 1 - \lambda_*$. For any two probability measures μ, ν on $\mathcal{X}$, the total variation distance between μ and ν is given by

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

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

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

For $g \in \ell^2(\pi)$, the mean of g with respect to π can be written as $\mathbb{E}_\pi(g) = \langle g, \mathbf{1} \rangle_\pi$. Similarly, the variance of g with respect to π is $\mathrm{Var}_\pi(g) = \langle g, g \rangle_\pi - \mathbb{E}_\pi^2(g)$. Finally, we recall that Fill [24, Theorem 2.1], in the finite state space case, obtained the following bound valid for all $n \in \mathbb{N}_0$:

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

where $\pi_{\min} = \min_{x \in \mathcal{X}} \pi(x)$ and $\sigma_*(P) = \sqrt{\lambda_*(\widehat{P}P)}$ 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 2.2.

Corollary 2.6. *Let $P \in \mathcal{S}$ with invariant distribution π, that is, $\pi P = \pi$, and assume that P is compact.*

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

$$\lambda_*^n \leqslant \|P^n - \pi\|_{\ell^2(\pi) \to \ell^2(\pi)} \leqslant \min\left(\sigma_*^n(P), \kappa(\Lambda)\lambda_*^n\right) = \sigma_*^n(P)\mathbb{1}_{\{n < n^*\}} + \kappa(\Lambda)\lambda_*^n\mathbb{1}_{\{n \geqslant n^*\}}, \tag{2.3}$$

   where

$$n^* = \left\lceil \frac{\ln \kappa(\Lambda)}{\ln \sigma_*(P) - \ln \lambda_*} \right\rceil,$$

   and

$$\kappa(\Lambda) = \|\Lambda\|_{\ell^2(\pi_Q) \to \ell^2(\pi)} \|\Lambda^{-1}\|_{\ell^2(\pi) \to \ell^2(\pi_Q)} \geqslant 1$$

   is the condition number of Λ. When $\mathcal{X}$ is a finite state space, a sufficient condition for which $\lambda_* < \sigma_*(P)$ is given by $\max_{i \in \mathcal{X}} P(i, i) > \lambda_*$. In this case, for n large enough, the convergence rate λ_* given (2.3) is strictly better than the reversibilization rate $\sigma_*(P)$.

(2) *Suppose that $\mathcal{X}$ is a finite state space. For any $n \in \mathbb{N}_0$,*

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

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

Remark 2.7. 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.* [10, Section 4.3]) $\|P^n - \pi\|_{\ell^2(\pi) \to \ell^2(\pi)} = \lambda_*^n$ and $\kappa(\Lambda) = 1$, which is a specific instance of item (1).

Remark 2.8. We also recall the discrete analogue of the notion of hypocoercivity introduced in [59], that is, there exists a constant $C < \infty$ and $\rho \in (0, 1)$ such that, for all $n \in \mathbb{N}$,

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

Note that, in general, these constants are not known explicitly. We observe that the upper bound in (2.3) reveals that the ergodic chains in $\mathcal{S}$ satisfy this hypocoercivity phenomenon. More interestingly, our similarity concept approach enables us to obtain both an explicit rate of convergence and a spectral interpretation of this rate. 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 of Patie and Savov [49], where a similar spectral interpretation of the hypocoercivity phenomenon is given for a class of non-self-adjoint Markov semigroups.

Remark 2.9. Here we provide an alternative expression for the upper bound of (2.3). Let $\ell^2(\mathcal{X})$ be the space of square summable functions on $\mathcal{X}$ equipped with the standard inner product $\langle \cdot, \cdot \rangle$. Suppose that $P \in \mathcal{S}(Q)$ with $QU = UD$, where D is a diagonal matrix and U is an isometry from $\ell^2(\mathcal{X})$ to $\ell^2(\pi_Q)$. Then $B = \Lambda U$ is an eigenbasis for P with $\kappa(B) = \kappa(\Lambda)$, where $\kappa(B) := \|B\|_{\ell^2(\mathcal{X}) \to \ell^2(\pi)} \|B^{-1}\|_{\ell^2(\pi) \to \ell^2(\mathcal{X})}$ is the condition number of B, so the upper bound in (2.3) can be written as

$$\|P^n - \pi\|_{\ell^2(\pi) \to \ell^2(\pi)} \leqslant \kappa(B)\lambda_*^n.$$

As a second application of Theorem 2.1, we first recall the celebrated eigentime identity studied by Aldous and Fill [1], Cui and Mao [16] and Miclo [45]. 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(a)), we can easily see that both P and Q share the same eigentime identity.

Corollary 2.10 (eigentime identity). *Suppose that $\mathcal{X}$ is a finite state space and $(Q_t)_{t \geqslant 0}$ (resp. $(P_t)_{t \geqslant 0}$) has generator G (resp. L) associated with the ergodic Markov chain $(X_t)_{t \geqslant 0}$ (resp. $(Y_t)_{t \geqslant 0}$). If $L \in \mathcal{S}(G)$, with G being a normal generator and common eigenvalues $(-\lambda_i)_{i \in [\![|\mathcal{X}|]\!]}$, then $(P_t)_{t \geqslant 0}$ and $(Q_t)_{t \geqslant 0}$ share the same eigentime identity. That is, if $\tau_y^Q := \inf\{t \geqslant 0; X_t = y\}$ (resp. $\tau_y^P := \inf\{t \geqslant 0; Y_t = y\}$), then*

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

2.1 Proof of Theorem 2.1

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

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

follows from Proposition 3.9 in [3]. Define $\overline{P} := \int_{\sigma(P)} \lambda \, dF_\lambda$. We have

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

so the desired spectral resolution of P follows, and 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 the spectral scalar-type operator; see *e.g.* [22, Theorem 1, Chapter XV.5, p. 1941]. We proceed to handle the case when P is compact. Let (g_k) denote the (orthogonal) eigenfunctions of the normal transition kernel Q. Since $f_k = \Lambda g_k$ and Λ is bounded, (f_k) is complete as (g_k) is a basis. As Λ is bounded from above and below, for any $n \in \mathbb{N}$ and arbitrary sequence $(c_k)_{k=1}^n$, we have

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

where we can take $A = \|\Lambda^{-1}\|^{-2}$ and $B = \|\Lambda\|^2$, so (2.1) is satisfied. It follows from [61, Theorem 9] that there exists the sequence (f_k^*) being the unique Riesz basis biorthogonal to (f_k), and any $f \in \ell^2(\pi)$ can be written as

$$f = \sum_{k \in \mathcal{X}} c_k f_k,$$

where $c_k = \langle f, f_k^* \rangle_\pi$. The 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 $g \in \ell^2(\pi)$,

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

which shows that $\widehat{Q} \overset{\widehat{\Lambda}}{\sim} \widehat{P}$. The opposite direction can be shown similarly.

For item (c), since Λ is unitary, the spectral measures of P and Q are related by $F_B = \Lambda E_B \widehat{\Lambda}$, so F_B is self-adjoint if and only if E_B is self-adjoint, which implies that P is normal if and only if Q is normal. If Q is self-adjoint, then item (b) yields $P \overset{\Lambda}{\sim} Q$ if and only if $Q \overset{\Lambda^{-1}}{\sim} \widehat{P}$, which implies that $\widehat{P} = P$ in $\ell^2(\pi)$. The opposite direction can be shown similarly.

Next we show item (d). If Λ is a permutation link, then it is trivial to see that Λ is an invertible Markov kernel. For the opposite direction, it is known (see *e.g.* [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 $\Lambda = \Lambda_\sigma$. Now let $Q \in \mathcal{M}$ 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 the inverse spectral theorem (see *e.g.* [23, Section 5.8]), one can always construct an ergodic Markov chain with transition matrix Q such that

$$Q = V D V^{-1}. \qquad \square$$

2.2 Proof of Corollary 2.6

We first show the upper bound in item (1). Define the synthesis operator $T^* : \ell^2 \to \ell^2(\pi)$ by $\alpha = (\alpha_i) \mapsto T^*(\alpha) = \sum_{i \in \mathcal{X}} \alpha_i f_i$, 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 $i \in \mathcal{X}$, we take $\alpha_i = \alpha_i(n) = \lambda_i^n \langle g, f_i^* \rangle_\pi$, and

denote (q_i) to be the orthonormal eigenfunctions of $Q \in \mathcal{N}$, where $f_i = \Lambda q_i$. Note that $\|T^*\|_{\mathrm{op}} \leqslant \|\Lambda\|_{\ell^2(\pi_Q) \to \ell^2(\pi)} < \infty$, since

$$\|T^*(\alpha)\|_{\ell^2 \to \ell^2(\pi)} = \left\| \sum_{i \in \mathcal{X}} \alpha_i \Lambda q_i \right\|_{\ell^2 \to \ell^2(\pi)}$$

$$\leqslant \|\Lambda\|_{\ell^2(\pi_Q) \to \ell^2(\pi)} \left\| \sum_{i \in \mathcal{X}} \alpha_i q_i \right\|_{\pi_Q}$$

$$\leqslant \|\Lambda\|_{\ell^2(\pi_Q) \to \ell^2(\pi)} \|\alpha\|_{\ell^2}.$$

For $g \in \ell_0^2(\pi)$, we also have

$$\sum_{i \in \mathcal{X}} |\langle g, f_i^* \rangle_\pi|^2 = \sum_{i \in \mathcal{X}} |\langle g, (\Lambda^*)^{-1} q_i \rangle_\pi|^2$$

$$= \sum_{i \in \mathcal{X}} |\langle \Lambda^{-1} g, q_i \rangle_{\pi_Q}|^2$$

$$= \|\Lambda^{-1} g\|_{\pi_Q}^2$$

$$\leqslant \|\Lambda^{-1}\|_{\ell^2(\pi) \to \ell^2(\pi_Q)}^2 \|g\|_\pi^2,$$

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

$$\|P^n g - \pi g\|_\pi^2 = \|T^*(\alpha)\|_{\ell^2 \to \ell^2(\pi)}^2$$

$$\leqslant \|\Lambda\|_{\ell^2(\pi_Q) \to \ell^2(\pi)}^2 \|\alpha\|_{l^2}^2$$

$$\leqslant \|\Lambda\|_{\ell^2(\pi_Q) \to \ell^2(\pi)}^2 \|\Lambda^{-1}\|_{\ell^2(\pi) \to \ell^2(\pi_Q)}^2 \lambda_*^{2n} \|g\|_\pi^2. \qquad (2.4)$$

The desired upper bound follows from (2.4) and

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

(see *e.g.* [24]). The lower bound in (1) follows readily from the well-known result that the nth power of the spectral radius λ_*^n is less than or equal to the norm of P^n on the reduced space $\ell_0^2(\pi)$. For the sufficient condition in item (1), that is, $\max_{i \in \mathcal{X}} P(i, i) > \lambda_*$ implies $\lambda_* < \sigma_*(P)$, it is a straightforward consequence of the Sing–Thompson theorem; see [58].

Next we show item (2). Using (2.4), we get

$$\mathrm{Var}_\pi(\widehat{P}^n g) \leqslant \kappa(\widehat{\Lambda})^2 \lambda_*^{2n} \mathrm{Var}_\pi(g) = \kappa(\Lambda)^2 \lambda_*^{2n} \mathrm{Var}_\pi(g), \quad n \in \mathbb{N}_0, \qquad (2.5)$$

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

$$\|\delta_x P^n - \pi\|_{TV}^2 = \frac{1}{4} \mathbb{E}_\pi^2 \left| \frac{\delta_x P^n}{\pi} - 1 \right|$$

$$\leqslant \frac{1}{4} \mathrm{Var}_\pi \left(\frac{\delta_x P^n}{\pi} \right)$$

$$= \frac{1}{4} \mathrm{Var}_\pi \left(\widehat{P}^n \frac{\delta_x}{\pi} \right)$$

$$\leqslant \frac{1}{4}\kappa(\Lambda)^2\lambda_*^{2n}\mathrm{Var}_\pi\left(\frac{\delta_x}{\pi}\right)$$

$$= \frac{1}{4}\kappa(\Lambda)^2\lambda_*^{2n}\frac{1-\pi(x)}{\pi(x)}$$

$$\leqslant \frac{1}{4}\kappa(\Lambda)^2\lambda_*^{2n}\frac{1-\pi_{\min}}{\pi_{\min}},$$

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

3. The $\mathcal{GMC}$ class and separation cutoff

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, a fundamental yet challenging issue. However, we manage to identify a set of sufficient conditions that we call the generalized monotonicity condition class $\mathcal{GMC}$, generalizing the $\mathcal{MC}$ class for skip-free chains as introduced in [13], such that the kernels in $\mathcal{GMC}$ have real and distinct eigenvalues and the time-reversal $\widehat{P}$ intertwines with a birth–death chain with the link kernel Λ being related to the Siegmund kernel $H_S(x, y) = \mathbb{1}_{\{x\leqslant y\}}$.

Definition 3.1. (the $\mathcal{GMC}$ class). We say that, for some $\mathfrak{r} \geqslant 3$, $P \in \mathcal{GMC}_\mathfrak{r}$ if $P \in \mathcal{M}$ with $\mathcal{X} = [\![0, \mathfrak{r}]\!]$ and for every $x \in [\![0, \mathfrak{r}-1]\!]$, its time-reversal $(X, \widehat{\mathbb{P}})$ satisfies

  (1) (stochastic monotonicity) $\widehat{\mathbb{P}}_{x+1}(X_1 \leqslant x) \leqslant \widehat{\mathbb{P}}_x(X_1 \leqslant x)$,
  (2) (strict stochastic monotonicity) $\widehat{\mathbb{P}}_{x+1}(X_1 \leqslant x-1) < \widehat{\mathbb{P}}_x(X_1 \leqslant x-1)$, $x \neq 0$, and
  (3) (strict stochastic monotonicity) $\widehat{\mathbb{P}}_{x+1}(X_1 \leqslant x+1) < \widehat{\mathbb{P}}_x(X_1 \leqslant x+1)$, $x \neq \mathfrak{r}-1$, and
  (4) (restricted downward jump) $\widehat{\mathbb{P}}_{x+1}(X_1 \leqslant x-k) = \widehat{\mathbb{P}}_x(X_1 \leqslant x-k)$, $k \in [\![2, x]\!]$, and
  (5) (restricted upward jump) $\widehat{\mathbb{P}}_{x+1}(X_1 \leqslant x+k) = \widehat{\mathbb{P}}_x(X_1 \leqslant x+k)$, $k \in [\![2, \mathfrak{r}-1-x]\!]$.

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

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

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 superscript plus sign in $\mathcal{GMC}^+$ means that this class has non-negative eigenvalues; see Remark 3.9 below.

Remark 3.2. Recall that in [13], if $P \in \mathcal{MC}$, that is, P is upward skip-free and satisfies (1), (3) and (5), then it is clear that $\mathcal{MC} \subset \mathcal{GMC}$, as items (2) and (4) in Definition 3.1 are automatically satisfied since the time-reversal $\widehat{P}$ is downward skip-free.

Next, we formally state that a transformation of $P \in \mathcal{GMC}$ is contained in the similarity orbit of an irreducible birth–death kernel. The proof can be found in Section 3.1. For any square matrix M on $[\![0, \mathfrak{r}]\!]$, we write $M^{[\![0, \mathfrak{r}-1]\!]}$ to denote the principal submatrix on $[\![0, \mathfrak{r}-1]\!]$.

Theorem 3.3. *Let $P \in \mathcal{GMC}$ and write $M := P\Lambda$ on $[\![0, \mathfrak{r}]\!]$, where $\Lambda = (H_S^T D_\pi)^{-1}$ and D_π is the diagonal matrix of π. Then $(\Lambda^{-1})^{[\![0, \mathfrak{r}-1]\!]}(P\Lambda)^{[\![0, \mathfrak{r}-1]\!]} \in \mathcal{S}(Q)$ with Q being an irreducible birth–death transition kernel on $[\![0, \mathfrak{r}-1]\!]$.*

Remark 3.4 (on the connection to the strong stationary duality theory of [17]). We would like to highlight the connection between Theorem 3.3 and the classical construction of strong stationary duality (SSD) proposed by Diaconis and Fill. In [17, Theorem 5.5], writing π_0 to be the initial distribution at time 0, if $\pi_0(x)/\pi(x)$ is decreasing in x and the time-reversal is stochastically monotone, then the SSD of a chain X can be derived as the Doob H-transform of the Siegmund dual of the time-reversal of X, with $H = H_S^T \pi$ being the cumulative distribution function of π. As we shall see in the proof of Theorem 3.3, our result bears a resemblance to the above construction by Diaconis and Fill, with Q in Theorem 3.3 being a Doob transform of the state-restriction of the Siegmund dual of the time-reversal. Note that both our result and the classical SSD construction require stochastic monotonicity of the time-reversal, yet our $\mathcal{GMC}$ class requires more conditions (namely items (2) to (4) of Definition 3.1). While we apply the same Doob H-transform to the Siegmund dual, we apply a further Doob $\widetilde{h}$-transform for the state-restricted and Doob H-transformed Siegmund dual restricted to $[\![0, \mathfrak{r} - 1]\!]$, where $\widetilde{h}$ is defined in (3.2) below.

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

Example 3.5.

$$\widehat{P} = \begin{pmatrix} 0.5 & 0.35 & 0.05 & 0.1 \\ 0.3 & 0.5 & 0.1 & 0.1 \\ 0.2 & 0.1 & 0.5 & 0.2 \\ 0.2 & 0.05 & 0.25 & 0.5 \end{pmatrix}, \quad P = \begin{pmatrix} 0.5 & 0.2629 & 0.1157 & 0.1213 \\ 0.3994 & 0.5 & 0.0660 & 0.0346 \\ 0.0864 & 0.1515 & 0.5 & 0.2621 \\ 0.1648 & 0.1444 & 0.1907 & 0.5 \end{pmatrix}$$

has eigenvalues $1, 0.54, 0.28, 0.18$, and satisfies (1)–(6) of Definition 3.1. Note that

$$(\Lambda^{-1})^{[\![0, \mathfrak{r} - 1]\!]} (P\Lambda)^{[\![0, \mathfrak{r} - 1]\!]} = \begin{pmatrix} 0.2 & 0.1 & 0 \\ 0.05 & 0.5 & 0.05 \\ 0 & 0.1 & 0.3 \end{pmatrix}.$$

We proceed to investigate the separation cutoff phenomenon for the $\mathcal{GMC}$ class. For birth–death chains such phenomena were studied in [20] and [11], and were recently extended to upward skip-free chains in [42] and [13]. In order to establish the famous 'spectral gap times mixing time' criterion (see *e.g.* [52]) for this class, we will build upon the result of [25] to first analyse the fastest strong stationary time of this class, followed by demonstrating that the proof in [20] carries over for this class of non-reversible chains.

To this end, we recall the definition of separation distance of Markov chains, which is used as a standard measure for convergence to equilibrium. For $n \in \mathbb{N}$, the maximum separation distance $s(n)$ is defined by

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

One of the nice features is its connection to strong stationary times, which we now describe. We say that a randomized stopping time T for a Markov chain X with stationary distribution π is a strong stationary time T, possibly depending on the initial starting position x if, for all $x, y \in E$,

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

It is well known (see *e.g.* [41, Lemma 6.11]) that the tail probability of a strong stationary time U provides an upper bound on the separation distance, that is,

$$s_x(n) \leqslant \mathbb{P}(U > n).$$

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

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

and

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

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

$$\lim_{n \to \infty} \frac{\mathrm{T}_n(\varepsilon)}{t_n} = 1.$$

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

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

We now proceed to discuss the main results of this section, with Theorem 3.6 addressing the case of a discrete-time family of Markov chains, and Theorem 3.8 discussing the continuous-time version. Recall that the notation $\mathcal{GMC}^+$ introduced in Definition 3.1 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 3.6 (resp. Theorem 3.8).

Theorem 3.6. *For $n \geqslant 1$, suppose that $P_n \in \mathcal{GMC}_{\mathfrak{r}_n}^+$ on the state space $\mathcal{X}_n = [\![0, \mathfrak{r}_n]\!]$ that started at 0. Let $(\theta_{n,i})_{i=1}^{\mathfrak{r}_n}$ be the non-zero eigenvalues of $I - P_n$, and let $(c_{n,i})_{i=0}^{\mathfrak{r}_n}$ be the mixture weights of the nth chain defined in (3.3) in Lemma 3.10. Define*

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

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

Remark 3.7. For discrete-time stochastically monotone birth–death chains that start at 0, we have $w_i = 1$ for $i \in [\![1, \mathfrak{r}_n]\!]$ and $c_{n,0} = 0$, and hence we recover [20, Theorem 5.2].

Theorem 3.8. *For $n \geqslant 1$, suppose that $L_n = P_n - I$ is the infinitesimal generator with $P \in \mathcal{GMC}_{\mathfrak{r}_n}^+$ on the state space $\mathcal{X}_n = [\![0, \mathfrak{r}_n]\!]$ that started at 0. Let $(\theta_{n,i})_{i=1}^{\mathfrak{r}_n}$ be the non-zero eigenvalues of $-L_n$, and let $(c_{n,i})_{i=0}^{\mathfrak{r}_n}$ be the mixture weights defined in (3.4) in Remark 3.13. Define*

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

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

We will only prove Theorem 3.6 as the proof of Theorem 3.8 is very similar and thus omitted.

3.1 Proof of Theorem 3.3

We write $\widetilde{P}$ for the so-called Siegmund dual (or H_S-dual) of $\widehat{P}$. That is, $\widetilde{P}^T = H_S^{-1}\widehat{P}H_S$ where $H_S = (H_S(x,y))_{x,y\in\mathcal{X}}$ is defined to be $H_S(x,y) = \mathbb{1}_{\{x\leqslant y\}}$ and its inverse $H_S^{-1} = (H_S^{-1}(x,y))_{x,y\in\mathcal{X}}$ is $H_S^{-1}(x,y) = \mathbb{1}_{\{x=y\}} - \mathbb{1}_{\{x=y-1\}}$; see [57]. Since $X \in \mathcal{GMC}$, then $\widehat{P}$ is stochastically monotone, hence from [4, Proposition 4.1] we have that $\widetilde{P}$ is a sub-Markovian kernel. For $x \in [\![0, \mathfrak{r}-2]\!]$, conditions (2) and (3) in $\mathcal{GMC}$ yield, respectively, $\widetilde{p}(x, x+1) > 0$, while for $x \in [\![1, \mathfrak{r}-1]\!]$, we have $\widetilde{p}(x, x-1) > 0$. Conditions (4) and (5) in $\mathcal{GMC}$ guarantee that $\widetilde{p}(x,y) = 0$ for each $x \in [\![0, \mathfrak{r}-3]\!]$ and $y \in [\![x+2, \mathfrak{r}-1]\!]$ and for each $x \in [\![2, \mathfrak{r}-1]\!]$ and $y \in [\![0, x-2]\!]$. That is, $\widetilde{P}$ is a (strictly substochastic) irreducible birth–death chain when restricted to the state space $[\![0, \mathfrak{r}-1]\!]$. Let $\widetilde{P}^{bd}$ denote the restriction of $\widetilde{P}$ to $[\![0, \mathfrak{r}-1]\!]$. By breaking off the last row and last column of $\widetilde{P}$, we can write

$$\widetilde{P} = \begin{pmatrix} \widetilde{P}^{bd} & \mathbf{v} \\ \mathbf{0} & 1 \end{pmatrix} = (H_S^{-1}\widehat{P}H_S)^T, \tag{3.1}$$

where $\mathbf{0}$ is a row vector of zero, and $\mathbf{v}$ is a column vector storing $\widetilde{p}(x, \mathfrak{r})$ for $x \in [\![0, \mathfrak{r}-1]\!]$. Considering the h-transform of $\widetilde{P}$ with $h = H_S^T\pi > \mathbf{0}$ (see *e.g.* [29, Theorem 2]), we see that

$$M = P\Lambda = \Lambda\widetilde{P},$$

where $\Lambda = (H_S^T D_\pi)^{-1}$ (D_π is the diagonal matrix of π). Observing that the last row of $\widetilde{P}$ is zero except the last entry, we have

$$M^{[\![0,\mathfrak{r}-1]\!]} = \Lambda^{[\![0,\mathfrak{r}-1]\!]}\widetilde{P}^{bd}.$$

Note that $\widetilde{P}^{bd}$ is a strictly substochastic matrix with $\mathfrak{r}$ as a killing boundary. Let $\widetilde{T}^{bd}$ denote the lifetime of Markov chain with transition kernel $\widetilde{P}^{bd}$. However, defining, with the obvious notation, for any $x \in [\![0, \mathfrak{r}-1]\!]$,

$$\widetilde{h}(x) = \mathbb{P}_x(\widetilde{T}^{bd}_{\mathfrak{r}-1} < \widetilde{T}^{bd}), \tag{3.2}$$

we have, according to [13, Theorem 3.1], that $\widetilde{h}$ is a harmonic function for $\widetilde{P}^{bd}$, *i.e.* $\widetilde{P}^{bd}\widetilde{h} = \widetilde{h}$. Hence, a standard result in Martin boundary theory (see *e.g.* [13, Theorem 2.2]) entails that the Markov chain with transition kernel Q, defined on $[\![0, \mathfrak{r}-1]\!] \times [\![0, \mathfrak{r}-1]\!]$ by

$$Q(x,y) = \frac{\widetilde{h}(y)}{\widetilde{h}(x)}\widetilde{P}^{bd}(x,y),$$

is an ergodic birth–death chain, which completes the proof. $\qquad\square$

Remark 3.9. Note that condition (6) in $\mathcal{GMC}^+$ guarantees that $\widetilde{P}$ is a lazy chain, *i.e.* $\widetilde{P}(x,x) \geqslant 1/2$ for all $x \in \mathcal{X}$, and hence the class $\mathcal{GMC}^+$ possesses non-negative eigenvalues.

3.2 Proof of Theorem 3.6

Following the plan as outlined above in Section 3, we first analyse the distribution of the fastest strong stationary time of the class $\mathcal{GMC}^+$ in Lemma 3.10, followed by detailing the proof of Theorem 3.6.

Lemma 3.10. *Suppose that X is an ergodic Markov chain on the state space $\mathcal{X} = [\![0, \mathfrak{r}]\!]$ (and $\mathfrak{r} \geqslant 3$) with transition matrix P and stationary distribution π which starts at 0. If $P \in \mathcal{GMC}^+$,*

then the fastest strong stationary time is distributed as the **c**-mixture of convolution of geometric $\sum_{k=1}^{\mathfrak{r}} c_k \mathcal{G}(\lambda_1, \ldots, \lambda_k)$, where $i, j, k \in [\![0, \mathfrak{r}]\!]$,

$$Q_k := \frac{(P - \lambda_1 I) \cdots (P - \lambda_k I)}{(1 - \lambda_1) \cdots (1 - \lambda_k)}, \quad \Gamma(i, j) := Q_i(0, j), \quad c_k := \frac{\Gamma(k, \mathfrak{r}) - \Gamma(k - 1, \mathfrak{r})}{\pi(\mathfrak{r})}, \quad (3.3)$$

$\{\lambda_k\}_{k=1}^{\mathfrak{r}}$ are the non-unit eigenvalues of P in non-decreasing order and $\mathcal{G}(\lambda_1, \ldots, \lambda_k)$ is the convolution of geometric distributions with success probabilities $1 - \lambda_1, \ldots, 1 - \lambda_k$ respectively.

Remark 3.11. We alert the readers that Q_k are the so-called spectral polynomials and $\Gamma(i, j)$ is $\Lambda(i, j)$ of [25, Theorem 5.2] (since Λ is used as the link kernel throughout this paper).

Remark 3.12 (on the fastest strong stationary time of P and the absorption time of Q). We would like to highlight the connection between the fastest strong stationary time T of P and the absorption time to $\mathfrak{r}$ of Q. According to Lemma 3.10, if $P \in \mathcal{GMC}^+$, then T is distributed as $\sum_{k=1}^{\mathfrak{r}} c_k \mathcal{G}(\lambda_1, \ldots, \lambda_k)$. On the other hand, according to Theorem 3.3, Q is an irreducible birth–death process on $[\![0, \tau - 1]\!]$; then, using [44, Theorem 1.1], there exists a probability measure such that the absorption time to τ of Q starting from 0 is distributed as, in our notations, $\sum_{k=1}^{\mathfrak{r}} a_k \mathcal{G}(\lambda_1, \ldots, \lambda_k)$. In these two distributions, the same eigenvalues appear as parameters in the geometric distributions. See [44] for further connections with the class of phase-type distribution.

Proof. Suppose that $P\Lambda = \Lambda Q$. In view of [25, Theorem 5.2], it suffices to show that $c_k \geqslant 0$. First, we show that $(Q - \lambda_1 I) \cdots (Q - \lambda_k I)$ are non-negative matrices, where Q is the Siegmund dual of $\widehat{P}$. We will prove this via induction on k. For $k = 1$, thanks to [43, Theorem 3.2], we have $Q^{BD} - \lambda_1 I \geqslant 0$, where $Q^{BD} := Q^{[\![0, \mathfrak{r}-1]\!]}$ is the restriction of Q except the last row and column, which leads to

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

Suppose that

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

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

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

which completes the induction by using [43, Theorem 3.2] again on $\prod_{i=1}^{k+1} (Q^{BD} - \lambda_i I)$. Define

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

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

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

which is true. $\qquad \square$

When we have a handle on the fastest strong stationary time, we can then analyse the separation cutoff phenomenon, and the rest of the proof follows the Chebyshev inequality framework introduced by Diaconis and Saloff-Coste [20]. More precisely, let P_n^k be the distribution of the nth chain at time k, π_n the stationary measure, and T_n the fastest strong stationary time of the nth chain. We note that $\mathbb{E}(T_n) = t_n$ and $\mathrm{Var}(T_n) = \rho_n^2$. The key to the proof is the following:

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

where we use $\theta_{n,i} \geqslant 0$, $\underline{\theta}_n/\theta_{n,i} \leqslant 1$ and $w_i \leqslant 1$ in the first inequality. The rest of the proof follows like that of [13, Theorem 8.1], which does not require reversibility of the chain.

Remark 3.13. The corresponding result of Lemma 3.10 in the continuous-time setting is stated as follows in order to prove Theorem 3.8. Suppose that X is a continuous-time ergodic Markov chain on the state space $\mathcal{X} = [\![0, \mathfrak{r}]\!]$ (and $\mathfrak{r} \geqslant 3$) with generator $L = P - I$ and stationary distribution π which starts at 0. If $P \in \mathcal{GMC}^+$, then the fastest strong stationary time is distributed as the **c**-mixture of convolution of exponential $\sum_{k=1}^{\mathfrak{r}} c_k \mathcal{E}(\theta_1, \ldots, \theta_k)$, where $i, j, k \in [\![0, \mathfrak{r}]\!]$,

$$Q_k := \frac{(L+\theta_1 I) \cdots (L+\theta_k I)}{\theta_1 \cdots \theta_k}, \quad \Gamma(i,j) := Q_i(0,j), \quad c_k := \frac{\Gamma(k,\mathfrak{r}) - \Gamma(k-1,\mathfrak{r})}{\pi(\mathfrak{r})}, \quad (3.4)$$

and $\{\theta_k\}_{k=1}^{\mathfrak{r}}$ are the non-zero eigenvalues of $-L$ in non-increasing order and $\mathcal{E}(\theta_1, \ldots, \theta_k)$ is the convolution of exponential distributions with mean $1/\theta_1, \ldots, 1/\theta_k$ respectively.

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

$$\lambda = \lambda(L) = \inf\{\langle -Lf, f\rangle_\pi ; f \in \mathrm{Dom}(L), \text{real-valued}, \mathbb{E}_\pi(f) = 0, \mathbb{E}_\pi(f^2) = 1\}. \quad (4.1)$$

This follows and generalizes the work of Chen, Hsu and Sheu [8] and Chen and Saloff-Coste [9, 10], who studied the L^2-cutoff phenomenon in the context of normal Markov processes. Adapting the notations therein, we proceed to provide a quick review of the notion of L^2-cutoff.

Definition 4.1. For $n \geqslant 1$, let $g_n \colon [0, \infty) \mapsto [0, \infty]$ be a non-increasing function vanishing at infinity. Assume that

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

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

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

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

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

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

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

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

Suppose that we have a family of measurable spaces $(\mathcal{X}_n, \mathcal{B}_n)_{n \in \mathbb{N}}$. For $n \in \mathbb{N}$, we let $p_n(t, \eta_n, \cdot)$, defined on $(\mathcal{X}_n, \mathcal{B}_n)$, be the transition function with initial probability law $\eta_n \ll \pi_n$ and $t \geqslant 0$. We denote f_n to be the L^2-density of η_n with respect to π_n. The family $\{p_n(t, \eta_n, \cdot) : t \in [0, \infty)\}$ has an L^2-cutoff (resp. (t_n, b_n) L^2-cutoff) if $\{g_n(t) = D_{n,2}(\eta_n, t) : n \geqslant 1\}$ has a cutoff (resp. (t_n, b_n)-cutoff) as in Definition 4.1, where $D_{n,2}(\eta_n, t)$ is the chi-squared distance of the nth chain.

Our main result in Theorem 4.2 gives the spectral criterion for L^2-cutoff for the family of processes 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 nth 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, for any $n \in \mathbb{N}$,

$$V_n(B) = \langle H_{n,B} f_n, f_n \rangle_{\pi_n}, \quad t_n(\delta) = \inf\{t : D_{n,2}(\eta_n, t) \leqslant \delta\},$$

$$\lambda_n(C) = \inf\{\lambda : V_n([\lambda_n, \lambda]) > C\}, \quad \tau_n(C) = \sup\left\{\frac{\log(1 + V_n([\lambda_n, \lambda]))}{2\lambda} : \lambda \geqslant \lambda_n(C)\right\},$$

$$\gamma_n = \lambda_n(C)^{-1} \quad \text{and} \quad b_n = \lambda_n(C)^{-1} \log(\lambda_n(C)\tau_n(C)).$$

Theorem 4.2. *Suppose that $L_n \in \mathcal{S}(G_n)$ for each member in the family $\{p_n(t, \eta_n, \cdot) : t \in [0, \infty)\}$, where G_n is a reversible generator. If $\pi_n(f_n^2) \to \infty$, then the following are equivalent:*

(1) *$\{p_n(t, \eta_n, \cdot) : t \in [0, \infty)\}$ has an L^2-cutoff,*
(2) *for some positive constants C, δ, ε,*

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

(3) *for some positive constants C, ε,*

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

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).

Remark 4.3. If L_n is reversible, then $H_{n,B} = F_{n,B} F_{n,B}^* = F_{n,B}^2 = F_{n,B}$ since $F_{n,B}$ is a self-adjoint projection in this case. The above result then retrieves exactly [10, Theorem 4.6].

4.1 Proof of Theorem 4.2

To prove Theorem 4.2 we need the following lemma, which 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 4.4. *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 $\lambda > 0$. Let $\{F_B : B \in \mathcal{B}(\mathbb{C})\}$ denote the non-self-adjoint spectral measure for $-L$, and we define, for $B \in \mathcal{B}(\mathbb{C})$,*

$$H_B = F_B F_B^*.$$

Then, for $t \geqslant 0$,

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

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

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

where the last equality follows from [30, second half of the proof of Lemma 3.19, p. 1542]. □

Lemma 4.4 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 4.2. By Lemma 4.4, 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 criterion in Theorems 3.5 and 3.8 of [10]. Precisely, the chi-squared distance is of the form

$$D_{n,2}(\eta_n, t)^2 = \int_{[\lambda_n, \infty)} e^{-2\gamma t} \, dV_n(\lambda).$$

This is exactly the form of function considered in [10, equation (3.1)], and consequently we can invoke [10, Theorems 3.5 and 3.8]. □

Remark 4.5. As mentioned in Remark 4.3, if L_n is reversible, then our Theorem 4.2 retrieves exactly Theorem 4.6 of [10], whose proof is a combination of the results of Theorems 3.5, 3.8 and 4.4 therein. The strategy of the proof is as follows: Theorem 4.4 claims that the chi-squared distance to stationarity of a reversible Markov process is a Laplace transform, thus the Laplace transform cutoff results of Theorems 3.5 and 3.8 can be applied. In our proof of Theorem 4.2, we follow the same strategy. We first show Lemma 4.1 (which is the parallel version of [10, Theorem 4.4]), which states that for $L \in \mathcal{S}(G)$ with G being reversible, the chi-squared distance to stationarity of L is also a Laplace transform. Consequently, the Laplace transform cutoff results of [10, Theorems 3.5 and 3.8] can be applied in our setting.

4.2 L^p-cutoff

We proceed by investigating the L^p-cutoff for fixed $p \in (1, \infty)$ for the class $\mathcal{S}$. Recall that Chen and Saloff-Coste [9, Theorems 4.2, 4.3] have shown that, for a family of *normal* ergodic transition kernels P_n, the max-L^p cutoff is equivalent to the *spectral gap times mixing time* going to infinity. We can extend their result to the case of the non-normal chains in $\mathcal{S}$ as follows, using techniques similar to those of [13] for the class of skip-free chains similar to birth–death chains.

Theorem 4.6 (max-L^p cutoff). *Suppose that, for each $n \geqslant 1$, $L_n \in \mathcal{S}(G_n)$ with G_n being a reversible generator, transition kernel $P_n^t = e^{tL_n} \overset{\Lambda_n}{\sim} Q_n^t = e^{tG_n}$ on $\mathcal{X}_n$ and spectral gap of G_n given by $\lambda_n = \lambda_n(G_n)$, where we recall the definition of spectral gap in (4.1). Assume that the condition numbers $\kappa(\Lambda_n)$ of the link kernels are uniformly bounded, that is,*

$$\sup_{n \geqslant 1} \kappa(\Lambda_n) < \infty.$$

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

$$f_n(t) = \sup_{x \in \mathcal{X}_n} \left(\int_{\mathcal{X}_n} |p_n(t, \delta_x, y) - 1|^p \, \pi_n(dy) \right)^{1/p}$$

and define

$$t_n = \inf\{t > 0; f_n(t) \leqslant \varepsilon\}, \quad \mathcal{F} = \{f_n; n = 1, 2, \ldots\}.$$

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

The proof in [9, Theorems 4.2, 4.3] works nicely as long as we have Lemma 4.7 below, which gives a two-sided control on the $L^p(\pi)$ norm of $P^t - \pi$. The following lemma is then the key to the proof.

Lemma 4.7. *Suppose that $L \in \mathcal{S}(G)$, with G being a reversible generator, transition kernel $P^t = e^{tL} \overset{\Lambda}{\sim} Q^t = e^{tG}$ and the spectral gap of G is $\lambda = \lambda(G)$, where we recall the definition of spectral gap in (4.1). Fix $p \in (1, \infty)$. Then, for any $t \geqslant 0$, we have*

$$2^{-1+\theta_p} e^{-\lambda t \theta_p} \leqslant \|P^t - \pi\|_{L^p(\pi) \to L^p(\pi)} \leqslant 2^{|1-2/p|} (\kappa(\Lambda) e^{-\lambda t})^{1-|1-2/p|}, \tag{4.2}$$

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

Proof. By the Riesz–Thorin interpolation theorem (see *e.g.* [9, equation 3.4]), we have

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

which when combined with Corollary 2.6 gives the upper bound of (4.2). Next, to show the lower bound in (4.2), we use another version of the Riesz–Thorin interpolation theorem (see *e.g.* [9, Lemma 4.1]) to get

$$\|P^t - \pi\|_{L^p(\pi) \to L^p(\pi)} \geqslant 2^{-1+\theta_p} \|P^t - \pi\|_{L^2(\pi) \to L^2(\pi)}^{\theta_p} \geqslant 2^{-1+\theta_p} e^{-\lambda t \theta_p},$$

where we use Corollary 2.6 in the second inequality. This completes the proof. $\qquad\square$

5. Non-asymptotic estimation error bounds for integral functionals

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

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

where T is a fixed time and f is a function such that the integral $\Gamma_T(f)$ is well-defined. This follows the line of work of [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 functional appears 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 this functional appears in the study of path-dependent derivatives in mathematical finance; see *e.g.* [12]. In practice, however, we often only have access to a sample path of the Markov process at a discrete time point. A natural estimator for $\Gamma_T(f)$, known as the Riemann sum estimator, is given by

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

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

For a stationary Markov process and $f \in L^2(\pi)$, both $\Gamma_T(f)$ and $\hat{\Gamma}_{T,n}(f)$ are π-a.s. defined everywhere in $L^2(\mathbb{P})$. If $L \in \mathcal{S}(G)$, we identify, by the 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 that $H_\lambda^* = F_\lambda^* F_\lambda$ is a self-adjoint operator and F_λ is the spectral measure of $-L$. For $s \geqslant 0$, we define the space $\mathcal{D}^s(A) = \mathrm{Dom}(A^s) \subset L^2(\pi)$ by functional calculus on A with the seminorm $\|f\|_{\mathcal{D}^s(A)} = \|A^{s/2} f\|_\pi$.

The main results are the following error bounds, in which the proof is similar to that of [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 $T \geqslant 0$, $0 \leqslant s \leqslant 1$, $f \in \mathcal{D}^s(A)$, $f_0 \in \mathrm{Dom}(A^{-1})$ with $f_0 = f - \int f \, d\pi$,*

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

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

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

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

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

Remark 5.2. When L is reversible, then A can be identified as $|L|^2$, where we can then retrieve the results of [2].

5.1 Proof of Theorem 5.1

We first state a lemma (see [30, first half of the proof of Lemma 3.19, p. 1542]) that will be used repeatedly in the proof.

Lemma 5.3. *For $f \in \mathcal{D}^s(A)$,*

$$\left| \int_{\sigma(L)} \lambda^s \, d\langle F_\lambda f, f \rangle_\pi \right| \leqslant \left(\int_{\sigma(L)} |\lambda|^{2s} \, d\langle H_\lambda^* f, f \rangle_\pi \right)^{1/2} \|f\|_\pi = \|f\|_{\mathcal{D}^s(A)} \|f\|_\pi. \tag{5.4}$$

Proof. For the sake of completeness, we repeat the arguments of [30, first half of the proof of Lemma 3.19, p. 1542]. Let $[\alpha, \beta]$ be a bounded interval and let $(\Delta_k)_{k=1}^n$ be a family of disjoint intervals whose union is $[\alpha, \beta]$. For every k, we choose $\lambda_k \in \Delta_k$. Fix $f \in L^2(\pi)$. For the Cauchy

sums defining the integrals using the triangle inequality and the Cauchy–Schwarz inequality, we have

$$\left| \sum_{k=1}^{n} \lambda_k^s \langle F_{\Delta_k} f, f \rangle_\pi \right| \leqslant \sum_{k=1}^{n} |\lambda_k|^s |\langle F_{\Delta_k} f, F_{\Delta_k}^* f \rangle_\pi|$$

$$\leqslant \left(\sum_{k=1}^{n} |\lambda_k|^{2s} \langle F_{\Delta_k} f, F_{\Delta_k} f \rangle_\pi \right)^{1/2} \left(\sum_{k=1}^{n} \langle F_{\Delta_k}^* f, F_{\Delta_k}^* f \rangle_\pi \right)^{1/2}$$

$$\leqslant \left(\sum_{k=1}^{n} |\lambda_k|^{2s} \langle H_{\Delta_k}^* f, f \rangle_\pi \right)^{1/2} \| f \|_\pi.$$

The inequality (5.4) holds on every finite interval and the desired result follows by taking limits. □

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

$$\| \Gamma_T(f) - \hat{\Gamma}_{T,n}(f) \|_{L^2(\mathbb{P})}^2$$

$$= \mathbb{E}\left[\left(\sum_{k=1}^{n} \int_{(k-1)\Delta_n}^{k\Delta_n} (f(X_r) - f(X_{(k-1)\Delta_n})) \, dr \right)^2 \right]$$

$$= \sum_{k,l=1}^{n} \int_{(k-1)\Delta_n}^{k\Delta_n} \int_{(l-1)\Delta_n}^{l\Delta_n} \mathbb{E}[(f(X_r) - f(X_{(k-1)\Delta_n}))(f(X_h) - f(X_{(l-1)\Delta_n}))] \, dr \, dh.$$

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

$$\mathbb{E}[(f(X_r) - f(X_{(k-1)\Delta_n}))(f(X_h) - f(X_{(k-1)\Delta_n}))]$$

$$= \langle (P_{h-r} - I)f + (I - P_{h-(k-1)\Delta_n})f + (I - P_{r-(k-1)\Delta_n})f, f \rangle_\pi,$$

so by symmetry in r and h we have

$$\sum_{k=1}^{n} \int_{(k-1)\Delta_n}^{k\Delta_n} \int_{(k-1)\Delta_n}^{k\Delta_n} \mathbb{E}[(f(X_r) - f(X_{(k-1)\Delta_n}))(f(X_h) - f(X_{(k-1)\Delta_n}))] \, dr \, dh$$

$$= 2n \left\langle \left(\int_0^{\Delta_n} \int_0^h (P_{h-r} - I) \, dr \, dh + \Delta_n \int_0^{\Delta_n} (I - P_h) \, dh \right) f, f \right\rangle_\pi$$

$$= \langle \Phi(L) f, f \rangle_\pi$$

$$= \int_{\sigma(L)} \Phi(\lambda) \, d\langle F_\lambda f, f \rangle_\pi,$$

where the last equality 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 [2, p. 15] we have $|\Phi(\lambda)| \leqslant 4n\Delta_n^{2+s} |\lambda|^s$ with fixed $0 \leqslant s \leqslant 1$. Now we apply Lemma 5.3 to arrive at

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

Next we bound the off-diagonal terms, in which

$$2\sum_{k>l}^{n}\int_{(k-1)\Delta_n}^{k\Delta_n}\int_{(l-1)\Delta_n}^{l\Delta_n}\mathbb{E}[(f(X_r)-f(X_{(k-1)\Delta_n}))(f(X_h)-f(X_{(l-1)\Delta_n}))]\,dr\,dh$$

$$=2\left\langle\left(\int_0^{\Delta_n}\int_0^{\Delta_n}\left(\sum_{k>l=1}^{n}P_{(k-l)\Delta_n-r}\right)(P_h-I)(I-P_r)\,dr\,dh\right)f,f\right\rangle_\pi$$

$$=\langle\tilde{\Phi}(L)f,f\rangle_\pi$$

$$=\int_{\sigma(L)}\tilde{\Phi}(\lambda)\,d\langle F_\lambda f,f\rangle_\pi,$$

where the last equality again follows 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 [2, (16)], there exists a universal constant $\tilde{C}<\infty$ such that $|\tilde{\Phi}(\lambda)|\leqslant\tilde{C}T\Delta_n^{1+s}|\lambda|^s$, and together with Lemma 5.3 yield

$$\left|\int_{\sigma(L)}\tilde{\Phi}(\lambda)\,d\langle F_\lambda f,f\rangle_\pi\right|$$

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

$$=\tilde{C}T\Delta_n^{1+s}\|f\|_\pi\|f\|_{\mathcal{D}^s(A)}.$$

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

$$\left\|T^{-1}\hat{\Gamma}_{T,n}(f)-\int f\,d\pi\right\|_{\mathrm{L}^2(\mathbb{P})}\leqslant T^{-1}\|\hat{\Gamma}_{T,n}(f)-\Gamma_T(f)\|_{\mathrm{L}^2(\mathbb{P})}+\left\|T^{-1}\Gamma_T(f)-\int f\,d\pi\right\|_{\mathrm{L}^2(\mathbb{P})}$$

$$\leqslant\frac{C}{\sqrt{T}}\sqrt{\|f\|_{\mathcal{D}^s(A)}\|f\|_\pi\Delta_n}+\|T^{-1}\Gamma_T(f_0)\|_{\mathrm{L}^2(\mathbb{P})}.$$

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

$$\|T^{-1}\Gamma_T(f_0)\|_{\mathrm{L}^2(\mathbb{P})}^2=2T^{-2}\int_0^T\int_0^h\langle P_{h-r}f_0,f_0\rangle_\pi\,dr\,dh$$

$$=\int_{\sigma(L)}\overline{\Phi}(\lambda)\,d\langle F_\lambda f_0,f_0\rangle_\pi,$$

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

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

and there exists a constant $\tilde{C}$ such that $|\overline{\Phi}(\lambda)| \leqslant \tilde{C}/(|\lambda|T)$. Using Lemma 5.3 gives

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

Finally, (5.3) follows via a standard change of measure argument.   □

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 as this result generalizes easily to this setting; see Remarks 1.3 and 2.5. More precisely, suppose we start with a reversible generator G with transition semigroup $(Q_t)_{t \geqslant 0}$; we would like to characterize the family of Markov chains with generator L associated with G under the similarity transformation $G\Lambda = \Lambda L$ with Λ being a bounded invertible Markov link. This idea allows us to generate Markov or contraction kernels 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 pure birth link in particular. While we consider univariate examples in the subsequent section, nonetheless we can still handle the orbits of multivariate reversible Markov chains (*e.g.* [28, 34, 35] and [62]) by considering the link kernel to be the tensor product from univariate link and analyse the corresponding tensorized orbits.

Before detailing the examples, we introduce the following notation, which will be used throughout. Let G be a reversible birth–death generator with respect to π_G on $\mathcal{X} = [\![0, \mathfrak{r}]\!]$. Let

$$
G(x, x-1) = \mathsf{d}_x, \quad G(x, x) = -(\mathsf{d}_x + \mathsf{b}_x) \quad \text{and} \quad G(x, x+1) = \mathsf{b}_x, \tag{6.1}
$$

where d_x (resp. b_x) is the death (resp. birth) rate at state x, and eigenvalues/eigenvectors denoted by $(-\lambda_j, \phi_j)_{j=0}^N$, where ϕ_j are orthonormal in $l^2(\pi_G)$. We assume that $\mathsf{d}_0 = \mathsf{b}_\mathfrak{r} = 0$. Write $(Q_t)_{t \geqslant 0}$ for its transition semigroup. Then the spectral decomposition of Q_t is given by

$$
Q_t(x, y) = \sum_{j=0}^{\mathfrak{r}} e^{-\lambda_j t} \phi_j(x) \phi_j(y) \pi_G(y). \tag{6.2}
$$

For further details on various birth–death models and their connections with orthogonal polynomials, we refer interested readers to [18, 32, 36, 55, 56, 62] and the references therein.

6.1 Pure birth link on finite state space

In this section we specialize to the case $\mathcal{X} = [\![0, \mathfrak{r}]\!]$, with the link being the pure birth link, as introduced by Fill [25] to study the distribution of hitting time and fastest strong stationary time, generated by birth–death processes with birth and death rates being b_x and d_x respectively. The particular pure birth link Λ_{pb} that we study is of the form

$$
\Lambda_{pb}(x, y) = 1/2, \quad x \in [\![0, \mathfrak{r}-1]\!], \ y \in \{x, x+1\}, \ \Lambda_{pb}(\mathfrak{r}, \mathfrak{r}) = 1, \tag{6.3}
$$
$$
\Lambda_{pb}(x, y) = 0 \quad \text{otherwise.} \tag{6.4}
$$

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 6.1 below. However, we give sufficient conditions on a birth–death generator G to guarantee L is a Markov generator.

Proposition 6.1. *Suppose that $G \overset{\Lambda_{pb}}{\sim} L$ and let $(P_t)_{t \geqslant 0}$ denote the transition semigroup associated with L. Note that $(P_t)_{t \geqslant 0}$ need not be Markov under Λ_{pb}. For any $t \geqslant 0$ and $j, x, y \in [\![0, \mathfrak{r}]\!]$, P_t admits the following spectral decomposition:*

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

where

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

$$f_j(x) = \sum_{k=x}^{\mathfrak{r}-1}(-2)^{k-x}\phi_j(k) + \phi_j(\mathfrak{r}),$$

and

$$\|P_t - \pi_L\|_{TV} \leqslant \frac{\kappa(\Lambda_{pb})\,e^{-\lambda_1 t}}{2}\sqrt{\frac{1-\pi_L^*}{\pi_L^*}}, \tag{6.5}$$

where

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

and recall that $\pi_L^* = \min_{y\in[\![0,\mathfrak{r}]\!]} \pi_L(y)$.

Moreover, note that for all $x \in [\![0, \mathfrak{r}]\!]$,

$$L(x, x) = -b_x - d_{\max\{x+1,\mathfrak{r}\}}\left(\frac{\mathbb{1}_{x+1\geqslant\mathfrak{r}}}{2} + \mathbb{1}_{x+1<\mathfrak{r}}\right) < 0,$$

for $x > y + 1$, $L(x, y) = 0$ and for $x < \mathfrak{r}$

$$L(x + 1, x) = (2\mathbb{1}_{x+1\neq\mathfrak{r}} + \mathbb{1}_{x+1=\mathfrak{r}})\frac{d_{x+1}}{2} > 0.$$

If $\mathfrak{r} \geqslant 4$, for $y \in [\![1, \mathfrak{r} - 1]\!]$,

$$L(y - 1, y) = -d_{y-1} + b_y + d_{y+1}\mathbb{1}_{y<\mathfrak{r}-1} + \frac{d_{y+1}}{2}\mathbb{1}_{y+1=\mathfrak{r}} \geqslant 0 \quad and$$

$$L(\mathfrak{r} - 1, \mathfrak{r}) = -d_{\mathfrak{r}-1} + b_{\mathfrak{r}-1} + d_{\mathfrak{r}} \geqslant 0,$$

for $x \in [\![0, y - 2]\!]$ and $y \in [\![2, \mathfrak{r} - 1]\!]$,

$$L(x, y) = (-1)^{x+y}\left(b_{y-2} + d_{y-1} - b_y - d_{y+1} + \frac{d_{\mathfrak{r}}}{2}\mathbb{1}_{y=\mathfrak{r}-1}\right) \geqslant 0,$$

and for $x \in [\![0, \mathfrak{r} - 2]\!]$,

$$L(x, \mathfrak{r}) = (-1)^{x+\mathfrak{r}}\left(b_{\mathfrak{r}-2} + d_{\mathfrak{r}-1} - b_{\mathfrak{r}-1} - \frac{d_{\mathfrak{r}}}{2}\right) \geqslant 0,$$

and then L is a Markov generator.

Remark 6.2. 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. We first observe that the inverse of Λ_{pb} is given by

$$\Lambda_{pb}^{-1}(x, y) = (-1)^{y-x}(2\mathbb{1}_{y \neq \mathfrak{r}} + \mathbb{1}_{y=\mathfrak{r}}) \quad \text{for } x \leqslant y, \ x, y \in [\![0, \mathfrak{r}]\!], \tag{6.6}$$

$$\Lambda_{pb}^{-1}(x, y) = 0 \quad \text{otherwise.} \tag{6.7}$$

Upon expanding $P_t = \Lambda_{pb}^{-1} Q_t \Lambda_{pb}$, we get

$$P_t(x, y) = \sum_{k=x}^{\mathfrak{r}} (-1)^{k-x}(2\mathbb{1}_{k \neq \mathfrak{r}} + \mathbb{1}_{k=\mathfrak{r}}) \left(Q_t(k, y-1)\left(\frac{\mathbb{1}_{y-1 \geqslant 0}}{2}\right) + Q_t(k, y)\left(\frac{\mathbb{1}_{y \neq \mathfrak{r}}}{2} + \mathbb{1}_{y=\mathfrak{r}}\right) \right)$$

$$= \sum_{j=0}^{\mathfrak{r}} e^{-\lambda_j t} \left(\sum_{k=x}^{\mathfrak{r}} (-1)^{k-x}(2\mathbb{1}_{k \neq \mathfrak{r}} + \mathbb{1}_{k=\mathfrak{r}})\phi_j(k) \right)$$

$$\times \left(\phi_j(y-1)\pi_G(y-1)\left(\frac{\mathbb{1}_{y-1 \geqslant 0}}{2}\right) + \phi_j(y)\pi_G(y)\left(\frac{\mathbb{1}_{y \neq \mathfrak{r}}}{2} + \mathbb{1}_{y=\mathfrak{r}}\right) \right),$$

where the second equality follows from substituting the spectral expansion of $(Q_t)_{t \geqslant 0}$. The bound (6.5) follows directly from Corollary 2.6. To show that L is a Markov generator under the proposed conditions on birth and death rates, we need to impose sufficient conditions such that $L(x, x) < 0$ for all $x \in \mathcal{X}$ and $L(x, y) \geqslant 0$ for all $x \neq y \in \mathcal{X}$; see *e.g.* [41, Chapter 20]. We proceed by calculating $G\Lambda_{pb}$, and the entries not mentioned below are all zero. We have

$$2G\Lambda_{pb}(x, x-1) = \mathsf{d}_x, \quad x \in [\![1, \mathfrak{r}]\!],$$

$$2G\Lambda_{pb}(x, x) = -\mathsf{b}_x - \mathsf{d}_\mathfrak{r}\mathbb{1}_{x=\mathfrak{r}}, \quad x \in [\![0, \mathfrak{r}]\!],$$

$$2G\Lambda_{pb}(x, x+1) = -\mathsf{d}_x + \mathsf{b}_{\mathfrak{r}-1}\mathbb{1}_{x=\mathfrak{r}-1}, \quad x \in [\![0, \mathfrak{r}-1]\!],$$

$$2G\Lambda_{pb}(x, x+2) = \mathsf{b}_x, \quad x \in [\![0, \mathfrak{r}-2]\!].$$

Using the form of $G\Lambda_{pb}$ described above, we first note that $L(x, x) < 0$ is automatically satisfied since

$$L(x, x) = -\mathsf{b}_x - \mathsf{d}_{\max\{x+1, \mathfrak{r}\}}\left(\frac{1}{2}\mathbb{1}_{x+1 \geqslant \mathfrak{r}} + \mathbb{1}_{x+1 < \mathfrak{r}}\right) < 0.$$

It remains to check $L(x, y) \geqslant 0$ for all $x \neq y$. Indeed, we have

$$L(x, y) = \sum_{k=\max\{x, y-2\}}^{\min\{y+1, \mathfrak{r}\}} (-1)^{k-x}(2\mathbb{1}_{k \neq \mathfrak{r}} + \mathbb{1}_{k=\mathfrak{r}})G\Lambda_{pb}(k, y).$$

For $x > y+1$, $L(x, y) = 0$. For $x = y+1$,

$$L(y+1, y) = (2\mathbb{1}_{y+1 \neq \mathfrak{r}} + \mathbb{1}_{y+1=\mathfrak{r}})\frac{1}{2}\mathsf{d}_{y+1} > 0.$$

Thus, it boils down to checking $L(x, y) \geqslant 0$ for $x \in [\![0, y-1]\!]$. For $y \in [\![1, \mathfrak{r}-1]\!]$ and $x = y-1$,

$$L(x, y) = -\mathsf{d}_{y-1} + \mathsf{b}_y + \mathsf{d}_{y+1}\mathbb{1}_{y < \mathfrak{r}-1} + \frac{\mathsf{d}_{y+1}}{2}\mathbb{1}_{y+1=\mathfrak{r}}.$$

For $y = \mathfrak{r}$ and $x = \mathfrak{r} - 1$,

$$L(\mathfrak{r}-1, \mathfrak{r}) = 2G\Lambda_{pb}(\mathfrak{r}-1, \mathfrak{r}) - G\Lambda_{pb}(\mathfrak{r}, \mathfrak{r}) = -\mathsf{d}_{\mathfrak{r}-1} + \mathsf{b}_{\mathfrak{r}-1} + \mathsf{d}_\mathfrak{r}.$$

For $y \in [\![2, \mathfrak{r} - 1]\!]$ and $x \in [\![0, y - 2]\!]$, since $\mathfrak{r} \geqslant 4$,

$$L(x, y) = (-1)^{x+y}(2G\Lambda_{pb}(y - 2, y) - 2G\Lambda_{pb}(y - 1, y) + 2G\Lambda_{pb}(y, y) - 2G\Lambda_{pb}(y + 1, y))$$

$$= (-1)^{x+y}\left(b_{y-2} + d_{y-1} - b_y - d_{y+1} + \frac{d_{\mathfrak{r}}}{2}\mathbb{1}_{y=\mathfrak{r}-1}\right),$$

and for $y = \mathfrak{r}$ and $x \in [\![0, y - 2]\!]$,

$$L(x, \mathfrak{r}) = (-1)^{x+\mathfrak{r}}(2G\Lambda_{pb}(\mathfrak{r} - 2, \mathfrak{r}) - 2G\Lambda_{pb}(\mathfrak{r} - 1, \mathfrak{r}) + G\Lambda_{pb}(\mathfrak{r}, \mathfrak{r}))$$

$$= (-1)^{x+\mathfrak{r}}\left(b_{\mathfrak{r}-2} + d_{\mathfrak{r}-1} - b_{\mathfrak{r}-1} - \frac{d_{\mathfrak{r}}}{2}\right). \qquad \square$$

Example 6.3. The pair

$$G = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 \\ 0.5 & -1 & 0.5 & 0 & 0 \\ 0 & 0.5 & -1 & 0.5 & 0 \\ 0 & 0 & 0.5 & -1 & 0.5 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}, \quad L = \begin{pmatrix} -1.5 & 1 & 0.5 & 0 & 0 \\ 0.5 & -1 & 0.5 & 0 & 0 \\ 0 & 0.5 & -1 & 0.5 & 0 \\ 0 & 0 & 0.5 & -1 & 0.5 \\ 0 & 0 & 0 & 0.5 & -0.5 \end{pmatrix}$$

satisfy the assumption Proposition 6.1, where L is a non-reversible Markov generator since $\pi_L = (0.0625, 0.1875, 0.25, 0.25, 0.25)$ and $\pi_L(0)L(0, 1) \neq \pi_L(1)L(1, 0)$.

Example 6.4. (pure birth variants of constant-rate birth–death processes with reflection at 0 and $\mathfrak{r}$). A more general example is that $b_x = d_x = \lambda$ for $x \in [\![1, \mathfrak{r} - 1]\!]$ and $b_0 = d_{\mathfrak{r}} = 2\lambda$ for some $\lambda > 0$. The stationary distribution π_G is $\pi_G(x) = 1/\mathfrak{r}$ for $x \in [\![1, \mathfrak{r} - 1]\!]$ and $\pi_G(0) = \pi_G(\mathfrak{r}) = 1/(2\mathfrak{r})$, and the associated eigenvalues and orthogonal polynomials are, for $j, x \in [\![0, \mathfrak{r}]\!]$,

$$\lambda_j = 2\lambda(1 - \cos(\theta_j))$$
$$\phi_j(x) = \cos(\theta_j x + c),$$

where $(\theta_j)_{j=0}^{\mathfrak{r}}$ and c are determined by the boundary values $\cos(\theta x + c) = \cos(\theta)\cos(c)$ and $\cos(\theta(N - 1) + c) = \cos(\theta)\cos(\theta N + c)$ and are arranged such that $(\cos(\theta_j))_{j=0}^{\mathfrak{r}}$ is in non-increasing order; see [19, Proposition 22] and [62]. We proceed to check that the conditions in Proposition 6.1 are fulfilled: for $y \in [\![1, \mathfrak{r} - 1]\!]$,

$$L(y - 1, y) = -d_{y-1} + b_y + d_{y+1}\mathbb{1}_{y<\mathfrak{r}-1} + \frac{d_{y+1}}{2}\mathbb{1}_{y+1=\mathfrak{r}} = \lambda\mathbb{1}_{y<\mathfrak{r}-1} + \frac{\lambda}{2}\mathbb{1}_{y+1=\mathfrak{r}} \geqslant 0,$$

and

$$L(\mathfrak{r} - 1, \mathfrak{r}) = -d_{\mathfrak{r}-1} + b_{\mathfrak{r}-1} + d_{\mathfrak{r}} = 2\lambda \geqslant 0.$$

For $x = 0$ and $y = 2$,

$$L(0, 2) = \lambda,$$

and otherwise for $y \in [\![3, \mathfrak{r} - 1]\!]$ and $x \in [\![0, y - 2]\!]$,

$$(-1)^{x+y}\left(b_{y-2} + d_{y-1} - b_y - d_{y+1} + \frac{d_{\mathfrak{r}}}{2}\mathbb{1}_{y=\mathfrak{r}-1}\right) = 0 \geqslant 0,$$

and for $y = \mathfrak{r}$ and $x \in [\![0, y - 2]\!]$,

$$(-1)^{x+\mathfrak{r}}\left(b_{\mathfrak{r}-2} + d_{\mathfrak{r}-1} - b_{\mathfrak{r}-1} - \frac{d_{\mathfrak{r}}}{2}\right) = 0 \geqslant 0,$$

so L is a Markov generator, with spectral decomposition given by

$$P_t(x, y) = \sum_{j=0}^{\mathfrak{r}} e^{-2\lambda(1-\cos(\theta_j))t} f_j(x) f_j^*(y) \pi_L(y),$$

$$\|P_t - \pi_L\|_{TV} \leqslant \frac{\kappa(\Lambda_{pb}) \, e^{-2\lambda(1-\cos(\theta_1))t}}{2} \sqrt{\frac{1 - \pi_L^*}{\pi_L^*}} = O(e^{-2\lambda(1-\cos(\theta_1))t}),$$

where

$$\pi_L^* = \frac{1}{4\mathfrak{r}},$$

$$f_j(x) = \sum_{k=x}^{\mathfrak{r}} (-1)^{k-x} (2\mathbb{1}_{k \neq \mathfrak{r}} + \mathbb{1}_{k=\mathfrak{r}}) \cos(\theta_j k + c),$$

$$f_j^*(y) \pi_L(y) = \cos(\theta_j(y-1) + c) \pi_G(y-1) \left(\frac{\mathbb{1}_{y-1 \geqslant 0}}{2} \right) + \cos(\theta_j y + c) \pi_G(y) \left(\frac{\mathbb{1}_{y \neq \mathfrak{r}}}{2} + \mathbb{1}_{y=\mathfrak{r}} \right).$$

Acknowledgements

References

[1] Aldous, D. and Fill, J. A. (2002) Reversible Markov chains and random walks on graphs. Unfinished monograph, recompiled 2014. http://www.stat.berkeley.edu/~aldous/RWG/book.html

[2] Altmeyer, R. and Chorowski, J. (2018) Estimation error for occupation time functionals of stationary Markov processes. *Stoch. Process. Appl.* **128** 1830–1848.

[3] Antoine, J.-P. and Trapani, C. (2013) Partial inner product spaces, metric operators and generalized Hermiticity. *J. Phys. A* **46** 025204.

[4] Asmussen, S. (2003) *Applied Probability and Queues*, second edition, Vol. 51 of Applications of Mathematics (New York): Stochastic Modelling and Applied Probability, Springer.

[5] Berman, A. and Plemmons, R. J. (1974) Matrix group monotonicity. *Proc. Amer. Math. Soc.* **46** 355–359.

[6] Bierkens, J. (2016) Non-reversible Metropolis–Hastings. *Statist. Comput.* **26** 1213–1228.

[7] Chafaï, D. and Joulin, A. (2013) Intertwining and commutation relations for birth–death processes. *Bernoulli* **19** 1855–1879.

[8] Chen, G.-Y., Hsu, J.-M. and Sheu, Y.-C. (2017) The L2-cutoffs for reversible Markov chains. *Ann. Appl. Probab.* **4** 2305–2341.

[9] Chen, G.-Y. and Saloff-Coste, L. (2008) The cutoff phenomenon for ergodic Markov processes. *Electron. J. Probab.* **13** 26–78.

[10] Chen, G.-Y. and Saloff-Coste, L. (2010) The L^2-cutoff for reversible Markov processes. *J. Funct. Anal.* **258** 2246–2315.

[11] Chen, G.-Y. and Saloff-Coste, L. (2015) Computing cutoff times of birth and death chains. *Electron. J. Probab.* **20** 76.

[12] Chesney, M., Jeanblanc-Picqué, M. and Yor, M. (1997) Brownian excursions and Parisian barrier options. *Adv. Appl. Probab.* **29** 165–184.

[13] Choi, M. and Patie, P. (2019) Skip-free Markov chains. *Trans. Amer. Math. Soc.* **371** 7301–7342.

[14] Clifford, P. and Sudbury, A. (1985) A sample path proof of the duality for stochastically monotone Markov processes. *Ann. Probab.* **13** 558–565.

[15] Cloez, B. and Delplancke, C. (2019) Intertwinings and Stein's magic factors for birth-death processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **55** 341–377.

[16] Cui, H. and Mao, Y.-H. (2010) Eigentime identity for asymmetric finite Markov chains. *Front. Math. China* **5** 623–634.

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

[18] Diaconis, P., Khare, K. and Saloff-Coste, L. (2008) Gibbs sampling, exponential families and orthogonal polynomials. *Statist. Sci.* **23** 151–178.

[19] Diaconis, P. and Miclo, L. (2015) On quantitative convergence to quasi-stationarity. *Ann. Fac. Sci. Toulouse Math. (6)* **24** 973–1016.

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

[21] Dunford, N. (1954) Spectral operators. *Pacific J. Math.* **4** 321–354.

[22] Dunford, N. and Schwartz, J. T. (1971) *Linear Operators, III: Spectral Operators*, Interscience (Wiley).

[23] Dym, H. and McKean, H. P. (1976) *Gaussian Processes, Function Theory, and the Inverse Spectral Problem*, Vol. 31 of Probability and Mathematical Statistics, Academic Press (Harcourt Brace Jovanovich).

[24] Fill, J. A. (1991) Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process. *Ann. Appl. Probab.* **1** 62–87.

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

[26] Fort, G., Meyn, S., Moulines, E. and Priouret, P. (2008) The ODE method for stability of skip-free Markov chains with applications to MCMC. *Ann. Appl. Probab.* **18** 664–707.

[27] Friedland, S. and Melkman, A. A. (1979) On the eigenvalues of nonnegative Jacobi matrices. *Linear Algebra Appl.* **25** 239–253.

[28] Griffiths, R. (2016) Multivariate Krawtchouk polynomials and composition birth and death processes. *Symmetry* **8** 33.

[29] Huillet, T. and Martinez, S. (2011) Duality and intertwining for discrete Markov kernels: Relations and examples. *Adv. Appl. Probab.* **43** 437–460.

[30] Inoue, A. and Trapani, C. (2014) Non-self-adjoint resolutions of the identity and associated operators. *Complex Anal. Oper. Theory* **8** 1531–1546.

[31] Jansen, S. and Kurt, N. (2014) On the notion(s) of duality for Markov processes. *Probab. Surv.* **11** 59–120.

[32] Karlin, S. and McGregor, J. (1959) Coincidence properties of birth and death processes. *Pacific J. Math.* **9** 1109–1140.

[33] Kendall, D. G. (1959) 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.

[34] Khare, K. and Mukherjee, N. (2013) Convergence analysis of some multivariate Markov chains using stochastic monotonicity. *Ann. Appl. Probab.* **23** 811–833.

[35] Khare, K. and Zhou, H. (2009) Rates of convergence of some multivariate Markov chains with polynomial eigenfunctions. *Ann. Appl. Probab.* **19** 737–777.

[36] Koekoek, R. and Swarttouw, R. F. (1998) The Askey-scheme of hypergeometric orthogonal polynomials and its *q*-analogue. arXiv:9602214.

[37] Kontoyiannis, I. and Meyn, S. P. (2003) Spectral theory and limit theorems for geometrically ergodic Markov processes. *Ann. Appl. Probab.* **13** 304–362.

[38] Kontoyiannis, I. and Meyn, S. P. (2005) Large deviations asymptotics and the spectral theory of multiplicatively regular Markov processes. *Electron. J. Probab.* **10** 61–123.

[39] Kontoyiannis, I. and Meyn, S. P. (2012) Geometric ergodicity and the spectral gap of non-reversible Markov chains. *Probab. Theory Related Fields* **154** 327–339.

[40] Ledermann, W. and Reuter, G. E. H. (1954) Spectral theory for the differential equations of simple birth and death processes. *Philos. Trans. Roy. Soc. London. Ser. A.* **246** 321–369.

[41] Levin, D. A., Peres, Y. and Wilmer, E. L. (2009) *Markov Chains and Mixing Times*, American Mathematical Society.

[42] Mao, Y. H., Zhang, C. and Zhang, Y. H. (2016) Separation cutoff for upward skip-free chains. *J. Appl. Probab.* **53** 299–306.

[43] Micchelli, C. A. and Willoughby, R. A. (1979) On functions which preserve the class of Stieltjes matrices. *Linear Algebra Appl.* **23** 141–156.

[44] Miclo, L. (2010) On absorption times and Dirichlet eigenvalues. *ESAIM Probab. Statist.* **14** 117–150.

[45] Miclo, L. (2015) An absorbing eigentime identity. *Markov Process. Related Fields* **21** 249–262.

[46] Miclo, L. (2019) On the Markov commutator. *Bull. Sci. Math.* 154 1–35.

[47] Miclo, L. (2018) On the Markovian similarity. Séminaire de Probabilités XLIX 375–403.

[48] Montenegro, R. and Tetali, P. (2006) Mathematical aspects of mixing times in Markov chains. *Found. Trends Theor. Comput. Sci.* **1** 237–354.

[49] Patie, P. and Savov, M. Spectral expansion of non-self-adjoint generalized Laguerre semigroups. To appear in *Mem. Amer. Math. Soc.*

[50] Patie, P. and Zhao, Y. (2017) Spectral decomposition of fractional operators and a reflected stable semigroup. *J. Diff. Equations* **262** 1690–1719.

[51] Paulin, D. (2015) Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electron. J. Probab.* **20** 79.

[52] Peres, Y. (2004) American Institute of Mathematics (AIM) research workshop 'Sharp Thresholds for Mixing Times' (Palo Alto, December 2004). Summary available at http://www.aimath.org/WWN/mixingtimes.

[53] Pike, J. (2013) Eigenfunctions for random walks on hyperplane arrangements. PhD thesis, University of Southern California.

[54] Rosenthal, J. S. and Rosenthal, P. (2015) Spectral bounds for certain two-factor non-reversible MCMC algorithms. *Electron. Commun. Probab.* **20** 91.

[55] Sasaki, R. (2009) Exactly solvable birth and death processes. *J. Math. Phys.* **50** 103509.

[56] Schoutens, W. (2000) *Stochastic Processes and Orthogonal Polynomials*, Vol. 146 of Lecture Notes in Statistics, Springer.

[57] Siegmund, D. (1976) The equivalence of absorbing and reflecting barrier problems for stochastically monotone Markov processes. *Ann. Probab.* **4** 914–924.

[58] Thompson, R. C. (1977) Singular values, diagonal elements, and convexity. *SIAM J. Appl. Math.* **32** 39–63.

[59] Villani, C. (2009) *Hypocoercivity*, Vol. 202, no. 950, of Memoirs of the American Mathematical Society, American Mathematical Society.

[60] Wermer, J. (1954) Commuting spectral measures on Hilbert space. *Pacific J. Math.* **4** 355–361.

[61] Young, R. M. (2001) *An Introduction to Nonharmonic Fourier Series*, first edition, Academic Press.

[62] Zhou, H. (2008) Examples of multivariate Markov chains with orthogonal polynomial eigenfunctions. PhD thesis, Stanford University.