SPECTRAL COMPUTATIONS FOR BIRTH AND DEATH CHAINS

GUAN-YU CHEN[1] AND LAURENT SALOFF-COSTE[2]

Abstract. We consider the spectrum of birth and death chains on a n-path. An iterative scheme is proposed to compute any eigenvalue with exponential convergence rate independent of n. This allows one to determine the whole spectrum in order n^2 elementary operations. Using the same idea, we also provide a lower bound on the spectral gap, which is of the correct order on some classes of examples.

1. Introduction

Let $G=(V,E)$ be the undirected finite path with vertex set $V=\{1,2,...n\}$ and edge set $E=\{\{i,i+1\}:i=1,2,...,n-1\}$. Given two positive measures π,ν on V,E with $\pi(V)=1$, the Dirichlet form and variance associated with ν and π are defined by

$$\mathcal{E}_\nu(f,g):=\sum_{i=1}^{n-1}[f(i)-f(i+1)][g(i)-g(i+1)]\nu(i,i+1)$$

and

$$\mathrm{Var}_\pi(f):=\pi(f^2)-\pi(f)^2,$$

where f,g are functions on V. When convenient, we set $\nu(0,1)=\nu(n,n+1)=0$. The spectral gap of G with respect to π,ν is defined as

$$\lambda_{\pi,\nu}^G:=\min\left\{\frac{\mathcal{E}_\nu(f,f)}{\mathrm{Var}_\pi(f)}\middle| f\text{ is non-constant}\right\}.$$

Let $M_{\pi,\nu}^G$ be a matrix given by $M_{\pi,\nu}^G(i,j)=0$ for $|i-j|>1$ and

$$M_{\pi,\nu}^G(i,j)=-\frac{\nu(i,j)}{\pi(i)},\ \forall|i-j|=1,\quad M_{\pi,\nu}^G(i,i)=\frac{\nu(i-1,i)+\nu(i,i+1)}{\pi(i)}.$$

Obviously, $\lambda_{\pi,\nu}^G$ is the smallest non-zero eigenvalue of $M_{\pi,\nu}^G$.

Undirected paths equipped with measures π,ν are closely related to birth and death chains. A birth and death chain on $\{0,1,2,...,n\}$ with birth rate p_i, death rate q_i and holding rate r_i is a Markov chain with transition matrix K given by

$$K(i,i+1)=p_i,\quad K(i,i-1)=q_i,\quad K(i,i)=r_i,\quad \forall 0\le i\le n, \tag{1.1}$$

where $p_i+q_i+r_i=1$ and $p_n=q_0=0$. Under the assumption of irreducibility, that is, $p_iq_{i+1}>0$ for $0\le i<n$, K has a unique stationary distribution π given by $\pi(i)=c(p_0\cdots p_{i-1})/(q_1\cdots q_i)$, where c is the positive constant such that

2000 *Mathematics Subject Classification.* 60J10,60J27.

Key words and phrases. Birth and death chains, spectrum.

[1]Partially supported by NSC grant NSC100-2115-M-009-003-MY2 and NCTS, Taiwan.

[2]Partially supported by NSF grant DMS-1004771.

$\sum_{i=0}^{n} \pi(i) = 1$. The smallest non-zero eigenvalue of $I - K$ is exactly the spectral gap of the path on $\{0, 1, ..., n\}$ with measures π, ν, where $\nu(i, i+1) = \pi(i)p_i = \pi(i+1)q_{i+1}$ for $0 \le i < n$.

Note that if $\mathbf{1}$ is the constant function of value 1 and ψ is a minimizer for $\lambda^G_{\pi,\nu}$, then $\psi - \pi(\psi)\mathbf{1}$ is an eigenvector of $M^G_{\pi,\nu}$. This implies that any minimizer ψ for $\lambda^G_{\pi,\nu}$ satisfying $\pi(\psi) = 0$ satisfies the Euler-Lagrange equation,

$$\lambda^G_{\pi,\nu}\pi(i)\psi(i) = [\psi(i) - \psi(i-1)]\nu(i-1, i) + [\psi(i) - \psi(i+1)]\nu(i, i+1), \tag{1.2}$$

for all $1 \le i \le n$. Assuming the connectedness of G (i.e., the superdiagonal and subdiagonal entries of $M^G_{\pi,\nu}$ are positive), the rank of $M^G_{\pi,\nu} - \lambda I$ is at least $n-1$. This implies that all eigenvalues of $M^G_{\pi,\nu}$ are simple. See Lemma A.3 for an illustration. Observe that, by (1.2), any non-trivial eigenvector of $M^G_{\pi,\nu}$ has mean 0 under π. This implies that all minimizers for the spectral gap are of the form $a\psi + b\mathbf{1}$, where a, b are constants and ψ is a nontrivial solution of (1.2). In 2009, Miclo obtained implicitly the following result.

Theorem 1.1. [15, Proposition 1] *If ψ is a minimizer for $\lambda^G_{\pi,\nu}$, then ψ must be monotonic, that is, either $\psi(i) \le \psi(i+1)$ for all $1 \le i < n$ or $\psi(i) \ge \psi(i+1)$ for all $1 \le i < n$.*

One aim of this paper is to provide a scheme to compute the spectrum of $M^G_{\pi,\nu}$, in particular, the spectral gap. Based on Miclo's observation, it is natural to consider the following algorithm.

(A1)

Choose two positive reals λ_0, a in advance and set, for $k = 0, 1, ...,$

1. $\psi_k(1) = -a$,
2. $\psi_k(i+1) = \psi_k(i) + \dfrac{\{[\psi_k(i) - \psi_k(i-1)]\nu(i-1, i) - \lambda_k\pi(i)\psi_k(i)\}^+}{\nu(i, i+1)}$, for $1 \le i < n$, where $t^+ = \max\{t, 0\}$,
3. $\lambda_{k+1} = \dfrac{\mathcal{E}_\nu(\psi_k, \psi_k)}{\mathrm{Var}_\pi(\psi_k)}$.

The following theorems discuss the behavior of λ_k.

Theorem 1.2 (Convergence to the exact value)**.** *Referring to* (A1)*, if $n = 2$, then $\lambda_k = \lambda^G_{\pi,\nu}$ for all $k \ge 1$. If $n \ge 3$, then the sequence (λ_k, ψ_k) satisfies*

(1) *If $\lambda_0 = \lambda^G_{\pi,\nu}$, then $\lambda_k = \lambda^G_{\pi,\nu}$ for all $k \ge 0$.*
(2) *If $\lambda_0 \ne \lambda^G_{\pi,\nu}$, then $\lambda_k > \lambda_{k+1} > \lambda^G_{\pi,\nu}$ for $k \ge 1$.*
(3) *Set $(\lambda^*, \psi^*) = \lim_{k\to\infty}(\lambda_k, \psi_k)$. Then, $\lambda^* = \mathcal{E}_\nu(\psi^*, \psi^*)/\mathrm{Var}_\pi(\psi^*) = \lambda^G_{\pi,\nu}$ and $\pi(\psi^*) = 0$.*

Theorem 1.3 (Rate of convergence)**.** *Referring to* Theorem 1.2*, there is a constant $\sigma \in (0, 1)$ independent of the choice of (λ_0, a) such that $0 \le \lambda_k - \lambda^G_{\pi,\nu} \le \sigma^{k-1}\lambda_1$ for all $k \ge 1$.*

By Theorem 1.3, we know that the sequence λ_k generated in (A1) converges to the spectral gap exponentially but the rate $(-\log\sigma)$ is undetermined. The following alternative scheme is based on using more information on the spectral gap and will provide convergence at a constant rate.

(A2) Choose $a > 0, L_0 < \lambda^G_{\pi,\nu} < U_0$ in advance and set, for $k = 0, 1, ...,$

1. $\psi_k(1) = -a$, $\lambda_k = \frac{1}{2}(L_k + U_k)$

2. $\psi_k(i+1) = \psi_k(i) + \dfrac{\{[\psi_k(i) - \psi_k(i-1)]\nu(i-1,i) - \lambda_k \pi(i)\psi_k(i)\}^+}{\nu(i,i+1)}$,
for $1 \le i < n$, where $t^+ = \max\{t, 0\}$,

3. $\begin{cases} L_{k+1} = L_k,\, U_{k+1} = \lambda_k & \text{if } \pi(\psi_k) > 0 \\ L_{k+1} = \lambda_k,\, U_{k+1} = U_k & \text{if } \pi(\psi_k) < 0\ . \\ L_{k+1} = U_{k+1} = \lambda_k & \text{if } \pi(\psi_k) = 0 \end{cases}$

Theorem 1.4 (Dichotomy method)**.** *Referring to* (A2)*, it holds true that*

$$0 \le \max\{U_k - \lambda^G_{\pi,\nu}, \lambda^G_{\pi,\nu} - L_k\} \le (U_0 - L_0)2^{-k}, \quad \forall k \ge 0.$$

In Theorem 1.4, the convergence to the spectral gap is exponentially fast with explicit rate, $\log 2$. See Remark 2.2 for a discussion on the choice of L_0 and U_0. For higher order spectra, Miclo has a detailed description of the shape of eigenvectors in [14] and this will motivate the definition of similar algorithms for every eigenvalue in spectrum. See (Di) and Theorem 3.4 for a generalization of (A2) and Theorem 3.14 for a localized version of Theorem 1.3.

The spectral gap is an important parameter in the quantitative analysis of Markov chains. The cutoff phenomenon, a sharp phase transition phenomenon for Markov chains, was introduced by Aldous and Diaconis in early 1980s. It is of interest in many applications. A heuristic conjecture proposed by Peres in 2004 says that the cutoff exists if and only if the product of the spectral gap and the mixing time tends to infinity. Assuming reversibility, this has been proved to hold for L^p-convergence with $1 < p \le \infty$ in [2]. For the L^1-convergence, Ding *et al.* [10] prove this conjecture for continuous time birth and death chains. In order to use Peres' conjecture in practice, the orders of the magnitudes of spectral gap and mixing time are required. The second aspect of this paper is to derive a theoretical lower bound on the spectral gap using only the birth and death rates. This lower bound is obtained using the same idea used to analyze the above algorithm. For estimates on the mixing time of birth and death chains, we refer the readers to the recent work [4] by Chen and Saloff-Coste. For illustration, we consider several examples of specific interest and show that the lower bound provided here is in fact of the correct order in these examples.

This article is organized as follows. In Section 2, the algorithms in (A1)-(A2) are explored and proofs for Theorems 1.2-1.4 are given. In Section 3, the spectrum of $M^G_{\pi,\nu}$ is discussed further and, based on Miclo's work [14], Algorithm (A2) is generalized to any specified eigenvalue of $M^G_{\pi,\nu}$. Our method is applicable for paths of infinite length (one-sided) and this is described in Section 4. For illustration, we consider some Metropolis chains and display numerical results of Algorithm (A2) in Section 5. In Section 6, we focus on uniform measures with bottlenecks and determine the correct order of the spectral gap using the theory in Sections 2-3. It is worthwhile to remark that the assumptions in Section 6 can be relaxed using the comparison technique in [7, 8]. As the work in this paper can also be regarded as a stochastic counterpart of theory of finite Jacobi matrices, we would like to refer the readers to [18, 19] for a complementary perspective.

2. Convergence to the spectral gap

This section is devoted to proving Theorems 1.2-1.4. First, we prove Theorem 1.1 in the following form.

Lemma 2.1. *Let $\lambda > 0$ and ψ be a non-constant function on V. Suppose (λ, ψ) solves (1.2) and ψ is monotonic. Then, ψ is strictly monotonic, that is, either $\psi(i) < \psi(i+1)$ for $1 \le i < n$ or $\psi(i) > \psi(i+1)$ for $1 \le i < n$.*

Proof. Obviously, (1.2) implies that $\pi(\psi) = 0$. Without loss of generality, it suffices to consider the case when $\psi(1) < 0$ and $\psi(n) > 0$. Since ψ is non-constant and $\lambda^G_{\pi,\nu} > 0$, we have $\psi(1) < \psi(2)$ and $\psi(n-1) < \psi(n)$. Note that if there are $1 < i < j < n$ such that $\psi(i-1) < \psi(i)$, $\psi(j) < \psi(j+1)$ and $\psi(k) = \psi(i) = \psi(j)$ for $i \le k \le j$, then (1.2) yields

$$\lambda^G_{\pi,\nu}\pi(i)\psi(i) = [\psi(i) - \psi(i-1)]\nu(i-1,i) + [\psi(i) - \psi(i+1)]\nu(i,i+1) > 0$$

and

$$\lambda^G_{\pi,\nu}\pi(j)\psi(j) = [\psi(j) - \psi(j-1)]\nu(j-1,j) + [\psi(j) - \psi(j+1)]\nu(j,j+1) < 0,$$

a contradiction. Thus, ψ is strictly increasing. □

We note the following corollary.

Corollary 2.2. *Let (λ, ψ) be a pair satisfying (1.2). Then, $\lambda = \lambda^G_{\pi,\nu}$ if and only if ψ is monotonic.*

Proof. One direction is obvious from Theorem 1.1. For the other direction, assume that ψ is monotonic and let ϕ be a minimizer for $\lambda^G_{\pi,\nu}$ with $\pi(\phi) = 0$. Since (λ, ψ) and $(\lambda^G_{\pi,\nu}, \phi)$ are solutions to (1.2), one has

$$\lambda\pi(\psi\phi) = \mathcal{E}_\nu(\psi, \phi) = \lambda^G_{\pi,\nu}\pi(\phi\psi).$$

By Lemma 2.1, ψ and ϕ are strictly monotonic and this implies $\mathcal{E}_\nu(\psi, \phi) \neq 0$. As a consequence of the above equations, we have $\lambda = \lambda^G_{\pi,\nu}$. □

The following proposition is the key to Theorem 1.2.

Proposition 2.3. *Suppose that (λ, ψ) satisfies $\lambda > 0$, $\psi(1) < 0$ and, for $1 \le i < n$,*

$$\psi(i+1) = \psi(i) + \frac{\{[\psi(i) - \psi(i-1)]\nu(i-1,i) - \lambda\pi(i)\psi(i)\}^+}{\nu(i,i+1)}, \tag{2.1}$$

where $t^+ = \max\{t, 0\}$. Then, the following are equivalent.

(1) $\mathcal{E}_\nu(\psi, \psi) = \lambda \mathrm{Var}_\pi(\psi)$.
(2) $\pi(\psi) = 0$.
(3) $\lambda = \lambda^G_{\pi,\nu}$.

Furthermore, if $n \ge 3$, then any of the above is equivalent to

(4) $\mathcal{E}_\nu(\psi, \psi) = \lambda^G_{\pi,\nu}\mathrm{Var}(\psi)$

Remark 2.1. For $n = 2$, it is an easy exercise to show that $\lambda^G_{\pi,\nu} = \nu(1,2)/(\pi(1)\pi(2))$. By following the formula in (2.1), one has $\psi(2) = \psi(1)[1 - \lambda\pi(1)/\nu(1,2)]$, which leads to $\mathcal{E}_\nu(\psi, \psi)/\mathrm{Var}_\pi(\psi) = \lambda^G_{\pi,\nu}$.

Proof of Proposition 2.3. Set $B = \{1 \le i \le n | \psi(i) = \psi(n)\}$ and $B^c = \{1, 2, ..., i_0\}$. Since $\psi(1) < 0$ and $\lambda > 0$, $\psi(1) < \psi(2)$ and B^c is nonempty. According to (2.1), ψ is non-decreasing. Note that if $\psi(i) = \psi(i+1)$, then $\psi(i) \ge 0$ and $\psi(i+2) = \psi(i+1)$. This implies ψ is strictly increasing on $\{1, 2, ..., i_0 + 1\}$ and, for $1 \le i \le i_0$,

$$\lambda\pi(i)\psi(i) = [\psi(i) - \psi(i+1)]\nu(i, i+1) + [\psi(i) - \psi(i-1)]\nu(i-1, i).$$

Multiplying $\psi(i)$ on both sides and summing over all i in B^c yields

$$\begin{aligned}\lambda \sum_{i=1}^{i_0} \psi(i)^2\pi(i) &= \sum_{i=1}^{i_0-1} [\psi(i) - \psi(i+1)]^2\nu(i, i+1) \\ &\quad + \psi(i_0)[\psi(i_0) - \psi(i_0+1)]\nu(i_0, i_0+1) \\ &= \mathcal{E}_\nu(\psi, \psi) + \psi(i_0+1)[\psi(i_0) - \psi(i_0+1)]\nu(i_0, i_0+1) \\ &= \mathcal{E}_\nu(\psi, \psi) + \lambda\psi(n) \sum_{i=1}^{i_0} \psi(i)\pi(i).\end{aligned}$$

This is equivalent to

$$\mathcal{E}_\nu(\psi, \psi) = \lambda \mathrm{Var}_\pi(\psi) + \lambda\pi(\psi)[\pi(\psi) - \psi(n)], \tag{2.2}$$

which proves (1)⇔(2).

If $\lambda = \lambda^G_{\pi,\nu}$, then ψ is an eigenvector for $M^G_{\pi,\nu}$ associated to $\lambda^G_{\pi,\nu}$. This proves (3)⇒(2). For (2)⇒(3), assume that $\pi(\psi) = 0$. In this case, ψ must be strictly increasing. Otherwise, $\psi(i) = \psi(n) > 0$ for $i \in B$ and, according to (2.1), this implies

$$\lambda \mathrm{Var}_\pi(\psi) > \lambda \sum_{i=1}^{n-1} \pi(i)\psi^2(i) \ge \sum_{i=1}^{n-1} [\psi(i) - \psi(i+1)]^2\nu(i, i+1) = \mathcal{E}(\psi, \psi),$$

which contradicts (1). As ψ is strictly increasing and $\pi(\psi) = 0$, (λ, ψ) solves (1.2). By Corollary 2.2, $\lambda = \lambda^G_{\pi,\nu}$.

To finish the proof, it remains to show (4)⇒(3) ((3)⇒(4) is obvious from the equivalence among (1), (2) and (3)). Assume that $\mathcal{E}_\nu(\psi, \psi) = \lambda^G_{\pi,\nu}\mathrm{Var}_\pi(\psi)$. By Lemma 2.1, ψ is strictly monotonic and this implies, for $1 \le i < n$,

$$\lambda\pi(i)\psi(i) = [\psi(i) - \psi(i+1)]\nu(i, i+1) + [\psi(i) - \psi(i-1)]\nu(i-1, i).$$

As ψ is a minimizer for $\lambda^G_{\pi,\nu}$, one has, for $1 \le i \le n$,

$$\lambda^G_{\pi,\nu}\pi(i)[\psi(i) - \pi(\psi)] = [\psi(i) - \psi(i+1)]\nu(i, i+1) + [\psi(i) - \psi(i-1)]\nu(i-1, i).$$

If $\lambda \ne \lambda^G_{\pi,\nu}$, the comparison of both systems yields

$$\psi(i) = \frac{\lambda^G_{\pi,\nu}\pi(\psi)}{\lambda^G_{\pi,\nu} - \lambda}, \quad \forall 1 \le i < n.$$

As $n \ge 3$, $\psi(1) = \psi(2)$, a contradiction! This forces $\lambda = \lambda^G_{\pi,\nu}$, as desired. □

The following is a simple corollary of Proposition 2.3, which plays an important role in proving Theorem 1.4.

Corollary 2.4. *Let $n \ge 3$. For $\lambda > 0$, let ϕ_λ be the vector generated by* (2.1) *with $\phi(1) < 0$. Then, $(\lambda - \lambda^G_{\pi,\nu})\pi(\phi_\lambda) > 0$ for $\lambda > 0$ and $\lambda \ne \lambda^G_{\pi,\nu}$.*

Proof. Without loss of generality, we fix $\phi_\lambda(1) = -1$ for all $\lambda > 0$. Set $T(\lambda) = \pi(\phi_\lambda)$. To prove this corollary, it suffices to show that

$$T(\lambda)\begin{cases} < 0 & \text{if } \lambda < \lambda^G_{\pi,\nu} \\ > 0 & \text{if } \lambda > \lambda^G_{\pi,\nu} \end{cases}.$$

For $\lambda > 0$, define $L(\lambda) := \mathcal{E}_\nu(\phi_\lambda, \phi_\lambda)/\mathrm{Var}_\pi(\phi_\lambda)$. By (2.2), one has

$$L(\lambda) - \lambda = \frac{\lambda T(\lambda)[\pi(\phi_\lambda) - \phi_\lambda(n)]}{\mathrm{Var}_\pi(\phi_\lambda)}. \tag{2.3}$$

Since ϕ_λ is non-constant, $\pi(\phi_\lambda) < \phi_\lambda(n)$. This implies $T(\lambda) < 0$ for $\lambda \in (0, \lambda^G_{\pi,\nu})$.

For $\lambda > \lambda^G_{\pi,\nu}$, set $I = (\lambda^G_{\pi,\nu}, \infty)$. By Proposition 2.3, $T(\lambda) = 0$ if and only if $\lambda = \lambda^G_{\pi,\nu}$. By the continuity of T, this implies either $T(I) \subset (-\infty, 0)$ or $T(I) \subset (0, \infty)$. In the case $T(I) \subset (-\infty, 0)$, one has $L(\lambda) > \lambda$ for $\lambda \in I$. As $L(I)$ is bounded, $L^k(\lambda)$ is convergent with limit $\widetilde{\lambda} > \lambda^G_{\pi,\nu}$ and this yields

$$0 = \lim_{k\to\infty}[L^{k+1}(\lambda) - L^k(\lambda)] = \frac{\widetilde{\lambda}T(\widetilde{\lambda})[\pi(\phi_{\widetilde{\lambda}}) - \phi_{\widetilde{\lambda}}(n)]}{\mathrm{Var}_\pi(\phi_{\widetilde{\lambda}})} > 0,$$

a contradiction. Hence, $T(\lambda) > 0$ for $\lambda > \lambda^G_{\pi,\nu}$. □

Proof of Theorem 1.2. The proof for $n = 2$ is obvious from a direct computation and we deal with the case $n \geq 3$, here. By the equivalence of Proposition 2.3 (3)-(4), if $\lambda_0 = \lambda^G_{\pi,\nu}$, then $\lambda_k = \lambda^G_{\pi,\nu}$ for all $k \geq 1$. If $\lambda_0 \neq \lambda^G_{\pi,\nu}$, then $\lambda_k > \lambda^G_{\pi,\nu}$ for $k \geq 1$. Note that (λ_k, ψ_k) solves the system in (2.1). By (2.2), this implies

$$\lambda_{k+1} - \lambda_k = \frac{\lambda_k \pi(\psi_k)[\pi(\psi_k) - \psi_k(n)]}{\mathrm{Var}_\pi(\psi_k)}, \quad \forall k \geq 0.$$

The strict monotonicity of λ_k in (2) comes immediately from Corollary 2.4. In (3), the continuity of (2.1) in λ implies that (λ^*, ψ^*) is a solution to (2.1) and $\mathcal{E}_\nu(\psi^*, \psi^*) = \lambda^*\mathrm{Var}(\psi^*)$. By Proposition 2.3, $\lambda^* = \lambda^G_{\pi,\nu}$ and $\pi(\psi^*) = 0$, as desired. □

Proof of Theorem 1.3. Recall the notation in the proof of Corollary 2.4: For $\lambda > 0$, let ϕ_λ be the function defined by (2.1) and $L(\lambda) = \mathcal{E}_\nu(\phi_\lambda, \phi_\lambda)/\mathrm{Var}_\pi(\phi_\lambda)$. By (2.2) and Corollary 2.4, $L(\lambda) \in (\lambda^G_{\pi,\nu}, \lambda)$ for $\lambda > \lambda^G_{\pi,\nu}$. As L is bounded, Theorem 1.3 follows from Lemma A.1. □

Proof of theorem 1.4. Immediate from Corollary 2.4. □

In the end of this section, we use the following proposition to find how the shape of the function ψ in (2.1) evolves with λ. In Proposition 2.5, we set $\phi_\lambda = \psi$ when ψ is given by (2.1). It is easy to see from (2.1) that ϕ_λ is strictly increasing before some constant, say $i_0 = i_0(\lambda)$, and then stays constant equal to $\phi_\lambda(i_0)$ after i_0. The proposition shows how the constant $i_0(\lambda)$ evolves.

Proposition 2.5. *For $\lambda > 0$, let ϕ_λ be the function generated by (2.1) with $\phi_\lambda(1) = -1$ and, for $1 \leq i \leq n$, set $T_i(\lambda) = \sum_{j=1}^i \phi_\lambda(i)\pi(i)$. For $1 \leq i < n$, let*

$$a_i(\lambda) = 1 + \pi(i+1)/\pi(i) - \lambda\pi(i+1)/\nu(i, i+1),$$

$$A_i(\lambda) = \begin{pmatrix} a_1(\lambda) & 1 & 0 & 0 & \cdots & 0 \\ \frac{\pi(3)}{\pi(2)} & a_2(\lambda) & 1 & 0 & & \vdots \\ 0 & \frac{\pi(4)}{\pi(3)} & a_3(\lambda) & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & a_{i-1}(\lambda) & 1 \\ 0 & \cdots & \cdots & 0 & \frac{\pi(i+1)}{\pi(i)} & a_i(\lambda) \end{pmatrix}, \tag{2.4}$$

and let $\lambda^{(i)}$ be the smallest root of $\det A_i(\lambda) = 0$. Then,

(1) $\lambda_{\pi,\nu}^G = \lambda^{(n-1)} < \lambda^{(n-2)} < \cdots < \lambda^{(1)}$.
(2) $\phi_\lambda(i) < \phi_\lambda(i+1) = \phi_\lambda(i+2)$ *for* $\lambda \in [\lambda^{(i)}, \lambda^{(i-1)})$ *and* $1 \le i \le n-2$, *where* $\lambda^{(0)} := \infty$.
(3) $\phi_\lambda(n-1) < \phi_\lambda(n)$ *for* $\lambda \in (0, \lambda^{(n-2)})$.

In particular, $T_{i+1}(\lambda) = -\pi(1)\det A_i(\lambda)$ *for* $\lambda \in (0, \lambda^{(i-1)})$ *and* $(\lambda - \lambda^{(i)})T_{i+1}(\lambda) > 0$ *for* $\lambda \in (0, \lambda^{(i)}) \cup (\lambda^{(i)}, \infty)$ *with* $1 \le i \le n-1$.

Proof. By Lemma A.2, $\lambda^{(1)} > \lambda^{(2)} > \cdots > \lambda^{(n-1)} > 0$ and, for $1 \le i \le n-1$,

$$\det A_i(\lambda) \begin{cases} > 0 & \forall \lambda \in (-\infty, \lambda^{(i)}) \\ < 0 & \forall \lambda \in (\lambda^{(i)}, \lambda^{(i-1)}) \end{cases}, \tag{2.5}$$

where $\lambda^{(0)} = \infty$. Note that if $T_i(\lambda) < 0$ for some $1 \le i \le n-1$, then

$$\phi_\lambda(j+1) = \phi_\lambda(j) + \frac{[\phi_\lambda(j) - \phi_\lambda(j-1)]\nu(j-1,j) - \lambda\pi(j)\phi_\lambda(j)}{\nu(j,j+1)}, \quad \forall 1 \le j \le i.$$

This implies

$$\phi_\lambda(\ell+1) = \phi_\lambda(\ell) - \frac{\lambda}{\nu(\ell,\ell+1)} \sum_{j=1}^{\ell} \pi(j)\phi_\lambda(j), \quad \forall 1 \le \ell \le i. \tag{2.6}$$

Multiplying $\pi(\ell+1)$ and adding up $T_\ell(\lambda)$ yields

$$T_{\ell+1}(\lambda) = a_\ell(\lambda)T_\ell(\lambda) - \frac{\pi(\ell+1)}{\pi(\ell)}T_{\ell-1}(\lambda), \quad \forall 1 \le \ell \le i.$$

From the above discussion, we conclude that if $T_i(\lambda) < 0$, then

$$T_{\ell+1}(\lambda) = -\pi(1)\det A_\ell(\lambda), \quad \forall 1 \le \ell \le i. \tag{2.7}$$

When $\ell = i-1$, (2.5) implies $\det A_{i-1}(\lambda) > 0$ for $\lambda < \lambda^{(i-1)}$. By the continuity of T_i and $\det A_{i-1}$, if there is some $\lambda < \lambda^{(i-1)}$ such that $T_i(\lambda) < 0$, then $T_i(\lambda) = -\pi(1)\det A_{i-1}(\lambda)$ for $\lambda < \lambda^{(i-1)}$. As a consequence of (2.7) with $\ell = i$, this will imply $T_{i+1}(\lambda) = -\pi(1)\det A_i(\lambda)$ for $\lambda < \lambda^{(i-1)}$. Hence, it remains to show that $T_i(\lambda) < 0$ for some $\lambda < \lambda^{(i-1)}$. To see this, according to Corollary 2.4, one can choose a constant $\widetilde{\lambda} < \min\{\lambda_{\pi,\nu}^G, \lambda^{(i-1)}\}$ such that $T_{n-1}(\widetilde{\lambda}) < 0$. Since $\phi_\lambda(i)$ is non-decreasing in i, we obtain $T_i(\widetilde{\lambda}) < 0$, as desired. This proves $T_{i+1}(\lambda) = -\pi(1)\det A_i(\lambda)$ for $\lambda < \lambda^{(i-1)}$. In particular, $T_n(\lambda) = -\pi(1)\det A_{n-1}(\lambda)$ for $\lambda < \lambda^{(n-2)}$. By Corollary 2.4, we have $\lambda^{(n-1)} = \lambda_{\pi,\nu}^G$. This proves Proposition 2.5 (1).

Next, observe that, for $\lambda \in (\lambda^{(i)}, \lambda^{(i-1)})$,

$$\sum_{j=1}^{i+1} \pi(j)\phi_\lambda(j) = T_{i+1}(\lambda) > 0, \quad \sum_{j=1}^{i} \pi(j)\phi_\lambda(j) = T_i(\lambda) < 0.$$

By (2.6), it is easy to see that $[\phi_\lambda(i+1) - \phi_\lambda(i)]\nu(i,i+1) = -\lambda T_i(\lambda) > 0$ and

$$\begin{aligned}&[\phi_\lambda(i+2) - \phi_\lambda(i+1)]\nu(i+1,i+2)\\ =&\{[\phi_\lambda(i+1) - \phi_\lambda(i)]\nu(i,i+1) - \lambda\pi(i+1)\phi_\lambda(i+1)\}^+\\ =&\{-\lambda T_{i+1}(\lambda)\}^+ = 0.\end{aligned}$$

This proves Proposition 2.5 (2). To prove Proposition 2.5 (3), we use (1) to derive

$$T_{n-1}(\lambda) = -\pi(1)\det A_{n-2}(\lambda) < 0, \quad \forall \lambda \in (0, \lambda^{(n-2)}).$$

Using (2.6), this implies $\phi_\lambda(n-1) < \phi_\lambda(n)$. The last part of Proposition 2.5 follows easily from (2.5) and the fact that

$$T_i(\lambda) \ge 0 \Rightarrow T_{i+1}(\lambda) > 0 \text{ and } T_i(\lambda) \le 0 \Rightarrow T_{i-1}(\lambda) < 0.$$

□

Remark 2.2. In Proposition 2.5, if $\lambda > \lambda^{(1)} = \nu(1,2)[\pi(1)^{-1}+\pi(2)^{-1}]$, then $\phi_\lambda(i) = \phi_\lambda(2)$ for $i = 2,...,n$. Note that, for $\lambda \ge \lambda^{(1)}$, $\phi_\lambda(2) = -1 + \lambda\pi(1)/\nu(1,2)$ and

$$\pi(\phi_\lambda) = -1 + \frac{\lambda\pi(1)(1-\pi(1))}{\nu(1,2)}, \quad \mathrm{Var}_\pi(\phi_\lambda) = \frac{\lambda^2\pi(1)^3(1-\pi(1))}{\nu(1,2)^2}.$$

By (2.3), this leads to $L(\lambda) = \nu(1,2)/[\pi(1)(1-\pi(1)]$ for $\lambda \ge \lambda^{(1)}$. In the case $n = 2$, it is clear that $\nu(1,2)/[\pi(1)(1-\pi(1)] = \nu(1,2)[\pi(1)^{-1} + \pi(2)^{-1}] = \lambda^G_{\pi,\nu}$.

3. Convergence to other eigenvalues

In this section, we generalize the algorithms (A1) and (A2) so that they can be applied for the computation to any specified eigenvalue.

3.1. **Basic setup and fundamental results.** Recall that G is a graph with vertex set $V = \{1,2,...,n\}$ and edge set $E = \{\{i,i+1\}|i = 1,2,...,n-1\}$. Given two positive measures π,ν on V, E with $\pi(V) = 1$, let $M^G_{\pi,\nu}$ be a n-by-n matrix defined in the introduction and given by

$$M^G_{\pi,\nu}(i,j) = \begin{cases} -\nu(i,j)/\pi(i) & \text{if } |i-j| = 1\\ [\nu(i-1,i)+\nu(i,i+1)]/\pi(i) & \text{if } j = i\\ 0 & \text{if } |i-j| > 1\end{cases}. \tag{3.1}$$

Since ν is positive everywhere and $M^G_{\pi,\nu}$ is tridiagonal, all eigenvalues of $M^G_{\pi,\nu}$ have algebraic multiplicity 1. Throughout this section, let $\{\lambda^G_0 < \lambda^G_1 < \cdots < \lambda^G_{n-1}\}$ denote the eigenvalues of $M^G_{\pi,\nu}$ with associated $L^2(\pi)$-normalized eigenvectors $\zeta_0 = \mathbf{1}, \zeta_2, ..., \zeta_{n-1}$. Clearly, $\lambda^G_0 = 0$, $\lambda^G_1 = \lambda^G_{\pi,\nu}$ and, for $1 \le k \le n$,

$$\lambda^G_i\zeta_i(k)\pi(k) = [\zeta_i(k) - \zeta_i(k-1)]\nu(k-1,k) + [\zeta_i(k) - \zeta_i(k+1)]\nu(k,k+1). \tag{3.2}$$

Let $1 \le i \le n-1$. As ζ_i is non-constant, it is clear that $\zeta_i(1) \ne \zeta_i(2)$ and $\zeta_i(n-1) \ne \zeta_i(n)$. Moreover, if $\zeta_i(k) = \zeta_i(k+1)$ for some $1 < k < n$, then $\zeta_i(k) \ne \zeta_i(k-1)$ and $\zeta_i(k+1) \ne \zeta_i(k+2)$. Gantmacher and Krein [13] showed that there are exactly i sign changes for ζ_i with $1 \le i \le n$. Miclo [14] gives a detailed description on the shape of ζ_i as follows.

Theorem 3.1. *For $1 \le i \le n-1$, let ζ_i be an eigenvector associated to the ith smallest non-zero eigenvalue of the matrix in* (3.1) *with $\zeta_i(1) < 0$. Then, there are $1 = a_1 < b_1 \le a_2 < b_2 \le \cdots \le a_i < b_i = n$ with $a_{j+1} - b_j \in \{0, 1\}$ such that ζ_i is strictly increasing on $[a_j, b_j]$ for odd j and is strictly decreasing on $[a_j, b_j]$ for even j, and $\zeta_i(a_{j+1}) = \zeta_i(b_j)$ for $1 \le j < i$.*

In the following, we make some analysis related to the Euler-Lagrange equations in (3.2).

Definition 3.1. Fix $n \ge 1$ and let f be a function on $\{1, 2, ..., n\}$. For $1 \le i \le n-1$, f is called "Type i" if there are $1 = a_1 < b_1 \le a_2 < b_2 \le \cdots \le a_i < b_i \le n$ satisfying $a_{j+1} - b_j \in \{0, 1\}$ such that

(1) f is strictly monotonic on $[a_j, b_j]$ for $1 \le j \le i$.
(2) $[f(a_j) - f(a_j + 1)][f(a_{j+1}) - f(a_{j+1} + 1)] < 0$ for $1 \le j < i$.
(3) $f(a_{j+1}) = f(b_j)$, for $1 \le j < i$, and $f(k) = f(b_i)$, for $b_i \le k \le n$.

The points a_j, b_j will be called "peak-valley points" in this paper.

Remark 3.1. Note that the difference between Definition 3.1 and Theorem 3.1 is the requirement $b_i \le n$, instead of $b_i = n$. By Theorem 3.1, any eigenvector associated to the ith smallest non-zero eigenvalue of the matrix in (3.1) must be of type i with $b_i = n$.

Definition 3.2. Let π, ν be positive measures on V, E with $\pi(V) = 1$. For $\lambda \in \mathbb{R}$, let ξ_λ be a function on $\{1, 2, ..., n\}$ defined by $\xi_\lambda(1) = -1$ and, for $1 \le k < n$,

$$\xi_\lambda(k+1) = \xi_\lambda(k) + \frac{[\xi_\lambda(k) - \xi_\lambda(k-1)]\nu(k-1,k) - \lambda\pi(k)\xi_\lambda(k)}{\nu(k,k+1)}.$$

Remark 3.2. Note that $\xi_0 = -\mathbf{1}$ and, for $\lambda < 0$, ξ_λ is strictly decreasing and of type 1. For $\lambda > 0$, if $\xi_\lambda(k-1) < \xi_\lambda(k) = \xi_\lambda(k+1)$, then $\xi_\lambda(k) > 0$ and this implies $\xi_\lambda(k+2) < \xi_\lambda(k+1)$. Similarly, if $\xi_\lambda(k-1) > \xi_\lambda(k) = \xi_\lambda(k+1)$, then $\xi_\lambda(k) < 0$ and $\xi_\lambda(k+2) > \xi_\lambda(k+1)$. Thus, ξ_λ must be of type i for some $1 \le i \le n-1$.

Lemma 3.2. *For $\lambda > 0$, let ξ_λ be the function in Definition 3.2. Suppose that ξ_λ is of type i with $1 \le i \le n-1$.*

(1) *If $\xi_\lambda(n-1) \ne \xi_\lambda(n)$, then there is $\epsilon > 0$ such that $\xi_{\lambda+\delta}$ is of type i for $-\epsilon < \delta < \epsilon$.*
(2) *If $\xi_\lambda(n-1) = \xi_\lambda(n)$, then there is $\epsilon > 0$ such that $\xi_{\lambda+\delta}$ is of type $i+1$ and $\xi_{\lambda-\delta}$ is of type i for $0 < \delta < \epsilon$.*

Proof. Let a_j, b_j be the peak-valley points of ξ_λ. By the continuity of ξ_λ in λ and Remark 3.2, one can choose $\epsilon > 0$ such that, for $\delta \in (-\epsilon, \epsilon)$, $\xi_{\lambda+\delta}$ remains strictly monotonic on $[a_j, b_j]$ for $j = 1, ..., i$ and

$$[\xi_{\lambda+\delta}(b_j - 1) - \xi_{\lambda+\delta}(b_j)][\xi_{\lambda+\delta}(a_{j+1} + 1) - \xi_{\lambda+\delta}(a_{j+1})] > 0,$$

for $1 \le j < i$. In (1), $b_i = n$. Fix $\delta \in (-\epsilon, \epsilon)$ and set $a_1' = a_1 = 1$, $b_i' = b_i = n$. For $1 < j < i$, set

$$\begin{cases} b_j' = a_{j+1}' = b_j & \text{if } [\xi_{\lambda+\delta}(b_j - 1) - \xi_{\lambda+\delta}(b_j)][\xi_{\lambda+\delta}(b_j) - \xi_{\lambda+\delta}(a_{j+1})] < 0 \\ b_j' = a_{j+1}' = a_{j+1} & \text{if } [\xi_{\lambda+\delta}(b_j - 1) - \xi_{\lambda+\delta}(b_j)][\xi_{\lambda+\delta}(b_j) - \xi_{\lambda+\delta}(a_{j+1})] > 0 \\ b_j' = b_j,\ a_{j+1}' = a_{j+1} & \text{if } [\xi_{\lambda+\delta}(b_j - 1) - \xi_{\lambda+\delta}(b_j)][\xi_{\lambda+\delta}(b_j) - \xi_{\lambda+\delta}(a_{j+1})] = 0 \end{cases}.$$

Clearly, $\xi_{\lambda+\delta}$ is of type i with peak-valley points a'_j, b'_j. This proves Lemma 3.2 (1).

For part (2), we consider $i \le n-2$ and $b_i = n-1$. By similar argument as before, one can choose $\epsilon > 0$ such that the restriction of $\xi_{\lambda+\delta}$ to $\{1, 2, ..., n-1\}$ is of type i for $\delta \in (-\epsilon, \epsilon)$. To finish the proof, it remains to compare $\xi_{\lambda+\delta}(n-1)$ and $\xi_{\lambda+\delta}(n)$. Recall that $T_j(\lambda) = \sum_{k=1}^{j} \xi_\lambda(k)\pi(k)$ as in the proof for Proposition 2.5. Using a similar reasoning as for (2.7), one shows that $T_{i+1}(\lambda) = -\pi(1)\det A_i(\lambda)$ for $1 \le i < n$, where $A_i(\lambda)$ is the matrix in (2.4). This implies that the non-zero eigenvalues of $M^G_{\pi,\nu}$, say $\lambda^G_1, ..., \lambda^G_{n-1}$, are the roots of $\det A_{n-1}(\lambda) = 0$. As a consequence of Lemma A.2, $\det A_{n-2}(\lambda) = 0$ has exactly $n-2$ distinct roots, say $\alpha_1 < \alpha_2 < \cdots < \alpha_{n-1}$, and they satisfy the interlacing property $\lambda^G_j < \alpha_j < \lambda^G_{j+1}$ for $1 \le j \le n-2$. Note that $\det A_{n-2}(\lambda)$ and $\det A_{n-1}(\lambda)$ tend to infinity as $-\lambda$ tends to infinity. This leads to the fact that if $\det A_{n-2}(\lambda) = 0$ and $\det A_{n-1}(\lambda) < 0$, then $\det A_{n-2}(\cdot)$ is strictly decreasing in a neighborhood of λ. If $\det A_{n-2}(\lambda) = 0$ and $\det A_{n-1}(\lambda) > 0$, then $\det A_{n-2}(\cdot)$ is strictly increasing in a neighborhood of λ.

Back to the proof of (2). Suppose that $\xi_\lambda(n-2) < \xi_\lambda(n-1)$. By Remark 3.2, it is easy to check that $T_{n-1}(\lambda) = 0$ and $T_n(\lambda) > 0$ or, equivalently, $\det A_{n-2}(\lambda) = 0$ and $\det A_{n-1}(\lambda) < 0$. According to the conclusion in the previous paragraph, we can find $\epsilon > 0$ such that $\det A_{n-2}(\cdot)$ is strictly decreasing on $(\lambda-\epsilon, \lambda+\epsilon)$, which yields

$$\xi_{\lambda+\delta}(n) = \xi_{\lambda+\delta}(n-1) - \frac{(\lambda+\delta)T_{n-1}(\lambda+\delta)}{\nu(n-1,n)} \begin{cases} < \xi_{\lambda+\delta}(n-1) & \text{if } 0 < \delta < \epsilon \\ > \xi_{\lambda+\delta}(n-1) & \text{if } -\epsilon < \delta < 0 \end{cases}.$$

This gives the desired property in Lemma 3.2 (2). The other case, $\xi_\lambda(n-2) > \xi_\lambda(n-1)$, can be proved in the same way and we omit the details. □

The following proposition characterizes the shape of ξ_λ for $\lambda > 0$.

Proposition 3.3. *For $\lambda > 0$, let ξ_λ be the function in* Definition 3.2. *Let $\lambda^G_1 < \cdots < \lambda^G_{n-1}$ be non-zero eigenvalues of $M^G_{\pi,\nu}$ in* (3.1) *and $\alpha_1 < \cdots < \alpha_{n-2}$ be zeros of* $\det A_{n-2}(\lambda)$*, where $A_{n-2}(\cdot)$ is the matrix in* (2.4)*. Then,*

(1) $\lambda^G_j < \alpha_j < \lambda^G_{j+1}$*, for* $1 \le j \le n-2$*.*
(2) *ξ_λ is of type j for $\lambda \in (\alpha_{j-1}, \alpha_j]$ and $1 \le j \le n-1$, where $\alpha_0 := 0$ and $\alpha_{n-1} := \infty$.*

Proof. (1) is immediate from Lemma A.2. For (2), note that α_i is an eigenvalue of the submatix of $M^G_{\pi,\nu}$ obtained by removing the nth row and column. This implies $\xi_{\alpha_i}(n-1) = \xi_{\alpha_i}(n)$ for $i = 1, ..., n-2$ and $\xi_\lambda(n-1) \ne \xi_\lambda(n)$ for $\lambda > 0$ and $\lambda \notin \{\alpha_1, ..., \alpha_{n-2}\}$. By Lemma 3.2, ξ_λ is of type i for $\alpha_{i-1} < \lambda \le \alpha_i$. □

Given $\lambda > 0$, the above proposition provides a simple criterion to determine to which of the intervals $(\alpha_j, \alpha_{j+1}]$ λ belongs to, that is, the type of ξ_λ. However, knowing the type of ξ_λ is not sufficient to determine whether λ is bigger or smaller than λ^G_i. We need the following remark.

Remark 3.3. Using the same argument as the proof of Proposition 2.5, one can show that $\pi(\xi_\lambda) = -\pi(1)\det A_{n-1}(\lambda)$, where $A_{n-1}(\lambda)$ is the matrix in (2.4). Clearly, $\pi(\xi_\lambda)$ has zeros $\lambda^G_1, ..., \lambda^G_{n-1}$ and tends to minus infinity as λ tends to minus infinity. This implies that $\pi(\xi_\lambda) < 0$, for $\lambda < \lambda^G_1$, and

$$\pi(\xi_\lambda) > 0 \quad \forall \lambda \in (\lambda^G_{2i-1}, \lambda^G_{2i}), \quad \pi(\xi_\lambda) < 0 \quad \forall \lambda \in (\lambda^G_{2i}, \lambda^G_{2i+1}),$$

for $i \ge 1$, where $\lambda_n^G := \infty$.

As a consequence of Proposition 3.3 and Remark 3.3, we obtain the following dichotomy algorithm, which is a generalization of (A2). Let $1 \le i \le n-1$.

Choose positive reals $L_0 < \lambda_i^G < U_0$ and set, for $\ell = 0, 1, ...,$

1. ξ_{λ_ℓ} be the function generated by $\lambda_\ell = (L_\ell + U_\ell)/2$ in Definition 3.2,
2. According to Definition 3.1, set

$$\text{(D}i\text{)}\qquad \begin{cases} L_{\ell+1} = L_\ell,\, U_{\ell+1} = \lambda_\ell & \text{if } \xi_{\lambda_\ell} \text{ is of type } j \text{ with } j > i, \\ & \text{or if } \xi_{\lambda_\ell} \text{ is of type } i \text{ and } (-1)^{i-1}\pi(\xi_{\lambda_\ell}) > 0 \\ U_{\ell+1} = U_\ell,\, L_{\ell+1} = \lambda_\ell & \text{if } \xi_{\lambda_\ell} \text{ is of type } j \text{ with } j < i, \\ & \text{or if } \xi_{\lambda_\ell} \text{ is of type } i \text{ and } (-1)^{i-1}\pi(\xi_{\lambda_\ell}) < 0 \\ L_{\ell+1} = U_{\ell+1} = \lambda_\ell & \text{if } \xi_{\lambda_\ell} \text{ is of type } i \text{ and } \pi(\xi_{\lambda_\ell}) > 0 \end{cases}.$$

Theorem 3.4. *Referring to* (Di),

$$0 \le \max\{U_\ell - \lambda_i^G, \lambda_i^G - L_\ell\} \le (U_0 - L_0)2^{-\ell}, \quad \forall \ell \ge 0.$$

Proof. Immediate from Proposition 3.3 and Remark 3.3. □

Proposition 3.3 (2) bounds the eigenvalues using the shape of ξ_λ generated from one end point. We now introduce some other criteria to bound eigenvalues using the shape of ξ_λ from either boundary point. Those results will be used to prove Theorem 6.1.

Proposition 3.5. *For $\lambda > 0$, let ξ_λ be the function in Definition 3.2 and $\widetilde{\xi}_\lambda$ be a function given by*

$$\widetilde{\xi}_\lambda(k-1) = \widetilde{\xi}_\lambda(k) + \frac{[\widetilde{\xi}_\lambda(k) - \widetilde{\xi}_\lambda(k+1)]\nu(k,k+1) - \lambda\pi(k)\widetilde{\xi}_\lambda(k)}{\nu(k-1,k)},$$

for $k = n, n-1, ..., 2$ with $\widetilde{\xi}_\lambda(n) = -1$. Let $\lambda_0^G < \cdots < \lambda_{n-1}^G$ be eigenvalues of $M_{\pi,\nu}^G$ in (3.1) *and let $f|_B$ be the restriction of f to a subset B of V. Suppose $1 \le k_0 \le n$.*

(1) *If $\xi_\lambda|_{\{1,...,k_0\}}$ is of type i with $(-1)^i\xi_\lambda(k_0) > 0$ and $\widetilde{\xi}_\lambda|_{\{k_0,...,n\}}$ is of type j with $(-1)^j\widetilde{\xi}_\lambda(k_0) > 0$, then $\lambda_{i+j-2}^G < \lambda < \lambda_{i+j-1}^G$.*

(2) *If $\xi_\lambda|_{\{1,...,k_0\}}$ is of type i with $(-1)^i\xi_\lambda(k_0) < 0$ and $\widetilde{\xi}_\lambda|_{\{k_0,...,n\}}$ is of type j with $(-1)^j\widetilde{\xi}_\lambda(k_0) < 0$, then $\lambda_{i+j-1}^G < \lambda < \lambda_{i+j+1}^G$.*

(3) *If $\xi_\lambda|_{\{1,...,k_0\}}$ is of type i with $(-1)^i\xi_\lambda(k_0) > 0$ and $\widetilde{\xi}_\lambda|_{\{k_0,...,n\}}$ is of type j with $(-1)^j\widetilde{\xi}_\lambda(k_0) < 0$, then $\lambda_{i+j-2}^G < \lambda < \lambda_{i+j}^G$.*

Proof. By Proposition 3.3, $\xi_\lambda(n)$ is a polynomial of degree $n-1$ satisfying

$$(-1)^{i+1}\xi_{\lambda_i^G}(n) > 0,\ \forall 0 \le i < n, \quad (-1)^{i+1}\xi_{\beta_i}(n) > 0,\ \forall 1 \le i < n-1.$$

This implies that there are $w_i \in (\beta_i, \lambda_{i+1}^G)$, $0 \le i \le n-2$, such that $(-1)^{i+1}\xi_\lambda(n) > 0$ for $\lambda \in (w_{i-1}, w_i)$ and $0 \le i \le n-1$ with $w_{-1} = -\infty$ and $w_{n-1} = \infty$.

The proofs for (1)-(3) in Proposition 3.5 are similar and we deal with (1) only. By the Euler-Lagrange equations in (3.2), it is easy to see that, for $1 \le l < n$, $\xi_{\lambda_l^G}$ and $\widetilde{\xi}_{\lambda_l^G}$ are eigenvectors of $M_{\pi,\nu}^G$ in (3.1) associated with λ_l^G, which implies $\xi_{\lambda_l^G} = -\xi_{\lambda_l^G}(n)\widetilde{\xi}_{\lambda_l^G}$. First, assume that $\lambda \le \lambda_{i+j-2}^G$. By Proposition 3.3, $\xi_{\lambda_{i+j-2}^G}|_{\{1,...,k_0\}}$

is of type at least i and $\widetilde{\xi}_{\lambda^G_{i+j-2}}|_{\{k_0,...,n\}}$ is of type at least j. This implies that the patching of $\xi_{\lambda^G_{i+j-2}}|_{\{1,...,k_0\}}$ and $-\xi_{\lambda^G_{i+j-2}}(n)\widetilde{\xi}_{\lambda^G_{i+j-2}}|_{\{k_0,...,n\}}$, which equals to $\xi_{\lambda^G_{i+j-2}}$, is of type at least $i+j-1$. This is a contradiction.

Next, assume that $\lambda \ge \lambda^G_{i+j-1}$. By Proposition 3.3, we may choose $a_1 < \lambda$ (resp. $a_2 < \lambda$) such that $\xi_\lambda|_{\{1,...,k_0\}}$ (resp. $\widetilde{\xi}_\lambda|_{\{k_0,...,n\}}$) changes the type at a_1 (resp. a_2). If $\lambda^G_{i+j-1} \le \min\{a_1, a_2\}$, then a similar reasoning as before implies that $\xi_{\lambda^G_{i+j-1}}$ is of type at most $i+j-2$, a contradiction. If $\min\{a_1,a_2\} < \lambda^G_{i+j-1} < \max\{a_1,a_2\}$, then exactly one of $\xi_{\lambda^G_{i+j-1}}|_{\{1,...,k_0\}}$ and $\widetilde{\xi}_{\lambda^G_{i+j-1}}|_{\{k_0,...,n\}}$ does not change its type. This implies that the gluing point k_0 can not be a local extremum and, thus, the patching function is of type at most $i+j-2$, another contradiction! According to the discussion in the first paragraph of this proof, if $\lambda^G_{i+j-1} \ge \max\{a_1,a_2\}$, then none of $\xi_{\lambda^G_{i+j-1}}|_{\{1,...,k_0\}}$ and $\widetilde{\xi}_{\lambda^G_{i+j-1}}|_{\{k_0,...,n\}}$ changes type nor, of course, the sign at k_0. Consequently, we obtain $(-1)^{i+j}\xi_{\lambda^G_{i+j-1}}(k_0)\widetilde{\xi}_{\lambda^G_{i+j-1}}(k_0) > 0$, which contradicts the fact $\xi_{\lambda^G_{i+j-1}} = -\xi_{\lambda^G_{i+j-1}}(n)\widetilde{\xi}_{\lambda^G_{i+j-1}}$. □

Proposition 3.6. *For $\lambda > 0$ and $1 \le k \le n-1$, let $s_k(\lambda)$ be the kth sign change of ξ_λ defined by $s_0 := 0$ and $s_{k+1}(\lambda) := \inf\{l > s_k(\lambda) | \xi_\lambda(l)\xi_\lambda(l-1) < 0$ or $\xi_\lambda(l) = 0\}$, where $\inf\emptyset := n+1$. Then, for $0 < \lambda_1 < \lambda_2$, $s_k(\lambda_1) \ge s_k(\lambda_2)$ for all $1 \le k \le n-1$.*

Proof. Let $1 \le k \le n-1$. If $s_k(\lambda_1) = n+1$, then it is clear that $s_k(\lambda_1) \ge s_k(\lambda_2)$. Suppose that $s_k(\lambda_1) = \ell \le n$. Obviously, $\xi_{\lambda_1}|_{\{1,...,\ell\}}$ is of type k. Referring to (2.4), let $\lambda^\ell_1, ..., \lambda^\ell_{\ell-1}$ be the roots of $\det A_{\ell-1}(\lambda) = 0$ and $\alpha^\ell_1, ..., \alpha^\ell_{\ell-2}$ be roots of $\det A_{\ell-2}(\lambda) = 0$. According to the first paragraph of the proof for Proposition 3.5, there are $w^\ell_i \in (\alpha^\ell_{i-1}, \lambda^\ell_i)$ with $1 \le i \le \ell-1$ such that $(-1)^{i+1}\xi_\lambda(\ell) > 0$ for $\lambda \in (w^\ell_i, w^\ell_{i+1})$ and $1 \le i \le \ell-1$, where $\alpha^\ell_0 := 0$. Since $\xi_{\lambda_1}(\ell)\xi_{\lambda^\ell_k}(\ell) \ge 0$, one has $w^\ell_k \le \lambda_1 < \alpha^\ell_k$. As it is assumed that $\lambda_2 > \lambda_1$, if $\lambda_2 > \alpha^\ell_k$, then $\xi_{\lambda_2}|_{\{1,...,\ell\}}$ is of type at least $k+1$ and, consequently, $s_k(\lambda_2) < \ell = s_k(\lambda_1)$. If $\lambda 1 < \alpha^\ell_k$, then $\xi_{\lambda_2}|_{\{1,...,\ell\}}$ is type k and $\xi_{\lambda_2}(\ell) < 0$. This implies $s_k(\lambda_2) \le \ell = s_k(\lambda_1)$, as desired. □

3.2. **Bounding eigenvalues from below.** Motivated by Theorem 3.1, we introduce another scheme generalizing (2.1) to bound the other eigenvalues of $M^G_{\pi,\nu}$ from below.

Definition 3.3. For $\lambda > 0$, let ξ_λ be a function in Definition 3.2. If ξ_λ is of type i, $1 \le i \le n-1$, with peak-valley points $1 = a_1 < b_1 \le a_2 < b_2 \le \cdots \le a_i < b_i \le n$, then define

$$\xi^{(j)}_\lambda(k) = \begin{cases} \xi_\lambda(k) & \text{for } k \le b_j \\ \xi_\lambda(k) = \xi_\lambda(b_j) & \text{for } k > b_j \end{cases}, \quad \forall 1 \le j < i$$

and set $\xi^{(j)}_\lambda = \xi_\lambda$ for $i \le j \le n-1$.

Remark 3.4. For $\lambda > 0$, if ξ_λ is of type i, then $\xi_\lambda^{(j)}$ is of type j for $j < i$. Moreover, for $k < b_j$,

$$\begin{aligned}\xi_\lambda^{(j)}(k+1) &= \xi_\lambda^{(j)}(k) + \frac{[\xi_\lambda^{(j)}(k) - \xi_\lambda^{(j)}(k-1)]\nu(k-1,k) - \lambda\pi(k)\xi_\lambda^{(j)}(k)}{\nu(k,k+1)} \\ &= \xi_\lambda^{(j)}(k) - \frac{\lambda[\pi(1)\xi_\lambda^{(j)}(1) + \cdots + \pi(k)\xi_\lambda^{(j)}(k)]}{\nu(k,k+1)},\end{aligned}$$

and, for $b_j \le k < n$,

$$\xi_\lambda^{(j)}(k+1) = \xi_\lambda^{(j)}(k) + \frac{F_j([\xi_\lambda^{(j)}(k) - \xi_\lambda^{(j)}(k-1)]\nu(k-1,k) - \lambda\pi(k)\xi_\lambda^{(j)}(k))}{\nu(k,k+1)},$$

where $F_j(t) = \max\{t, 0\}$ if j is odd, and $F_j(t) = \min\{t, 0\}$ if j is even. Note that $\xi_\lambda^{(1)}$ is exactly ϕ_λ in Proposition 2.5.

Thereafter, let $\mathcal{L}$ and $\mathcal{L}^{(i)}$ be functions on $(0, \infty)$ defined by

$$\mathcal{L}(\lambda) = \frac{\mathcal{E}_\nu(\xi_\lambda, \xi_\lambda)}{\mathrm{Var}_\pi(\xi_\lambda)}, \quad \mathcal{L}^{(i)}(\lambda) = \frac{\mathcal{E}_\nu(\xi_\lambda^{(i)}, \xi_\lambda^{(i)})}{\mathrm{Var}_\pi(\xi_\lambda^{(i)})}, \quad \forall 1 \le i \le n-1, \tag{3.3}$$

where ξ_λ and $\xi_\lambda^{(i)}$ are functions in Definitions 3.2-3.3.

Remark 3.5. Note that $\mathcal{L} = \mathcal{L}^{(n-1)}$. By a similar reasoning as in the proof for (2.2), one can show that, for $\lambda > 0$,

$$\mathcal{L}(\lambda) = \lambda + \frac{\lambda\pi(\xi_\lambda)[\pi(\xi_\lambda) - \xi_\lambda(n)]}{\mathrm{Var}_\pi(\xi_\lambda)}, \quad \mathcal{L}^{(i)}(\lambda) = \lambda + \frac{\lambda\pi(\xi_\lambda^{(i)})[\pi(\xi_\lambda^{(i)}) - \xi_\lambda^{(i)}(n)]}{\mathrm{Var}_\pi(\xi_\lambda^{(i)})}.$$

From Proposition 3.3, it followss immediately that $\mathcal{L}(\lambda) = \mathcal{L}^{(i)}(\lambda)$ for $\lambda \in (0, \alpha_i]$.

To explore further $\mathcal{L}$ and $\mathcal{L}^{(i)}$, we need more information of $\pi(\xi_\lambda)$, $\pi(\xi_\lambda^{(i)})$, $\pi(\xi_\lambda) - \xi_\lambda(n)$ and $\pi(\xi_\lambda^{(i)}) - \xi_\lambda^{(i)}(n)$.

Lemma 3.7. *Let ξ_λ be the function in* Definition 3.2 *and λ_i^G, α_i be constants in* Proposition 3.3. *Then, $\pi(\xi_\lambda) - \xi_\lambda(n) = 0$ has $n-1$ distinct roots, say $\beta_0 < \beta_1 < \cdots < \beta_{n-2}$, which satisfy $\beta_0 = 0$ and $\alpha_i < \beta_i < \lambda_{i+1}^G$ for $1 \le i \le n-2$. Furthermore, $\pi(\xi_\lambda) - \xi_\lambda(n) > 0$ for $\lambda \in (\beta_{2i-1}, \beta_{2i})$ and $\pi(\xi_\lambda) - \xi_\lambda(n) < 0$ for $\lambda \in (\beta_{2i}, \beta_{2i+1})$, with $\beta_{-1} = -\infty$ and $\beta_{n-1} = \infty$.*

Proof. Set $u(\lambda) := \pi(\xi_\lambda) - \xi_\lambda(n)$. According to Definition 3.2, $u(\lambda)$ is a polynomial of degree $n-1$ and satisfies $u(0) = 0$. Note that $\pi(\xi_\lambda) = 0$ for $\lambda \in \{\lambda_1^G, ..., \lambda_{n-1}^G\}$. If i is odd, then $\xi_{\lambda_i^G}(n-1) < \xi_{\lambda_i^G}(n)$. This implies $\xi_{\lambda_i^G}(n) > 0$ and, hence, $u(\lambda_i^G) < 0$. Similarly, if i is even, then $u(\lambda_i^G) > 0$.

By Lemma 3.2 and Proposition 3.3, if $\lambda = \alpha_i$ with odd i, then ξ_{α_i} is of type i with $\xi_{\alpha_i}(n-1) = \xi_{\alpha_i}(n)$. This implies $\xi_{\alpha_i}(n) > 0$ and $\pi(\xi_{\alpha_i}) = \pi(n)\xi_{\alpha_i}(n)$, which yields $u(\alpha_i) < 0$. Similarly, one can show that $u(\alpha_i) > 0$ if i is even. □

Remark 3.6. We consider the sign of $\pi(\xi_\lambda^{(i)})$ and $\pi(\xi_\lambda^{(i)}) - \xi_\lambda^{(i)}(n)$ in this remark. By Proposition 3.3, $\xi_\lambda^{(i)} = \xi_\lambda$ for $\lambda \le \alpha_i$. If $\lambda > \alpha_i$ with $1 \le i \le n-2$, then ξ_λ is of type j with $j > i$. Fix $1 \le i \le n-2$ and set $k_0 = k_0(\lambda) = \min\{k|\xi_\lambda^{(i)}(j) = \xi_\lambda^{(i)}(n), \forall k \le j \le n\}$. Clearly, $k_0(\lambda) \le n-1$ for $\lambda > \alpha_i$. Observe that, for $\lambda > \alpha_i$ with odd i, $\xi_\lambda(k_0 - 1) < \xi_\lambda(k_0) \ge \xi_\lambda(k_0 + 1)$, which implies

$\sum_{k=1}^{k_0-1}\pi(k)\xi_\lambda(k) < 0$ and $\sum_{k=1}^{k_0}\pi(k)\xi_\lambda(k) \geq 0$. A similar reasoning for the case of even i gives $\sum_{k=1}^{k_0-1}\pi(k)\xi_\lambda(k) > 0$ and $\sum_{k=1}^{k_0}\pi(k)\xi_\lambda(k) \leq 0$. Consequently, we obtain

$$(-1)^{i-1}\pi(\xi_\lambda^{(i)}) > 0, \quad (-1)^i[\pi(\xi_\lambda^{(i)}) - \xi_\lambda^{(i)}(n)] > 0, \tag{3.4}$$

for $\lambda > \alpha_i$ and $1 \leq i \leq n-2$. Note that, by Proposition 3.3, $\xi_\lambda^{(i)} = \xi_\lambda$ for $\lambda \leq \alpha_i$. In addition with Remark 3.3, Lemma 3.7 and the continuity of $\xi_\lambda^{(i)}$, the first inequality of (3.4) holds for $\lambda > \lambda_i^G$ and the second inequalities of (3.4) hold for $\lambda > \beta_{i-1}$.

According to Lemma 3.7 and Remark 3.6, we derive a generalized version of Proposition 2.3 in the following.

Proposition 3.8. *Let $n \geq 3$ and $1 \leq i \leq n-1$. For $\lambda > 0$, let $\xi_\lambda, \xi_\lambda^{(i)}$ be the functions in* Definition 3.2 *and β_i be the constants in* Lemma 3.7*.*

(1) *For $\lambda > \beta_{i-1}$, the following are equivalent.*
  (1-1) $\mathcal{E}_\nu(\xi_\lambda^{(i)}, \xi_\lambda^{(i)}) = \lambda \mathrm{Var}_\pi(\xi_\lambda^{(i)})$.
  (1-2) $\pi(\xi_\lambda^{(i)}) = 0$.
  (1-3) $\lambda = \lambda_i^G$.

(2) *For $\beta_{i-1} < \lambda < \beta_i$, the following are equivalent.*
  (2-1) $\mathcal{E}_\nu(\xi_\lambda, \xi_\lambda) = \lambda \mathrm{Var}_\pi(\xi_\lambda)$.
  (2-2) $\pi(\xi_\lambda) = 0$.
  (2-3) $\lambda = \lambda_i^G$.

Proof. The proof for Proposition 3.8 (2) is similar to the proof for Proposition 3.8 (1) and we deal only with the latter. By Lemma 3.7 and Remark 3.6, one has

$$\pi(\xi_\lambda^{(i)})[\pi(\xi_\lambda^{(i)}) - \xi_\lambda^{(i)}(n)] \begin{cases} < 0 & \text{for } \lambda > \lambda_i^G \\ > 0 & \text{for } \beta_{i-1} < \lambda < \lambda_i^G \end{cases}.$$

This proves the equivalence of (1-1) and (1-2). Under the assumption of (1-2) and using Remark 3.3, one has $\lambda \leq \alpha_i$. This implies $\xi_\lambda^{(i)} = \xi_\lambda$ is an eigenvector for $M_{\pi,\nu}^G$ with associated eigenvalue λ. As $\lambda \in (\beta_{i-1}, \alpha_i]$, it must be the case $\lambda = \lambda_i^G$. This gives (1-3), while (1-3)⇒(1-2) is obvious and omitted. □

Remark 3.7. It is worthwhile to note that if (1-1) and (2-1) of Proposition 3.8 are removed, then the equivalence in (1) holds for $\lambda > \lambda_{i-1}^G$ and the equivalence in (2) holds for $\lambda \in (\lambda_{i-1}^G, \lambda_{i+1}^G)$. Once λ_{i-1}^G is known, we can determine λ_i^G using the sign of $\pi(\xi_\lambda^{(i)})$. See Theorem 3.9 for details.

Remark 3.8. Note that condition (4) of Proposition 2.3 is not included in Proposition 3.8. In fact, the equivalence may fail, that is, there may exist some $\lambda \in (\beta_{i-1}, \beta_i) \setminus \{\lambda_i^G\}$ such that $\mathcal{E}_\nu(\xi_\lambda, \xi_\lambda)/\mathrm{Var}_\pi(\xi_\lambda) = \lambda_i^G$. See Example 3.2 for a counterexample.

As Proposition 3.8 focuses on the characterization of zeros of $\mathcal{L}(\lambda) - \lambda$, the following theorem concerns the sign of $\mathcal{L}(\lambda) - \lambda$.

Theorem 3.9. *Let $\lambda_i^G, \alpha_i, \beta_i$ be the constants in* Proposition 3.3 *and* Lemma 3.7, *and $\mathcal{L}$ be the function in* (3.3)*. Then, $\lambda_1^G, ..., \lambda_{n-1}^G, \beta_1, ..., \beta_{n-2}$ are fixed points of $\mathcal{L}$ and, for $1 \leq i \leq n-2$,*

(1) *$\mathcal{L}(\lambda) < \lambda$ for $\lambda \in (\lambda_i^G, \beta_i)$.*

(2) $\mathcal{L}(\lambda) > \lambda$ *for* $\lambda \in (\beta_i, \lambda_{i+1}^G)$.
(3) $\mathcal{L}^{(i)}(\lambda) < \lambda$ *for* $\lambda \in (\lambda_i^G, \infty)$.

Proof. Immediate from Lemma 3.7 and Remarks 3.5-3.6. □

By Theorem 3.9, we obtain a lower bound on any specified eigenvalue of $M_{\pi,\nu}^G$.

Corollary 3.10. *Let* $1 \le i \le n-1$ *and* $\lambda_0 > \lambda_i^G$. *Consider the sequence* $\lambda_{\ell+1} = \mathcal{L}^{(i)}(\lambda_\ell)$ *with* $\ell \ge 0$ *and set*

$$\lambda^* = \begin{cases} \lim_{\ell\to\infty} \lambda_\ell & \text{if } \lambda_\ell \text{ converges} \\ \sup_{\ell\in I} \lambda_\ell & \text{if } \lambda_\ell \text{ diverges} \end{cases},$$

where $I = \{\ell | \lambda_{\ell-1} > \lambda_\ell < \lambda_{\ell+1}\}$. *Then,* $\lambda^* \le \lambda_i^G$.

It is not clear yet whether the sequence λ_ℓ in Corollary 3.10 is convergent, even locally. This subject will be discussed in the next subsection. Now, we establish some relations between the roots of $\det A_i(\lambda) = 0$ and the shape of $\xi_\lambda^{(i)}$. This is a generalization of Proposition 2.5.

Proposition 3.11. *For* $1 \le i \le n-1$, *let* $A_i(\lambda)$ *be the matrix in* (2.4), $\theta_1^{(i)} < \cdots < \theta_i^{(i)}$ *be zeros of* $\det A_i(\lambda) = 0$ *and set* $\theta_i^{(i-1)} := \infty$. *Referring to the notation in* Proposition 3.3, *it holds true that, for* $1 \le i \le n-1$,

(1) $\lambda_i^G = \theta_i^{(n-1)} < \alpha_i = \theta_i^{(n-2)} < \cdots < \theta_i^{(i)}$.
(2) $\xi_\lambda^{(i)}(j) \neq \xi_\lambda^{(i)}(j+1) = \cdots = \xi_\lambda^{(i)}(n)$ *for* $\lambda \in [\theta_i^{(j)}, \theta_i^{(j-1)})$ *and* $i \le j \le n-2$.
(3) $\xi_\lambda^{(i)}(n-1) \neq \xi_\lambda^{(i)}(n)$ *for* $\lambda \in (\theta_{i-1}^{(n-2)}, \theta_i^{(n-2)})$ *and* $i \le n-1$.

Proof. The order in (1) is a simple application of Lemma A.3. For (2), fix $1 \le i \le n-1$ and set $\gamma(\lambda) = \min\{j | \xi_\lambda^{(i)}(k) = \xi_\lambda^{(i)}(n), \forall j \le k \le n\}$ and $B(\lambda) = \{1, 2, ..., \gamma(\lambda)\}$, $B^+(\lambda) = B(\lambda) \cup \{\gamma(\lambda)+1\}$. Clearly, $i+1 \le \gamma(\lambda) \le n$. We use the notation $\xi_\lambda|_C$ to denote the restriction of ξ_λ to a set C. Suppose that i is odd. By Remark 3.4, $\xi_\lambda^{(i)} = \xi_\lambda$ on $B(\lambda)$ and $\xi_\lambda|_{B(\lambda)}$ is of type i with

$$\xi_\lambda(\gamma(\lambda)-1) < \xi_\lambda(\gamma(\lambda)) \ge \xi_\lambda(\gamma(\lambda)+1).$$

By Lemma 3.2(1), if $\xi_\lambda(\gamma(\lambda)+1) < \xi_\lambda(\gamma(\lambda))$, then there is $\epsilon > 0$ such that, for $|\delta| < \epsilon$, $\xi_{\lambda+\delta}|_{B(\lambda)}$ is of type i and

$$\xi_{\lambda+\delta}(\gamma(\lambda)-1) < \xi_{\lambda+\delta}(\gamma(\lambda)) > \xi_{\lambda+\delta}(\gamma(\lambda)+1).$$

This implies $\gamma(\lambda+\delta) = \gamma(\lambda)$ for $\delta \in (-\epsilon, \epsilon)$. By Lemma 3.2(2), if $\xi_\lambda(\gamma(\lambda)+1) = \xi_\lambda(\gamma(\lambda))$, then there is $\epsilon > 0$ such that, for $\delta \in (-\epsilon, 0)$, $\xi_{\lambda+\delta}|_{B^+(\lambda)}$ is of type i with

$$\xi_{\lambda+\delta}(\gamma(\lambda)-1) < \xi_{\lambda+\delta}(\gamma(\lambda)) < \xi_{\lambda+\delta}(\gamma(\lambda)+1),$$

and, for $\delta \in (0, \epsilon)$, $\xi_{\lambda+\delta}|_{B^+(\lambda)}$ is of type $i+1$ with

$$\xi_{\lambda+\delta}(\gamma(\lambda)-1) < \xi_{\lambda+\delta}(\gamma(\lambda)) > \xi_{\lambda+\delta}(\gamma(\lambda)+1).$$

This yields $\gamma(\lambda+\delta) = \gamma(\lambda)$ for $\delta \in (0, \epsilon)$ and $\gamma(\lambda+\delta) = \gamma(\lambda)+1$ for $\delta \in (-\epsilon, 0)$. The proof for the case of even i is similar and we conclude from the above that $\gamma(\lambda)$ is a non-increasing and right-continuous function taking values on $\{i+1, ..., n\}$. Let $c_{i+1} > \cdots > c_{n-1}$ be the discontinuous points of $\gamma(\lambda)$ such that $\gamma(c_j) = j$ for $i+1 \le j \le n-1$. As a consequence of the above discussion, $\xi_{c_j}|_{\{1,...,j\}}$ is of type i with $\xi_{c_j}(j) = \xi_{c_j}(j+1)$ and this implies $\sum_{k=1}^j \pi(k)\xi_{c_j}(k) = 0$. That means c_j is a root of $\det A_{j-1}(\lambda) = 0$ for $j = i+1, ..., n-1$. By Proposition 3.3 and the second equality in (1), $\gamma(\lambda) = n$ for $\theta_{i-1}^{(n-2)} < \lambda < \theta_i^{(n-2)}$ and, thus, $c_j \geq \theta_i^{(n-2)}$ for $j \geq i+1$. As a consequence of the interlacing relationship $\theta_i^{(\ell)} < \theta_i^{(\ell-1)} < \theta_{i+1}^{(\ell)}$, it must be $c_j = \theta_i^{(j+1)}$ for $i+1 \leq j \leq n-1$. This finishes the proof. □

Remark 3.9. For $1 \leq i \leq n-1$, $\theta_1^{(i)}, ..., \theta_i^{(i)}$ are also non-zero eigenvalues of the $(i+1) \times (i+1)$ principal submatrix of (3.1) indexed by $1, ..., i+1$.

Remark 3.10. In fact, by Proposition 2.5, $\xi_\lambda^{(1)}(n-1) \neq \xi_\lambda^{(1)}(n)$ for $\lambda \in (0, \theta_1^{(n-2)})$, which is better than Proposition 3.11(3).

3.3. **Local convergence of $\mathcal{L}$.** This subsection is dedicated to the local convergence of $\mathcal{L}$ in (3.3). Let $\alpha_i, \beta_i, \lambda_i^G$ be the constants in Proposition 3.3 and Lemma 3.7. As before, let $\zeta_0 = \mathbf{1}, ..., \zeta_{n-1}$ denote the $L^2(\pi)$-normalized eigenvectors of $M_{\pi,\nu}^G$ associated with $\lambda_0^G, ..., \lambda_{n-1}^G$. Clearly, $\xi_{\lambda_i^G} = -\zeta_i/\zeta_i(1)$ and $\xi_\lambda = \sum_{i=0}^{n-1} \rho_i(\lambda)\zeta_i$, where $\rho_i(\lambda) = \pi(\xi_\lambda \zeta_i)$ for $0 \leq i \leq n-1$. Note that $\rho_i(\lambda)$ is a polynomial of degree $n-1$ and satisfies $\rho_i(\lambda_j) = -\delta_i(j)/\zeta_i(1)$ for $i, j \in \{0, 1, ..., n-1\}$. This implies

$$(3.5) \qquad \rho_0(\lambda) = -\prod_{j=1}^{n-1} \frac{\lambda_j^G - \lambda}{\lambda_j^G}, \quad \rho_i(\lambda) = -\frac{\lambda}{\zeta_i(1)\lambda_i^G} \prod_{j=1, j\neq i}^{n-1} \frac{\lambda_j^G - \lambda}{\lambda_j^G - \lambda_i^G},$$

for all $1 \leq i \leq n-1$. Moreover, by multiplying (3.2) with $\xi_\lambda(k)$ and summing up k, we obtain $\mathcal{E}_\nu(\xi_\lambda, \zeta_i) = \lambda_i^G \rho_i(\lambda)$. In the same spirit, one can show that $\mathcal{E}_\nu(\xi_\lambda, \zeta_i) = \lambda[\rho_i(\lambda) - \zeta_i(n)\rho_0(\lambda)]$ using Definition 3.2. Putting both equations together yields

$$(3.6) \qquad \rho_i(\lambda) = \frac{\lambda \zeta_i(n)}{\lambda - \lambda_i^G} \rho_0(\lambda), \quad \forall 0 \leq i \leq n-1.$$

As a consequence of Remark 3.5, this gives

$$(3.7) \qquad \mathcal{L}(\lambda) = \frac{\sum_{i=1}^{n-1} \lambda_i^G \rho_i^2(\lambda)}{\sum_{i=1}^{n-1} \rho_i^2(\lambda)} = \lambda + \frac{\sum_{i=1}^{n-1} (\lambda_i^G - \lambda)^{-1} \zeta_i^2(n)}{\sum_{i=1}^{n-1} (\lambda_i^G - \lambda)^{-2} \zeta_i^2(n)},$$

for $\lambda \notin \{\lambda_0^G, ..., \lambda_{n-1}^G\}$. The next proposition follows immediately from the second equation in (3.5) and (3.6).

Proposition 3.12. *Let $\lambda_1^G, ..., \lambda_{n-1}^G$ be the non-zero eigenvalues of $M_{\pi,\nu}^G$ in* (3.1) *and $\zeta_1, ..., \zeta_{n-1}$ be the corresponding $L^2(\pi)$-normalized eigenvectors. Then,*

$$\zeta_i(1)\zeta_i(n) = -\prod_{j=1, j\neq i}^{n-1} \frac{\lambda_j^G}{\lambda_j^G - \lambda_i^G}, \quad \forall 1 \leq i \leq n-1.$$

Set $u(\lambda) = \sum_{j=1}^{n-1} (\lambda_j^G - \lambda)^{-1} \zeta_j^2(n)$. By Theorem 3.9, $\beta_1, ..., \beta_{n-2}$ are zeros of $u(\lambda) \prod_{j=1}^{n-1} (\lambda_j^G - \lambda)$, which is a polynomial of degree $n-2$. This implies

$$u(\lambda) = C \left(\prod_{j=1}^{n-1} \frac{1}{\lambda_j^G - \lambda} \right) \left(\prod_{j=1}^{n-2} (\beta_j - \lambda) \right),$$

where $C = \frac{\lambda_1 \cdots \lambda_{n-1}}{\beta_1 \cdots \beta_{n-2}} \sum_{j=1}^{n-1} \zeta_j^2(n)/\lambda_j^G$. Putting this back to $\mathcal{L}$ yields

$$(3.8) \qquad \frac{1}{\mathcal{L}(\lambda) - \lambda} = \frac{u'(\lambda)}{u(\lambda)} = \sum_{j=1}^{n-1} \frac{1}{\lambda_j^G - \lambda} - \sum_{j=1}^{n-2} \frac{1}{\beta_j - \lambda},$$

for $\lambda \notin \{\lambda_0^G, ..., \lambda_{n-1}^G, \beta_1, ..., \beta_{n-2}\}$.

Proposition 3.13. *Let $\mathcal{L}$ be the function in* (3.3), *λ_i^G be the eigenvalue of $M_{\pi,\nu}^G$ and β_i be the constant in* Lemma 3.7. *Let $D_i = \sum_{j=1}^{n-2}(\beta_j - \lambda_i^G)^{-1} - \sum_{j=1,j\neq i}^{n-1}(\lambda_j^G - \lambda_i^G)^{-1}$ with $1 \le i \le n-1$. Then, for $2 \le i \le n-2$,*

(1) *If $D_i < 0$, then there is $\tau \in (\lambda_i^G, \beta_i)$ such that $\mathcal{L}$ is strictly increasing on $(\beta_{i-1}, \lambda_i^G) \cup (\tau, \beta_i)$ and strictly decreasing on (λ_i^G, τ).*
(2) *If $D_i > 0$, then there is $\eta \in (\beta_{i-1}, \lambda_i^G)$ such that $\mathcal{L}$ is strictly increasing on $(\beta_{i-1}, \eta) \cup (\lambda_i^G, \beta_i)$ and strictly increasing on (η, λ_i^G).*
(3) *If $D_i = 0$, then $\mathcal{L}$ is strictly increasing on (β_{i-1}, β_i).*

Proof. Using (3.7) and (3.8), one can show that $\mathcal{L}'(\lambda_i^G) = 0$ and

$$\mathcal{L}''(\lambda_i^G) = \sum_{j=1,j\neq i}^{n-1} \frac{\zeta_i^2(n)}{\lambda_j^G - \lambda_i^G} = 2\left[\sum_{j=1}^{n-2} \frac{1}{\beta_j - \lambda_i^G} - \sum_{j=1,j\neq i}^{n-1} \frac{1}{\lambda_j^G - \lambda_i^G}\right] = 2D_i. \tag{3.9}$$

To prove (1) and (2), it suffices to show that if $\mathcal{L}'(\tau) = 0$ for some $\tau \in (\lambda_i^G, \beta_i)$, then τ is a local minimum of $\mathcal{L}$, and if $\mathcal{L}'(\eta) = 0$ for some $\eta \in (\beta_{i-1}, \lambda_i^G)$, then η is a local maximum of $\mathcal{L}$. We discuss the first case, whereas the second case is similar and is omitted. Recall that $u(\lambda) = \sum_{j=1}^{n-1}(\lambda_j^G - \lambda)^{-1}\zeta_j^2(n)$. As τ is a critical point for $\mathcal{L}$, one has $2(u'(\tau))^2 = u(\tau)u''(\tau)$. This implies

$$\mathcal{L}''(\tau) = \frac{u(\tau)[3(u''(\tau))^2 - 2u'(\tau)u'''(\tau)]}{2(u'(\tau))^3} > 0,$$

where the last inequality uses the fact that $u(\lambda) < 0$, for $\lambda \in (\lambda_i^G, \beta_i)$, and

$$3(u''(\lambda))^2 - 2u'(\lambda)u'''(\lambda) = -12 \sum_{1\le i<j\le n-1} \left[\frac{(\lambda_i^G - \lambda_j^G)\zeta_i(n)\zeta_j(n)}{(\lambda_i^G - \lambda)^2(\lambda_j^G - \lambda)^2}\right]^2 < 0.$$

This proves (1) and (2).

To see (3), we assume that $D_i = 0$. Computations show that

$$\begin{aligned}\frac{\mathcal{L}(\lambda) - \lambda_i^G}{\mathcal{L}(\lambda) - \lambda} &= (\lambda - \lambda_i^G)\left[\sum_{j=1,j\neq i}^{n-1} \frac{1}{\lambda_j^G - \lambda} - \sum_{j=1}^{n-2} \frac{1}{\beta_j - \lambda}\right] \\ &= (\lambda - \lambda_i^G)^2\left[\sum_{j=1,j\neq i}^{n-1} \frac{1}{(\lambda_j^G - \lambda)(\lambda_j^G - \lambda_i^G)} - \sum_{j=1}^{n-1} \frac{1}{(\beta_j - \lambda)(\beta_j - \lambda_i^G)}\right] < 0,\end{aligned}$$

for $\lambda \in (\beta_{i-1}, \lambda_i^G) \cup (\lambda_i^G, \beta_i)$, where the last inequality uses the fact that $(\lambda_j^G - \lambda)(\lambda_j^G - \lambda_i^G) > (\beta_j - \lambda)(\beta_j - \lambda_i^G)$ for $j < i$ and $(\lambda_j^G - \lambda)(\lambda_j^G - \lambda_i^G) > (\beta_{j-1} - \lambda)(\beta_{j-1} - \lambda_i^G)$ for $j > i$. By Theorem 3.9, this implies $\mathcal{L}(\lambda) > \lambda_i^G$ for $\lambda \in (\lambda_i^G, \beta_i)$ and $\mathcal{L}(\lambda) < \lambda_i^G$ for $\lambda \in (\beta_{i-1}, \lambda_i^G)$. The desired property comes immediate from the discussion in the previous paragraph. □

Remark 3.11. Note that $D_1 > 0$ and $D_{n-1} < 0$. Using the same proof as above, this implies that $\mathcal{L}(\lambda)$ is strictly increasing on $(\lambda_1^G, \beta_1) \cup (\beta_{n-2}, \lambda_{n-1}^G)$. Moreover, by (3.7), one may compute

$$(u'(\lambda))^2\mathcal{L}'(\lambda) = -2\sum_{i<j} \frac{(\lambda_i^G - \lambda_j^G)^2}{(\lambda_i^G - \lambda)^3(\lambda_j^G - \lambda)^3} < 0, \quad \forall \lambda \in (0, \lambda_1^G) \cup (\lambda_{n-1}^G, \infty).$$

This implies $\mathcal{L}(\lambda)$ is strictly decreasing on $(0, \lambda_1^G) \cup (\lambda_{n-1}^G, \infty)$ and

$$\lim_{\lambda\to 0} \mathcal{L}(\lambda) = \frac{\sum_{i=1}^{n-1} \zeta_i^2(n)/\lambda_i^G}{\sum_{i=1}^{n-1} \zeta_i^2(n)/(\lambda_i^G)^2}, \quad \lim_{\lambda\to\infty} \mathcal{L}(\lambda) = \left(\frac{1}{\pi(n)} - 1\right) \sum_{i=1}^{n-1} \lambda_i^G \zeta_i^2(n).$$

The following local convergence is a simple corollary of Theorem 3.9 and Proposition 3.13.

Theorem 3.14 (Local convergence)**.** *Let* $\lambda_0 > 0$ *and set* $\lambda_{\ell+1} = \mathcal{L}(\lambda_\ell)$ *for* $\ell \geq 0$. *Then, there is* $\epsilon > 0$ *such that the sequence* $(\lambda_\ell)_{\ell=1}^\infty$ *is monotonic and converges to* λ_i^G *for* $\lambda_0 \in (\lambda_i^G - \epsilon, \lambda_i^G + \epsilon)$ *and* $1 \leq i \leq n-1$.

We use the following examples to illustrate the different cases in Proposition 3.13.

Example 3.1 (Simple random walks). Let $n > 1$. A simple random walk on $\{1, 2, ..., n\}$ with reflecting probability $1/2$ at the boundary is a birth and death chain with transition matrix given by $K(i,j) = K(1,1) = K(n,n) = 1/2$ for $|i-j| = 1$. It is easy to see that the uniform probability is the stationary distribution of K. In the setting of graph, we have $\nu(i, i+1) = 1/(2n)$ and $\pi(i) = 1/n$. One may apply the method in [11] to obtain the following spectral information.

$$\lambda_j^G = 1 - \cos\frac{j\pi}{n}, \quad \zeta_j(k) = \frac{1}{\sqrt{\lambda_j^G}}\left(\sin\frac{jk\pi}{n} - \sin\frac{j(k-1)\pi}{n}\right), \quad \forall 1 \leq j < n.$$

See, e.g., [3, Section 7]. By (3.9), we get

$$D_i = \frac{1}{2}\sum_{j=1, j\neq i}^{n-1} \frac{\sin^2(j\pi/n)}{\lambda_j^G(\lambda_j^G - \lambda_i^G)} = \sum_{j=1, j\neq i}^{n-1} \frac{1 + \cos(j\pi/n)}{\cos(i\pi/n) - \cos(j\pi/n)}.$$

Clearly, $D_1 > 0$ and $D_{n-1} < 0$. If n is even, then $D_{n/2} < 0$.

Example 3.2 (Ehrenfest chains). An Ehrenfest chain on $V = \{0, 1, ..., n\}$ is a Markov chain with transition matrix K given by $K(i, i+1) = 1 - i/n$ and $K(i+1, i) = (i+1)/n$ for $i = 0, ..., n-1$. The associated stationary distribution is the unbiased binomial distribution on V, that is, $\pi(i) = \binom{n}{i} 2^{-n}$ for $i \in V$. To the Ehrenfest chain, the measure ν is defined by $\nu(i, i+1) = \binom{n-1}{i} 2^{-n}$ for $i = 0, ..., n-1$. Using the group representation for the binary group $\{0,1\}^n$, one may compute

$$\lambda_j = \frac{2j}{n}, \quad \zeta_j(k) = \binom{n}{j}^{-1/2} \sum_{\ell=0}^{j} (-1)^\ell \binom{k}{\ell}\binom{n-k}{j-\ell}, \quad \forall 1 \leq j \leq n.$$

Plugging this back into (3.9) yields

$$D_i = \frac{n}{4}\sum_{j=1, j\neq i}^{n} \frac{\binom{n}{j}}{j-i} \begin{cases} > 0 & \text{for } i < n/2 \\ = 0 & \text{for } i = n/2 \\ < 0 & \text{for } i > n/2. \end{cases}.$$

This example points out the possibility of different signs in $\{D_i | i = 1, ..., n-1\}$ including 0.

3.4. **A remark on the separation for birth and death chains.** In this subsection, we give a new proof of a result, Theorem 3.15, which deals with convergence in separation distance for birth and death chains. Let $(X_m)_{m=0}^{\infty}$ be a birth and death chain with transition matrix K given by (1.1). In the continuous time setting, we consider the process $Y_t = X_{N_t}$, where N_t is a Poisson process with parameter 1 independent of X_m. Given the initial distribution μ, which is the distribution of X_0, the distributions of X_m and Y_t are respectively μK^m and $\mu e^{-t(I-K)}$, where $e^A := \sum_{l=0}^{\infty} A^l/l!$. Briefly, we write $H_t = e^{-t(I-K)}$. It is well-known that if K is irreducible, then μH_t converges to π as $t \to \infty$. If K is irreducible and $r_i > 0$ for some i, then μK^m converges to π as $m \to \infty$. Concerning the convergence, we consider the separations of X_m, Y_t with respect to π, which are defined by

$$d_{\mathrm{sep}}(\mu, m) = \max_{0 \le x \le n} \left\{ 1 - \frac{\mu K^m(x)}{\pi(x)} \right\}, \quad d^c_{\mathrm{sep}}(\mu, t) = \max_{0 \le x \le n} \left\{ 1 - \frac{\mu H_t(x)}{\pi(x)} \right\}.$$

The following theorem is from [9].

Theorem 3.15. *Let K be an irreducible birth and death chain on $\{0, 1, ..., n\}$ with eigenvalues $\lambda_0 = 0 < \lambda_1 < \cdots < \lambda_n$.*

(1) *For the discrete time chain, if $p_i + q_{i+1} \le 1$ for all $0 \le i < n$, then*

$$d_{\mathrm{sep}}(0, m) = d_{\mathrm{sep}}(n, m) = \sum_{j=1}^{n} \left(\prod_{i=1, i \ne j}^{n} \frac{\lambda_i}{\lambda_i - \lambda_j} \right) (1 - \lambda_j)^m.$$

(2) *For the continuous time chain, it holds true that*

$$d^c_{\mathrm{sep}}(0, t) = d^c_{\mathrm{sep}}(n, t) = \sum_{j=1}^{n} \left(\prod_{i=1, i \ne j}^{n} \frac{\lambda_i}{\lambda_i - \lambda_j} \right) e^{-\lambda_j t}.$$

Diaconis and Fill [6, 12] introduce the concept of dual chain to express the separations in Theorem 3.15 as the probability of the first passage time. Brown and Shao [1] characterize the first passage time using the eigenvalues of K for a special class of continuous time Markov chains including birth and death chains. The idea in [1] is also applicable for discrete time chains and this leads to the formula above. See [9] for further discussions. Here, we use Proposition 3.12 and Lemma 3.16 to prove this result directly.

Lemma 3.16. *Let K be the transition matrix in (1.1) with stationary distribution π. Suppose that μ is a probability distribution satisfying $\mu(i)/\pi(i) \le \mu(i+1)/\pi(i+1)$ for all $0 \le i \le n-1$.*

(1) *For the discrete time chain, if $p_i + q_{i+1} \le 1$ for all $0 \le i < n$, then $\mu K^m(i)/\pi(i) \le \mu K^m(i+1)/\pi(i+1)$ for all $0 \le i < n$ and $m \ge 0$.*
(2) *For the continuous time chain, $\mu H_t(i)/\pi(i) \le \mu H_t(i+1)/\pi(i+1)$ for all $0 \le i < n$ and $t \ge 0$.*

Proof. Note that (2) follows from (1) if we write $H_t = \exp\{-2t(I - \frac{I+K}{2})\}$. For the proof of (1), observe that

$$\frac{\mu K^{m+1}(i)}{\pi(i)} = \frac{\mu K^m(i-1)}{\pi(i-1)} q_i + \frac{\mu K^m(i)}{\pi(i)} r_i + \frac{\mu K^m(i+1)}{\pi(i+1)} p_i, \quad \forall i.$$

By induction, if $\mu K^m(i)/\pi(i) \le \mu K^m(i+1)/\pi(i+1)$ for $0 \le i < n$, then

$$\begin{aligned}\frac{\mu K^{m+1}(i+1)}{\pi(i+1)} &= \frac{\mu K^m(i)}{\pi(i)}q_{i+1} + \frac{\mu K^m(i+1)}{\pi(i+1)}r_{i+1} + \frac{\mu K^m(i+2)}{\pi(i+2)}p_{i+1} \\ &\ge \frac{\mu K^m(i)}{\pi(i)}q_{i+1} + \frac{\mu K^m(i+1)}{\pi(i+1)}(1-q_{i+1}) \\ &\ge \frac{\mu K^m(i)}{\pi(i)}(1-p_i) + \frac{\mu K^m(i+1)}{\pi(i+1)}p_i \ge \frac{\mu K^{m+1}(i)}{\pi(i)}.\end{aligned}$$

□

Remark 3.12. Lemma 3.16 is also developed in [10] in which it is shown that, for any non-negative function f, $K^m f$ is non-decreasing if f is non-decreasing for all $m \ge 0$. Consider the adjoint chain K^* of K in $L^2(\pi)$. As birth and death chains are reversible, one has $K^* = K$. Using the identity $\mu K/\pi = K^*(\mu/\pi)$, it is easy to see that the above proof is consistent with the proof in [10].

Proof of Theorem 3.15. Assume that K is irreducible and let $\lambda_0 = 0 < \lambda_1 < \cdots < \lambda_n$ be the eigenvalues of $I - K$ with $L^2(\pi)$-normalized eigenvector $\zeta_0 = \mathbf{1}, ..., \zeta_n$. By Lemma 3.16, if μ satisfies $\mu(i)/\pi(i) \ge \mu(i+1)/\pi(i+1)$ for $0 \le i < n$, then

$$d^c_{\mathrm{sep}}(\mu, t) = 1 - \frac{\mu H_t(n)}{\pi(n)} = \sum_{j=1}^n \mu(\zeta_j)\zeta_j(n)e^{-\lambda_j t},$$

where $\mu(\zeta_j) = \sum_{i=0}^n \zeta_j(i)\mu(i)$. If K satisfies $p_i + q_{i+1} \le 1$ for all $0 \le i < n$, then

$$d_{\mathrm{sep}}(\mu, m) = 1 - \frac{\mu K^m(n)}{\pi(n)} = \sum_{j=1}^n \mu(\zeta_j)\zeta_j(n)(1-\lambda_j)^m.$$

By Proposition 3.12, setting μ to be one of the dirac measure δ_0, δ_n leads to the desired identities. □

4. Paths of infinite length

In this section, the graph $G = (V, E)$ under consideration is infinite with $V = \{1, 2, ...\}$ and $E = \{\{i, i+1\} | i = 1, 2, ...\}$. As before, let π, ν be positive measures on V, E satisfying $\pi(V) = 1$. The Dirichlet form and the variance are defined in a similar way as in the introduction and the spectral gap of G with respect to π, ν is given by

$$\lambda^G_{\pi,\nu} = \inf\left\{ \frac{\mathcal{E}_\nu(f,f)}{\mathrm{Var}_\pi(f)} \middle| f \text{ is non-constant and } \pi(f^2) < \infty \right\}.$$

For $n \ge 2$, let $G_n = (V_n, E_n)$ be the subgraph of G with $V_n = \{1, 2, ..., n\}$, $E_n = \{\{i, i+1\} | 1 \le i < n\}$ and let π_n, ν_n be normalized restrictions of π, ν to V_n, E_n. That is, $\pi_n(i) = c_n\pi(i)$, $\nu_n(i, i+1) = c_n\nu(i, i+1)$ with $c_n = 1/[\pi(1) + \cdots + \pi(n)]$. As before, let $M^G_{\pi,\nu}$ be an infinite matrix indexed by V and defined by

$$(4.1) \quad M^G_{\pi,\nu}(i,j) = -\frac{\nu(i,j)}{\pi(i)}, \quad \forall |i-j| = 1, \quad M^G_{\pi,\nu}(i,i) = \frac{\nu(i-1,i) + \nu(i,i+1)}{\pi(i)}.$$

Clearly, $M^{G_n}_{\pi_n,\nu_n}$ is the principal submatrix of $M^G_{\pi,\nu}$ indexed by $V_n \times V_n$.

Lemma 4.1. *Referring to the above setting,* $\lambda^{G_{n+1}}_{\pi_{n+1},\nu_{n+1}} < \lambda^{G_n}_{\pi_n,\nu_n}$ *for* $n > 1$ *and* $\lambda^G_{\pi,\nu} = \lim_{n\to\infty} \lambda^{G_n}_{\pi_n,\nu_n}$.

Proof. Briefly, we write λ for $\lambda^G_{\pi,\nu}$ and λ_n for $\lambda^{G_n}_{\pi_n,\nu_n}$. Note that λ_n is the smallest non-zero eigenvalue of the principal submatrix of $M^G_{\pi,\nu}$ indexed by $V_n \times V_n$. As a consequence of Proposition 3.11(1) and Remark 3.9, $\lambda_{n+1} < \lambda_n$. For $n > 1$, let ϕ_n be a minimizer for λ_n and define $\psi_n(i) = \mathbf{1}_{V_n}(i)\phi_n(i)$ for $i \geq 1$. Clearly, one has $\mathcal{E}_{\nu_n}(\phi_n, \phi_n) = c_n \mathcal{E}_\nu(\psi_n, \psi_n)$ and $\mathrm{Var}_{\pi_n}(\phi_n) = c_n \mathrm{Var}_\pi(\psi_n)$. This implies $\lambda \leq \lambda_n$ for $n \geq 2$. Let $\lambda^* = \lim_{n\to\infty} \lambda_n$. Note that it remains to show $\lambda^* = \lambda$. For $\epsilon > 0$, choose a function f on V such that $\mathcal{E}_\nu(f, f) < (\lambda + \epsilon/2)\mathrm{Var}_\pi(f)$ with $\pi(f^2) < \infty$. For $\delta > 0$, we choose $N > 0$ such that $\mathrm{Var}_{\pi_N}(g) > (1-\delta)\mathrm{Var}_\pi(f)$ and $\mathcal{E}_{\nu_N}(g, g) < (1+\delta)\mathcal{E}_\nu(f, f)$, where $g = f|_{V_N}$, the restriction of f to V_N. This implies

$$\lambda^* \leq \lambda_N \leq \frac{\mathcal{E}_{\nu_N}(g, g)}{\mathrm{Var}_{\pi_N}(g)} \leq \frac{(1+\delta)\mathcal{E}_\nu(f, f)}{(1-\delta)\mathrm{Var}_\pi(f)}.$$

Letting $\delta \to 0$ and then $\epsilon \to 0$ yields $\lambda^* \leq \lambda$, as desired. □

Remark 4.1. Silver [17] contains a discussion of the (weak*) convergence of the spectral measure for G_n to the spectral measure for G in a very general setting. Lemma 4.1 can also be proved using Theorem 4.3.4 in [17].

Proposition 4.2. *For $\lambda > 0$, let $\phi_\lambda(1) = -1$ and*

$$\phi_\lambda(i+1) = \phi_\lambda(i) + \frac{\{[\phi_\lambda(i) - \phi_\lambda(i-1)]\nu(i-1, i) - \lambda\pi(i)\phi_\lambda(i)\}^+}{\nu(i, i+1)}, \ \forall i \geq 1.$$

Set $\lambda_1 = \infty$ and $\lambda_n = \lambda^{G_n}_{\pi_n,\nu_n}$ for $n \geq 2$.

(1) *For $i \geq 2$ and $\lambda \in [\lambda_i, \lambda_{i-1})$, $\phi_\lambda(i-1) < \phi_\lambda(i) = \phi_\lambda(i+1)$.*
(2) *For $\lambda \in (0, \lambda^G_{\pi,\nu}]$, $\phi_\lambda(i) < \phi_\lambda(i+1)$ for all $i \geq 1$.*

Proof. Immediate from Proposition 3.11 and Remarks 3.9-3.10. □

Remark 4.2. By Proposition 4.2, one may generate a dichotomy algorithm for $\lambda^G_{\pi,\nu}$ using the shape of ϕ_λ. See (Di).

The following theorem extends Theorem 1.1 to infinite paths.

Theorem 4.3. *If $\lambda^G_{\pi,\nu} > 0$ and $\mathcal{E}_\nu(\psi, \psi)/\mathit{Var}_\pi(\psi) = \lambda^G_{\pi,\nu}$ for some function ψ on V with $\pi(\psi) = 0$, then ψ is strictly monotonic and satisfies*

$$\lambda^G_{\pi,\nu}\pi(i)\psi(i) = [\psi(i) - \psi(i+1)]\nu(i, i+1) + [\psi(i) - \psi(i-1)]\nu(i-1, i), \ \forall i \geq 1.$$

Theorem 4.4. *For $\lambda > 0$, let ϕ_λ be the function in Proposition 4.2 and set $L(\lambda) = \mathcal{E}_\pi(\phi_\lambda, \phi_\lambda)/\mathrm{Var}_\pi(\phi_\lambda)$. Then,*

(1) *$\lambda^G_{\pi,\nu} < L(\lambda) < \lambda$ for $\lambda \in (\lambda^G_{\pi,\nu}, \infty)$.*
(2) *$L^n(\lambda) \to \lambda^G_{\pi,\nu}$ as $n \to \infty$ for $\lambda \in (\lambda^G_{\pi,\nu}, \infty)$.*

Proof. Let $\lambda > \lambda^G_{\pi,\nu}$. By Lemma 4.1, $\lambda_i \leq \lambda < \lambda_{i-1}$ for some $i \geq 2$. By Proposition 4.2 (1), one has $\phi_\lambda(i-1) < \phi_\lambda(i) = \phi_\lambda(i+1)$. As in (2.2), we obtain

$$L(\lambda) = \lambda + \lambda\frac{\pi(\phi_\lambda)[\pi(\phi_\lambda) - \phi_\lambda(i)]}{\mathrm{Var}_\pi(\phi_\lambda)}, \quad \sum_{j=1}^{i} \phi_\lambda(j)\pi(j) \geq 0.$$

This leads to $\pi(\phi_\lambda) > 0$ and $\pi(\phi_\lambda) < \phi_\lambda(i)$, which implies $L(\lambda) < \lambda$. That means L has no fixed point on $(\lambda^G_{\pi,\nu}, \infty)$. The lower bound of (1) follows immediately from Theorem 4.3. For (2), set $\lambda^* = \lim_{n\to\infty} L^n(\lambda) \geq \lambda^G_{\pi,\nu}$. As a consequence of (1), L is continuous on $(\lambda^G_{\pi,\nu}, \infty)$. If $\lambda^* > \lambda^G_{\pi,\nu}$, then λ^* is a fixed point of L, a contradiction! Hence, $\lambda^* = \lambda^G_{\pi,\nu}$. □

5. A numerical experiment

In this section, we illustrate the algorithm (A2) on a specific Metropolis chain. The Metropolis algorithm introduced by Metropolis *et al.* in 1953 is a widely used construction that produces a Markov chain with a given stationary distribution π. Let π be a positive probability measure on V and K be an irreducible Markov transition matrix on V. For simplicity, we assume that $K(x,y)=K(y,x)$ for all $x,y\in V$. The Metropolis chain evolves in the following way. Given the initial state x, select a state, say y, according to $K(x,\cdot)$ and compute the ratio $A(x,y)=\pi(y)/\pi(x)$. If $A(x,y)\ge 1$, then move to y. If $A(x,y)<1$, then flip a coin with probability $A(x,y)$ on heads and move to y if the head appears. If the coin lands on tails, stay at x. Accordingly, if M is the transition matrix of the Metropolis chain, then

$$M(x,y)=\begin{cases} K(x,y) & \text{if } A(x,y)\ge 1,\ x\ne y\\ K(x,y)A(x,y) & \text{if } A(x,y)<1\\ K(x,x)+\sum\limits_{z:A(x,z)<1}K(x,z)(1-A(x,z)) & \text{if } x=y\end{cases}.$$

It is easy to check $\pi(x)M(x,y)=\pi(y)M(y,x)$. As K is irreducible, M is irreducible. Moreover, if π is not uniform, then $M(x,x)>0$ for some $x\in V$. This implies that M is aperiodic and, consequently, $M^t(x,y)\to\pi(y)$ and $e^{-t(I-M)}(x,y)\to\pi(y)$ as $t\to\infty$. For further information on Metropolis chains, see [5] and the references therein.

For $n\ge 1$, let $G_n=(V_n,E_n)$ be a graph with $V_n=\{0,\pm 1,...,\pm n\}$ and $E_n=\{\{i,i+1\}:i=-n,...,n-1\}$. Suppose that K_n is the transition matrix of the simple random walk on V_n, that is, $K_n(-n,-n)=K_n(n,n)=1/2$ and $K_n(i,i+1)=K_n(i+1,i)=1/2$ for all $-n\le i<n$. For $a>0$, let $\check\pi_{n,a},\hat\pi_{n,a}$ be probabilities on $V_n=\{0,\pm 1,...,\pm n\}$ given by

$$\check\pi_{n,a}(i)=\check c_{n,a}(|i|+1)^a,\quad \hat\pi_{n,a}(i)=\hat c_{n,a}(n-|i|+1)^a,$$

where $\check c_{n,a}$ and $\hat c_{n,a}$ are normalizing constants. It is easy to compute that

$$c_{n,a}/2\le 1/\hat c_{n,a}<1/\check c_{n,a}\le 2c_{n,a}, \tag{5.1}$$

where

$$c_{n,a}=\frac{(n+1)^{a+1}}{a+1}+(n+1)^a.$$

The Metropolis chains, $\check K_{n,a}$ and $\hat K_{n,a}$, for $\check\pi_{n,a}$ and $\hat\pi_{n,a}$ based on the simple random walk K_n have transition matrices given by

$$\check K_{n,a}(i,j)=\check K_{n,a}(-i,-j),\quad \hat K_{n,a}(i,j)=\hat K_{n,a}(-i,-j)$$

and

$$\check K_{n,a}(i,j)=\begin{cases}\frac12 & \text{if } j=i+1, i\in[0,n-1]\\ \frac{i^a}{2(i+1)^a} & \text{if } j=i-1, i\in[1,n]\\ \frac{(i+1)^a-i^a}{2(i+1)^a} & \text{if } j=i, i\notin\{0,n\}\\ 1-\frac{n^a}{2(n+1)^a} & \text{if } i=j=n\end{cases}$$

and

$$\hat{K}_{n,a}(i,j) = \begin{cases} \frac{1}{2} & \text{if } j = i-1, i \in [1,n] \\ \frac{(n-i)^a}{2(n-i+1)^a} & \text{if } j = i+1, i \in [0,n-1] \\ \frac{(n-i+1)^a-(n-i)^a}{2(n-i+1)^a} & \text{if } j = i \neq 0 \\ 1 - \frac{n^a}{(n+1)^a} & \text{if } i = j = 0 \end{cases}.$$

Saloff-Coste [16] discussed the above chains and obtained the correct order of the spectral gaps. Let $\check{\lambda}_{n,a}, \hat{\lambda}_{n,a}$ denote the spectral gaps of $\check{K}_{n,a}, \hat{K}_{n,a}$. Referring to the recent work in [4], one has

$$1/(4C) \le \lambda \le 1/C,$$

where (λ, C) is any of $(\check{\lambda}_{n,a}, \check{C}_n(a))$ and $(\hat{\lambda}_{n,a}, \hat{C}_n(a))$, and

$$\check{C}_n(a) = 2 \max_{1\le i\le n} \left(\sum_{j=0}^{i-1}(j+1)^{-a}\right)\left(\sum_{j=i}^{n}(j+1)^{a}\right),$$

and

$$\hat{C}_n(a) = 2 \max_{1\le i\le n} \left(\sum_{j=0}^{i-1}(j+1)^{a}\right)\left(\sum_{j=i-1}^{n-1}(j+1)^{-a}\right).$$

Theorem 5.1. *Let $\check{\lambda}_{n,a}, \hat{\lambda}_{n,a}$ be spectral gaps for $\check{K}_{n,a}, \hat{K}_{n,a}$. Then,*

$$\frac{1}{8\eta_{-a}(1,n)\eta_a(2,n+1)} \le \check{\lambda}_{n,a} \le \frac{2}{\eta_{-a}(1,n)\eta_a(2,n+1)},$$

and

$$\frac{1}{64\eta_a(1,\lceil n/2\rceil)\eta_{-a}(\lceil n/2\rceil,n)} \le \hat{\lambda}_{n,a} \le \frac{1}{2\eta_a(1,\lceil n/2\rceil)\eta_{-a}(\lceil n/2\rceil,n)}.$$

where $\eta_a(k,l) = \sum_{i=k}^{l} i^a$.

Proof of Theorem 5.1. The bound for $\check{\lambda}_{n,a}$ follows immediately from the fact

$$\frac{\eta_{-a}(1,n)\eta_a(2,n+1)}{2} \le \check{C}_n(a) \le 2\eta_{-a}(1,n)\eta_a(2,n+1).$$

For $\hat{\lambda}_{n,a}$, note that

$$\hat{C}_n(a) = 2 \max_{n/2\le i\le n} \left(\sum_{j=0}^{i-1}(j+1)^{a}\right)\left(\sum_{j=i-1}^{n-1}(j+1)^{-a}\right).$$

Taking $i = \lceil n/2\rceil$ yields the upper bound. For the lower bound, we write

$$\hat{C}_n(a) = 2 \max_{n/2\le i\le n} \left(\sum_{j=0}^{i-1}\left(1-\frac{j}{i}\right)^{a}\right)\left(\sum_{j=0}^{n-i}\left(1-\frac{j}{i+j}\right)^{a}\right).$$

For $i \ge n/2$, it is clear that

$$\sum_{j=0}^{i-1}\left(1-\frac{j}{i}\right)^{a} \ge \sum_{j=0}^{i-1}\left(1-\frac{2j}{n}\right)^{a} \ge \frac{1}{2}\sum_{j=0}^{n-1}\left(1-\frac{j}{n}\right)^{a}.$$

Observe that, for $a > 0$,

$$\frac{C'_{i,n}(a)}{2} \le \sum_{j=0}^{n-i} \left(1 - \frac{j}{i+j}\right)^a \le C'_{i,n}(a), \tag{5.2}$$

where

$$C'_{i,n}(a) = 1 + \begin{cases} i\frac{(i/n)^{a-1}-1}{1-a} & \text{if } a \neq 1 \\ i \log \frac{n}{i} & \text{if } a = 1 \end{cases}.$$

It is clear that, for $i \ge n/2$, $C'_{i,n}(a) \le 2C'_{\lceil n/2\rceil,n}(a)$ and this leads to

$$\sum_{j=0}^{n-i} \left(1 - \frac{j}{i+j}\right)^a \le 4 \sum_{j=0}^{n-\lceil n/2\rceil} \left(1 - \frac{j}{\lceil n/2\rceil + j}\right)^a.$$

Summarizing all above gives the desired lower bound. □

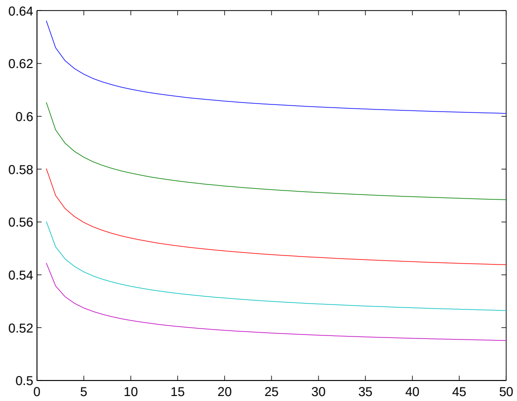


FIGURE 1. These curves display the mapping $m \mapsto \check{\lambda}_{100m,a}\eta_{-a}(1, 100m)\eta_a(2, 100m + 1)$ in Theorem 5.1 in order from the top $a = 0.8, 0.9, 1.0, 1.1$ and 1.2. The right most point corresponds to a path of length $n = 5000$.

TABLE 1. These numbers denote $\check{\lambda}_{n,a}\eta_{-a}(1, n)\eta_a(2, n+1)$ in Theorem 5.1.

n	10000	20000	30000	40000	50000
a=0.8	0.5983	0.5960	0.5948	0.5941	0.5935
a=0.9	0.5652	0.5625	0.5610	0.5601	0.5594
a=1.0	0.5405	0.5377	0.5362	0.5353	0.5345
a=1.1	0.5235	0.5210	0.5197	0.5189	0.5183
a=1.2	0.5128	0.5109	0.5099	0.5093	0.5088

Remark 5.1. Comparing with [16, Theorem 9.5], the bounds for $\check{\lambda}_{n,a}$ given in Theorem 5.1 have a similar lower bound and an improved upper bound by a multiple of about 1/4. For $\hat{\lambda}_{n,a}$, observe that

$$\frac{C_i''(a)}{2} \le \sum_{j=0}^{i-1} \left(1 - \frac{j}{i}\right)^a \le C_i''(a),$$

where

$$C_i''(a) = 1 + \frac{i - i^{-a}}{1+a}.$$

Recall the constant $C'_{i,n}(a)$ in the proof of Theorem 5.1. Note that

$$\frac{n+a}{2(1+a)} \le C''_{\lceil n/2 \rceil}(a) \le \frac{2(n+a)}{(1+a)},$$

and, for $a > 0$, $a \neq 1$ and $n \geq 3$,

$$C'_{\lceil n/2 \rceil, n}(a) \le 1 + \frac{n+1}{2(1+a)} \sup_{a>0, a\neq 1} \frac{(2^{1-a} - 1)(1+a)}{1-a} \le \frac{3(n+a)}{1+a},$$

where the last inequality is obtained by considering the subcases $a < 2$ and $a \geq 2$. The above computation also applies for $a = 1$ and $n \in \{1, 2\}$. In the same spirit, one can show that $C'_{\lceil n/2 \rceil, n}(a) \geq \frac{n+a}{6(1+a)}$. This yields

$$\frac{(n+a)^2}{6(1+a)^2} \le \hat{C}_{n,a} \le \frac{12(n+a)^2}{(1+a)^2}, \quad \forall n \geq 1. \tag{5.3}$$

Hence, we have $\hat{\lambda}_{n,a} \asymp (1+a)^2/(n+a)^2$. As a consequence of (5.1) and (5.2), we obtain that, uniformly for $a > 0$,

$$1/\check{\lambda}_{n,a} \asymp n^a \left(\left(1 + \frac{1}{n}\right)^a + \frac{n}{1+a} \right) (1 + v(n,a)) \quad \text{as } n \to \infty,$$

where $v(n, 1) = \log n$ and $v(n, a) = (n^{1-a} - 1)/(1-a)$ for $a \neq 1$.

Remark 5.2. Note that the lower bound in Theorem 6.1 provides the correct order of the spectral gap for the chain $\check{K}_{n,a}$ uniformly in a but not for $\hat{K}_{n,a}$. For instance, if a grows with n, say $a = n$, then Theorem 6.1 implies $1/\hat{\lambda}_{n,n} = O(n)$, while (5.3) gives $1/\hat{\lambda}_{n,n} \asymp 1$.

Remark 5.3. Consider the chain in Theorem 5.1. A numerical experiment of Algorithm (A2) is implemented and the data is collected in Figure 1 and Table 1. One may conjecture that $\check{\lambda}_{n,a}\eta_{-a}(1,n)\eta_a(2,n+1) \to c(a)$ as $n \to \infty$, where $c(a)$ is a constant depending on a.

6. Spectral gaps for uniform measures with bottlenecks

In this section, we discuss some examples of special interests and show how the theory developed in the previous sections can be used to bound the spectral gap. In the first subsection, we develop a lower bound on the spectral gap in a very general setting using the theory in Section 3. In the second subsection, we focuses on the case of one bottleneck, where a precise estimation on the spectral gap is presented. Those computations are based on the theoretical work in Section 2. In the third subsection, we consider the case of multiple bottlenecks in which the exact order of the spectral gap is determined for some special classes of chains.

In what follows, we will use the notation $\pi(A)$ to represent the summation $\sum_{i\in A}\pi(i)$ for any measure π on V and any set $A\subset V$. Given two sequences of positive reals a_n, b_n, we write $a_n=O(b_n)$ if a_n/b_n is bounded. If $a_n=O(b_n)$ and $b_n=O(a_n)$, we write $a_n\asymp b_n$. If $a_n/b_n\to 1$, we write $a_n\sim b_n$.

6.1. **A lower bound on the spectral gap.** In this subsection, we give a lower bound on the spectral gap in the general case.

Theorem 6.1. *Let $G=(V,E)$ be a graph with vertex set $V=\{0,1,...,n\}$ and edge set $E=\{\{i,i+1\}|i=0,...,n-1\}$. Let π,ν be positive measures on V,E with $\pi(V)=1$. Then,*

$$\lambda_{\pi,\nu}^G\ge\max_{0\le i\le n}\left\{\left(\sum_{j=0}^{i-1}\frac{\pi([0,j])}{\nu(j,j+1)}\right)^{-1}\wedge\left(\sum_{j=i+1}^{n}\frac{\pi([j,n])}{\nu(j-1,j)}\right)^{-1}\right\},$$

where $a\wedge b:=\min\{a,b\}$.

Remark 6.1. Let C be the lower of the spectral gap in Theorem 6.1. Note that, for any positive reals, $(a+b)/2\le\max\{a,b\}\le a+b$. Using this fact, it is easy to see that $C'\le C\le 2C'$, where

$$C'=\max_{0\le i\le n}\left(\sum_{j=0}^{i-1}\frac{\pi([0,j])}{\nu(j,j+1)}+\sum_{j=i+1}^{n}\frac{\pi([j,n])}{\nu(j-1,j)}\right)^{-1}.$$

In particular, if i_0 is the median of π, that is, $\pi([0,i_0])\ge 1/2$ and $\pi([i_0,n])\ge 1/2$, then

$$C'=\left(\sum_{j=0}^{i_0-1}\frac{\pi([0,j])}{\nu(j,j+1)}+\sum_{j=i_0+1}^{n}\frac{\pi([j,n])}{\nu(j-1,j)}\right)^{-1}.$$

Remark 6.2. Let $(X_m)_{m=0}^\infty$ be an irreducible birth and death chain on $\{0,1,...,n\}$ with birth rate p_i, death rate q_i and holding rate r_i as in (1.1). For $0\le i\le n$, set $\tau_i=\min\{m\ge 0|X_m=i\}$ as the first passage time to state i. By the strong Markov property, the expected hitting time to i started at 0 can be expressed as

$$\mathbb{E}_0\tau_i=\sum_{j=0}^{i-1}\frac{\pi([0,j])}{p_j\pi(j)},\quad \mathbb{E}_n\tau_i=\sum_{j=i+1}^{n}\frac{\pi([j,n])}{q_j\pi(j)},$$

where π is the stationary distribution of $(X_m)_{m=0}^\infty$. Let λ be the spectral gap for $(X_m)_{m=0}^\infty$. Then, $\lambda=\lambda_{\pi,\nu}^G$, where G is the path with vertex set $\{0,...,n\}$ and $\nu(i,i+1)=p_i\pi(i)=q_{i+1}\pi(i+1)$ for $0\le i<n$. The conclusion of Theorem 6.1 can be written as $1/\lambda\le\min_{0\le i\le n}\{\mathbb{E}_0\tau_i\vee\mathbb{E}_n\tau_i\}$.

Remark 6.3. The lower bound in Theorem 6.1 is not necessary the right order of the spectral gap. See Remark 5.2.

Proof of Theorem 6.1. For $\lambda>0$, let ξ_λ be the function in Definition 3.2. That is, $\xi_\lambda(0)=-1$ and, for $i\ge 0$,

$$[\xi_\lambda(i+1)-\xi_\lambda(i)]\nu(i,i+1)=[\xi_\lambda(i)-\xi_\lambda(i-1)]\nu(i-1,i)-\lambda\pi(i)\xi_\lambda(i).$$

Inductively, one can show that if $1/\lambda > \sum_{j=0}^{\ell-1}[\pi([0,j])/\nu(j,j+1)]$, then

$$\begin{cases} 0 < \xi_\lambda(i+1) - \xi_\lambda(i) \le \lambda\pi([0,i])/\nu(i,i+1), \\ -1 \le \xi_\lambda(i+1) \le -1 + \lambda\sum_{j=1}^{i}[\pi([0,j])/\nu(j,j+1)] < 0, \end{cases}$$

for $0 \le i \le \ell-1$. One may do a similar computation from the other end point and, by Proposition 3.5, this implies

$$1/\lambda_{\pi_n,\nu_n}^{G_n} \le \max\left\{\sum_{j=0}^{\ell-1}\frac{\pi([0,j])}{\nu(j,j+1)}, \sum_{j=1}^{n-\ell}\frac{\pi([n-j+1,n])}{\nu(n-j,n-j+1)}\right\}.$$

Taking the minimum over $1 \le \ell \le n$ gives the desired inequality. □

6.2. **One bottleneck.** For $n \ge 1$, let $G_n = (V_n, E_n)$ be the path on $\{0,1,...,n\}$ and set $\pi_n \equiv 1/(n+1)$ and $\nu_n \equiv 1/(n+1)$ with $C > 0$. Using Feller's method in [11, Chapter XVI.3], one can show that the eigenvalues of $M_{\pi_n,\nu_n}^{G_n}$ are $2(1-\cos\frac{i\pi}{n+1})$ for $0 \le i \le n$.

Theorem 6.2. *For $n \ge 1$, let $\epsilon_n > 0$, $1 \le x_n \le \lceil n/2\rceil$ and set $\pi_n \equiv 1/(n+1)$,*

$$\nu_n^{x_n}(x_n-1,x_n) = \frac{\epsilon_n}{n+1}, \quad \nu_n^{x_n}(i-1,i) = \frac{1}{n+1}, \quad \forall i \ne x_n. \tag{6.1}$$

Then, the spectral gap are bounded by

$$\frac{1}{n^2/4 + x_n/\epsilon_n} \le \lambda_{\pi_n,\nu_n^{x_n}}^{G} \le \min\left\{2\left(1-\cos\frac{\pi}{n-x_n+1}\right), \frac{\epsilon_n}{x_n}\right\}.$$

In particular, $\lambda_{\pi_n,\nu_n^{x_n}}^{G_n} \asymp \min\{1/n^2, \epsilon_n/x_n\}$.

Proof of Theorem 6.2. The lower bound is immediate from Theorem 6.1 by choosing $i = \lceil n/2\rceil$ in the computation of the maximum. For the upper bound, we set $\lambda_n = 1-\cos\frac{\pi}{n+1}$ and let f_n be the function on V_{n-x_n} defined by $f_n(0) = -1$ and, for $0 \le i \le n-x_n-1$,

$$f_n(i+1) = f_n(i) + \frac{[f_n(i) - f_n(i-1)]\nu_{n-x_n}(i-1,i) - 2\lambda_{n-x_n}\pi_{n-x_n}(i)f_n(i)}{\nu_{n-x_n}(i,i+1)}.$$

By Proposition 2.3, $\mathcal{E}_{\nu_{n-x_n}}(f_n,f_n) = 2\lambda_{n-x_n}\mathrm{Var}_{\pi_{n-x_n}}(f_n)$ and $\pi_{n-x_n}(f_n) = 0$. Let g_n be the function on V_n defined by $g_n(n-i) = f_n(i)$ for $0 \le i \le n-x_n$ and $g_n(i) = f_n(n-x_n)$ for $0 \le i < x_n$. A direct computation shows that

$$(n+1)\mathcal{E}_{\nu_n^{x_n}}(g_n,g_n) = (n-x_n+1)\mathcal{E}_{\nu_{n-x_n}}(f_n,f_n)$$

and

$$(n+1)\mathrm{Var}_{\pi_n}(g_n,g_n) = (n-x_n+1)\mathrm{Var}_{\pi_{n-x_n}}(f_n) + \frac{x_n(n-x_n+1)}{n+1}f_n^2(n-x_n).$$

This implies $\lambda_{\pi_n,\nu_n^{x_n}}^{G_n} \le 2\lambda_{n-x_n}$. On the other hand, using the test function, $h_n(i) = n-x_n+1$ for $0 \le i < x_n$ and $h_n(i) = -x_n$ for $x_n \le i \le n$, one has $\mathcal{E}_{\nu_n^{x_n}}(h_n,h_n)/\mathrm{Var}_{\pi_n}(h_n) = \epsilon_n(n+1)/[x_n(n-x_n+1)] \le \epsilon_n/x_n$. This finishes the proof. □

The next theorem has a detailed description on the coefficient of the spectral gap. The proof is based on Section 3, particularly Proposition 3.11 and Remark 3.10, and is given in the appendix.

Theorem 6.3. *For $n \geq 1$, let $x_n, \epsilon_n, \pi_n, \nu_n^{x_n}$ be as in Theorem 6.2. Suppose $x_n/(\epsilon_n n^2) \to a \in [0,\infty]$ and $x_n/n \to b \in [0,1/2]$.*

(1) *If $a < \infty$ and $b = 0$, then $\lambda_{\pi_n,\nu_n^{x_n}}^{G_n} \sim \min\{\pi^2, a^{-2}\}n^{-2}$.*
(2) *If $a < \infty$ and $b \in (0,1/2]$, then $\lambda_{\pi_n,\nu_n^{x_n}}^{G_n} \sim Cn^{-2}$, where C is the unique positive solution of the following equation.*

$$1 + 4\log 2 - \frac{\pi^2}{6} - \frac{\pi^2 aC}{1-b} - bC\sum_{i=1}^{\infty} \frac{(1-b)i^2 - bC}{(i^2 - C)[(1-b)^2 i^2 - b^2 C]} = 0.$$

(3) *If $a = \infty$, then $\lambda_{\pi_n,\nu_n^{x_n}}^{G_n} \sim \epsilon_n/x_n$.*

6.3. **Multiple bottlenecks.** In this subsection, we consider paths with multiple bottlenecks. As before, $G_n = (V_n, E_n)$ with $V_n = \{0, 1, ..., n\}$ and $E_n = \{\{i, i+1\} | i = 0, ..., n-1\}$. Let k be a positive integer and $x_n = (x_{n,1}, ..., x_{n,k})$ be a k-vector satisfying $x_{n,i} \in V_n$ and $x_{n,1} \geq 1$ and $x_{n,i} < x_{n,i+1}$ for $1 \leq i < k$. Let $\epsilon_n = (\epsilon_{n,1}, ..., \epsilon_{n,k})$ be a vector with positive entries and $\nu_n^{x_n}$ be the measure on E_n given by

$$\nu_n^{x_n}(i-1, i) = \begin{cases} 1/(n+1) & \text{if } i \notin \{x_{n,1}, ..., x_{n,k}\} \\ \epsilon_{n,j}/(n+1) & \text{if } i = x_{n,j},\ 1 \leq j \leq k \end{cases}. \tag{6.2}$$

Theorem 6.4. *Let $G_n = (V_n, E_n)$ be the path on $\{0, ..., n\}$. For $0 \leq k \leq n$, let π_n be the uniform probability on V_n and $\nu_n^{x_n}$ be the measure on E_n given by (6.2). Then,*

$$\min\{1/(4n^2), C_{n,1}/2\} \leq \lambda_{\pi_n,\nu_n^{x_n}}^{G_n} \leq \min\left\{2\left(1 - \cos\frac{\pi}{n-k+1}\right), C_{n,2}\right\},$$

where

$$C_{n,1} = \left(\frac{n^2}{4} + \sum_{i=1}^{k} \min\{x_{n,i}, n - x_{n,i} + 1\}\left(\frac{1}{\epsilon_{n,i}} - 1\right)\right)^{-1}$$

and

$$C_{n,2} = \min_{0 \leq m_1 \leq m_2 \leq n} \left\{ \frac{(n+1)\sum_{i=m_1}^{m_2} 1/\epsilon_{n,i}}{\sum_{m_1 \leq i \leq j \leq m_2} x_{n,i}(n - x_{n,j} + 1)/(\epsilon_{n,i}\epsilon_{n,j})} \right\}.$$

Remark 6.4. Observe that, in Theorem 6.4, $1 - \cos\frac{2\pi}{n-k+1} \asymp n^{-2}$ and

$$\begin{aligned} C_{n,2} &\leq \min_{1 \leq j \leq k}\left\{\frac{\epsilon_{n,j}}{\min\{x_{n,j}, n - x_{n,j} + 1\}}\right\} \\ &= \min\left\{\min_{j: x_{n,j} \leq \frac{n}{2}} \frac{\epsilon_{n,j}}{x_{n,j}}, \min_{j: x_{n,j} > \frac{n}{2}} \frac{\epsilon_{n,j}}{n - x_{n,j} + 1}\right\}. \end{aligned}$$

Proof of Theorem 6.4. We first prove the upper bound. Let f_1 be a function on $\{0, 1, ..., n\}$ satisfying $f(x_{n,j} - 1) = f(x_{n,j})$ for $1 \leq i \leq k$ and f_2 be a function on $\{0, ..., n-k\}$ obtained by identifying points $x_{n,i} - 1$ and $x_{n,i}$ for $1 \leq i \leq k$. By setting f_2 as a minimizer for $\lambda^{G_{n-k}}_{\pi_{n-k},\nu_{n-k}}$ with $\pi_n(f_1)=0$, we obtain

$$2\left(1-\cos\frac{2\pi}{n-k+1}\right)=\frac{\mathcal{E}_{\nu_{n-k}}(f_2,f_2)}{\mathrm{Var}_{\pi_{n-k}}(f_2)}\geq\frac{\mathcal{E}_{\nu_{n-k}}(f_2,f_2)}{\pi_{n-k}(f_2^2)}$$
$$\geq\frac{\mathcal{E}_{\nu_n}(f_1,f_1)}{\pi_n(f_1^2)}=\frac{\mathcal{E}_{\nu_n}(f_1,f_1)}{\mathrm{Var}_{\pi_n}(f_1)}.$$

To see the other upper bound, let f_j be the function on V_n satisfying $g_j(i)=-(n-x_{n,j}+1)$ for $0\leq i\leq x_{n,j}-1$ and $g_j(i)=x_{n,j}$ for $x_{n,j}\leq i\leq n$. Computations show that $\pi_n(g_j)=0$, $\pi_n(g_ig_j)=x_{n,i}(n-x_{n,j}+1)$ for $i\leq j$, and $\mathcal{E}_{\nu_n}(g_j,g_j)=\epsilon_{n,j}(n+1)$. Set $g=\sum_{j=1}^k a_jg_j$. As a consequence of the above discussion, we obtain

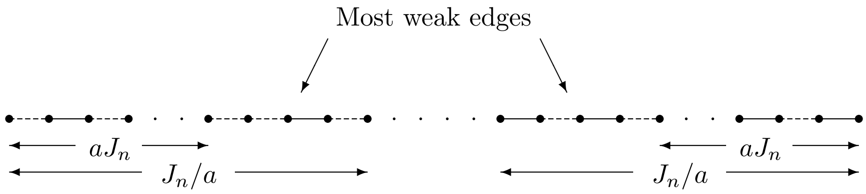
$$\frac{\mathcal{E}_{\nu_n}(g,g)}{\mathrm{Var}_{\pi_n}(g)}=\frac{(n+1)\sum_{i=1}^k a_i^2\epsilon_{n,i}}{2\sum_{i<j}a_ia_jx_{n,i}(n-x_{n,j}+1)+\sum_{i=1}^k a_i^2x_{n,i}(n-x_{n,i}+1)}.$$

Taking $a_i=1/\epsilon_{n,i}$ for $m_1\leq i\leq m_2$ and $a_i=0$ otherwise gives the bound $C_{n,2}$.

The lower bound is immediate from Theorem 6.1 and Remark 6.1. □

Finally, we discuss some special cases illustrating Theorem 6.4.

Figure 2. The dashed lines denote the weak edges of ν in Theorem 6.5.



Theorem 6.5. *For $n\geq 1$, let $\pi_n\equiv 1/(n+1)$ and ν_n be the measure in (6.2) with k_n bottlenecks satisfying $n-k_n\asymp n$. Suppose there are $I_n\subset\{1,...,k_n\}$, $a\in(0,1)$ and $J_n>0$ such that $|I_n|$ is bounded and, for $i\notin I_n$, $aJ_n\leq\min\{x_{n,i},n-x_{n,i}+1\}\leq J_n/a$. Then,*

$$\lambda^{G_n}_{\pi_n,\nu_n}\asymp\min\left\{\frac{1}{n^2},\min_{i\in I_n}\frac{\epsilon_{n,i}}{\min\{x_{n,i},n-x_{n,i}+1\}},\frac{\left(\sum_{i=1,i\notin I_n}^{k_n}1/\epsilon_{n,i}\right)^{-1}}{J_n}\right\}.$$

Proof. It is easy to get the lower bound from Theorem 6.4, while the upper bound is the minimum of $C_{n,2}$ over all connected components of $\{1,...,\ell\}\setminus I_n$ and $\{\ell+1,...,k_n\}\setminus I_n$. □

See Figure 2 for a reference on the bottlenecks. The following are immediate corollaries of Theorems 6.4-6.5.

Corollary 6.6 (Finitely many bottlenecks)**.** *Referring to* Theorem 6.5*, if k_n is bounded, then*

$$\lambda^{G_n}_{\pi_n,\nu_n}\asymp\min\left\{\frac{1}{n^2},\min_{1\leq i\leq k_n}\frac{\epsilon_{n,i}}{\min\{x_{n,i},n-x_{n,i}+i\}}\right\}.$$

Corollary 6.7 (Bottlenecks far away the boundary)**.** *Referring to* Theorem 6.5*, if* $n-k_n \asymp n$ *and there are* $a \in (0,1)$ *and* $J_n > 0$ *such that* $aJ_n < \min\{x_{n,i}, n-x_{n,i}+1\} < J_n/a$ *for* $1 \le i \le k_n$*, then*

$$\lambda_{\pi_n,\nu_n}^{G_n} \asymp \min\left\{\frac{1}{n^2}, \frac{\left(\sum_{j=1}^{k_n} 1/\epsilon_{n,i}\right)^{-1}}{J_n}\right\}.$$

Corollary 6.8 (Uniformly distributed bottlenecks)**.** *Referring to* Theorem 6.5*, if* $\min_i \epsilon_{n,i} \asymp \max_i \epsilon_{n,i}$ *and* $x_{n,i} = \lfloor in/k_n \rfloor$ *with* $k_n \le n/2$*, then*

$$\lambda_{\pi_n,\nu_n}^{G_n} \asymp \min\left\{\frac{1}{n^2}, \frac{\epsilon_{n,1}}{nk_n}\right\}.$$

Remark 6.5. Note that the assumption of the uniformity of π and ν, except at the bottlenecks, can be relaxed by using a comparison argument.

Appendix A. Techniques and proofs

We start with an elementary lemma.

Lemma A.1. *Let* $a > 0$ *and* $f : [a,\infty) \to \mathbb{R}$ *be a continuous function satisfying* $f(a) = a$ *and* $f(x) \in [a,x)$ *for* $x > a$*. For* $b > a$*, set* $C_b = \sup_{a \le x \le b}\{(f(x)-a)/(x-a)\}$*. Then,* $C_b < 1$ *and* $a \le f^n(b) \le a + C_b^n(b-a)$ *for* $n \ge 0$*. Moreover, if* f *is bounded on* $[a,\infty)$*, then* $a \le f^n(x) \le a + C^n(x-a)$ *for* $n \ge 0$ *and* $x \ge a$ *with* $C = \sup_{a \le t < \infty}\{(f(t)-a)/(t-a)\} < 1$*.*

Lemma A.2. *Let* $(a_i, b_i, c_i)_{i=1}^\infty$ *be sequences of reals with* $b_i > 0$ *and* $c_i > 0$*. For* $n \ge 1$ *and* $t \in \mathbb{R}$*, let*

$$M_n(t) = \begin{pmatrix} a_1 - c_1 t & 1 & 0 & 0 & \cdots & 0 \\ b_1 & a_2 - c_2 t & 1 & 0 & & \vdots \\ 0 & b_2 & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & a_{n-1} - c_{n-1} t & 1 \\ 0 & \cdots & \cdots & 0 & b_{n-1} & a_n - c_n t \end{pmatrix}.$$

Then, there are n *distinct real roots for* $\det M_n(t) = 0$*, say* $t_1^{(n)} < \cdots < t_n^{(n)}$*, and*

$$t_j^{(n+1)} < t_j^{(n)} < t_{j+1}^{(n+1)}, \quad \forall 1 \le j \le n,\ n \ge 1.$$

Furthermore, if $a_1 \ge 1$ *and* $a_{i+1} \ge 1 + b_i$*, then* $t_1^{(n)} > 0$ *for all* $n \ge 1$*.*

To prove Lemma A.2, we need the following statement.

Lemma A.3. *Fix $n > 0$ and, for $i \le 1 \le n$, let a_i, b_i, d_i be reals with $b_i > 0$ and $d_i \ne 0$. Consider the following matrix*

$$M = \begin{pmatrix} a_1 & d_1 & 0 & 0 & \cdots & 0 \\ d_1^{-1}b_1 & a_2 & d_2 & 0 & & \vdots \\ 0 & d_2^{-1}b_2 & a_3 & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & a_{n-1} & d_{n-1} \\ 0 & \cdots & \cdots & 0 & d_{n-1}^{-1}b_{n-1} & a_n \end{pmatrix}. \tag{A.1}$$

Then, the eigenvalues of M are distinct reals and independent of $d_1, ..., d_{n-1}$. Furthermore, if $a_1 \ge 1$ and $a_{i+1} \ge 1 + b_i$, then all eigenvalues of M are positive.

Proof of Lemma A.3. Let X, Y be diagonal matrices with $X_{11} = Y_{11} = 1$, $X_{ii} = d_1 d_1 \cdots d_{i-1}$ and $Y_{ii} = (b_1 b_2 \cdots b_{i-1})^{-1/2}(d_1 d_2 \cdots d_{i-1})$ for $i > 1$. One can show that

$$XMX^{-1} = \begin{pmatrix} a_1 & 1 & 0 & 0 & \cdots & 0 \\ b_1 & a_2 & 1 & 0 & & \vdots \\ 0 & b_2 & a_3 & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & a_{n-1} & 1 \\ 0 & \cdots & \cdots & 0 & b_{n-1} & a_n \end{pmatrix}.$$

Since XMX^{-1} is independent of the choice of $d_1, ..., d_{n-1}$, the eigenvalues of M are independent of $d_1, ..., d_{n-1}$. Note that YMY^{-1} is Hermitian. This implies that the eigenvalues of M are all real. As M is tridiagonal with non-zero entries in the superdiagonal, the rank of $M - \lambda I$ is either $n-1$ or n. This implies that the eigenvalues of M are all distinct.

Next, assume that $a_1 \ge 1$ and $a_{i+1} \ge 1 + b_i$. Let $(YMY^{-1})_i$ be the leading $i \times i$ principal matrices of YMY^{-1}. By induction, one can prove that $\det(YMY^{-1})_i = \prod_{j=1}^{i} \ell_j$, where $\ell_1 = a_1$ and $\ell_{j+1} = a_{j+1} - b_j/\ell_j$ for $1 \le j < n$. By the assumption at the beginning of this paragraph, $\ell_j \ge 1$ for all $1 \le j < n$ and $\det(YMY^{-1})_i > 0$ for all $1 \le i \le n$. As the leading principal matrices have positive determinants, (YMY^{-1}) is positive definite. This proves that all eigenvalues of M are positive. □

Proof of Lemma A.2. We prove this lemma by induction. For $n = 1$, it is clear that $t_1^{(1)} = a_1/c_1$ is the root for $\det M_1(t)$. For $n = 2$, note that $\det M_2(t)$ is a quadratic function that tends to infinity as $|t| \to \infty$. Since $\det M_2(t_1^{(1)}) = -b_1 < 0$, the polynomial, $\det M_2(t)$, has two real roots, say $t_1^{(2)} < t_2^{(2)}$, satisfying $t_1^{(2)} < t_1^{(1)} < t_2^{(2)}$. Now, we assume that, for some $n \ge 1$, $\det M_n(t)$ and $\det M_{n+1}(t)$ have reals roots $(t_i^{(n)})_{i=1}^n$ and $(t_i^{(n+1)})_{i=1}^{n+1}$ satisfying $t_i^{(n+1)} < t_i^{(n)} < t_{i+1}^{(n+1)}$ for $1 \le i \le n$. Clearly, $\det M_n(t) \to \infty$ as $t \to -\infty$. This implies

$$\det M_n(t_{2k+2}^{(n+1)}) < 0 < \det M_n(t_{2k+1}^{(n+1)}), \quad \forall k \ge 0.$$

Observe that $\det M_{n+2}(t) = (a_{n+2} - c_{n+2}t)\det M_{n+1}(t) - b_{n+1}\det M_n(t)$. Replacing t with $t_i^{(n+1)}$ yields

$$\det M_{n+2}(t_{2k+2}^{(n+1)}) > 0 > \det M_{n+2}(t_{2k+1}^{(n+1)}), \quad \forall k \geq 0.$$

This proves that $\det M_{n+2}(t)$ has $(n+2)$ distinct real roots with the desired interlacing property.

For the second part, assume that $a_1 \geq 1$ and $a_{i+1} \geq 1 + b_i$ for all $i \geq 1$. For $n = 1$, it is obvious that $t_1^{(1)} > 0$. Suppose $t_1^{(n)} > 0$. According to the first part, we have $t_2^{(n+1)} > t_1^{(n)} > 0$. By Lemma A.3, $\det M_{n+1}(0) > 0$, which implies $t_1^{(n+1)} \neq 0$. As it is known that $\det M_{n+1}(t) < 0$ for $t \in (t_1^{(n+1)}, t_2^{(n+1)})$, it must be the case $t_1^{(n+1)} > 0$. Otherwise, there will be another root for $\det M_{n+1}(t)$ between $t_1^{(n+1)}$ and 0, which is a contradiction. □

Proof of Theorem 6.3. For convenience, we set $\lambda_n^m = 1 - \cos\frac{m\pi}{n+1}$ for $1 \leq m \leq n$ and let $A_i(\lambda)$ be the i-by-i tridiagonal matrix with entries $(A_i(\lambda))_{kl} = 1$ for $|k-l| = 1$ and $(A_i(\lambda))_{kk} = 2 - \lambda$. For $1 \leq j \leq i$, let $B_i^j(\lambda)$ be the matrix equal to A_i except the (j,j)-entry, which is defined by $(B_i^j(\lambda,\epsilon))_{jj} = 2 - \lambda/\epsilon$. By Remark 3.9, $\lambda_{\pi_n,\nu_n^{x_n}}^{G_n}$ is the smallest root of $\det B_n^{x_n}(\lambda,\epsilon_n) = 0$ and $(\lambda_{n,m})_{m=1}^n$ are roots of $\det A_n(\lambda) = 0$. Note that, for $1 \leq j \leq n$,

$$\frac{\det B_n^j(\lambda,\epsilon)}{\det A_{j-1}(\lambda)\det A_{n-j}(\lambda)} = \Delta_n^j(\lambda,\epsilon) = 2 - \lambda/\epsilon - R_{j-1}(\lambda) - R_{n-j}(\lambda),$$

where $\det A_0(\lambda) := 1$, $\det A_{-1}(\lambda) := 0$ and

$$R_j(\lambda) = \frac{\det A_{j-1}(\lambda)}{\det A_j(\lambda)} = \frac{\prod_{i=1}^{j-1}(2\lambda_{j-1}^i - \lambda)}{\prod_{i=1}^{j}(2\lambda_j^i - \lambda)}.$$

To prove this theorem, one has to determine the sign of $\Delta_n^j(\lambda,\epsilon)$.

Let $\ell_n = \delta_n/n^2$ with $\delta_n \to 0$. As $n \to \infty$,

$$\log\frac{2\lambda_n^i - \ell_n}{2\lambda_n^i} = -\frac{\delta_n}{2\lambda_n^i n^2}(1 + o(1)),$$

where $o(1)$ is uniform for $1 \leq i \leq n$. Note that $\prod_{i=1}^j (2\lambda_j^i) = \det A_j(0) = j+1$. This implies

$$\begin{aligned}\log R_n(\ell_n) &= \log\frac{n}{n+1} + \left(\sum_{i=1}^n \frac{1}{\lambda_n^i n^2} - \sum_{i=1}^{n-1}\frac{1}{\lambda_{n-1}^i (n-1)^2}\right)\frac{\delta_n(1+o(1))}{2} \\ &= \log\frac{n}{n+1} + O\left(\frac{\delta_n}{n}\right).\end{aligned}$$

By a similar reasoning, one can prove that $\log R_j(\ell_n) = \log\frac{j}{j+1} + O(\delta_n/n)$ for bounded j. This shows that, for $j_n \in \{1, ..., n\}$ and $\ell_n = o(j_n^{-2})$,

$$R_{j_n}(\ell_n) = 1 - \frac{1}{j_n+1} + O(j_n\ell_n), \quad \text{as } n \to \infty. \tag{A.2}$$

Next, we compute $R_{j_n}(2C_n\lambda^1_{j_n})$ with $C_n \to C \in (0,1)$ and $j_n \to \infty$. Note that, for n large enough,

$$\text{(A.3)} \quad \begin{aligned} \log R_{j_n}(2C_n\lambda^1_{j_n}) = &\sum_{i=1}^{j_n-1} \frac{\lambda^i_{j_n-1} - \lambda^i_{j_n}}{\lambda^i_{j_n}} - \frac{1}{2}\sum_{i=1}^{j_n-1}\left(\frac{\lambda^i_{j_n-1} - \lambda^i_{j_n}}{\lambda^i_{j_n}}\right)^2 \\ &+ C_n \sum_{i=1}^{j_n-1} \frac{\lambda^1_{j_n}(\lambda^i_{j_n-1} - \lambda^i_{j_n})}{(\lambda^i_{j_n} - C_n\lambda^1_{j_n})\lambda^i_{j_n}} - \log 4 + O(j_n^{-2}). \end{aligned}$$

Calculus shows that

$$\begin{aligned} \sum_{i=1}^{j_n-1}\left(\frac{\lambda^i_{j_n-1} - \lambda^i_{j_n}}{\lambda^i_{j_n}}\right)^2 &= \frac{1}{\pi j_n}\int_0^\pi \frac{\theta^2 \sin^2\theta}{(1-\cos\theta)^2} d\theta + O(j_n^{-2}) \\ &= \frac{8\log 2 - \pi^2/3}{j_n} + O(j_n^{-2}) \end{aligned}$$

and

$$\sum_{i=1}^{j_n-1} \frac{\lambda^1_{j_n}(\lambda^i_{j_n-1} - \lambda^i_{j_n})}{(\lambda^i_{j_n} - C\lambda^1_{j_n})\lambda^i_{j_n}} = \frac{2}{j_n}\sum_{i=1}^{\infty}\frac{1}{i^2 - C} + O(j_n^{-2}).$$

Observe that, as $n \to \infty$,

$$\log \frac{j_n}{j_n+1} = \log R_{j_n}(0) = \sum_{i=1}^{j_n-1} \frac{\lambda^i_{j_n-1} - \lambda^i_{j_n}}{\lambda^1_{j_n}} - \log 4 + O(j_n^{-2}).$$

Putting this back into (A.3) implies

$$\text{(A.4)} \quad R_{j_n}(2C_n\lambda^1_{j_n}) = 1 + \left(-1 - 4\log 2 + \frac{\pi^2}{6} + C_n\sum_{i=1}^{\infty}\frac{1}{i^2 - C_n}\right)\frac{1}{j_n} + O(j_n^{-2}).$$

We consider the following two cases.

Case 1: $x_n = O(\epsilon_n n^2)$**.** In this case, Theorem 6.2 implies that $\lambda^{G_n}_{\pi_n,\nu_n^{x_n}} \asymp n^{-2}$. We assume further that $x_n/(\epsilon_n n^2) \to a$ and $x_n/n \to b$ with $a \in [0,\infty)$ and $b \in [0, 1/2]$. Let $C_n \to C \in (0,1)$. Replacing j_n with $x_n - 1$ in (A.2) and with $n - x_n$ in (A.4) yields that, for $b = 0$,

$$\Delta_n^{x_n}(2C_n\lambda^1_{n-x_n}, \epsilon_n) = \frac{(1-\pi^2 aC)(1+o(1))}{x_n}$$

and, for $b \in (0, 1/2]$,

$$\Delta_n^{x_n}(2C_n\lambda^1_{n-x_n}, \epsilon_n) = \left(1 + 4\log 2 - \frac{\pi^2}{6} - \frac{\pi^2 aC}{1-b} - bC\kappa_b(C)\right)\frac{(1+o(1))}{b(1-b)n},$$

where $\kappa_t(c) = \sum_{i=1}^{\infty}\frac{(1-t)i^2 - tc}{(i^2-c)[(1-t)^2 i^2 - t^2 c]}$. This proves (1) and (2).

Case 2: $\epsilon n^2 = o(x_n)$**.** This is exactly (3) and the result is immediate from Theorem 6.2. □

References

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

[2] Guan-Yu Chen and Laurent Saloff-Coste. The cutoff phenomenon for ergodic markov processes. *Electron. J. Probab.*, 13:26–78, 2008.

[3] Guan-Yu Chen and Laurent Saloff-Coste. The L^2-cutoff for reversible Markov processes. *J. Funct. Anal.*, 258(7):2246–2315, 2010.
[4] Guan-Yu Chen and Laurent Saloff-Coste. On the mixing time and spectral gap for birth and death chains. In preparation, 2012.
[5] P. Diaconis and L. Saloff-Coste. What do we know about the Metropolis algorithm? *J. Comput. System Sci.*, 57(1):20–36, 1998. 27th Annual ACM Symposium on the Theory of Computing (STOC'95) (Las Vegas, NV).
[6] Persi Diaconis and James Allen Fill. Strong stationary times via a new form of duality. *Ann. Probab.*, 18(4):1483–1522, 1990.
[7] Persi Diaconis and Laurent Saloff-Coste. Comparison techniques for random walk on finite groups. *Ann. Probab.*, 21(4):2131–2156, 1993.
[8] Persi Diaconis and Laurent Saloff-Coste. Comparison theorems for reversible Markov chains. *Ann. Appl. Probab.*, 3(3):696–730, 1993.
[9] Persi Diaconis and Laurent Saloff-Coste. Separation cut-offs for birth and death chains. *Ann. Appl. Probab.*, 16(4):2098–2122, 2006.
[10] Jian Ding, Eyal Lubetzky, and Yuval Peres. Total variation cutoff in birth-and-death chains. *Probab. Theory Related Fields*, 146(1-2):61–85, 2010.
[11] William Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons Inc., New York, 1968.
[12] James Allen Fill. Strong stationary duality for continuous-time Markov chains. I. Theory. *J. Theoret. Probab.*, 5(1):45–70, 1992.
[13] F. R. Gantmacher and M. G. Krein. Sur les matrices complétement non négatives et oscillatoires. *Compositio Math.*, 4:445–470, 1937.
[14] Laurent Miclo. On eigenfunctions of Markov processes on trees. *Probab. Theory Related Fields*, 142(3-4):561–594, 2008.
[15] Laurent Miclo. Monotonicity of the extremal functions for one-dimensional inequalities of logarithmic Sobolev type. In *Séminaire de Probabilités XLII*, volume 1979 of *Lecture Notes in Math.*, pages 103–130. Springer, Berlin, 2009.
[16] L. Saloff-Coste. Simple examples of the use of Nash inequalities for finite Markov chains. In *Stochastic geometry (Toulouse, 1996)*, volume 80 of *Monogr. Statist. Appl. Probab.*, pages 365–400. Chapman & Hall/CRC, Boca Raton, FL, 1999.
[17] Jeffrey Scott Silver. *Weighted Poincare and exhaustive approximation techniques for scaled Metropolis-Hastings algorithms and spectral total variation convergence bounds in infinite commutable Markov chain theory.* ProQuest LLC, Ann Arbor, MI, 1996. Thesis (Ph.D.)–Harvard University.
[18] Gerald Teschl. Oscillation theory and renormalized oscillation theory for Jacobi operators. *J. Differential Equations*, 129(2):532–558, 1996.
[19] Gerald Teschl. *Jacobi operators and completely integrable nonlinear lattices*, volume 72 of *Mathematical Surveys and Monographs.* American Mathematical Society, Providence, RI, 2000.

[1]Department of Applied Mathematics, National Chiao Tung University, Hsinchu 300, Taiwan
E-mail address: gychen@math.nctu.edu.tw

[2]Malott Hall, Department of Mathematics, Cornell University, Ithaca, NY 14853-4201
E-mail address: lsc@math.cornell.edu

ON THE MIXING TIME AND SPECTRAL GAP FOR BIRTH AND DEATH CHAINS

GUAN-YU CHEN[1] AND LAURENT SALOFF-COSTE[2]

Abstract. For birth and death chains, we derive bounds on the spectral gap and mixing time in terms of birth and death rates. Together with the results of Ding *et al.* in [15], this provides a criterion for the existence of a cutoff in terms of the birth and death rates. A variety of illustrative examples are treated.

1. Introduction

Let Ω be a countable set and (Ω, K, π) be an irreducible Markov chain on Ω with transition matrix K and stationary distribution π. Let I be the identity matrix indexed by Ω and

$$H_t = e^{-t(I-K)} = \sum_{i=0}^{\infty} e^{-t} t^i K^i / i!$$

be the associated semigroup which describes the corresponding natural continuous time process on Ω. For $\delta \in (0,1)$, set

$$K_\delta = \delta I + (1-\delta)K. \tag{1.1}$$

Clearly, K_δ is similar to K but with an additional holding probability depending of δ. We call K_δ the δ-lazy walk or δ-lazy chain of K. It is well-known that if K is irreducible with stationary distribution π, then

$$\lim_{m\to\infty} K_\delta^m(x,y) = \lim_{t\to\infty} H_t(x,y) = \pi(y), \quad \forall x,y \in \Omega,\ \delta \in (0,1).$$

In this paper, we consider convergence in total variation. The total variation between two probabilities μ, ν on Ω is defined by $\|\mu - \nu\|_{\mathrm{TV}} = \sup\{\mu(A) - \nu(A) | A \subset \Omega\}$. For any irreducible K with stationary distribution π, the (maximum) total variation distance is defined by

$$d_{\mathrm{TV}}(m) = \sup_{x\in\Omega} \|K^m(x,\cdot) - \pi\|_{\mathrm{TV}}, \tag{1.2}$$

and the corresponding mixing time is given by

$$T_{\mathrm{TV}}(\epsilon) = \inf\{m \ge 0 | d_{\mathrm{TV}}(m) \le \epsilon\}, \quad \forall \epsilon \in (0,1). \tag{1.3}$$

We write $d_{\mathrm{TV}}^{(c)}, T_{\mathrm{TV}}^{(c)}$ for the total variation distance and mixing time for the continuous semigroup and $d_{\mathrm{TV}}^{(\delta)}, T_{\mathrm{TV}}^{(\delta)}$ for the δ-lazy walk.

2000 *Mathematics Subject Classification.* 60J10,60J27.
Key words and phrases. Birth and death chains, Cutoff phenomenon.
[1]Partially supported by NSC grant NSC100-2115-M-009-003-MY2.
[2]Partially supported by NSF grant DMS-1004771.

A birth and death chain on $\{0,1,...,n\}$ with birth rate p_i, death rate q_i and holding rate r_i is a Markov chain with transition matrix K given by

$$K(i,i+1)=p_i,\quad K(i,i-1)=q_i,\quad K(i,i)=r_i,\quad \forall 0\le i\le n,$$

where $p_i+q_i+r_i=1$ and $p_n=q_0=0$. It is obvious that K is irreducible if and only if $p_iq_{i+1}>0$ for $0\le i<n$. Under the assumption of irreducibility, the unique stationary distribution π of K is given by $\pi(i)=c(p_0\cdots p_{i-1})/(q_1\cdots q_i)$, where c is a positive constant such that $\sum_{i=0}^n\pi(i)=1$. The following theorem provides a bound on the mixing time using the birth and death rates and is treated in Theorems 3.1 and 3.5.

Theorem 1.1. *Let K be an irreducible birth and death chain on $\{0,1,...,n\}$ with birth, death and holding rates p_i,q_i,r_i. Let i_0 be a state satisfying $\pi([0,i_0])\ge 1/2$ and $\pi([i_0,n])\ge 1/2$, where $\pi(A)=\sum_{i\in A}\pi(i)$, and set*

$$t=\max\left\{\sum_{k=0}^{i_0-1}\frac{\pi([0,k])}{\pi(k)p_k},\sum_{k=i_0+1}^{n}\frac{\pi([k,n])}{\pi(k)q_k}\right\}.$$

Then, for any $\delta\in[1/2,1)$,

$$\min\left\{T_{\rm TV}^{(c)}(1/10),T_{\rm TV}^{(\delta)}(1/20)\right\}\ge\frac{t}{6},$$

and

$$\max\left\{T_{\rm TV}^{(c)}(\epsilon),T_{\rm TV}^{(\delta)}(\epsilon)\right\}\le\frac{18t}{\epsilon^2},\quad\forall\epsilon\in(0,1).$$

The authors of [15] derive a similar upper bound. Note that if $(X_m)_{m=0}^\infty$ is a Markov chain on Ω_n with transition matrix K and $\tau_i:=\min\{m\ge 0|X_m=i\}$, then $t=\max\{\mathbb{E}_0\tau_{i_0},\mathbb{E}_n\tau_{i_0}\}$, where $\mathbb{E}_i$ denotes the conditional expectation given $X_0=i$. See Lemma 3.2 for details.

A sharp transition phenomenon, known as cutoff, was observed by Aldous and Diaconis in early 1980s. See e.g. [10, 5] for an introduction and a general review of cutoffs. In total variation, a family of irreducible Markov chains $(\Omega_n,K_n,\pi_n)_{n=1}^\infty$ is said to present a cutoff if

$$\lim_{n\to\infty}\frac{T_{n,\rm TV}(\epsilon)}{T_{n,\rm TV}(\eta)}=1,\quad\forall 0<\epsilon<\eta<1. \tag{1.4}$$

The family is said to present a (t_n,b_n) cutoff if $b_n=o(t_n)$ and

$$|T_{n,\rm TV}(\epsilon)-t_n|=O(b_n),\quad\forall 0<\epsilon<1.$$

The cutoff for the associated continuous semigroups is defined in a similar way. Given a family $\mathcal{F}$ of irreducible Markov chains, we write $\mathcal{F}_c$ and $\mathcal{F}_\delta$ for the families of corresponding continuous time chain and δ-lazy discrete time chains.

Let $\mathcal{F}=\{(\Omega_n,K_n,\pi_n)|n=1,2,...\}$ be a family of birth and death chains, where $\Omega_n=\{0,1,...,n\}$ and K_n has birth rate $p_{n,i}$, death rate $q_{n,i}$ and holding rate $r_{n,i}$. Suppose that K_n is irreducible with stationary distribution π_n. For the family $\{(\Omega_n,K_n,\pi_n)|n=1,2,...\}$, Ding *et al.* [15] showed that, in the discrete time case and assuming $\inf_{i,n}r_{n,i}>0$, the cutoff in total variation exists if and only if the product of the total variation mixing time and the spectral gap, i.e. the smallest non-zero eigenvalue of $I-K$, tends to infinity. There is also a similar version for the continuous time case. In [6], we use the results of [13, 15] to provide another criterion on the cutoff using the eigenvalues of K_n. In both cases, the spectral gap is needed to determine if there is a cutoff. The following theorem provides a bound on the spectral gap using the birth and death rates.

Theorem 1.2. *Consider an irreducible birth and death chain K on $\{0,1,...,n\}$ with birth, death and holding rates, p_i, q_i, r_i. Let π and λ be the stationary distribution and spectral gap of K and set*

$$\ell = \max\left\{\max_{j:j<i_0}\sum_{k=j}^{i_0-1}\frac{\pi([0,j])}{\pi(k)p_k}, \max_{j:j>i_0}\sum_{k=i_0+1}^{j}\frac{\pi([j,n])}{\pi(k)q_k}\right\},$$

where i_0 is a state such that $\pi([0,i_0]) \geq 1/2$ and $\pi([i_0,n]) \geq 1/2$. Then,

$$\frac{1}{4\ell} \leq \lambda \leq \frac{2}{\ell}.$$

The above theorem is motivated by [16], where the author considers the spectral gap of birth and death chains on $\mathbb{Z}$. We refer the reader to [16] and the references therein for more information. Note that if t, ℓ are the constants in Theorem 1.1-1.2, then $t \geq \ell$. Based on the results in [15], we obtain a theorem regarding cutoffs for birth and death chains.

Theorem 1.3. *Consider a family of irreducible birth and death chains*

$$\mathcal{F} = \{(\Omega_n, K_n, \pi_n) | n = 1, 2, ...\},$$

where $\Omega_n = \{0,1,...,n\}$ and K_n has birth, death and holding rates, $p_{n,i}, q_{n,i}, r_{n,i}$. For $n \geq 1$, let $i_n \in \{0,...,n\}$ be a state satisfying $\pi_n([0,i_n]) \geq 1/2$ and $\pi_n([i_n+1,n]) \geq 1/2$ and set

$$t_n = \max\left\{\sum_{k=0}^{i_n-1}\frac{\pi_n([0,k])}{\pi_n(k)p_{n,k}}, \sum_{k=i_n+1}^{n}\frac{\pi_n([k,n])}{\pi_n(k)q_{n,k}}\right\}.$$

and

$$\ell_n = \max\left\{\max_{j:j<i_n}\sum_{k=j}^{i_n-1}\frac{\pi_n([0,j])}{\pi_n(k)p_{n,k}}, \max_{j:j>i_n}\sum_{k=i_n+1}^{j}\frac{\pi_n([j,n])}{\pi_n(k)q_{n,k}}\right\},$$

Then, for any $\epsilon \in (0,1/2)$ and $\delta \in (0,1)$, there is a constant $C = C(\epsilon,\delta) > 1$ such that

$$C^{-1}t_n \leq \min\{T^{(c)}_{n,\mathrm{TV}}(\epsilon), T^{(\delta)}_{n,\mathrm{TV}}(\epsilon)\} \leq \max\{T^{(c)}_{n,\mathrm{TV}}(\epsilon), T^{(\delta)}_{n,\mathrm{TV}}(\epsilon)\} \leq Ct_n,$$

for n large enough. Moreover, the following are equivalent.

1. *$\mathcal{F}_c$ has a total variation cutoff.*
2. *For $\delta \in (0,1)$, $\mathcal{F}_\delta$ has a total variation cutoff.*
3. *$t_n\ell_n \to \infty$.*

The above theorem is immediate from Theorems 1.1, 1.2, 2.2 and 2.3. The selection of i_n can be relaxed. See Theorem 3.6 for a precise statement. By the results in [6], Theorem 1.3 also holds when t_n is replaced by the following constant

$$s_n = \frac{1}{\lambda_{n,1}} + \cdots + \frac{1}{\lambda_{n,n}},$$

where $\lambda_{n,1}, ..., \lambda_{n,n}$ are nonzero eigenvalues of $I - K_n$. Furthermore, Theorem 1.3 also holds in separation with $\delta \in [1/2, 1)$. We will use Theorem 1.3 to study the cutoff of several examples including the following theorem which concerns random walks with bottlenecks. It is a special case of Theorem 4.8.

Theorem 1.4. *For $n \geq 1$, let $\Omega_n = \{0, 1, ..., n\}$, $\pi_n \equiv 1/(n+1)$ and K_n be an irreducible birth and death chain on Ω_n satisfying*

$$K_n(i-1, i) = K_n(i, i-1) = \begin{cases} 1/2 & \text{for } i \notin \{x_{n,1}, ..., x_{n,k_n}\} \\ \epsilon_n & \text{for } i = x_{n,j},\ 1 \leq j \leq k_n \end{cases},$$

where $0 \leq k_n \leq n$, $\epsilon_n \in (0, 1/2]$, $x_{n,1}, ..., x_{n,k_n} \in \Omega_n$ are distinct and the holding rate at i is adjusted accordingly. Set $t_n = n^2 + a_n/\epsilon_n$, where

$$a_n = \sum_{i=1}^{k_n} \min\{x_{n,i}, n + 1 - x_{n,i}\},$$

and set

$$b_n = \max_{j:j \leq n/2} \{(j+1) \times |\{1 \leq i \leq k_n : j < x_{n,i} \leq n - j\}|\}.$$

Then, for any $\epsilon \in (0, 1/2)$ and $\delta \in (0, 1)$, there is $C = C(\epsilon, \delta) > 1$ such that

$$C^{-1} t_n \leq \min\{T^{(c)}_{n,\mathrm{TV}}(\epsilon), T^{(\delta)}_{n,\mathrm{TV}}(\epsilon)\} \leq \max\{T^{(c)}_{n,\mathrm{TV}}(\epsilon), T^{(\delta)}_{n,\mathrm{TV}}(\epsilon)\} \leq C t_n,$$

for n large enough.

Moreover, the following are equivalent.

(1) *$\mathcal{F}_c$ has a total variation cutoff.*
(2) *For $\delta \in (0, 1)$, $\mathcal{F}_\delta$ has a total variation cutoff.*
(3) *$a_n/(n^2 \epsilon_n) \to \infty$ and $a_n/b_n \to \infty$.*

The remaining of this article is organized as follows. In Section 2, the concepts of cutoffs and mixing times and fundamental results are reviewed. In Section 3, we give a proof for Theorems 1.1 and 1.2. For illustration, we consider several nontrivial examples in Section 4, where the mixing time and cutoff are determined. Note that the assumption regarding birth and death rates in Sections 3 and 4 can be relaxed using the comparison technique in [11, 12].

2. Backgrounds

Throughout this paper, for any two sequences s_n, t_n of positive numbers, we write $s_n = O(t_n)$ if there are $C > 0, N > 0$ such that $|s_n| \leq C|t_n|$ for $n \geq N$. If $s_n = O(t_n)$ and $t_n = O(s_n)$, we write $s_n \asymp t_n$. If $t_n/s_n \to 1$ as $n \to \infty$, we write $t_n \sim s_n$.

2.1. **Cutoffs and mixing time.** Consider the following definitions.

Definition 2.1. Referring to the notation in (1.2), a family $\mathcal{F} = \{(\Omega_n, K_n, \pi_n) | n = 1, 2, ...\}$ is said to present a total variation

(1) precutoff if there is a sequence t_n and $B > A > 0$ such that

$$\lim_{n\to\infty} d_{n,\mathrm{TV}}(\lceil Bt_n \rceil) = 0, \quad \liminf_{n\to\infty} d_{n,\mathrm{TV}}(\lfloor At_n \rfloor) > 0.$$

(2) cutoff if there is a sequence t_n such that, for all $\epsilon > 0$,

$$\lim_{n\to\infty} d_{n,\mathrm{TV}}(\lceil (1+\epsilon)t_n \rceil) = 0, \quad \lim_{n\to\infty} d_{n,\mathrm{TV}}(\lfloor (1-\epsilon)t_n \rfloor) = 1.$$

In definition 2.1(2), t_n is called a cutoff time. The definition of a cutoff for continuous semigroups is similar with $\lceil\cdot\rceil$ and $\lfloor\cdot\rfloor$ deleted.

Remark 2.1. In Definition 2.1, if $t_n \to \infty$ (or equivalently $T_{n,\mathrm{TV}}(\epsilon) \to \infty$ for some $\epsilon \in (0,1)$), then the cutoff is consistent with (1.4). This is also true for cutoffs in continuous semigroups without the assumption $t_n \to \infty$. See [4, 5] for further discussions on cutoffs.

It is well-known that the mixing time can be bounded below by the reciprocal of the spectral gap up to a multiple constant. We cite the bound in [6] as follows.

Lemma 2.1. *Let K be an irreducible transition matrix on a finite set Ω with stationary distribution π. For $\delta \in (0,1)$, let K_δ be the δ-lazy walk given by* (1.1)*. Suppose (π, K) is reversible, that is, $\pi(x)K(x,y) = \pi(y)K(y,x)$ for all $x, y \in \Omega$ and let λ be the smallest non-zero eigenvalue of $I - K$. Then, for $\epsilon \in (0, 1/2)$,*

$$T^{(c)}_{\mathrm{TV}}(\epsilon) \ge \frac{-\log(2\epsilon)}{\lambda}, \quad T^{(\delta)}_{\mathrm{TV}}(\epsilon) \ge \left\lfloor \frac{-\log(2\epsilon)}{2\max\{1-\delta, \log(2/\delta)\}\lambda} \right\rfloor,$$

where the second inequality requires $|\Omega| \ge 2/\delta$.

2.2. **Cutoffs for birth and death chains.** Consider a family of irreducible birth and death chains

$$\mathcal{F} = \{(\Omega_n, K_n, \pi_n) | n = 1, 2, ...\},$$

where $\Omega_n = \{0, 1, ..., n\}$ and K_n has birth rate $p_{n,i}$, death rate $q_{n,i}$ and holding rate $r_{n,i}$. We write $\mathcal{F}_c, \mathcal{F}_\delta$ as families of the corresponding continuous time chains and δ-lazy discrete time chains in $\mathcal{F}$. A criterion on total variation cutoffs for families of birth and death chains was introduced in [15], which say that, for $\delta \in (0,1)$, $\mathcal{F}_c, \mathcal{F}_\delta$ have total variation cutoffs if and only if the product of the mixing time and the spectral gap tends to infinity. As the total variation distance is comparable with the separation distance, the authors of [15] identify cutoffs in total variation and separation, where a criterion on separation cutoffs was proposed in [13]. In the recent work [6], the cutoffs for $\mathcal{F}_c$ and $\mathcal{F}_\delta$ are proved to be equivalent and this leads to the following theorems.

Theorem 2.2. [6, Section 4] *Let $\mathcal{F} = \{(\Omega_n, K_n, \pi_n) | n = 1, 2, ...\}$ be a family of irreducible birth and death chain with $\Omega_n = \{0, 1, ..., n\}$. For $n \ge 1$, let $\lambda_{n,1}, ..., \lambda_{n,n}$ be nonzero eigenvalues of $I - K_n$ and set*

$$\lambda_n = \min_{1 \le i \le n} \lambda_{n,i}, \quad s_n = \frac{1}{\lambda_{n,1}} + \cdots + \frac{1}{\lambda_{n,n}}.$$

Then, the following are equivalent.

1. *$\mathcal{F}_c$ has a total variation cutoff.*
2. *$\mathcal{F}_\delta$ has a total variation cutoff.*
3. *$\mathcal{F}_c$ has a total variation precutoff.*
4. *$\mathcal{F}_\delta$ has a total variation precutoff.*
5. *$T^{(c)}_{n,\mathrm{TV}}(\epsilon)\lambda_n \to \infty$ for some $\epsilon \in (0,1)$.*
6. *$T^{(\delta)}_{n,\mathrm{TV}}(\epsilon)\lambda_n \to \infty$ for some $\epsilon \in (0,1)$.*
7. *$s_n \lambda_n \to \infty$.*

Theorem 2.3. [6, Section 4] *Referring to Theorem 2.2, it holds true that, for $\epsilon, \eta \in (0, 1/2)$ and $\delta \in (0,1)$,*

$$T^{(c)}_{n,\mathrm{TV}}(\epsilon) \asymp T^{(\delta)}_{n,\mathrm{TV}}(\eta).$$

Further, if there is $\epsilon_0 \in (0,1/2)$ *such that* $T^{(c)}_{n,\mathrm{TV}}(\epsilon_0)\lambda_n$ *or* $T^{(\delta)}_{n,\mathrm{TV}}(\epsilon_0)\lambda_n$ *is bounded, then, for any* $\epsilon \in (0,1/2)$ *and* $\delta \in (0,1)$,

$$T^{(c)}_{n,\mathrm{TV}}(\epsilon) \asymp T^{(\delta)}_{n,\mathrm{TV}}(\epsilon) \asymp \lambda_n^{-1}.$$

2.3. **A remark on the precutoff.** Note that if there is no cutoff in total variation, the approximation in Theorem 2.3 may fail for $\epsilon \in (1/2,1)$. This means that, for $0<\epsilon<1/2<\eta<1$, the orders of $T^{(c)}_{n,\mathrm{TV}}(\epsilon)$ and $T^{(c)}_{n,\mathrm{TV}}(\eta)$ can be different. Consider the following example. For $n \ge 3$, let $\Omega_n = \{0,1,...,n\}$, $M_n = \lfloor n/2 \rfloor$ and

$$\begin{cases} K_n(i,i+1) = K_n(i+1,i) = 1/2 \quad \text{for } 0 \le i < n,\ i \ne M_n \\ K_n(M_n, M_n+1) = K_n(M_n+1, M_n) = \epsilon_n \\ K_n(0,0) = K_n(n,n) = 1/2 \\ K_n(M_n,M_n) = K_n(M_n+1,M_n+1) = 1/2 - \epsilon_n \end{cases}, \tag{2.1}$$

with $\epsilon_n \le 1/2$. Assume that $\epsilon_n = o(n^{-2})$. By Theorem 1.4, we have

$$T^{(c)}_{n,\mathrm{TV}}(\epsilon) \asymp T^{(\delta)}_{n,\mathrm{TV}}(\epsilon) \asymp n/\epsilon_n, \quad \forall \epsilon \in (0,1/2),\ \delta \in (0,1).$$

Next, we consider the δ-lazy discrete time case with $\delta = 1/2$. Let $K_{n,1/2} = (I+K_n)/2$ and K'_n be the 1/2-lazy simple random walk on $\{0,1,...,M_n\}$, that is,

$$\begin{cases} K'_n(i,i+1) = K'_n(i+1,i) = 1/4, \quad \forall 0 \le i < M_n \\ K'_n(i,i) = 1/2, \quad \forall 0 < i < M_n \\ K'_n(0,0) = K'_n(M_n,M_n) = 3/4 \end{cases}.$$

For $n \ge 3$, set

$$c_n = \min_{0 \le i,j \le M_n} \frac{K^{m_n}_{n,1/2}(i,j)}{(K'_n)^{m_n}(i,j)}, \quad C_n = \max_{0 \le i,j \le M_n} \frac{K^{m_n}_{n,1/2}(i,j)}{(K'_n)^{m_n}(i,j)}.$$

Proposition 2.4. *If* $m_n \asymp n^2$, *then*

$$c_n \to 1, \quad C_n \to 1, \quad \text{as } n \to \infty.$$

Proof. For $\ell \ge 1$, let $(i_0, i_1, ..., i_\ell)$ be a path in $\{0,1,...,M_n\}$. Note that

$$\prod_{k=1}^{\ell} K_{n,1/2}(i_{k-1}, i_k) \ge \left(\frac{3/4 - \epsilon_n/2}{3/4}\right)^{\ell} \prod_{k=1}^{\ell} K'_n(i_{k-1}, i_k).$$

This implies $c_n \ge (1 - 2\epsilon_n/3)^{m_n} \sim 1$ as $n \to \infty$. To see an upper bound of C_n, one may use Lemma 4.4 in [15] to conclude that, for $0 \le i \le n$ and $\ell \ge 0$,

$$\begin{cases} K^{\ell}_{n,1/2}(i,j) \ge K^{\ell}_{n,1/2}(i,j-1) \quad \forall 1 \le j \le 0 \\ K^{\ell}_{n,1/2}(i,j) \ge K^{\ell}_{n,1/2}(i,j+1) \quad \forall i \le j < n \end{cases},$$

and, for $0 \le i \le M_n$ and $\ell \ge 0$,

$$\begin{cases} (K'_n)^{\ell}(i,j) \ge (K'_n)^{\ell}(i,j-1) \quad \forall 1 \le j \le i \\ (K'_n)^{\ell}(i,j) \ge (K'_n)^{\ell}(i,j+1) \quad \forall i \le j < M_n \end{cases}.$$

By the induction, the above observation implies that, for any probabilities μ, ν on $\{0,...,n\}, \{0,...,M_n\}$ satisfying $\mu(i) = \nu(i)$ for $0 \le i \le M_n$,

$$\mu K^{\ell}_{n,1/2}(j) \le \nu (K'_n)^{\ell}(j), \quad \forall 0 \le j \le M_n,\ \ell \ge 0.$$

This yields $C_n \le 1$ for all $n \ge 3$. □

For $\epsilon \in (0,1)$, let $T'_{n,\mathrm{TV}}(\epsilon)$ be the total variation mixing time for K'_n. It is well-known that, for $\epsilon \in (0,1)$, $T'_{n,\mathrm{TV}}(\epsilon) \asymp n^2$. Let $d^{(1/2)}_{n,\mathrm{TV}}, d'_{n,\mathrm{TV}}$ be the total variation distance for $K_{n,1/2}, K'_n$. As a consequence of the above discussion, we obtain, for $\epsilon \in (0,1)$,

$$\limsup_{n\to\infty} d^{(1/2)}_{n,\mathrm{TV}}(T'_{n,\mathrm{TV}}(\epsilon)) \le \frac{1}{2}\left(1+\limsup_{n\to\infty} d'_{n,\mathrm{TV}}(T'_{n,\mathrm{TV}}(\epsilon)\right) \le \frac{1+\epsilon}{2}.$$

Thus, for $\epsilon \in (1/2,1)$, $T^{(1/2)}_{n,\mathrm{TV}}(\epsilon) = O(n^2)$. Note that, for $m_n = o(n^2)$,

$$\lim_{n\to\infty}\sum_{i\le an} K^{m_n}_{n,1/2}(0,i) = 1, \quad \forall a>0.$$

This yields $n^2 = O(T^{(1/2)}_{n,\mathrm{TV}}(\epsilon))$ for $\epsilon > 0$. The above discussion is also valid for the continuous time case and any δ-lazy discrete time case. We summarizes the results in the following theorem.

Theorem 2.5. *Let $\mathcal{F} = \{(\Omega_n, K_n, \pi_n)|n = 1,2,...\}$ be the family of birth and death chains in* (2.1) *and $\delta \in (0,1)$. Suppose that $\epsilon_n = o(n^{-2})$. Then, there is no total variation cutoff for $\mathcal{F}_c$ and $\mathcal{F}_\delta$. Furthermore, for $\epsilon \in (0,1/2)$,*

$$T^{(c)}_{n,\mathrm{TV}}(\epsilon) \asymp T^{(\delta)}_{n,\mathrm{TV}}(\epsilon) \asymp n/\epsilon_n,$$

and, for $\epsilon \in (1/2,1)$,

$$T^{(c)}_{n,\mathrm{TV}}(\epsilon) \asymp T^{(\delta)}_{n,\mathrm{TV}}(\epsilon) \asymp n^2.$$

Remark 2.2. Figure 1 displays the total variaton distances of the birth and death chains on $\{1,2,...,100\}$ with transition matrices K_1 and K_2 given by

$$\begin{cases} K_1(i,i) = 1/2, & \text{for } i \notin \{1,50,51,100\} \\ K_1(i,i+1) = K_1(i+1,i) = 1/4, & \text{for } i<50 \text{ or } i>51 \\ K_1(i,i) = 3/4 & \text{for } k \in \{1,100\} \\ K_1(i,i+1) = K_1(i+1,i) = 10^{-3} & \text{for } k = 50 \\ K_1(i,i) = K_1(i,i) = 3/4 - 10^{-3} & \text{for } i \in \{50,51\} \\ K_1(i,j) = 0 & \text{otherwise} \end{cases}$$

and

$$\begin{cases} K_2(i,i+1) = K_2(i+1,i) = 10^{-2} & \text{for } i = 25 \\ K_2(i,i) = 3/4 - 10^{-2} & \text{for } i \in \{25,26\} \\ K_2(i,j) = K_1(i,j) & \text{otherwise} \end{cases}.$$

Note that each curve has only one sharp transition for $d_{\mathrm{TV}}(t) \le 1/2$. This is consistent with Theorem 1.3. These examples show that multiple sharp transitions may occur for $d_{\mathrm{TV}}(t) > 1/2$. Note also that the flat part of the curves occupy very large time regions. For instance, the left most curve stays near the value 1/2 for t between 10^3 and 10^6.

3. Bounds for mixing time and spectral gap

This section is dedicated to proving Theorems 1.1 and 1.2. In the first two subsections, we treat respectively the upper and lower bounds of the total variation mixing time. This leads to Theorem 1.1. In the third subsection, we provide a

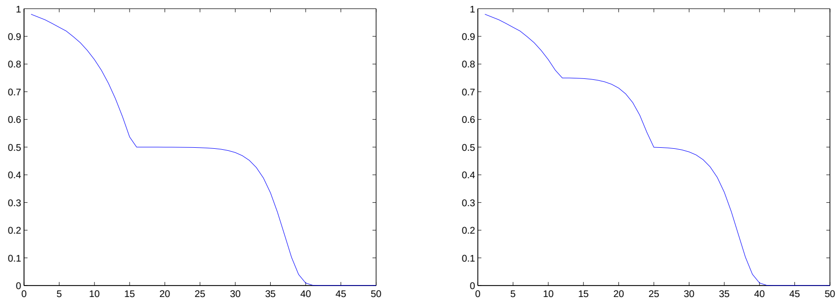


FIGURE 1. The curves display the total variation distance of the chains in Remark 2.2, where the left most curve is for K_1 and the right most curve is for K_2. The curve consists of the points $(m, d_{\mathrm{TV}}(100^{\lfloor 0.1\times m\rfloor}))$ with $m = 1, 2, ..., 50$. The right most point of each curve corresponds to $d_{\mathrm{TV}}(t)$ with $t = 10^{10}$.

relaxation of the choice of i_n in Theorem 1.3. In the last subsection, we introduce a bound on the spectral gap which includes Theorem 1.2.

3.1. **An upper bound of the mixing time.** Let (Ω, K, π) be an irreducible birth and death chain, where $\Omega = \{0, 1, ..., n\}$ and K has birth rate p_i, death rate q_i and holding rate r_i. Let $(X_m)_{m=0}^{\infty}$ be a realization of the discrete time chain. Obviously, if N_t is a Poisson process with parameter 1 and independent of $(X_m)_{m=0}^{\infty}$, then $(X_{N_t})_{t\geq 0}$ is a realization of the continuous time chain. For $\delta \in [0, 1)$, if $(B_m^{(\delta)})_{m=1}^{\infty}$ is a sequence of independent Bernoulli$(1-\delta)$ trials which are independent of $(X_m)_{m=0}^{\infty}$, then $Y_m^{(\delta)} = X_{B_1^{(\delta)}+\cdots+B_m^{(\delta)}}$ is a realization of the δ-lazy chain. For $0 \leq i \leq n$, we define the first passage time to i by

$$\widetilde{\tau}_i := \inf\{t \geq 0 | X_{N_t} = i\}, \quad \tau_i^{(\delta)} := \min\{m \geq 0 | Y_m = i\}, \tag{3.1}$$

and simply put $\tau_i := \tau_i^{(0)} = \min\{m \geq 0 | X_m = i\}$. Briefly, we write $\mathbb{P}_i(\cdot)$ for $\mathbb{P}(\cdot | X_0 = i)$ and write $\mathbb{E}_i, \mathrm{Var}_i$ as the expectation and variance under $\mathbb{P}_i$. The main result of this subsection is as follows.

Theorem 3.1 (Upper bound)**.** *Let (Ω, K, π) be an irreducible birth and death chain with $\Omega = \{0, 1, ..., n\}$. Let $\tau_i = \tau_i^{(0)}$ be the first passage time to i defined in* (3.1)*. For $\epsilon \in (0, 1)$ and $\delta \in [1/2, 1)$,*

$$\max\left\{T_{\mathrm{TV}}^{(c)}(\epsilon), (1-\delta)T_{\mathrm{TV}}^{(\delta)}(\epsilon)\right\} \leq \frac{9(\mathbb{E}_0\tau_{i_0} + \mathbb{E}_n\tau_{i_0})}{\epsilon^2}, \tag{3.2}$$

where $i_0 \in \{0, ..., n\}$ satisfies $\pi([0, i_0 - 1]) \leq 1/2$ and $\pi([i_0 + 1, n]) \leq 1/2$.

Remark 3.1. The authors of [6] obtain a slightly improved upper bound similar to (3.2), which says that

$$\max\left\{T_{\mathrm{TV}}^{(c)}(\epsilon), (1-\delta)T_{\mathrm{TV}}^{(\delta)}(\epsilon)\right\} \leq \frac{(\sqrt{\epsilon} + \sqrt{1-\epsilon})(\mathbb{E}_0\tau_{i_0} + \mathbb{E}_n\tau_{i_0})}{\sqrt{\epsilon}}.$$

Comparing with (3.2), the above inequality has an improved dependence on ϵ.

To understand the right side of (3.2), we introduce the following lemma.

Lemma 3.2. *Referring to the setting in* (3.1)*, it holds true that, for* $i < j$*,* $\mathbb{E}_i(\tau_j^{(\delta)}) = \mathbb{E}_i(\tau_j)/(1-\delta)$ *and* $\mathbb{E}_i(\tau_j) = \mathbb{E}_i(\widetilde{\tau}_j) = \sum_{k=i}^{j-1} \pi([0,k])/(p_k\pi(k))$.

Proof. The proof is based on the strong Markov property. See [2, Proposition 2] for a reference on the discrete time case, whereas the continuous time case is an immediate result of the fact $\{\widetilde{\tau}_i > t\} = \{\tau_i > N_t\}$. □

Remark 3.2. By Theorem 3.1 and Lemma 3.2, the total variation mixing time for the continuous time and the δ-lazy, with $\delta \geq 1/2$, discrete time birth and death chain on $\{0, 1, ..., n\}$ are bounded above by the following term up to a multiple constant.

$$\sum_{k=0}^{i_0-1} \frac{\pi([0,k])}{p_k\pi(k)} + \sum_{k=i_0+1}^{n} \frac{\pi([k,n])}{q_k\pi(k)},$$

where $i_0 \in \{0, ..., n\}$ satisfies $\pi([0, i_0 - 1]) \leq 1/2$ and $\pi([i_0 + 1, n]) \leq 1/2$.

Remark 3.3. In Theorem 3.1, i_0 is unique if $\pi([0,i]) \neq 1/2$ for all $0 \leq i \leq n$. If $\pi([0,j]) = 1/2$, then i_0 can be j or $j+1$, but the right side of (3.2) is the same in either case using Lemma 3.2.

Remark 3.4. Let K be an irreducible birth and death chain with birth, death and holding rates p_i, q_i, r_i and stationary distribution π. Let λ be the spectral gap of K. As a consequence of Lemma 2.1 and theorem 3.1, we obtain, for $\epsilon \in (0, 1/2)$,

$$\lambda \geq \frac{\epsilon^2 \log(1/(2\epsilon))}{9} \left(\sum_{k=0}^{i_0-1} \frac{\pi([0,k])}{p_k\pi(k)} + \sum_{k=i_0+1}^{n} \frac{\pi([k,n])}{q_k\pi(k)} \right)^{-1},$$

where i_0 is such that $\pi([0, i_0 - 1]) \leq 1/2$ and $\pi([i_0 + 1, n]) \leq 1/2$. The maximum of $\epsilon^2 \log(1/(2\epsilon))$ on $(0, 1/2)$ is attained at $\epsilon = 1/(2\sqrt{e})$ and equal to $1/(8e)$. A similar lower bound of the spectral gap is also derived in [7] with improved constant.

As a simple application of Lemma 3.2, we have

Corollary 3.3. *Referring to Lemma 3.2, for* $i \leq j$*,*

$$\mathbb{E}_i\tau_j \leq \left(\frac{1}{\pi([j,n])} - 1 \right) \mathbb{E}_n\tau_i.$$

Proof. By Lemma 3.2, one has

$$\mathbb{E}_i\tau_j = \sum_{k=i}^{j-1} \frac{\pi([0,k])}{p_k\pi(k)}, \quad \mathbb{E}_n(\tau_i) = \sum_{k=i}^{n-1} \frac{\pi([k+1,n])}{q_{k+1}\pi(k+1)} = \sum_{k=i}^{n-1} \frac{\pi([k+1,n])}{p_k\pi(k)}.$$

The inequality is then given by the fact $\pi([0,k])/\pi([k+1,n]) = 1/\pi([k+1,n]) - 1 \leq 1/\pi([j,n]) - 1$ for $k < j$. □

The following proposition is the main technique used to prove Theorem 3.1.

Proposition 3.4. *Referring to the setting in* (3.1)*, it holds true that, for* $j < k$*,*

$$d_{\mathrm{TV}}^{(c)}(i,t) \leq \mathbb{P}_i(\max\{\widetilde{\tau}_j, \widetilde{\tau}_k\} > t) + 1 - \pi([j,k]),$$

and

$$d_{\mathrm{TV}}^{(1/2)}(i,t) \leq \mathbb{P}_i(\max\{\tau_j^{(1/2)}, \tau_k^{(1/2)}\} > t) + 1 - \pi([j,k]),$$

In particular,

$$d_{\mathrm{TV}}^{(c)}(t) \leq \frac{\mathbb{E}_0\widetilde{\tau}_k + \mathbb{E}_n\widetilde{\tau}_j}{t} + 1 - \pi([j,k])$$

and

$$d_{\rm TV}^{(1/2)}(t)\le \frac{2(\mathbb{E}_0\tau_k^{(1/2)}+\mathbb{E}_n\tau_j^{(1/2)})}{t}+1-\pi([j,k]).$$

In the above proposition, the discrete time case is discussed in Lemma 2.3 in [15]. Our method to prove this proposition is to construct a no-crossing coupling. We give the proof of the continuous time case for completeness and refer to [15] for the discrete time case, where a heuristic idea on the construction of no-crossing coupling is proposed.

Proof of Proposition 3.4. Let $(Y_t)_{t\ge 0}$ be another process corresponding to H_t with $Y_0\overset{d}{=}\pi$. Set $T:=\inf\{t\ge 0|X_t=Y_t\}$ and $Z_t:=Y_t\mathbf{1}_{\{t\le T\}}+X_t\mathbf{1}_{\{t>T\}}$. Clearly, $(X_t,Z_t)_{t\ge 0}$ is a coupling for the semigroup H_t and must be no-crossing according to the continuous time setting. Note that $T=\inf\{t\ge 0|X_t=Z_t\}$ is the coupling time of X_t and Z_t. The classical coupling statement implies that

$$d_{\rm TV}^{(c)}(i,t)\le \mathbb{P}_i(T>t). \tag{3.3}$$

See e.g. [1] for a reference. Note that $X_{\tau_j}=j$, $X_{\tau_k}=k$ and

$$\mathbb{P}_i(X_{\widetilde{\tau}_j}\le Y_{\widetilde{\tau}_j})=\pi([j,n]),\quad \mathbb{P}_i(X_{\widetilde{\tau}_k}\ge Y_{\widetilde{\tau}_k})=\pi([0,k]).$$

As X_t,Y_t can not cross each other without coalescing in advance, this implies

$$\begin{aligned}\mathbb{P}_i(T\le \max\{\widetilde{\tau}_j,\widetilde{\tau}_k\})&\ge \mathbb{P}_i(\min\{\widetilde{\tau}_j,\widetilde{\tau}_k\}\le T\le \max\{\widetilde{\tau}_j,\widetilde{\tau}_k\})\\&\ge \mathbb{P}_i(X_{\widetilde{\tau}_j}\le Y_{\widetilde{\tau}_j},X_{\widetilde{\tau}_k}\ge Y_{\widetilde{\tau}_k})\ge \pi([j,k]).\end{aligned}$$

Putting this back to (3.3) gives the desired result.

For the last part, note that if $i\le j$, then $\widetilde{\tau}_j<\widetilde{\tau}_k$ and, by Markov's inequality, this implies

$$\mathbb{P}_i(\max\{\widetilde{\tau}_j,\widetilde{\tau}_k\}>t)\le \mathbb{P}_0(\widetilde{\tau}_k>t)\le \mathbb{E}_0\widetilde{\tau}_k/t.$$

Similarly, for $i\ge k$, one can show that

$$\mathbb{P}_i(\max\{\widetilde{\tau}_j,\widetilde{\tau}_k\}>t)\le \mathbb{P}_n(\widetilde{\tau}_j>t)\le \mathbb{E}_n\widetilde{\tau}_j/t.$$

For $j<i<k$, we have

$$\mathbb{P}_i(\max\{\widetilde{\tau}_j,\widetilde{\tau}_k\}>t)\le \mathbb{P}_i(\widetilde{\tau}_j>t)+\mathbb{P}_i(\widetilde{\tau}_k>t)\le \frac{\mathbb{E}_n\widetilde{\tau}_j+\mathbb{E}_0\widetilde{\tau}_k}{t}.$$

□

Proof of Theorem 3.1. Set $j_\epsilon=\min\{i\ge 0|\pi([0,i])\ge \epsilon/3\}$ and $k_\epsilon=\min\{i\ge 0|\pi([0,i])\ge 1-\epsilon/3\}$. By Proposition 3.4 and Lemma 3.2, the choice of $j=j_\epsilon$ and $k=k_\epsilon$ implies that

$$T_{\rm TV}^{(c)}(\epsilon)\le \frac{3(\mathbb{E}_0\tau_{k_\epsilon}+\mathbb{E}_n\tau_{j_\epsilon})}{\epsilon}.$$

By Corollary 3.3, one has

$$\mathbb{E}_0\tau_{k_\epsilon}=\mathbb{E}_0\tau_{i_0}+\mathbb{E}_{i_0}\tau_{k_\epsilon}\le \mathbb{E}_0\tau_{i_0}+\left(\frac{3}{\epsilon}-1\right)\mathbb{E}_n\tau_{i_0}$$

and

$$\mathbb{E}_n\tau_{j_\epsilon}=\mathbb{E}_n\tau_{i_0}+\mathbb{E}_{i_0}\tau_{j_\epsilon}\le \mathbb{E}_n\tau_{i_0}+\left(\frac{3}{\epsilon}-1\right)\mathbb{E}_0\tau_{i_0}.$$

Adding up both terms gives the upper bound in continuous time case. The proof for the (1/2)-lazy discrete time case is similar and, by Proposition 3.4, we obtain

$T_{\rm TV}^{(1/2)}(\epsilon) \le 18(\mathbb{E}_0\tau_{i_0} + \mathbb{E}_n\tau_{i_0})/\epsilon^2$. For $\delta \in (1/2, 1)$, note that $K_\delta = (K_{2\delta-1})_{1/2}$. Since the birth and death rates of $K_{2\delta-1}$ are $2(1-\delta)p_i$ and $2(1-\delta)q_i$, the above result and Lemma 3.2 lead to $T_{\rm TV}^{(\delta)}(\epsilon) \le 9(\mathbb{E}_0\tau_{i_0} + \mathbb{E}_n\tau_{i_0})/((1-\delta)\epsilon^2)$. □

3.2. **A lower bound of the mixing time.** The goal of this subsection is to establish a lower bound on the total variation mixing time for birth and death chains. Recall the notations in the previous subsection. Let $(X_m)_{m=0}^\infty$ be an irreducible birth and death chain with transition matrix K and stationary distribution π. Let N_t be a Poisson process of parameter 1 that is independent of X_m. For $0 \le i \le n$, let $\tau_i = \min\{m \ge 0 | X_m = i\}$ and $\widetilde{\tau}_i = \inf\{t \ge 0 | X_{N_t} = i\}$. Then, the total variation mixing time satisfies

$$d_{\rm TV}(0,t) \ge K^t(0,[0,i-1]) - \pi([0,i-1]) \ge \mathbb{P}_0(\tau_i > t) - \pi([0,i-1]) \tag{3.4}$$

and

$$d_{\rm TV}^{(c)}(0,t) \ge H_t(0,[0,i-1]) - \pi([0,i-1]) \ge \mathbb{P}_0(\widetilde{\tau}_i > t) - \pi([0,i-1]). \tag{3.5}$$

Brown and Shao discuss the distribution of $\widetilde{\tau}_i$ in [3], of which proof also works for the discrete time case. In detail, if $-1 < \beta_1 < \cdots < \beta_i < 1$ are the eigenvalues of the submatrix of K indexed by $\{0, ..., i-1\}$ and $\lambda_j = 1 - \beta_j$, then

$$\mathbb{P}_0(\tau_i > t) = \sum_{j=1}^i \left(\prod_{k \ne j} \frac{\lambda_k}{\lambda_k - \lambda_j} \right) (1-\lambda_j)^t \tag{3.6}$$

and

$$\mathbb{P}_0(\widetilde{\tau}_i > t) = \sum_{j=1}^i \left(\prod_{k \ne j} \frac{\lambda_k}{\lambda_k - \lambda_j} \right) e^{-t\lambda_j}. \tag{3.7}$$

Note that, under $\mathbb{P}_0$, $\widetilde{\tau}_i$ is the sum of independent exponential random variables with parameters $\lambda_1, ..., \lambda_i$. If $\beta_1 > 0$, then τ is the sum of independent geometric random variables with parameters $\lambda_1, ..., \lambda_i$. In discrete time case, the requirement $\beta_1 > 0$ holds automatically for the δ-lazy chain with $\delta \ge 1/2$. The above formula leads to the following theorem.

Theorem 3.5 (Lower bound). *Let K be the transition matrix of an irreducible birth and death chain on $\{0, 1, ..., n\}$. Let $\tau_i = \tau_i^{(0)}$ be the first passage time to i defined in* (3.1). *For $\delta \in [1/2, 1)$,*

$$\min\{T_{\rm TV}^{(c)}(1/10), 2(1-\delta)T_{\rm TV}^{(\delta)}(1/20)\} \ge \frac{\max\{\mathbb{E}_0\tau_{i_0}, \mathbb{E}_n\tau_{i_0}\}}{6},$$

where $i_0 \in \{0, ..., n\}$ satisfies $\pi([0, i_0-1]) \le 1/2$ and $\pi([i_0+1, n]) \le 1/2$.

Proof of Theorem 3.5. First, we consider the continuous time case. Let $\lambda_1, ..., \lambda_i$ be eigenvalues of the submatrix of $I - K$ indexed by $0, ..., i-1$ and $\widetilde{\tau}_{i,1}, ..., \widetilde{\tau}_{i,i}$ be independent exponential random variables with parameters $\lambda_1, ..., \lambda_i$. By (3.7), $\widetilde{\tau}_i$ and $\widetilde{\tau}_{i,1} + \cdots + \widetilde{\tau}_{i,i}$ are identically distributed under $\mathbb{P}_0$ and, by (3.5), this implies

$$d_{\rm TV}^{(c)}(0,t) \ge \mathbb{P}(\widetilde{\tau}_{i,1} + \cdots + \widetilde{\tau}_{i,i} > t) - \pi([0,i-1]).$$

It is easy to see that

$$\mathbb{E}_0\widetilde{\tau}_i = \frac{1}{\lambda_1} + \cdots + \frac{1}{\lambda_i}, \quad \mathrm{Var}_0(\widetilde{\tau}_i) = \frac{1}{\lambda_1^2} + \cdots + \frac{1}{\lambda_i^2}.$$

Let $a \in (0,1)$ and consider the following two cases. If $1/\lambda_j > a\mathbb{E}_0\widetilde{\tau}_i$ for some $1 \le j \le i$, then

$$\mathbb{P}_0(\widetilde{\tau}_i > t) \ge \mathbb{P}(\widetilde{\tau}_{i,j} > t) > e^{-t/(a\mathbb{E}_0\widetilde{\tau}_i)}.$$

If $1/\lambda_j \le a\mathbb{E}_0\widetilde{\tau}_i$ for all $1 \le j \le i$, then $\mathrm{Var}_0(\widetilde{\tau}_i) \le a(\mathbb{E}_0\widetilde{\tau}_i)^2$ and, by the one-sided Chebyshev inequality, we have

$$\mathbb{P}_0(\widetilde{\tau}_i > t) \ge \frac{(t-\mathbb{E}_0\widetilde{\tau}_i)^2}{\mathrm{Var}_0(\widetilde{\tau}_i)+(t-\mathbb{E}_0\widetilde{\tau}_i)^2} \ge \frac{(t-\mathbb{E}_0\widetilde{\tau}_i)^2}{a(\mathbb{E}_0\widetilde{\tau}_i)^2+(t-\mathbb{E}_0\widetilde{\tau}_i)^2} = \frac{(1-b)^2}{a+(1-b)^2},$$

for $t = b\mathbb{E}_0\widetilde{\tau}_i$ with $b \in (0,1)$. Combining both cases and setting $i = i_0$ in (3.5) yields that, for $a, b \in (0,1)$,

$$d_{\mathrm{TV}}^{(c)}(0, b\mathbb{E}_0\widetilde{\tau}_{i_0}) \ge \min\left\{e^{-b/a}, \frac{(1-b)^2}{a+(1-b)^2}\right\} - \frac{1}{2}. \tag{3.8}$$

Putting $a = 1/3$ and $b = 1/6$ gives $T_{\mathrm{TV}}^{(c)}(0, 1/10) \ge \mathbb{E}_0\widetilde{\tau}_{i_0}/6$.

For the discrete time case, note that the eigenvalues of the submatrix of $I - K_{1/2} = \frac{1}{2}(I-K)$ indexed by $0, ..., i-1$ are $\lambda_1/2, ..., \lambda_i/2$. Let $\tau_{i,1}, ..., \tau_{i,i}$ be independent geometric random variables with success probabilities $\lambda_1/2, ..., \lambda_i/2$. Replacing K with $K_{1/2}$ in (3.4), we obtain

$$d_{\mathrm{TV}}^{(1/2)}(0,t) \ge \mathbb{P}_0(\tau_{i,1} + \cdots + \tau_{i,i} > t) - \pi([0, i-1]).$$

Note that, under $\mathbb{P}_0$, $\tau_i^{(1/2)}$ has the same distribution as $\tau_{i,1} + \cdots + \tau_{i,i}$ and this implies

$$\mathbb{E}_0\tau_i^{(1/2)} = \frac{2}{\lambda_1} + \cdots + \frac{2}{\lambda_i}, \quad \mathrm{Var}_0(\tau_i^{(1/2)}) = \sum_{j=1}^{i} \frac{4(1-\lambda_j/2)}{\lambda_j^2} \le \sum_{j=1}^{i} \frac{4}{\lambda_j^2}.$$

Using the same analysis as before, one may derive, for $1/\mathbb{E}_0\tau_i^{(1/2)} < a < 1$ and $t < \mathbb{E}_0\tau_i^{(1/2)}$,

$$\mathbb{P}_0(\tau_i^{(1/2)} > t) \ge \min\left\{\left(1 - \frac{1}{a\mathbb{E}_0\tau_i^{(1/2)}}\right)^t, \frac{\left(t-\mathbb{E}_0\tau_i^{(1/2)}\right)^2}{a\left(\mathbb{E}_0\tau_i^{(1/2)}\right)^2 + \left(t-\mathbb{E}_0\tau_i^{(1/2)}\right)^2}\right\}.$$

By Lemma 3.2, $\mathbb{E}_0\tau_i^{(1/2)} \ge 2i$. Obviously, if $i_0 = 0$, then $T_{\mathrm{TV}}^{(1/2)}(0, 1/20) \ge 0 = \mathbb{E}_0\tau_{i_0}^{(1/2)}$. For $i_0 \ge 1$, $\mathbb{E}_0\tau_{i_0}^{(1/2)} \ge 2$ and the setting, $a = 2/3$ and $t = \left\lfloor \mathbb{E}_0\tau_{i_0}^{(1/2)}/12 \right\rfloor$, implies

$$d_{\mathrm{TV}}^{(1/2)}\left(0, \left\lfloor \mathbb{E}_0\tau_{i_0}^{(1/2)}/12 \right\rfloor\right) \ge \min\left\{2^{-1/3}, \frac{(11/12)^2}{2/3+(11/12)^2}\right\} - \frac{1}{2} > \frac{1}{20},$$

where the first inequality use the fact that $s\log(1-3/(2s))$ is increasing on $[2,\infty)$. Hence, we have $T_{\mathrm{TV}}^{(1/2)}(0, 1/20) \ge \mathbb{E}_0\tau_{i_0}^{(1/2)}/12 = \mathbb{E}_0\tau_{i_0}/6$. For $\delta > 1/2$, the combination of the above result and the observation $K_\delta = (K_{2\delta-1})_{1/2}$ implies that $T_{\mathrm{TV}}^{(\delta)}(0, 1/20) \ge \mathbb{E}_0\tau_{i_0}/(12(1-\delta))$.

The analysis from the other end point gives the other lower bound. This finishes the proof. □

3.3. **Relaxation of the median condition.** In some cases, it is not easy to determine the value of i_n in Theorem 1.3. Let t_n be the constants in Theorem 3.1. For $c \in (0,1)$, let $i_n(c) \in \{0,...,n\}$ be the state such that $\pi_n([0, i_n(c)-1]) \le c$, $\pi_n([i_n(c)+1, n]) \le 1-c$ and let $t_n(c)$ be the following constant

$$t_n(c) = \sum_{k=0}^{i_n(c)-1} \frac{\pi_n([0,k])}{\pi_n(k)p_{n,k}} + \sum_{k=i_n(c)+1}^{n} \frac{\pi_n([k,n])}{\pi_n(k)q_{n,k}}.$$

Assume that $c \ge 1/2$. In this case, if i_n is the smallest median, then $i_n \le i_n(c)$ and

$$\sum_{k=i_n}^{i_n(c)-1} \frac{\pi([0,k])}{\pi_n(k)p_{n,k}} = \sum_{k=i_n+1}^{i_n(c)} \frac{\pi_n([0,k-1])}{\pi_n(k)q_{n,k}}.$$

Note that, for $i_n < k \le i_n(c)$,

$$\frac{1}{2} \le \pi_n([0,i_n]) \le \frac{\pi_n([0,k-1])}{\pi_n([k,n])} \le \frac{1}{\pi_n([i_n(c),n])} \le \frac{1}{1-c}.$$

This implies $t_n/2 \le t_n(c) \le t_n/(1-c)$. Similarly, for $c \le 1/2$, one can show that $t_n/2 \le t_n(c) \le t_n/c$. Combining both cases gives

$$t_n/2 \le t_n(c) \le t_n/\min\{c, 1-c\}. \tag{3.9}$$

As a consequence of the above discussion, we obtain the following theorem.

Theorem 3.6. *Referring to Theorem 1.3. For $n \ge 1$, let $j_n \in \{0,1,...,n\}$ and set*

$$t_n' = \max\left\{ \sum_{k=0}^{j_n-1} \frac{\pi_n([0,k])}{\pi_n(k)p_{n,k}}, \sum_{k=j_n+1}^{n} \frac{\pi_n([k,n])}{\pi_n(k)q_{n,k}} \right\}.$$

Suppose that

$$0 < \liminf_{n\to\infty} \pi_n([0,j_n]) \le \limsup_{n\to\infty} \pi_n([0,j_n]) < 1.$$

Then, Theorem 1.3 remains true if t_n is replaced by t_n'.

Proof. The proof comes immediately from (3.9) with $c = \pi_n([0,j_n])$. □

We use this observation to bound the cutoff time in the following theorem.

Theorem 3.7. *Referring to Theorem 1.3. Suppose that $\mathcal{F}_c$ has a total variation cutoff. Then, for any $\epsilon \in (0,1)$,*

$$\frac{2\log 2}{5} \le \liminf_{n\to\infty} \frac{T_{n,\mathrm{TV}}^{(c)}(\epsilon)}{t_n} \le \limsup_{n\to\infty} \frac{T_{n,\mathrm{TV}}^{(c)}(\epsilon)}{t_n} \le 2$$

Proof of Theorem 3.7. The upper bound is given by Remark 3.1 and the fact, $\max\{s,t\} \ge (s+t)/2$, whereas the lower bound is obtained by applying $a = 2/5$ and $b = a\log(2/(1+2\epsilon))$ in (3.8) with $\epsilon \to 0$. □

3.4. **Bounding the spectral gap.** This subsection is devoted to poviding bounds on the specral gap for birth and death chains. As the graph associated with a birth and death chain is a path, weighted Hardy's inequality can be used to bound the spectral gap. We refer to the Appendix for a detailed discussion of the following results. See Theorems A.1-A.3.

Theorem 3.8. *Consider an irreducible birth and death chain on* $\{0,...,n\}$ *with birth, death and holding rates* p_i, q_i, r_i *and stationary distribution* π*. Let* λ *be the spectral gap and set, for* $0 \le i \le n$*,*

$$C(i) = \max\left\{\max_{j:j<i}\sum_{k=j}^{i-1}\frac{\pi([0,j])}{\pi(k)p_k}, \max_{j:j>i}\sum_{k=i+1}^{j}\frac{\pi([j,n])}{\pi(k)q_k}\right\}.$$

Then, for $0 \le m \le n$*,*

$$\frac{1}{4C(m)} \le \lambda \le \frac{1}{\min\{\pi([0,m]),\pi([m,n])\}C(m)}.$$

In particular, if M *is a median of* π*, that is,* $\pi([0,M]) \ge 1/2$ *and* $\pi([M,n]) \ge 1/2$*, then*

$$\frac{1}{4C(M)} \le \lambda \le \frac{2}{C(M)}.$$

Theorem 3.9. *Consider an irreducible birth and death chain on* $\{0,...,n\}$ *with birth, death and holding rates* p_i, q_i, r_i *and stationary distribution* π*. Let* λ *be the spectral gap and set* $N = \lceil n/2 \rceil$*. Suppose that* $p_i = q_{n-i}$ *for* $0 \le i \le n$*. Then,*

$$\frac{1}{4C} \le \lambda \le \frac{1}{C},$$

where

$$C = \max_{0\le i\le N-1}\left\{\pi([0,i])\sum_{j=i}^{N-1}\frac{1}{\pi(j)p_j}\right\} \quad \text{if } n \text{ is even},$$

and

$$C = \max_{0\le i\le N-1}\left\{\pi([0,i])\left(\sum_{j=i}^{N-2}\frac{1}{\pi(j)p_j} + \frac{1}{2\pi(N-1)p_{N-1}}\right)\right\} \quad \text{if } n \text{ is odd}.$$

Remark 3.5. In [18], the author also obtained bounds similar to Theorem 3.9 for the case $\pi(i) \ge \pi(i+1)$ with $0 \le i < n/2$ using the path technique. For more information on path techniques, see [11, 12, 14] and the references therein.

4. Examples

In this section, we will apply the theory developed in the previous section to examples of special interest. First, we give a criterion on the cutoff using the birth and death rates.

Theorem 4.1 (Cutoffs from birth and death rates)**.** *Let* $\mathcal{F} = \{(\Omega_n, K_n, \pi_n)|n = 1,2,...\}$ *be a family of irreducible birth and death chains on* $\Omega_n = \{0,1,...,n\}$ *with* *birth rate, $p_{n,i}$, death rate $q_{n,i}$ and holding rate $r_{n,i}$. Let λ_n be the spectral gap of K_n. For $n \ge 1$, let $j_n \in \{0,...,n\}$ and set*

$$t_n = \max\left\{\sum_{k=0}^{j_n-1}\frac{\pi_n([0,k])}{\pi_n(k)p_{n,k}}, \sum_{k=j_n+1}^{n}\frac{\pi_n([k,n])}{\pi_n(k)q_{n,k}}\right\}$$

and

$$\ell_n = \max\left\{\max_{j:j<j_n}\sum_{k=j}^{j_n-1}\frac{\pi_n([0,j])}{\pi_n(k)p_{n,k}}, \max_{j:j>j_n}\sum_{k=j_n+1}^{j}\frac{\pi_n([j,n])}{\pi_n(k)q_{n,k}}\right\}.$$

Suppose that

$$0 < \liminf_{n\to\infty}\pi_n([0,j_n]) \le \limsup_{n\to\infty}\pi_n([0,j_n]) < 1.$$

Then, for $\epsilon \in (0,1/2)$ and $\delta \in (0,1)$,

$$\lambda_n \asymp 1/\ell_n, \quad T^{(c)}_{n,\mathrm{TV}}(\epsilon) \asymp t_n \asymp T^{(\delta)}_{n,\mathrm{TV}}(\epsilon).$$

Furthermore, the following are equivalent.

(1) *$\mathcal{F}_c$ has a cutoff in total variation.*
(2) *For $\delta \in (0,1)$, $\mathcal{F}_\delta$ has a cutoff in total variation.*
(3) *$\mathcal{F}_c$ has precutoff in total variation.*
(4) *For $\delta \in (0,1)$, $\mathcal{F}_\delta$ has a precutoff in total variation.*
(5) *$t_n/\ell_n \to \infty$.*

The above theorem is obvious from Theorems 2.2, 3.6 and 3.8. We use two classical examples, simple random walks and Ehrenfest chains, to illustrate how to apply Theorem 4.1 to determine the total variation cutoff and mixing times.

Example 4.1 (Simple random walks on finite paths). For $n \ge 1$, the simple random walk on $\{0,...,n\}$ is a birth and death chain with $p_{n,i} = q_{n,i+1} = 1/2$ for $0 \le i < n$ and $r_{n,0} = r_{n,n} = 1/2$. It is clear that K_n is irreducible and aperiodic with uniform stationary distribution. Let t_n, ℓ_n be the constants in Theorem 4.1. It is an easy exercise to show that $\ell_n \asymp n^2 \asymp t_n$. By Theorem 4.1, neither $\mathcal{F}_c$ nor $\mathcal{F}_\delta$ has total variation precutoff, but $T^{(c)}_{n,\mathrm{TV}}(\epsilon) \asymp n^2 \asymp T^{(\delta)}_{n,\mathrm{TV}}(\epsilon)$ for $\epsilon \in (0,1/2)$ and $\delta \in (0,1)$. In fact, one may use a hitting time statement to prove that the mixing time has order at least n^2, when $\epsilon \in [1/2,1)$. This implies that the above approximation of mixing time holds for $\epsilon \in (0,1)$.

Example 4.2 (Ehrenfest chains). Consider the Ehrenfest chain on $\{0,...,n\}$, which is a birth and death chain with rates $p_{n,i} = 1 - i/n$ and $q_{n,i} = i/n$. It is obvious that K_n is irreducible and periodic with stationary distribution $\pi_n(i) = 2^{-n}\binom{n}{i}$. An application of the representation theory shows that, for $0 \le i \le n$, $2i/n$ is an eigenvalue of $I - K_n$. Let λ_n, s_n be the constants in Theorem 2.2. Clearly, $\lambda_n = 2/n$ and $s_n \asymp n\log n$ and, by Theorem 2.2, both $\mathcal{F}_c$ and $\mathcal{F}_\delta$ have a total variation cutoff. Note that, as a simple corollary, one obtains the non-trivial estimates

$$\sum_{i=0}^{\lceil\frac{n}{2}\rceil-1}\frac{\binom{n}{0}+\cdots+\binom{n}{i}}{\binom{n}{i}} \asymp n\log n, \quad \max_{0\le i<n/2}\sum_{j=0}^{i}\binom{n}{j}\times\sum_{j=i}^{\lceil\frac{n}{2}\rceil-1}\binom{n}{i}^{-1} \asymp n.$$

For a detailed computation on the total variation and the L^2-distance, see e.g. [9].

In the next subsections, we consider birth and death chains of special types.

4.1. **Chains with valley stationary distributions.** In this subsection, we consider birth and death chains with valley stationary distribution. For $n \ge 1$, let $\Omega_n = \{0, 1, ..., n\}$ and K_n be an irreducible birth and death chain on Ω_n with birth, death and holding rates, $p_{n,i}, q_{n,i}, r_{n,i}$. Suppose that there is $j_n \in \Omega_n$ such that

$$p_{n,i} \le q_{n,i+1},\ \forall i < j_n, \quad p_{n,i} \ge q_{n,i+1},\ \forall i \ge j_n. \tag{4.1}$$

Obviously, the stationary distribution π_n of K_n satisfies $\pi_n(i) \ge \pi_n(i+1)$ for $i < j_n$ and $\pi_n(i) \le \pi_n(i+1)$ for $i \ge j_n$.

Let t_n, ℓ_n be the constants in Theorem 4.1 and write

$$\ell_n = \max\left\{ \max_{j:j<j_n} \sum_{k=j+1}^{j_n} \frac{\pi_n([0,j])}{\pi_n(k)q_{n,k}}, \max_{j:j>j_n} \sum_{k=j_n}^{j-1} \frac{\pi_n([j,n])}{\pi_n(k)p_{n,k}} \right\}.$$

Set

$$M_L = \max_{0<i\le j_n} q_{n,i},\ m_L = \min_{0<i\le j_n} q_{n,i},\ M_R = \max_{j_n\le i<n} p_{n,i},\ m_R = \min_{j_n\le i<n} p_{n,i}.$$

Clearly,

$$\ell_n \le \max\left\{ \frac{\pi_n([0,j_n])}{m_L} \sum_{i=0}^{j_n} \frac{1}{\pi_n(i)}, \frac{\pi_n([j_n,n])}{m_R} \sum_{i=j_n}^{n} \frac{1}{\pi_n(i)} \right\}.$$

Let j_n' be such that $\pi_n([0,j_n']) \ge \pi_n([0,j_n])/2$ and $\pi_n([j_n',j_n]) \ge \pi_n([0,j_n])/2$. Note that if $j_n \ge 1$, then $j_n \ge \max\{2j_n', j_n'+1\}$. By (4.1), this implies

$$\sum_{k=j_n'+1}^{j_n} \frac{\pi_n([0,j_n'])}{\pi_n(k)} \ge \frac{\pi_n([0,j_n])}{4} \sum_{k=j_n'}^{j_n} \frac{1}{\pi_n(k)} \ge \frac{\pi_n([0,j_n])}{8} \sum_{k=0}^{j_n} \frac{1}{\pi_n(k)}.$$

One can derive a similar inequality from the other end point and this yields

$$\ell_n \ge \frac{1}{8} \min\left\{ \frac{\pi_n([0,j_n])}{M_L} \sum_{i=0}^{j_n} \frac{1}{\pi_n(i)}, \frac{\pi_n([j_n,n])}{M_R} \sum_{i=j_n}^{n} \frac{1}{\pi_n(i)} \right\}.$$

For t_n, note that

$$\frac{\pi_n([0,j_n-1])}{2} \sum_{k=0}^{j_n-1} \frac{1}{\pi_n(k)} \le \sum_{k=0}^{j_n-1} \frac{\pi_n([0,k])}{\pi_n(k)} \le \pi_n([0,j_n-1]) \sum_{k=0}^{j_n-1} \frac{1}{\pi_n(k)}$$

and

$$\frac{\pi_n([j_n+1,n])}{2} \sum_{k=j_n+1}^{n} \frac{1}{\pi_n(k)} \le \sum_{k=j_n+1}^{n} \frac{\pi_n([k,n])}{\pi_n(k)} \le \pi_n([j_n+1,n]) \sum_{k=j_n+1}^{n} \frac{1}{\pi_n(k)}$$

This implies

$$t_n \le \max\left\{ \frac{\pi_n([0,j_n])}{m_L} \sum_{i=0}^{j_n} \frac{1}{\pi_n(i)}, \frac{\pi_n([j_n,n])}{m_R} \sum_{i=j_n}^{n} \frac{1}{\pi_n(i)} \right\}$$

and

$$t_n \ge \frac{1}{8} \max\left\{ \frac{\pi_n([0,j_n])}{M_L} \sum_{i=0}^{j_n} \frac{1}{\pi_n(i)}, \frac{\pi_n([j_n,n])}{M_R} \sum_{i=j_n}^{n} \frac{1}{\pi_n(i)} \right\}$$

The following theorem is an immediate consequence of the above discussion and Theorem 4.1.

Theorem 4.2. *Let $\mathcal{F} = \{(\Omega_n, K_n, \pi_n) | n = 1, 2, ...\}$ be a family of birth and death chains satisfying* (4.1). *Assume that $\pi_n([0, j_n]) \asymp \pi_n([j_n, n])$ and*

$$\max_{0<i\le j_n} q_{n,i} \asymp \min_{0<i\le j_n} q_{n,i}, \quad \max_{j_n\le i<n} p_{n,i} \asymp \min_{j_n\le i<n} p_{n,i}.$$

Then, there is no cutoff for $\mathcal{F}_c, \mathcal{F}_\delta$ and, for $\epsilon \in (0, 1/2)$ and $\delta \in (0, 1)$,

$$T^{(c)}_{n,\mathrm{TV}}(\epsilon) \asymp T^{(\delta)}_{n,\mathrm{TV}}(\epsilon) \asymp \frac{1}{\lambda_n} \asymp \max\left\{ \frac{1}{q_{n,j_n}} \sum_{i=0}^{j_n} \frac{1}{\pi_n(i)}, \frac{1}{p_{n,j_n}} \sum_{i=j_n}^{n} \frac{1}{\pi_n(i)} \right\}.$$

For an illustration of the above theorem, we consider the following Markov chains. For $n \ge 1$, let $\Omega_n = \{0, 1, ..., n\}$, π_n be a non-uniform probability distribution on Ω_n satisfying (4.1) and M_n be a transition matrix given by

$$M_n(i,j) = \begin{cases} 1/2 & \text{for } j = i-1, i \le j_n, \\ 1/2 & \text{for } j = i+1, i \ge j_n, \\ \pi_n(i+1)/(2\pi_n(i)) & \text{for } j = i+1, i < j_n, \\ \pi_n(i-1)/(2\pi_n(i)) & \text{for } j = i-1, i > j_n, \\ 1/2 - \pi_n(i+1)/(2\pi_n(i)) & \text{for } j = i < j_n, \\ 1/2 - \pi_n(i-1)/(2\pi_n(i)) & \text{for } j = i > j_n. \end{cases} \tag{4.2}$$

Note that M_n is the Metropolis chain for π_n associated to the simple random walk on Ω_n. For more information on the Metropolis chain, see [8] and the references therein. The next theorem is a corollary of Theorem 4.2.

Theorem 4.3. *Let $\mathcal{F} = \{(\Omega_n, M_n, \pi_n) | n = 1, 2, ..\}$ be the family of Metropolis chains satisfying* (4.1)-(4.2). *Suppose $\pi_n([0, j_n]) \asymp \pi_n([j_n, n])$. Then, neither $\mathcal{F}_c$ nor $\mathcal{F}_\delta$ has a total variation precutoff but, for $\epsilon \in (0, 1/2)$ and $\delta \in (0, 1)$,*

$$T^{(c)}_{n,\mathrm{TV}}(\epsilon) \asymp \sum_{i=0}^{n} \frac{1}{\pi_n(i)} \asymp T^{(\delta)}_{n,\mathrm{TV}}(\epsilon).$$

Example 4.3. Let $a > 0$ and $\check{\pi}_{n,a}, \hat{\pi}_{n,a}$ be probability measures on $\{0, \pm 1, ..., \pm n\}$ given by

$$\check{\pi}_{n,a}(i) = \check{c}_{n,a}(|i| + 1)^a, \quad \hat{\pi}_{n,a}(i) = \hat{c}_{n,a}(n - |i| + 1)^a, \tag{4.3}$$

where $\check{c}_{n,a}, \hat{c}_{n,a}$ are normalizing constants. Let $\check{\mathcal{F}}, \hat{\mathcal{F}}$ be families of the Metropolis chains for $\check{\pi}_{n,a}, \hat{\pi}_{n,a}$ associated to the simple random walks on $\{0, \pm 1, ..., \pm n\}$, that is,

$$\check{M}_{n,a}(i,j) = \check{M}_{n,a}(-i,-j), \quad \hat{M}_{n,a}(i,j) = \hat{M}_{n,a}(-i,-j)$$

and

$$\check{M}_{n,a}(i,j) = \begin{cases} \frac{1}{2} & \text{if } j = i+1, i \in [0, n-1] \\ \frac{i^a}{2(i+1)^a} & \text{if } j = i-1, i \in [1, n] \\ \frac{(i+1)^a - i^a}{2(i+1)^a} & \text{if } j = i, i \notin \{0, n\} \\ 1 - \frac{n^a}{2(n+1)^a} & \text{if } i = j = n \end{cases}$$

and

$$\hat{M}_{n,a}(i,j) = \begin{cases} \frac{1}{2} & \text{if } j = i-1, i \in [1, n] \\ \frac{(n-i)^a}{2(n-i+1)^a} & \text{if } j = i+1, i \in [0, n-1] \\ \frac{(n-i+1)^a - (n-i)^a}{2(n-i+1)^a} & \text{if } j = i \ne 0 \\ 1 - \frac{n^a}{(n+1)^a} & \text{if } i = j = 0 \end{cases}.$$

Let $\check{\lambda}_{n,a}, \hat{\lambda}_{n,a}$ and $\check{T}_{n,a}, \hat{T}_{n,a}$ be the spectral gaps and total variation mixing times of $\check{M}_{n,a}, \hat{M}_{n,a}$. It has been proved in [7, 18] that there is $C > 1$ such that, for all $a > 0$ and $n \geq 1$,

$$\frac{1}{C\check{\lambda}_{n,a}} \asymp n^a \left(\left(1 + \frac{1}{n}\right)^a + \frac{n}{1+a} \right) (1 + v(n,a)) \leq \frac{C}{\check{\lambda}_{n,a}}$$

and

$$\frac{1}{C\hat{\lambda}_{n,a}} \leq \frac{(n+a)^2}{(1+a)^2} \leq \frac{C}{\hat{\lambda}_{n,a}},$$

where $v(n,1) = \log n$ and $v(n,a) = (n^{1-a} - 1)/(1-a)$ for $a \neq 1$. By Theorem 4.2, $\check{\mathcal{F}}_c$ and $\check{\mathcal{F}}_\delta$ have no cutoff in total variation but, for fixed $a > 0$, $\epsilon \in (0, 1/2)$ and $\delta \in (0,1)$,

$$\check{T}^{(c)}_{n,a}(\epsilon) \asymp \check{T}^{(\delta)}_{n,a}(\epsilon) \asymp \begin{cases} n^2 & \text{if } a \in (0,1) \\ n^2 \log n & \text{if } a = 1 \\ n^{1+a} & \text{if } a \in (1,\infty) \end{cases}.$$

The above result in continuous time case is also obtained in [18].

To see the cutoff for $\hat{\mathcal{F}}$, let

$$t_n = \sum_{k=0}^{n-1} \frac{\hat{\pi}_{n,a}([-n, -n+k])}{\hat{\pi}_{n,a}(-n+k)} = \sum_{k=1}^{n} k^{-a} \sum_{j=1}^{k} j^a.$$

By Theorems 3.1-3.5, we have

$$\frac{2t_n}{3} \leq \hat{T}^{(c)}_{n,a}(1/10) \leq 3600 t_n.$$

Note that, for $k \geq 1$ and $a > 0$,

$$\frac{k^a(k+a)}{2(1+a)} \leq \sum_{j=1}^{k} j^a \leq \frac{2k^a(k+a)}{1+a}.$$

This implies

$$\frac{n(n+a)}{6(1+a)} \leq \hat{T}^{(c)}_{n,a}(1/10) \leq \frac{14400 n(n+a)}{1+a}.$$

We collect the above results in the following theorem.

Theorem 4.4. *For $n \geq 1$, let $a_n > 0$ and $\check{\pi}_{n,a_n}, \hat{\pi}_{n,a_n}$ be probability measures given by (4.3). Let $\check{\mathcal{F}}, \hat{\mathcal{F}}$ be the families of Metropolis chains for $\check{\pi}_{n,a_n}, \hat{\pi}_{n,a_n}$ as above with total variation mixing time $\check{T}_{n,\mathrm{TV}}, \hat{T}_{n,\mathrm{TV}}$. Then, for $\epsilon \in (0,1/2)$ and $\delta \in (0,1)$,*

$$\hat{T}^{(c)}_{n,\mathrm{TV}}(\epsilon) \asymp \hat{T}^{(\delta)}_{n,\mathrm{TV}}(\epsilon) \asymp \frac{n(n+a_n)}{1+a_n}$$

and

$$\check{T}^{(c)}_{n,\mathrm{TV}}(\epsilon) \asymp \check{T}^{(\delta)}_{n,\mathrm{TV}}(\epsilon) \asymp n^{a_n} \left(\left(1 + \frac{1}{n}\right)^{a_n} + \frac{n}{1+a_n} \right) (1 + v(n,a_n)),$$

where $v(n,1) = \log n$ and $v(n,a) = (n^{1-a} - 1)/(1-a)$ for $a \neq 1$.

Moreover, neither $\check{\mathcal{F}}_c$ nor $\check{\mathcal{F}}_\delta$ has a total variation cutoff. Also, $\hat{\mathcal{F}}_c$ and $\hat{\mathcal{F}}_\delta$ have a total variation cutoff if and only if $a_n \to \infty$.

4.2. **Chains with monotonic stationary distributions.** In this subsection, we consider birth and death chains with monotonic stationary distributions. For $n \ge 1$, let $\Omega_n = \{0, 1, ..., n\}$ and K_n be a birth and death chain on Ω_n with birth, death and holding rates, $p_{n,i}, q_{n,i}, r_{n,i}$. Suppose that

$$p_{n,i} \ge q_{n,i+1}, \quad \forall 0 \le i < n. \tag{4.4}$$

If K_n is irreducible, then the stationary distribution π_n satisfying $\pi_n(i) \le \pi_n(i+1)$ for $0 \le i < n$. Let $j_n \in \Omega_n$ and t_n, ℓ_n be the constants in Theorem 4.1. Assume that $\pi_n([0, j_n]) \asymp \pi_n([j_n, n])$ and

$$\max_{0 \le i < j_n} p_{n,i} \asymp \min_{0 \le i < j_n} p_{n,i}, \quad \max_{j_n \le i < n} p_{n,i} \asymp \min_{j_n \le i < n} p_{n,i}. \tag{4.5}$$

Using a discussion similar to that in front of Theorem 4.2, one can show that

$$t_n \asymp \max \left\{ \frac{1}{p_{n,1}} \sum_{k=0}^{j_n - 1} \frac{\pi_n([0, k])}{\pi_n(k)}, \frac{1}{p_{n,j_n}} \sum_{k=j_n}^{n} \frac{1}{\pi_n(k)} \right\}$$

and

$$\ell_n \asymp \max \left\{ \frac{1}{p_{n,1}} \max_{0 \le j < j_n} \sum_{k=j}^{j_n - 1} \frac{\pi_n([0, j])}{\pi_n(k)}, \frac{1}{p_{n,j_n}} \sum_{k=j_n}^{n} \frac{1}{\pi_n(k)} \right\}.$$

This leads to the following theorem.

Theorem 4.5. *Let $\mathcal{F} = \{(\Omega_n, K_n, \pi_n) | n = 1, 2, ...\}$ be a family of irreducible birth and death chains with $\Omega_n = \{0, 1, ..., n\}$ and birth, death and holding rates $p_{n,i}, q_{n,i}, r_{n,i}$. Let $\lambda_n, T_{n,\mathrm{TV}}$ be the spectral gap and total variation mixing time of K_n and set*

$$u_n = \sum_{k=0}^{j_n - 1} \frac{\pi_n([0, k])}{\pi_n(k)}, \quad v_n = \max_{0 \le j < j_n} \sum_{k=j}^{j_n - 1} \frac{\pi_n([0, j])}{\pi_n(k)}, \quad w_n = \sum_{k=j_n}^{n} \frac{1}{\pi_n(k)}.$$

Assume that $\pi_n([0, j_n]) \asymp \pi_n([j_n, n])$ and (4.5) holds. Then, for $\epsilon \in (0, 1/2)$ and $\delta \in (0, 1)$,

$$\lambda_n^{-1} \asymp \max \left\{ \frac{v_n}{p_{n,1}}, \frac{w_n}{p_{n,j_n}} \right\}, \quad T_{n,\mathrm{TV}}^{(c)}(\epsilon) \asymp T_{n,\mathrm{TV}}^{(\delta)}(\epsilon) \asymp \max \left\{ \frac{u_n}{p_{n,1}}, \frac{w_n}{p_{n,j_n}} \right\}.$$

Moreover, $\mathcal{F}_c$ and $\mathcal{F}_\delta$ have a total variation cutoff if and only if

$$u_n / v_n \to \infty, \quad (u_n p_{n,j_n}) / (w_n p_{n,1}) \to \infty.$$

For $n \ge 1$, let f_n be a non-decreasing function on $[0, n]$ and set $F_n(x) = \int_0^x f_n(t) dt$ and $G_n(x, m) = \int_x^m 1/f_n(t) dt$. Note that if there is $C > 1$ such that

$$C^{-1} f_n(i) \pi_n(0) \le \pi_n(i) \le C f_n(i) \pi_n(0), \quad \forall 0 \le i \le n, \, n \ge 1,$$

then

$$\frac{1}{2C^2} \left(\frac{F_n(k)}{f_n(k)} + 1 \right) \le \frac{\pi_n([0, k])}{\pi_n(k)} \le C^2 \left(\frac{F_n(k)}{f_n(k)} + 1 \right)$$

and

$$\frac{1}{2C} \left(G_n(j, j_n) + \frac{1}{f_n(j)} \right) \le \pi_n(0) \sum_{k=j}^{j_n - 1} \frac{1}{\pi_n(k)} \le C \left(G_n(j, j_n) + \frac{1}{f_n(j)} \right).$$

This implies

$$\pi_n([0,j])\sum_{k=j}^{j_n-1}\frac{1}{\pi_n(k)}\le C^2\left(G_n(j,j_n)+\frac{1}{f_n(j)}\right)(F_n(j)+f_n(j))$$

and

$$\pi_n([0,j])\sum_{k=j}^{j_n-1}\frac{1}{\pi_n(k)}\ge \frac{1}{4C^2}\left(G_n(j,j_n)+\frac{1}{f_n(j)}\right)(F_n(j)+f_n(j)).$$

Let u_n, v_n, w_n be the constants in Theorem 4.5 and assume that

$$\min_{0\le i<n} p_{n,i}\asymp \max_{0\le i<n} p_{n,i}\asymp 1.$$

Consider the following cases.

Case 1: $f_n(x)=\exp\{\alpha_n x^{\beta_n}\}$ with $\inf_n\alpha_n>0$ and $\inf_n\beta_n\ge 1$. In this case, $F_n(x)=O(f_n(x))$ and $G_n(x,m)=O(1/f_n(x))$ for $1\le x<m$. By setting $j_n=n$, we obtain

$$\pi_n([0,j_n])\asymp\pi_n([j_n,n]),\quad u_n\asymp n,\quad v_n\asymp w_n\asymp 1.$$

By Theorem 4.5, $\lambda_n\asymp 1$ and, for $\epsilon\in(0,1/2)$ and $\delta\in(0,1)$,

$$T_{n,\mathrm{TV}}^{(c)}(\epsilon)\asymp T_{n,\mathrm{TV}}^{(\delta)}(\epsilon)\asymp n.$$

There is a total variation cutoff for $\mathcal{F}_c$ or $\mathcal{F}_\delta$.

Case 2: $f_n(x)=\exp\{\alpha_n x^{\beta_n}\}$ with $0<\inf_n\alpha_n\le\sup_n\alpha_n<\infty$ and $0<\inf_n\beta_n\le\sup_n\beta_n<1$. Note that, for $\alpha\in\mathbb{R}$ and $\beta\in(0,1)$,

$$\frac{d}{dx}\left(x^{1-\beta}e^{\alpha x^\beta}\right)=\left(\alpha\beta+(1-\beta)x^{-\beta}\right)e^{\alpha x^\beta}.$$

This implies that, uniformly for $n/2\le x$ and $1+x\le m\le n$,

$$F_n(x)\asymp x^{1-\beta_n}f_n(x),\quad G_n(x,m)\asymp\left(\frac{x^{1-\beta_n}}{f_n(x)}-\frac{m^{1-\beta_n}}{f_n(m)}\right).$$

Letting $j_n=\lfloor n-n^{1-\beta_n}\rfloor$ yields

$$\pi_n([0,j_n])\asymp\pi_n([j_n,n]),\quad u_n\asymp n^{2-\beta_n},\quad v_n\asymp n^{2-2\beta_n}\asymp w_n.$$

By Theorem 4.5, $\mathcal{F}_c$ and $\mathcal{F}_\delta$ have a total variation cutoff and

$$\lambda_n\asymp n^{2\beta_n-2},\quad T_{n,\mathrm{TV}}^{(c)}(\epsilon)\asymp T_{n,\mathrm{TV}}^{(\delta)}(\epsilon)\asymp n^{2-\beta_n},\quad\forall\epsilon\in(0,1/2),\,\delta\in(0,1).$$

Case 3: $f_n(x)=\exp\{\alpha_n[\log(x+1)]^{\beta_n}\}$ with $0<\inf_n\alpha_n\le\sup_n\alpha_n<\infty$ and $1<\inf_n\beta_n\le\sup_n\beta_n<\infty$. Note that, for $\alpha\in\mathbb{R}$ and $\beta>1$,

$$\frac{d}{dx}\left(\frac{(x+1)e^{\alpha[\log(x+1)]^\beta}}{[\log(x+1)]^{\beta-1}}\right)=\left(\alpha\beta+\frac{1-(\beta-1)/\log(x+1)}{[\log(x+1)]^{\beta-1}}\right)e^{\alpha[\log(x+1)]^\beta}.$$

This implies that, uniformly for $n/2\le x<m\le n$,

$$F_n(x)\asymp\frac{(x+1)}{[\log(x+1)]^{\beta_n-1}}e^{\alpha_n[\log(x+1)]^{\beta_n}}$$

and

$$G_n(x,m)\asymp\left(\frac{(x+1)e^{-\alpha_n[\log(x+1)]^{\beta_n}}}{[\log(x+1)]^{\beta_n-1}}-\frac{(m+1)e^{-\alpha_n[\log(m+1)]^{\beta_n}}}{[\log(m+1)]^{\beta_n-1}}\right).$$

Set $j_n = n[1-(\log n)^{1-\beta_n}]$. The above computation leads to

$$\pi_n([0,j_n]) \asymp \pi_n([j_n,n]), \quad u_n \asymp n^2(\log n)^{1-\beta_n}, \quad v_n \asymp n^2(\log n)^{2-2\beta_n} \asymp w_n.$$

By Theorem 4.5, both $\mathcal{F}_c$ and $\mathcal{F}_\delta$ have a total variation cutoff and, for $\epsilon \in (0,1/2)$ and $\delta \in (0,1)$,

$$\lambda_n \asymp n^{-2}(\log n)^{2\beta_n-2}, \quad T^{(c)}_{n,\mathrm{TV}}(\epsilon) \asymp n^2(\log n)^{1-\beta_n} \asymp T^{(\delta)}_{n,\mathrm{TV}}(\epsilon).$$

Case 4: $f_n(x) = \exp\{\alpha_n[\log(x+1)]^{\beta_n}\}$ with $\sup_n \alpha_n < \infty$ and $\sup_n \beta_n \le 1$. Observe that, for $\alpha > 0$ and $\beta \le 1$,

$$\frac{d}{dx}\left((x+1)e^{\alpha[\log(x+1)]^\beta}\right) = \left(1+\alpha\beta[\log(x+1)]^{\beta-1}\right)e^{\alpha[\log(x+1)]^\beta}.$$

This implies that, uniformly for $n/4 \le i < m \le n$,

$$F_n(i) \asymp (i+1)e^{\alpha_n[\log(i+1)]^{\beta_n}},\ G_n(i,m) \asymp \left(\frac{i+1}{e^{\alpha_n[\log(i+1)]^{\beta_n}}} - \frac{m+1}{e^{\alpha_n[\log(m+1)]^{\beta_n}}}\right).$$

Letting $j_n = \lfloor n/2 \rfloor$ implies

$$\pi_n([0,j_n]) \asymp \pi_n([j_n,n]), \quad u_n \asymp v_n \asymp w_n \asymp n^2.$$

By Theorem 4.5, we have

$$T^{(c)}_{n,\mathrm{TV}}(\epsilon) \asymp T^{(\delta)}_{n,\mathrm{TV}}(\epsilon) \asymp \lambda_n^{-1} \asymp n^2, \quad \forall \epsilon \in (0,1/2),\ \delta \in (0,1),$$

and there is no total variation cutoff for $\mathcal{F}_c$ or $\mathcal{F}_\delta$.

4.3. **Chains with symmetric stationary distributions.** This subsection is dedicated to the study of birth and death chains with symmetric stationary distributions. Let K be an irreducible birth and death chain on $\{0,...,n\}$ with stationary distribution π. Note that π is symmetric at $n/2$, that is, $\pi(n-i) = \pi(i)$ for $0 \le i \le n/2$, if and only if

$$p_i p_{n-i-1} = q_{i+1} q_{n-i}, \quad \forall 0 \le i \le n/2.$$

By the symmetry of π, we will fix $j_n = \lfloor n/2 \rfloor$ when applying Theorem 4.1.

Consider a family of irreducible birth and death chains, $\mathcal{F} = \{(\Omega_n, K_n, \pi_n) | n = 1,2,...\}$ with $\Omega_n = \{0,1,...,n\}$. Let $p_{n,i}, q_{n,i}, r_{n,i}$ be respectively the birth, death and holding rates of K_n and t_n, ℓ_n be constants in Theorem 4.1. Assume that π_n is symmetric at $n/2$. Continuously using the fact $(a+b)/2 \le \max\{a,b\} \le a+b$ for $a \ge 0, b \ge 0$, we obtain

$$t_n \asymp \sum_{k:k\le n/2} \frac{\pi_n([0,k])}{\pi_n(k)\min\{p_{n,k}, q_{n,n-k}\}}$$

and

$$\ell_n \asymp \max_{j:j\le n/2} \sum_{k:j\le k\le n/2} \frac{\pi_n([0,j])}{\pi_n(k)\min\{p_{n,k}, q_{n,n-k}\}}.$$

Theorem 4.1 can be rewritten as follows.

Theorem 4.6. *Let $\mathcal{F} = \{(\Omega_n, K_n, \pi_n) | n = 1,2,...\}$ be a family of irreducible birth and death chains with $\Omega_n = \{0,1,...,n\}$. Let λ_n and $p_{n,i}, q_{n,i}, r_{n,i}$ be the spectral gap and the birth, death and holding rates of K_n. Assume that*

$$p_{n,i} p_{n,n-i-1} = q_{n,i+1} q_{n,n-i}, \quad \forall 0 \le i \le n/2.$$

Then, for $\epsilon \in (0,1/2)$ *and* $\delta \in (0,1)$,

$$\lambda_n \asymp 1/\ell_n, \quad T^{(c)}_{n,\mathrm{TV}}(\epsilon) \asymp T^{(\delta)}_{n,\mathrm{TV}}(\epsilon) \asymp t_n,$$

where

$$t_n = \sum_{k:k\le n/2} \frac{\pi_n([0,k])}{\pi_n(k)\min\{p_{n,k}, q_{n,n-k}\}}$$

and

$$\ell_n = \max_{j:j\le n/2} \left\{ \pi_n([0,j]) \sum_{k:j\le k\le n/2} \frac{1}{\pi_n(k)\min\{p_{n,k}, q_{n,n-k}\}} \right\}.$$

Moreover, the following are equivalent.

(1) $\mathcal{F}_c$ *has a cutoff in total variation.*
(2) *For* $\delta \in (0,1)$, $\mathcal{F}_\delta$ *has a cutoff in total variation.*
(3) $\mathcal{F}_c$ *has a precutoff in total variation.*
(4) *For* $\delta \in (0,1)$, $\mathcal{F}_\delta$ *has a precutoff in total variation.*
(5) $t_n/\ell_n \to \infty$.

The next theorem considers a perturbation of birth and death chains which has the same stationary distribution as the original chains. The new chains keep the order of mixing time and spectral gap unchanged.

Theorem 4.7. *Consider the family in Theorem 4.6 and assume that*

$$p_{n,i}p_{n,n-i-1} = q_{n,i+1}q_{n,n-i}, \quad \forall 0 \le i \le n/2.$$

For $n \ge 1$, *let* $A_n \subset \{0,...,n-1\}$, $c_{n,i} \in [0,1]$ *for* $i \in A_n$ *and* $\widetilde{K}_n$ *be a birth and death chain on* Ω_n *with birth and death rates,* $\widetilde{p}_{n,i}, \widetilde{q}_{n,i}$, *satisfying*

$$\begin{cases} \widetilde{p}_{n,i} = c_{n,i}p_{n,i} + (1-c_{n,i})\min\{p_{n,i}, q_{n,n-i}\} & \text{for } i \in A_n, \\ \widetilde{q}_{n,i+1} = q_{n,i+1}\widetilde{p}_{n,i}/p_{n,i} & \text{for } i \in A_n, \\ \widetilde{p}_{n,i} = p_{n,i}, \quad \widetilde{q}_{n,i+1} = q_{n,i+1} & \text{for } i \notin A_n. \end{cases}$$

Let $\lambda_n, \widetilde{\lambda}_n$ *and* $T_{n,\mathrm{TV}}(\epsilon), \widetilde{T}_{n,\mathrm{TV}}(\epsilon)$ *be the spectral gaps and total variation mixing times of* $K_n, \widetilde{K}_n$. *Then, given* $\epsilon \in (0,1/2)$ *and* $\delta \in (0,1)$,

$$\widetilde{\lambda}_n \asymp \lambda_n, \quad \widetilde{T}^{(c)}_{n,\mathrm{TV}}(\epsilon) \asymp T^{(c)}_{n,\mathrm{TV}}(\epsilon) \asymp \widetilde{T}^{(\delta)}_{n,\mathrm{TV}}(\epsilon) \asymp T^{(\delta)}_{n,\mathrm{TV}}(\epsilon),$$

where the approximation is uniform on the choice of $A_n, c_{n,i}$.

Proof. The approximation of the spectral gap and the total variation mixing time is immediate from Theorem 4.6, whereas the uniformity of the approximation is given by Theorems 3.1, 3.5 and 3.8. □

Example 4.4. For $n \ge 1$, let K_n be a birth and death chain on $\{0,1,...,2n\}$ given by

$$K_n(i,i+1) = K_n(i+1,i) = \begin{cases} 1/2 & \text{for even } i \\ 1/(2n) & \text{for odd } i \end{cases}.$$

By Theorem 4.7, the mixing time and spectral gap of K_n are comparable with those of $\widetilde{K}_n$, where $\widetilde{K}_n(i,i+1) = \widetilde{K}_n(i+1,i) = 1/(2n)$ for $0 \le i < 2n$. Let $\mathcal{F}$ be the family consisting of K_n. By Theorem 4.6, neither $\mathcal{F}_c$ nor $\mathcal{F}_\delta$ has a total variation precutoff and $T^{(c)}_{n,\mathrm{TV}}(\epsilon) \asymp T^{(\delta)}_{n,\mathrm{TV}}(\epsilon) \asymp \lambda_n^{-1} \asymp n^3$ for all $\epsilon \in (0,1/2)$ and $\delta \in (0,1)$, which is nontrivial.

Next, we consider simple random walks on finite paths with bottlenecks. For $n \ge 1$, let $k_n \le n$ and $x_{n,1},...,x_{n,k_n}$ be positive integers satisfying $1 \le x_{n,i} < x_{n,i+1} \le n$ for $i = 1,...,k_n - 1$. Let K_n be the birth and death chain on $\{0,1,...,n\}$ of which birth, death and holding rates are given by

$$p_{n,i-1} = q_{n,i} = \begin{cases} 1/2 & \text{for } i \notin \{x_{n,1},...,x_{n,k_n}\} \\ \epsilon_{n,j} & \text{for } i = x_{n,j},\ 1 \le j \le k_n \end{cases}, \tag{4.6}$$

where $\epsilon_{n,j} \in (0,1/2]$ for $1 \le j \le k_n$. Clearly, K_n is irreducible and the stationary distribution, say π_n, is uniform on $\{0,1,...,n\}$. The following theorem is immediate from Theorems 4.6.

Theorem 4.8. *Let $\mathcal{F}$ be a family of birth and death chains given by* (4.6) *and λ_n be the spectral gap of K_n. For $n \ge 1$, set*

$$t_n = n^2 + \sum_{i=1}^{k_n} \frac{\min\{x_{n,i}, n+1-x_{n,i}\}}{\epsilon_{n,i}}$$

and

$$\ell_n = n^2 + \max_{j:j\le n/2} \left\{ \sum_{i:|x_{n,i}-n/2|\le j} \frac{n/2+1-j}{\epsilon_{n,i}} \right\}.$$

Then, for all $\epsilon \in (0,1/2)$ and $\delta \in (0,1)$,

$$T_{n,\mathrm{TV}}^{(c)}(\epsilon) \asymp T_{n,\mathrm{TV}}^{(\delta)}(\epsilon) \asymp t_n, \quad \lambda_n \asymp 1/\ell_n.$$

Furthermore, the following are equivalent.

(1) *$\mathcal{F}_c$ has a cutoff in total variation.*
(2) *For $\delta \in (0,1)$, $\mathcal{F}_\delta$ has a cutoff in total variation.*
(3) *$\mathcal{F}_c$ has precutoff in total variation.*
(4) *For $\delta \in (0,1)$, $\mathcal{F}_\delta$ has a precutoff in total variation.*
(5) *$t_n/\ell_n \to \infty$.*

Remark 4.1. Let t_n, ℓ_n be the constants in Theorem 4.8. Then,

$$t_n \asymp n^2 + \sum_{j\in L_n} \frac{x_{n,j}}{\epsilon_{n,j}} + \sum_{j\in R_n} \frac{n+1-x_{n,j}}{\epsilon_{n,j}}$$

and

$$\ell_n \asymp n^2 + \max_{i\in L_n} \sum_{j\in L_n:j\ge i} \frac{x_{n,i}}{\epsilon_{n,j}} + \max_{i\in R_n} \sum_{j\in R_n:j\le i} \frac{n+1-x_{n,i}}{\epsilon_{n,j}}.$$

where $L_n = \{i : x_{n,i} \le n/2\}$ and $R_n = \{i : x_{n,i} > n/2\}$.

Theorem 1.4 considers a special case of Theorem 4.8 with $\epsilon_{n,i} = \epsilon_n$ for $1 \le i \le k_n$. It is clear from Theorem 1.4 that if k_n is bounded, then no cutoff exists for $\mathcal{F}_c$ or $\mathcal{F}_\delta$. The following example shows a case of cutoffs for the family in Theorem 1.4.

Example 4.5. Let $\mathcal{F}$ be the family in Theorem 1.4, with $k_n = \lfloor n^{1/3} \rfloor - 1$ and

$$x_{n,i} = \left\lfloor \frac{n^{5/6}}{n^{1/3} - i} \right\rfloor, \quad \forall 1 \le i \le k_n.$$

Clearly, for n large enough, $x_{n,i} \ne x_{n,j}$ when $i \ne j$. Let a_n, b_n be the constant in Theorem 1.4. It is not hard to show that

$$a_n \asymp n^{5/6} \log n, \quad b_n \asymp n^{5/6}.$$

By Theorem 1.4, $\mathcal{F}_c$ and $\mathcal{F}_\delta$, with $\delta \in (0,1)$, have a total variation cutoff if and only if $\epsilon_n = o(n^{-7/6}\log n)$. Furtheromre, if $\epsilon_n = o(n^{-7/6}\log n)$, then

$$T_{n,\mathrm{TV}}^{(c)}(\epsilon) \asymp \frac{n^{5/6}\log n}{\epsilon_n} \asymp T_{n,\mathrm{TV}}^{(\delta)}(\epsilon), \quad \forall \epsilon, \delta \in (0,1).$$

The following two theorems treat special cases of Theorem 4.8.

Theorem 4.9. *Let $\mathcal{F}$ be a family of birth and death chains satisfying* (4.6)*. Let N be a positive constant. Suppose, for $n \geq 1$, there are constants $J_1^{(n)}, ..., J_N^{(n)}$ and a partition of $\{1, ..., k_n\}$, say $I_1^{(n)}, ..., I_N^{(n)}$, such that, for $1 \leq k \leq N$,*

$$\max_{i \in I_k^{(n)}} \{x_{n,i} \wedge (n+1-x_{n,i})\} \asymp \min_{i \in I_k^{(n)}} \{x_{n,i} \wedge (n+1-x_{n,i})\} \asymp J_k^{(n)},$$

where $a \wedge b = \min\{a,b\}$. Then, neither $\mathcal{F}_c$ nor $\mathcal{F}_\delta$ has a total variation cutoff. Moreover,

$$T_{n,\mathrm{TV}}^{(c)}(\epsilon) \asymp T_{n,\mathrm{TV}}^{(\delta)}(\epsilon) \asymp \lambda_n^{-1} \asymp t_n, \quad \forall \epsilon \in (0,1/2),\ \delta \in (0,1)$$

where

$$t_n = n^2 + \max_{1 \leq k \leq N} \left\{ J_k^{(n)} \sum_{l \in I_k^{(n)}} \frac{1}{\epsilon_{n,l}} \right\}.$$

The next theorem gives an example that no total variation cutoff exists for $\mathcal{F}_c, \mathcal{F}_\delta$ even when the constant N in Theorem 4.9 tends to infinity.

Theorem 4.10. *Let $\mathcal{F}$ be a family of birth and death chains satisfying* (4.6)*. Suppose that $\min_j \epsilon_{n,j} \asymp \max_j \epsilon_{n,j}$ and $x_{n,i} = \lfloor in/k_n \rfloor$ with $k_n \leq n/2$, then neither $\mathcal{F}_c$ nor $\mathcal{F}_\delta$ has a total variation cutoff, but*

$$T_{n,\mathrm{TV}}^{(c)}(\epsilon) \asymp T_{n,\mathrm{TV}}^{(\delta)}(\epsilon) \asymp \lambda_n^{-1} \asymp \max\{n^2, nk_n/\epsilon_{n,1}\}, \quad \forall \epsilon \in (0,1/2),\ \delta \in (0,1).$$

Remark 4.2. Note that the assumption regarding the birth and death rates in this section can be relaxed using the comparison technique in [11, 12].

Appendix A. Spectral gaps of finite paths

This section is devoted to finding the correct order of spectral gaps of finite paths. Let $G = (V, E)$ be the undirected finite graph with vertex set $V = \{0, 1, 2, ...n\}$ and edge set $E = \{\{i, i+1\} : i = 0, 1, ..., n-1\}$. Given two positive measures π, ν on V, E with $\pi(V) = 1$, the Dirichlet form and variance associated with ν and π are defined by

$$\mathcal{E}_\nu(f,g) := \sum_{i=1}^{n-1} [f(i) - f(i+1)][g(i) - g(i+1)]\nu(i, i+1)$$

and

$$\mathrm{Var}_\pi(f) := \pi(f^2) - \pi(f)^2,$$

where f, g are functions on V. The spectral gap of G with respect to π, ν is defined as

$$\lambda_{\pi,\nu}^G := \min \left\{ \frac{\mathcal{E}_\nu(f,f)}{\mathrm{Var}_\pi(f)} \middle| f \text{ is non-constant} \right\}.$$

To bound the spectral gap, we need the following setting. Let $C_+(i)$ and $C_-(i)$ be constants defined by

$$C_+(i) = \max_{j:j>i} \sum_{k=i+1}^{j} \frac{\pi([j,n])}{\nu(k-1,k)}, \quad C_-(i) = \max_{j:j<i} \sum_{k=j}^{i-1} \frac{\pi([0,j])}{\nu(k,k+1)}, \tag{A.1}$$

where $\max\emptyset := 0$.

Theorem A.1. *Let $G = (V,E)$ be a path on $\{0,1,...,n\}$ and π,ν be positive measures on V,E with $\pi(V) = 1$. Referring to* (A.1)*, set $C(m) = \max\{C_+(m), C_-(m)\}$. Then, for $0 \le m \le n$,*

$$\frac{1}{4C(m)} \le \lambda_{\pi,\nu}^G \le \frac{1}{\min\{\pi([0,m]),\pi([m,n])\}C(m)}.$$

In particular, if M is a median of π, that is, $\pi([0,M]) \ge 1/2$ and $\pi([M,n]) \ge 1/2$, then

$$\frac{1}{4C(M)} \le \lambda_{\pi,\nu}^G \le \frac{2}{C(M)}.$$

Remark A.1. Referring to the setting in Theorem A.1, the authors in [7] obtained $\lambda_{\pi,\nu}^G \ge 1/C'$, where

$$C' = \min_{0\le j\le n} \max\left\{\sum_{k=0}^{j-1} \frac{\pi([0,k])}{\nu(k,k+1)}, \sum_{k=j+1}^{n} \frac{\pi([k,n])}{\nu(k-1,k)}\right\}. \tag{A.2}$$

Let $C_+(m), C_-(m)$ be constants in (A.1) and $C(m)$ be the constant in Theorem A.1. Then, for $0 \le j \le n$,

$$\sum_{k=0}^{j-1} \frac{\pi([0,k])}{\nu(k,k+1)} \ge C_-(j) \quad , \sum_{k=j+1}^{n} \frac{\pi([k,n])}{\nu(k-1,k)} \ge C_+(j).$$

This yields $C' \ge \min_m C(m)$. In particular, if M is a median of π, then

$$C' \ge \frac{1}{2}\left(\sum_{k=0}^{M-1} \frac{\pi([0,k])}{\nu(k,k+1)} + \sum_{k=M+1}^{n} \frac{\pi([k,n])}{\nu(k-1,k)}\right) \ge \frac{C(M)}{2}.$$

The lower bound of Theorem A.1 is at least of the same order than $/C'$ and sometimes significantly better.

The proof of Theorem A.1 is based on the following proposition, which is related to weighted Hardy's inequality on $\{1,...,n\}$.

Proposition A.2. *Fix $n \ge 1$. Let μ,π be positive measures on $\{1,...,n\}$ and A be the smallest constant such that*

$$\sum_{i=1}^{n} \left(\sum_{j=1}^{i} g(j)\right)^2 \pi(i) \le A \sum_{i=1}^{n} g^2(i)\mu(i), \quad \forall g \ne \mathbf{0}. \tag{A.3}$$

Then, $B \le A \le 4B$, where

$$B = \max_{1\le i\le n}\left\{\pi([i,n])\sum_{j=1}^{i}\frac{1}{\mu(j)}\right\}.$$

Remark A.2. Miclo [16] discussed the infinity case $\{1, 2, ...\}$ using the method in [17], which was introduced by Muckenhoupt to study the continuous case $[0, \infty)$. For more information on the weighted Hardy inequality, see [16] and the references therein.

Proof of Theorem A.1. We first consider the lower bound of $\lambda^G_{\pi,\nu}$. Let f be any function defined on V and set $f_+ = [f - f(m)]\mathbf{1}_{\{m,...,n\}}$ and $f_- = [f - f(m)]\mathbf{1}_{\{0,...,m\}}$. Then,

$$\frac{\mathcal{E}_\nu(f,f)}{\mathrm{Var}_\pi(f)} \geq \frac{\mathcal{E}_\nu(f,f)}{\pi(f - f(m))^2} = \frac{\mathcal{E}_\nu(f_+,f_+) + \mathcal{E}_\nu(f_-,f_-)}{\pi(f_+^2) + \pi(f_-^2)} \tag{A.4}$$

Set $g(j) = f(m+j) - f(m+j-1)$ for $1 \leq j \leq n-m$ and $h(i) = f(m-i) - f(m-i+1)$ for $1 \leq i \leq m$. Note that

$$\mathcal{E}_\nu(f_+,f_+) = \sum_{j=1}^{n-m} g^2(j)\nu(m+j-1, m+j),\ \pi(f_+^2) = \sum_{j=1}^{n-m}\left(\sum_{k=1}^{j} g(k)\right)^2 \pi(m+j),$$

and

$$\mathcal{E}_\nu(f_-,f_-) = \sum_{i=1}^{m} h^2(i)\nu(m-i, m-i+1),\ \pi(f_-^2) = \sum_{j=1}^{m}\left(\sum_{k=1}^{j} h(k)\right)^2 \pi(m-j).$$

By Proposition A.2, the above computation implies that

$$\frac{\mathcal{E}_\nu(f_+,f_+)}{\pi(f_+^2)} \geq \frac{1}{4C_+(m)},\quad \frac{\mathcal{E}_\nu(f_-,f_-)}{\pi(f_-^2)} \geq \frac{1}{4C_-(m)}.$$

Putting this back to (A.4) gives the desired lower bound.

For the upper bound, we first consider the case $C = C_+(m)$. By Proposition A.2, $C_+(m) \leq A$, where A is the smallest constant A such that, for any function ϕ defined on $\{1, 2, ..., n-m+1\}$,

$$\sum_{j=1}^{n-m}\left(\sum_{k=1}^{j}\phi(k)\right)^2 \pi(m+j) \leq A \sum_{j=1}^{n-m}\phi^2(j)\nu(m+j-1, m+j).$$

Let ϕ be a minimizer for A, which must exist, and define ψ by setting

$$\psi(i) = \begin{cases} \phi(1) + \cdots + \phi(i-m) & \text{for } m < i \leq n \\ 0 & \text{for } 0 \leq i \leq m \end{cases}.$$

Clearly, $1/C_+(m) \geq 1/A = \mathcal{E}_\nu(\psi,\psi)/\pi(\psi^2)$. Without loss of generality, we may assume further that ϕ is nonnegative. Note that $\pi(\{\psi = 0\}) \geq \pi([0, m])$. By the Cauchy-Schwartz inequality, this implies $\pi(\psi)^2 \leq \pi(\{\psi > 0\})\pi(\psi^2) \leq \pi([m+1, n])\pi(\psi^2)$ and, then, $\mathrm{Var}_\pi(\psi) \geq \pi([0, m])\pi(\psi^2)$. This leads to $1/C = 1/C_+(m) \geq \pi([0, m])\lambda^G_{\pi,\nu}$. Similarly, if $C = C_-(m)$, one can prove that $1/C \geq \pi([m, n])\lambda^G_{\pi,\nu}$. This yields the upper bound of the spectral gap. □

Proof of Proposition A.2. The proofs of Theorem A.1 and Proposition A.2 are very similar to those in [16]. Note that A is attained at functions of the same sign and we assume that g is non-negative. As A is attainable, the minimizer g for A satisfies the following Euler-Lagrange equations.

$$Ag(i)\mu(i) = \sum_{j=i}^{n}(g(1) + \cdots + g(j))\pi(j), \quad \forall 1 \leq i \leq n. \tag{A.5}$$

This is equivalent to the following system of equations.

$$A[g(i)\mu(i) - g(i+1)\mu(i+1)] = (g(1) + \cdots + g(i))\pi(i), \quad \forall 1 \le i \le n,$$

with the convention that $\mu(n+1) := 0$. Inductively, one can show that $g > 0$. Summing up (A.5) over $\{1, ..., \ell\}$ yields

$$\begin{aligned} A\sum_{i=1}^{\ell} g(i) &= \sum_{i=1}^{\ell} \frac{1}{\mu(i)} \sum_{j=i}^{n} (g(1) + \cdots + g(j))\pi(j) \\ &\ge \sum_{i=1}^{\ell} \sum_{j=\ell}^{n} \frac{(g(1) + \cdots + g(j))\pi(j)}{\mu(i)} \\ &\ge \left(\sum_{i=1}^{\ell} g(i)\right)\left(\sum_{i=1}^{\ell} \frac{1}{\mu(i)}\right)\pi([\ell, n]). \end{aligned}$$

This leads to $A \ge B$.

To see the upper bound, we use Miclo's method in [16]. Set $N(j) = \sum_{i=1}^{j} 1/\mu(i)$. By the Cauchy inequality, the left side of (A.3) is bounded above by

$$\sum_{i=1}^{n} \pi(i) \sum_{j=1}^{i} g^2(j)\mu(j)N^{1/2}(j) \sum_{l=1}^{i} \frac{1}{\mu(l)N^{1/2}(l)}.$$

Note that, for $s > 0, t > 0$, $t^{1/2} - s^{1/2} \ge (t-s)/(2t^{1/2})$. This implies $2(N^{1/2}(l) - N^{1/2}(l-1)) \ge 1/(\mu(l)N^{1/2}(l))$ with the convention that $N(0) := 0$. Consequently, we have

$$\sum_{l=1}^{i} \frac{1}{\mu(l)N^{1/2}(l)} \le 2N^{1/2}(i) \le \left(\frac{4B}{\pi([i,n])}\right)^{1/2},$$

and, thus,

$$\begin{aligned} \sum_{i=1}^{n} \left(\sum_{j=1}^{i} g(j)\right)^2 \pi(i) &\le \sqrt{4B} \sum_{i=1}^{n} \frac{\pi(i)}{\pi([i,n])^{1/2}} \sum_{j=1}^{i} g^2(j)\mu(j)N^{1/2}(j) \\ &\le \sqrt{4B} \sum_{j=1}^{n} g^2(j)\mu(j)N^{1/2}(j) \sum_{i=j}^{n} \frac{\pi(i)}{\pi([i,n])^{1/2}}. \end{aligned}$$

Again, the inequality for s, t implies

$$\sum_{i=j}^{n} \frac{\pi(i)}{\pi([i,n])^{1/2}} \le 2\pi([j,n])^{1/2} \le \frac{\sqrt{4B}}{N^{1/2}(j)}.$$

This gives the desired upper bound. □

Next, we consider a special case. Let π, ν are measures on $V = \{0, 1, ..., n\}, E = \{\{i, i+1\} | 0 \le i < n\}$ with $\pi(V) = 1$. Suppose

(A.6) $$\pi(i) = \pi(n-i), \quad \nu(i, i+1) = \nu(n-i-1, n-i), \quad \forall 0 \le i \le n/2.$$

By the symmetry of π and ν, if ψ is a minimizer for $\lambda_{\pi,\nu}^G$ with $\pi(\psi) = 0$, then ψ is either symmetric or anti-symmetric at $n/2$. The former is set aside because ψ is known to be monotonic and this leads to the case $\psi(n-i) = -\psi(i)$ for $0 \le i \le n/2$. If n is even with $n = 2k$, then $\psi(k) = 0$ and this implies

$$\lambda^G_{\pi,\nu} = \inf\left\{ \frac{\sum_{i=1}^k (f(i)-f(i-1))^2\nu(i-1,i)}{\sum_{i=0}^{k-1} f^2(i)\pi(i)} \middle| f(k) = 0, f \ne \mathbf{0} \right\}.$$

Equivalently, if one sets $g(i) = f(k-i) - f(k-i+1)$ and $\mu(i) = \nu(k-i, k-i+1)$ for $1 \le i \le k$, then $1/\lambda^G_{\pi,\nu}$ is the smallest constant A such that

$$\text{(A.7)} \qquad \sum_{i=1}^k \left(\sum_{j=1}^i g(j) \right)^2 \pi(k-i) \le A \sum_{i=1}^k g^2(i)\mu(i), \quad \forall g \ne \mathbf{0}.$$

Similarly, if n is odd with $n = 2k-1$, one has

$$\lambda^G_{\pi,\nu} = \min\left\{ \frac{\sum_{i=1}^{k-1} (f(i)-f(i-1))^2\nu(i-1,i) + 2f^2(k-1)\nu(k-1,k)}{\sum_{i=0}^{k-1} f^2(i)\pi(i)} \middle| f \ne \mathbf{0} \right\},$$

and this leads to (A.7) with $g(1) = f(k-1)$, $\mu(1) = 2\nu(k-1,k)$ and, for $2 \le i \le k$, $g(i) = f(k-i) - f(k-i+1)$ and $\mu(i) = \nu(k-i, k-i+1)$. A direct application of Proposition A.2 implies the following theorem.

Theorem A.3. *Let $G = (V, E)$ be the graph with $V = \{0, 1, ..., n\}$, $E = \{\{i, i+1\} | i = 0 \le i < n\}$ and let π, ν be positive measures on V, E satisfying $\pi(V) = 1$ and* (A.6). *Set $N = \lceil n/2 \rceil$. Then, $1/(4C) \le \lambda^G_{\pi,\nu} \le 1/C$, where*

$$C = \max_{0 \le i < N} \left\{ \pi([0,i]) \sum_{j=i}^{N-1} \frac{1}{\nu(j,j+1)} \right\} \quad \textit{if } n \textit{ is even},$$

and

$$C = \max_{0 \le i < N} \left\{ \pi([0,i]) \left(\sum_{j=i}^{N-2} \frac{1}{\nu(j,j+1)} + \frac{1}{2\nu(N-1,N)} \right) \right\} \quad \textit{if } n \textit{ is odd.}$$

Remark A.3. The symmetry of π, ν in Theorems A.3 can be relaxed using the comparison technique.

References

[1] David Aldous. Random walks on finite groups and rapidly mixing Markov chains. In *Seminar on probability, XVII*, volume 986 of *Lecture Notes in Math.*, pages 243–297. Springer, Berlin, 1983.

[2] J. Barrera, O. Bertoncini, and R. Fernández. Abrupt convergence and escape behavior for birth and death chains. *J. Stat. Phys.*, 137(4):595–623, 2009.

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

[4] Guan-Yu Chen. *The cutoff phenomenon for finite Markov chains.* PhD thesis, Cornell University, 2006.

[5] Guan-Yu Chen and Laurent Saloff-Coste. The cutoff phenomenon for ergodic markov processes. *Electron. J. Probab.*, 13:26–78, 2008.

[6] Guan-Yu Chen and Laurent Saloff-Coste. Comparison of cutoffs between lazy walks and markovian semigroups. In preparation, 2012.

[7] Guan-Yu Chen and Laurent Saloff-Coste. Spectral computations for birth and death chains. In preparation, 2012.

[8] P. Diaconis and L. Saloff-Coste. What do we know about the Metropolis algorithm? *J. Comput. System Sci.*, 57(1):20–36, 1998. 27th Annual ACM Symposium on the Theory of Computing (STOC'95) (Las Vegas, NV).

[9] Persi Diaconis. *Group representations in probability and statistics*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 11. Institute of Mathematical Statistics, Hayward, CA, 1988.

[10] Persi Diaconis. The cutoff phenomenon in finite Markov chains. *Proc. Nat. Acad. Sci. U.S.A.*, 93(4):1659–1664, 1996.

[11] Persi Diaconis and Laurent Saloff-Coste. Comparison techniques for random walk on finite groups. *Ann. Probab.*, 21(4):2131–2156, 1993.

[12] Persi Diaconis and Laurent Saloff-Coste. Comparison theorems for reversible Markov chains. *Ann. Appl. Probab.*, 3(3):696–730, 1993.

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

[14] Persi Diaconis and Daniel Stroock. Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.*, 1(1):36–61, 1991.

[15] Jian Ding, Eyal Lubetzky, and Yuval Peres. Total variation cutoff in birth-and-death chains. *Probab. Theory Related Fields*, 146(1-2):61–85, 2010.

[16] L. Miclo. An example of application of discrete Hardy's inequalities. *Markov Process. Related Fields*, 5(3):319–330, 1999.

[17] Benjamin Muckenhoupt. Hardy's inequality with weights. *Studia Math.*, 44:31–38, 1972. Collection of articles honoring the completion by Antoni Zygmund of 50 years of scientific activity, I.

[18] L. Saloff-Coste. Simple examples of the use of Nash inequalities for finite Markov chains. In *Stochastic geometry (Toulouse, 1996)*, volume 80 of *Monogr. Statist. Appl. Probab.*, pages 365–400. Chapman & Hall/CRC, Boca Raton, FL, 1999.

[1]Department of Applied Mathematics, National Chiao Tung University, Hsinchu 300, Taiwan

E-mail address: gychen@math.nctu.edu.tw

[2]Malott Hall, Department of Mathematics, Cornell University, Ithaca, NY 14853-4201

E-mail address: lsc@math.cornell.edu

COMPARISON OF CUTOFFS BETWEEN LAZY WALKS AND MARKOVIAN SEMIGROUPS

GUAN-YU CHEN[1] AND LAURENT SALOFF-COSTE[2]

Abstract. We make a connection between the continuous time and lazy discrete time Markov chains through the comparison of cutoffs and mixing time in total variation distance. For illustration, we consider finite birth and death chains and provide a criterion on cutoffs using eigenvalues of the transition matrix.

1. Introduction

Let Ω be a countable set and (Ω, K, π) be an irreducible Markov chain on Ω with transition matrix K and stationary distribution π. Let

$$H_t = e^{-t(I-K)} = \sum_{i=0}^{\infty} e^{-t} t^i K^i / i!$$

be the associated semigroup which describes the corresponding natural continuous time process on Ω. For $\delta \in (0, 1)$, set

$$K_\delta = \delta I + (1-\delta)K, \tag{1.1}$$

where I is the identity matrix indexed by Ω. Clearly, K_δ is similar to K but with an additional holding probability depending of δ. We call K_δ the δ-lazy walk or δ-lazy chain of K. It is well-known that if K is irreducible with stationary distribution π, then

$$\lim_{m\to\infty} K_\delta^m(x, y) = \lim_{t\to\infty} H_t(x, y) = \pi(y), \quad \forall x, y \in \Omega,\ \delta \in (0, 1).$$

In this paper, we consider convergence in total variation. The total variation between two probabilities μ, ν on Ω is defined by $\|\mu-\nu\|_{\mathrm{TV}} = \sup\{\mu(A)-\nu(A)|A \subset \Omega\}$. For any irreducible K with stationary distribution π, the (maximum) total variation distance is defined by

$$d_{\mathrm{TV}}(m) = \sup_{x\in\Omega} \|K^m(x, \cdot) - \pi\|_{\mathrm{TV}}, \tag{1.2}$$

and the corresponding mixing time is given by

$$T_{\mathrm{TV}}(\epsilon) = \inf\{m \geq 0 | d_{\mathrm{TV}}(m) \leq \epsilon\}. \tag{1.3}$$

We write the total variation distance and mixing time as $d^{(c)}_{\mathrm{TV}}, T^{(c)}_{\mathrm{TV}}$ for the continuous semigroup and as $d^{(\delta)}_{\mathrm{TV}}, T^{(\delta)}_{\mathrm{TV}}$ for the δ-lazy walk.

A sharp transition phenomenon, known as cutoff, was introduced by Aldous and Diaconis in early 1980s. See e.g. [8, 5] for an introduction and a general review of

2000 *Mathematics Subject Classification.* 60J10,60J27.
Key words and phrases. Markov chains, cutoff phenomenon.
[1]Partially supported by NSC grant NSC100-2918-I-009-010.
[2]Partially supported by NSF grant DMS-1004771.

cutoffs. In total variation, a family of irreducible Markov chains $(\Omega_n, K_n, \pi_n)_{n=1}^\infty$ is said to present a cutoff if

$$\lim_{n\to\infty} \frac{T_{n,\mathrm{TV}}(\epsilon)}{T_{n,\mathrm{TV}}(\eta)} = 1, \quad \forall 0 < \epsilon < \eta < 1. \tag{1.4}$$

The family is said to present a (t_n, b_n) cutoff if $b_n = o(t_n)$ and

$$|T_{n,\mathrm{TV}}(\epsilon) - t_n| = O(b_n), \quad \forall 0 < \epsilon < 1.$$

The cutoff for the associated continuous semigroups is defined in a similar way. This paper contains the following general result.

Theorem 1.1. *Consider a family of irreducible and positive recurrent Markov chains $\mathcal{F} = \{(\Omega_n, K_n, \pi_n)|n = 1, 2, ...\}$. For $\delta \in (0, 1)$, let $\mathcal{F}_\delta$ be the family of associated δ-lazy walks and let $\mathcal{F}_c$ be the family of associated continuous semigroups. Suppose $T^{(c)}_{n,\mathrm{TV}}(\epsilon_0) \to \infty$ for some $\epsilon_0 \in (0, 1)$. Then, the following are equivalent.*

1. *$\mathcal{F}_\delta$ has a cutoff in total variation.*
2. *$\mathcal{F}_c$ has a cutoff in total variation.*

Furthermore, if $\mathcal{F}_c$ has a cutoff, then

$$\lim_{n\to\infty} \frac{T^{(c)}_{n,\mathrm{TV}}(\epsilon)}{T^{(\delta)}_{n,\mathrm{TV}}(\epsilon)} = 1 - \delta, \quad \forall \epsilon \in (0, 1).$$

Theorem 1.2. *Let $\mathcal{F}$ be the family in Theorem 1.1. Assume that $t_n \to \infty$. Then, the following are equivalent.*

1. *$\mathcal{F}_c$ has a (t_n, b_n) cutoff.*
2. *For $\delta \in (0, 1)$, $\mathcal{F}_\delta$ has a $(t_n/(1 - \delta), b_n)$ cutoff.*

We refer the readers to Theorems 3.1, 3.4, 3.5 and 3.7 for more detailed discussions.

For an illustration, we consider finite birth and death chains. For $n \geq 1$, let $\Omega_n = \{0, 1, ..., n\}$ and K_n be the transition kernel of a birth and death chain on Ω_n with birth rate $p_{n,i}$, death rate $q_{n,i}$ and holding rate $r_{n,i}$, where $p_{n,n} = q_{n,0} = 0$ and $p_{n,i} + q_{n,i} + r_{n,i} = 1$. Suppose that K_n is irreducible with stationary distribution π_n. For the family $\{(\Omega_n, K_n, \pi_n)|n = 1, 2, ...\}$, Ding *et al.* [10] showed that, in the discrete time case, if $\inf_{i,n} r_{n,i} > 0$, then the cutoff in total variation exists if and only if the product of the total variation mixing time and the spectral gap, which is defined to be the smallest non-zero eigenvalue of $I - K$, tends to infinity. There is also a similar version for the continuous time case. The next theorem is an application of the above result and Theorem 1.1, which is summarized from Theorem 4.10.

Theorem 1.3. *Let $\mathcal{F} = \{(\Omega_n, K_n, \pi_n)|n = 1, 2, ...\}$ be a family of irreducible birth and death chains as above. For $n \geq 1$, let $0, \lambda_{n,1}, ..., \lambda_{n,n}$ be eigenvalues of $I - K_n$ and set*

$$\lambda_n = \min_{1\leq i\leq n} \lambda_{n,i}, \quad s_n = \sum_{i=1}^n \lambda_{n,i}^{-1}.$$

Then, the following are equivalent.

1. *$\mathcal{F}_c$ has a total variation cutoff.*
2. *For $\delta \in (0, 1)$, $\mathcal{F}_\delta$ has a total variation cutoff.*
3. *$s_n\lambda_n \to \infty$.*

The remaining of this article is organized as follows. In Section 2, the concepts of cutoffs and mixing times are introduced and fundamental results are reviewed. In Section 3, a detailed comparison of the cutoff time and window size is made between the continuous time and lazy discrete time cases, where the state space is allowed to be infinite. In Section 4, we focus on finite birth and death chains and provide a criterion on total variation cutoffs using the eigenvalues of the transition matrices.

2. Cutoffs in total variation

Throughout this paper, for any two sequences s_n, t_n of positive numbers, we write $s_n = O(t_n)$ if there are $C > 0, N > 0$ such that $|s_n| \le C|t_n|$ for $n \ge N$. If $s_n = O(t_n)$ and $t_n = O(s_n)$, we write $s_n \asymp t_n$. If $t_n/s_n \to 1$ as $n \to \infty$, we write $t_n \sim s_n$.

Consider the following definitions.

Definition 2.1. Referring to the notation in (1.2), a family $\mathcal{F} = \{(\Omega_n, K_n, \pi_n)|n = 1, 2, ...\}$ is said to present a total variation

(1) precutoff if there is a sequence t_n and $B > A > 0$ such that

$$\lim_{n\to\infty} d_{n,\mathrm{TV}}(\lceil Bt_n\rceil) = 0, \quad \liminf_{n\to\infty} d_{n,\mathrm{TV}}(\lfloor At_n\rfloor) > 0.$$

(2) cutoff if there is a sequence t_n such that, for all $\epsilon > 0$,

$$\lim_{n\to\infty} d_{n,\mathrm{TV}}(\lceil(1+\epsilon)t_n\rceil) = 0, \quad \lim_{n\to\infty} d_{n,\mathrm{TV}}(\lfloor(1-\epsilon)t_n\rfloor) = 1.$$

(3) (t_n, b_n) cutoff if $b_n = o(t_n)$ and

$$\lim_{c\to\infty} \overline{F}(c) = 0, \quad \lim_{c\to-\infty} \underline{F}(c) = 1,$$

where

$$\overline{F}(c) = \limsup_{n\to\infty} d_{n,\mathrm{TV}}(\lceil t_n + cb_n\rceil), \quad \underline{F}(c) = \liminf_{n\to\infty} d_{n,\mathrm{TV}}(\lfloor t_n + cb_n\rfloor).$$

In definition 2.1, t_n is called a cutoff time and b_n is called a window for t_n. The cutoffs for continuous semigroups is the same except the deletion of $\lceil\cdot\rceil$ and $\lfloor\cdot\rfloor$.

Remark 2.1. In Definition 2.1, if $t_n \to \infty$ (or equivalently $T_{n,\mathrm{TV}}(\epsilon) \to \infty$ for some $\epsilon \in (0, 1)$), then the cutoff is consistent with (1.4). This is also true for cutoffs in continuous semigroups without the assumption $t_n \to \infty$.

The following lemma characterizes the total variation convergence using specific subsequences of indices and events, which is useful in proving and disproving cutoffs.

Lemma 2.1. *Consider a family of irreducible and positive recurrent Markov chains $\{(\Omega_n, K_n, \pi_n)|n = 1, 2, ...\}$. Let t_n be a sequence of nonnegative integers. Then, the following are equivalent.*

(1) $d_{n,\mathrm{TV}}(t_n) \to 0$.

(2) *For any increasing sequence of positive integers n_k, any $A_{n_k} \subset \Omega_{n_k}$ and any $x_{n_k} \in \Omega_{n_k}$, there is a subsequence m_k such that*

$$\lim_{k\to\infty} \left| K_{m_k}^{t_{m_k}}(x_{m_k}, A_{m_k}) - \pi_{m_k}(A_{m_k}) \right| = 0.$$

Proof of Lemma 2.1. (1)⇒(2) is obvious. For (2)⇒(1), choose $A_n \subset \Omega_n$ and $x_n \in \Omega_n$ such that $d_{n,\mathrm{TV}}(t_n) \le 2|K_n^{t_n}(x_n, A_n) - \pi_n(A_n)|$. Let n_k be an increasing sequence of positive integers and choose a subsequence m_k such that

$$\lim_{k\to\infty} \left| K_{m_k}^{t_{m_k}}(x_{m_k}, A_{m_k}) - \pi_{m_k}(A_{m_k}) \right| = 0.$$

This implies $d_{m_k,\mathrm{TV}}(t_{m_k}) \to 0$, as desired. □

Remark 2.2. Lemma 2.1 also holds in continuous time under the release of t_n to positive real numbers. See [4, 5] for further discussions on cutoffs.

3. Comparisons of cutoffs

In this section, we establish the relation of cutoffs between lazy walks and continuous semigroups. Let Ω be a countable set and K be a transition matrix indexed by Ω. In the notation of (1.1), the δ-lazy walk evolves in accordance with

$$(K_\delta)^t = \sum_{i=0}^{t} \binom{t}{i} \delta^{t-i}(1-\delta)^i K^i, \quad \forall \delta \in (0,1),\ t \ge 0,$$

whereas the continuous time chain follows

$$H_t = e^{-t(I-K)} = \sum_{i=0}^{\infty} \left(e^{-t}\frac{t^i}{i!} \right) K^i.$$

Observe that $I - K = (I - K_\delta)/(1-\delta)$. This implies

$$d_{\mathrm{TV}}^{(c)}(t) \le e^{-t/(1-\delta)} \sum_{i=0}^{m} \frac{[t/(1-\delta)]^i}{i!} + d_{\mathrm{TV}}^{(\delta)}(m). \tag{3.1}$$

Concerning the cutoff times and windows, we discuss each of them in detail.

3.1. Cutoff times.

Theorem 3.1. *Let $\mathcal{F} = \{(\Omega_n, K_n, \pi_n)|n = 1, 2, ...\}$ be a family of irreducible Markov chains on countable state spaces with stationary distributions. For $\delta \in (0,1)$, let $\mathcal{F}_\delta = \{(\Omega_n, K_{n,\delta}, \pi_n)|n = 1, 2, ...\}$ and $\mathcal{F}_c = \{(\Omega_n, H_{n,t}, \pi_n)|n = 1, 2, ...\}$. Suppose there is $\epsilon_0 > 0$ such that $T_{n,\mathrm{TV}}^{(\delta)}(\epsilon_0) \to \infty$ or $T_{n,\mathrm{TV}}^{(c)}(\epsilon_0) \to \infty$. Then, the following are equivalent.*

1. *$\mathcal{F}_\delta$ has a cutoff (resp. precutoff) in total variation.*
2. *$\mathcal{F}_c$ has a cutoff (resp. precutoff) in total variation.*

Furthermore, if $\mathcal{F}_c$ has a cutoff, then

$$\lim_{n\to\infty} \frac{T_{n,\mathrm{TV}}^{(c)}(\epsilon)}{T_{n,\mathrm{TV}}^{(\delta)}(\epsilon)} = 1 - \delta, \quad \forall \epsilon \in (0,1).$$

The above theorem is in fact a simple corollary of the following proposition.

Proposition 3.2. *Let $\mathcal{F}_\delta, \mathcal{F}_c$ be families in Theorem 3.1 and t_n, r_n be sequences tending to infinity. Fix $\delta \in (0,1)$.*

1. *If $d_{n,\mathrm{TV}}^{(\delta)}(\lceil t_n \rceil) \to 0$, then*

$$\lim_{n\to\infty} d_{n,\mathrm{TV}}^{(c)}((1-\delta)t_n + cb_n) = 0,$$

for all $c > 0$ and for any sequence b_n satisfying $\sqrt{t_n} = o(b_n)$.

(2) *If* $d^{(c)}_{n,\mathrm{TV}}(r_n)\to 0$, *then*

$$\lim_{n\to\infty} d^{(\delta)}_{n,\mathrm{TV}}(\lceil r_n/(1-\delta)+cb_n\rceil)=0,$$

for all $c>0$ *and for any sequence* b_n *satisfying* $\sqrt{r_n}=o(b_n)$.

(3) *If* $d^{(c)}_{n,\mathrm{TV}}(r_n)\to 1$, *then*

$$\lim_{n\to\infty} d^{(\delta)}_{n,\mathrm{TV}}(\lfloor r_n/(1-\delta)\rfloor)=1.$$

(4) *If* $d^{(\delta)}_{n,\mathrm{TV}}(\lfloor t_n\rfloor)\to 1$, *then*

$$\lim_{n\to\infty} d^{(c)}_{n,\mathrm{TV}}((1-\delta)t_n)=1.$$

Proof. We prove (1), while (2) goes in a similar way and is omitted. Suppose $d^{(\delta)}_{n,\mathrm{TV}}(\lceil t_n\rceil)\to 0$. Since $\sqrt{t_n}=o(b_n)$, it is clear that

$$\lim_{n\to\infty} d^{(\delta)}_{n,\mathrm{TV}}(\lceil t_n+cb_n+c'\sqrt{t_n}\rceil)=0,\quad \forall c>0,\ c'\in\mathbb{R}. \tag{3.2}$$

Fix $c>0$ and let $x_n\in\Omega_n, A_n\subset\Omega_n$. Given any increasing sequence n_l, we may choose, according to Lemma 3.8, a subsequence m_l such that $\pi_{m_l}(A_{m_l})\to\alpha\in[0,1]$ and, for all $c'\in\mathbb{R}$,

$$\lim_{l\to\infty} K^{\lceil t_{m_l}+cb_{m_l}+c'\sqrt{t_{m_l}}\rceil}_{m_l,\delta}(x_{m_l},A_{m_l})=\frac{1}{\sqrt{2\pi\delta}}\int_{-\infty}^{\infty}e^{-(x-c')^2/(2\delta)}f(x)dx,$$

and

$$\lim_{l\to\infty} H_{m_l,(1-\delta)(t_{m_l}+cb_{m_l})}(x_{m_l},A_{m_l})=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-x^2/2}f(x)dx,$$

where f is nonnegative and bounded by 1. By (3.2) and Lemma 3.9, f equals to α almost everywhere and, by Lemma 2.1, this implies $d^{(c)}_{n,\mathrm{TV}}((1-\delta)t_n+cb_n)\to 0$ as $n\to\infty$ for all $c>0$.

The proofs for (3) and (4) are similar and we only give the details for (4). First, we choose sequences $x_n\in\Omega_n$ and $A_n\subset\Omega_n$ such that

$$\lim_{n\to\infty}\pi_n(A_n)=1,\quad \lim_{n\to\infty}K^{\lfloor t_n\rfloor}_{n,\delta}(x_n,A_n)=0.$$

Let n_l be a sequence tending to infinity. Applying Lemma 3.8 with $c=0$ and $a_{n,m}=K^m_n(x_n,A_n)$, we may choose a subsequence, say m_l, such that

$$\lim_{l\to\infty} H_{m_l,(1-\delta)t_{m_l}}(x_{m_l},A_{m_l})=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-x^2/2}g(x)dx$$

and

$$\lim_{l\to\infty} K^{\lfloor t_{m_l}\rfloor}_{m_l,\delta}(x_{m_l},A_{m_l})=\frac{1}{\sqrt{2\pi\delta}}\int_{-\infty}^{\infty}e^{-x^2/(2\delta)}g(x)dx,$$

where g is nonnegative measurable function bounded by 1. This leads to $g=0$ almost everywhere and

$$\lim_{l\to\infty} d^{(c)}_{m_l,\mathrm{TV}}((1-\delta)t_{m_l})=1.$$

□

The following is a simple corollary of Proposition 3.2 (1)-(2).

Corollary 3.3. *Let* $\mathcal{F}_\delta,\mathcal{F}_c$ *be families in Theorem 3.1 and* t_n,r_n *be sequences tending to infinity. Fix* $\delta\in(0,1)$.

(1) *If* $d^{(\delta)}_{n,\mathrm{TV}}(\lceil t_n\rceil) \to 0$*, then*

$$\lim_{n\to\infty} d^{(c)}_{n,\mathrm{TV}}((1+\epsilon)(1-\delta)t_n) = 0, \quad \forall \epsilon > 0.$$

(2) *If* $d^{(c)}_{n,\mathrm{TV}}(r_n) \to 0$*, then*

$$\lim_{n\to\infty} d^{(\delta)}_{n,\mathrm{TV}}(\lceil (1+\epsilon)r_n/(1-\delta)\rceil) = 0, \quad \forall \epsilon > 0.$$

Proof of Theorem 3.1. Set $r_n = T^{(\delta)}_{n,\mathrm{TV}}(\epsilon_0)$ and $s_n = T^{(c)}_{n,\mathrm{TV}}(\epsilon_0)$. Suppose $r_n \to \infty$. By Corollary 3.3 (2), if

$$\liminf_{n\to\infty} d^{(c)}_{n,\mathrm{TV}}((1-\delta)r_n/2) = 0,$$

then

$$\liminf_{n\to\infty} d^{(\delta)}_{n,\mathrm{TV}}(\lceil (1+\epsilon)r_n/2\rceil) = 0, \quad \forall \epsilon > 0.$$

But, taking $\epsilon = 1/2$ implies that, for n large enough,

$$d^{(\delta)}_{n,\mathrm{TV}}(\lceil (1+\epsilon)r_n/2\rceil) \ge d^{(\delta)}_{n,\mathrm{TV}}(r_n - 1) > \epsilon_0 > 0.$$

This makes a contradiction and, hence, if $r_n \to \infty$, then

$$\liminf_{n\to\infty} d^{(c)}_{n,\mathrm{TV}}((1-\delta)r_n/2) > 0.$$

In a similar way, if $s_n \to \infty$, then Corollary 3.3 (1) implies

$$\liminf_{n\to\infty} d^{(\delta)}_{n,\mathrm{TV}}(\lceil s_n\rceil) > 0.$$

This proves the following equivalence.

$$T^{(\delta)}_{n,\mathrm{TV}}(\epsilon_0) \to \infty \quad \text{for some } \epsilon_0 > 0 \quad \Leftrightarrow \quad T^{(c)}_{n,\mathrm{TV}}(\epsilon_0) \to \infty \quad \text{for some } \epsilon_0 > 0.$$

For the equivalence of (1) and (2), the proof for precutoffs is given by Corollary 3.3 (1)-(2), while the proof for cutoffs also uses Proposition 3.2 (3)-(4). □

3.2. **Cutoff windows.** This section is devoted to the comparison of cutoff windows introduced in Definition 2.1.

Theorem 3.4. *Let $\mathcal{F}$ be a family of irreducible positive recurrent Markov chains and $\mathcal{F}_\delta, \mathcal{F}_c$ be associated families of lazy walks and continuous semigroups. Let t_n, b_n be sequences of positive reals and assume that $t_n \to \infty$. If $\mathcal{F}_\delta$ (resp. $\mathcal{F}_c$) presents a (t_n, b_n) cutoff in total variation, then $\sqrt{t_n} = O(b_n)$.*

Remark 3.1. There are examples with cutoffs but the order of any window size must be bigger than $\sqrt{t_n}$. Consider the Ehrenfest chain on $\{0, ..., n\}$, which is a birth and death chain with rates $p_{n,i} = 1 - i/n$, $q_{n,i} = i/n$ and $r_{n,i} = 0$. It is obvious that K_n is irreducible and periodic with stationary distribution $\pi_n(i) = 2^{-n}\binom{n}{i}$. An application of the representation theory shows that, for $0 \le i \le n$, $2i/n$ is an eigenvalue of $I - K_n$. Let $\lambda_n = 2/n$ and $s_n = \sum_{i=1}^n n/(2i) = \frac{1}{2} n \log n + O(n)$. By Theorem 4.1, since $\lambda_n s_n$ tends to infinity, both $\mathcal{F}_c$ and $\mathcal{F}_\delta$ have a total variation cutoff. For a detailed computation on the total variation and the L^2-distance, see e.g. [7]. It is well-known that $\mathcal{F}_c$ has a $(\frac{1}{4} n \log n, n)$ total variation cutoff. By Theorem 3.5, $\mathcal{F}_\delta$ has a $(\frac{n \log n}{4(1-\delta)}, n)$ total variation cutoff for $\delta \in (0,1)$, which is nontrivial. For the continuous time Ehrenfest chains, Theorem 3.4 says that the window size is at least $\sqrt{n \log n}$, while n is the correct order.

Proof of Theorem 3.4. We prove the continuous time case. The lazy discrete time case can be treated similarly. Assume the inverse that the sequence $\sqrt{t_n}/b_n$ is not bounded. By considering the subsequence of $\sqrt{t_n}/b_n$ which tends to infinity, it loses no generality to assume that $b_n = o(\sqrt{t_n})$. According to the definition of cutoffs, we may choose $C > 0$, $x_n \in \Omega_n$ and $A_n \subset \Omega_n$ such that

$$\liminf_{n\to\infty} |H_{n,t_n+Cb_n}(x_n, A_n) - \pi_n(A_n)| > 0.$$

By Lemma 3.8, one may choose a sequence n_l tending to infinity such that $\pi_{n_l}(A_{n_l})$ converges to $\alpha \in [0, 1]$ and

$$\lim_{l\to\infty} H_{n_l,t_{n_l}+Cb_{n_l}}(x_{n_l}, A_{n_l}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} f(x)dx \neq \alpha,$$

where f is positive and bounded by 1. Let $c \in \mathbb{R}$. For any $\epsilon > 0$, choose $N > 0$ such that, for $n \geq N$,

$$\left| H_{n,t_n+cb_n}(x_n, A_n) - \sum_{i:|i-t_n|\leq N\sqrt{t_n}} \left(e^{-(t_n+cb_n)} \frac{(t_n + cb_n)^i}{i!} \right) K_n^i(x_n, A_n) \right| < \epsilon.$$

Note that

$$e^{-(t_n+cb_n)} \frac{(t_n + cb_n)^i}{i!} = e^{-(t_n+Cb_n)} \frac{(t_n + Cb_n)^i}{i!} (1 + o(1)) \quad \text{as } n \to \infty,$$

where $o(1)$ is uniform for $|i - t_n| \leq N\sqrt{t_n}$. This implies

$$\lim_{l\to\infty} H_{n_l,t_{n_l}+cb_{n_l}}(x_{n_l}, A_{n_l}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} f(x)dx, \quad \forall c \in \mathbb{R}.$$

Since $\mathcal{F}_c$ presents a (t_n, b_n) cutoff, the right-side integral is equal to α, a contradiction. □

Theorem 3.5. *Let $\mathcal{F}_\delta, \mathcal{F}_c$ be families in Theorem 3.4 and $t_n \to \infty$. Then, the following are equivalent.*

1. *$\mathcal{F}_\delta$ has a (t_n, b_n) cutoff.*
2. *$\mathcal{F}_c$ has a $((1 - \delta)t_n, b_n)$ cutoff.*

To prove this theorem, we need the following proposition.

Proposition 3.6. *Let $\mathcal{F}_\delta, \mathcal{F}_c$ be as in Theorem 3.5 and t_n, r_n be sequences tending to infinity.*

1. *If $\mathcal{F}_\delta$ has a (t_n, b_n) cutoff, then $\mathcal{F}_c$ has a $((1 - \delta)t_n, d_n)$ cutoff for any sequence satisfying $d_n = o(t_n)$ and $b_n = o(d_n)$.*
2. *If $\mathcal{F}_c$ has a (r_n, b_n) cutoff, then $\mathcal{F}_\delta$ has a $(r_n/(1 - \delta), d_n)$ cutoff for any sequence satisfying $d_n = o(r_n)$ and $b_n = o(d_n)$.*

Proof. Immediately from Theorem 3.4 and Proposition 3.2. □

Proof of Theorem 3.5. We prove (1)⇒(2), while the reasoning for (2)⇒(1) is similar. Suppose that $\mathcal{F}_\delta$ has a (t_n, b_n) cutoff with $t_n \to \infty$. Fix $\epsilon \in (0, 1)$ and set $c_n = |T^{(c)}_{n\mathrm{TV}}(\epsilon) - (1 - \delta)t_n|$. By [5, Proposition 2.3], it remains to show that $c_n = O(b_n)$. Assume the inverse, that is, there is a subsequence $\xi = \{n_l | l = 1, 2, ...\}$ such that $c_{n_l}/b_{n_l} \to \infty$ as $l \to \infty$. Let $\mathcal{F}_\delta(\xi), \mathcal{F}_c(\xi)$ be families of $\mathcal{F}_\delta, \mathcal{F}_c$ restricted to ξ. This implies $\mathcal{F}_\delta(\xi)$ has a (t_{n_l}, b_{n_l}) cutoff, but $\mathcal{F}_c(\xi)$ has no $\big((1 - \delta)t_{n_l}, \sqrt{b_{n_l}c_{n_l}}\big)$ cutoff, a contradiction with Proposition 3.6. □

3.3. **Chains with specified initial states.** For any probability μ on a countable set Ω, we write (μ, Ω, K, π) as an irreducible Markov chain on Ω with transition matrix K, stationary distribution π and initial distribution μ. The total variation distances for the associated δ-lazy walk and continuous time chain are defined by

$$d^{(\delta)}_{\mathrm{TV}}(\mu, n) = \|\mu K^n_\delta - \pi\|_{\mathrm{TV}}, \quad d^{(c)}_{\mathrm{TV}}(\mu, t) = \|\mu H_t - \pi\|_{\mathrm{TV}}. \tag{3.3}$$

Denoted by $T^{(\delta)}_{\mathrm{TV}}(\mu, \epsilon), T^{(c)}_{\mathrm{TV}}(\mu, \epsilon)$ are the corresponding mixing times and the concept of cutoffs can be defined similarly as Definition 2.1 according to (3.3). It is an easy exercise to achieve a similar version of Lemma 2.1 for cutoffs with specified initial distributions. The proofs for Propositions 3.2-3.6 and Corollary 3.3 can be adapted to the case when the initial distribution is prescribed. This gives the following theorems.

Theorem 3.7. *Let $\mathcal{F} = \{(\mu_n, \Omega_n, K_n, \pi_n) | n = 1, 2, ...\}$ be a family of irreducible Markov chains and $\mathcal{F}_\delta, \mathcal{F}_c$ be families of associated δ-lazy walks and continuous time chains.*

(1) *$\mathcal{F}_\delta$ has a cutoff (resp. precutoff) iff $\mathcal{F}_c$ has a cutoff (resp. precutoff).*
(2) *If $\mathcal{F}_\delta$ has a cutoff, then $T^{(c)}_{n,\mathrm{TV}}(\mu_n, \epsilon) \sim (1-\delta) T^{(\delta)}_{n,\mathrm{TV}}(\mu_n, \epsilon)$ as n tends to ∞ for all $\epsilon \in (0, 1)$.*

Let $t_n \to \infty$ and $b_n > 0$.

(3) *$\mathcal{F}_\delta$ has a (t_n, b_n) cutoff iff $\mathcal{F}_c$ has a $((1-\delta)t_n, b_n)$ cutoff.*
(4) *If $\mathcal{F}_\delta$ has a (t_n, b_n) cutoff, then $\sqrt{t_n} = O(b_n)$.*

3.4. **Proofs.** This subsection collects required techniques for the proof of theorems in Sections 3.1-3.2.

Lemma 3.8. *Let $a_{n,m} \in [0, 1]$, $t_n > 0$ and $c \in \mathbb{R}$. Suppose that $t_n \to \infty$. Then, there is a subsequence n_k of positive integers and a nonnegative measurable function f bounded by 1 such that*

$$\lim_{k\to\infty} \sum_{m=0}^{\infty} \left(e^{-t_{n_k} - c\sqrt{t_{n_k}}} \frac{(t_{n_k} + c\sqrt{t_{n_k}})^m}{m!} \right) a_{n_k,m} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-c)^2/2} f(x) dx,$$

for all $c \in \mathbb{R}$, and

$$\lim_{k\to\infty} \sum_{m\ge 0} \binom{[(t_{n_k} + c\sqrt{t_{n_k}})/(1-\delta)]}{m} (1-\delta)^m \delta^{[(t_{n_k} + c\sqrt{t_{n_k}})/(1-\delta)] - m} a_{n_k,m}$$
$$= \frac{1}{\sqrt{2\pi\delta}} \int_{-\infty}^{\infty} e^{-(x-c)^2/(2\delta)} f(x) dx,$$

for all $c \in \mathbb{R}$, $\delta \in (0, 1)$, where $[z]$ is any of $\lceil z \rceil, \lfloor z \rfloor$.

Proof. For $n \ge 1$ and any Borel set $A \subset \mathbb{R}$, set

$$\mu_n(A) = \frac{1}{\sqrt{t_n}} \sum_{m: m - t_n/\sqrt{t_n} \in A} a_{n,m}.$$

Let n_k be a subsequence of $\mathbb{N}$ such that

$$\lim_{k\to\infty} \mu_{n_k}((a, b]) = \mu((a, b]), \quad \forall a, b \in \mathbb{Q},\ a < b. \tag{3.4}$$

Clearly, $\mu((a, b]) \le b - a$ for $a < b$ and $a, b \in \mathbb{Q}$. This implies the convergence in (3.4) holds for all $a < b$ and $\mu((a, b]) \le b - a$. As a consequence of the Carathéodory extension theorem, μ can be extended to a measure on $\mathbb{R}$. It is obvious that μ is absolutely continuous with respect to the Lebesgue measure and we write f as the Radon-Nykodym derivative.

Let $\epsilon > 0$ and choose $M > 0$ such that, for $n \geq M$,

$$\sum_{m:|m-t_n|/\sqrt{t_n}\notin(-M,M]} e^{-t_n-c\sqrt{t_n}}\frac{(t_n+c\sqrt{t_n})^m}{m!} < \epsilon.$$

For any integer $N > 1$, set $x_i = iM/N$ and $A_{n,i} = \{m \geq 0 | |m - t_n|/\sqrt{t_n} \in (x_i, x_{i+1}]\}$. By Stirling's formula, it is easy to see that

$$e^{-t_n-c\sqrt{t_n}}\frac{(t_n+c\sqrt{t_n})^m}{m!} = \frac{1+o(1)}{\sqrt{2\pi t_n}}\exp\left\{-\frac{1}{2}\left(\frac{m-t_n}{\sqrt{t_n}}-c\right)^2\right\} \quad \text{as } n\to\infty,$$

where $o(1)$ is uniformly for $m \in A_{n,i}$ and $-N \leq i < N$. This implies

$$\sum_{m\in A_{n,i}}\left(e^{-t_n-c\sqrt{t_n}}\frac{(t_n+c\sqrt{t_n})^m}{m!}\right)a_{n,m}\begin{cases}\leq M_i\mu_n(A_{n,i})/\sqrt{2\pi}+o(1)\\ \geq m_i\mu_n(A_{n,i})/\sqrt{2\pi}+o(1)\end{cases},$$

where $U_i = \sup\{e^{-(x-c)^2/2}|x \in (x_i, x_{i+1}]$ and $L_i = \inf\{e^{-(x-c)^2/2}|x \in (x_i, x_{i+1}]\}$. Summing up i and replacing n with n_k yields

$$\limsup_{k\to\infty}\sum_{m=0}^{\infty}\left(e^{-t_{n_k}-c\sqrt{t_{n_k}}}\frac{(t_{n_k}+c\sqrt{t_{n_k}})^m}{m!}\right)a_{n_k,m} \leq \frac{1}{\sqrt{2\pi}}\sum_{i=-N}^{N-1}M_i\mu((x_i,x_{i+1}])+\epsilon$$

and

$$\liminf_{k\to\infty}\sum_{m=0}^{\infty}\left(e^{-t_{n_k}-c\sqrt{t_{n_k}}}\frac{(t_{n_k}+c\sqrt{t_{n_k}})^m}{m!}\right)a_{n_k,m} \geq \frac{1}{\sqrt{2\pi}}\sum_{i=-N}^{N-1}m_i\mu((x_i,x_{i+1}])-\epsilon.$$

Letting $N \to \infty$ and then $\epsilon \to 0$ gives the desired limit. The proof of the second limit is similar and omitted. ☐

Lemma 3.9. *Let f be a bounded nonnegative measurable function and set $F(t) = \int_{-\infty}^{\infty} e^{-(x-t)^2}f(x)dx$. If F is constant, then f is constant almost everywhere.*

Proof. Set $A = F(t)$, $B^{-1} = \int_{-\infty}^{\infty} e^{-x^2/2}f(x)dx$ and write

$$e^{-(x-t/2)^2}f(x) = B^{-1}\sqrt{2\pi}e^{t^2/4}\left(\frac{1}{\sqrt{2\pi}}e^{-(t-x)^2/2}\right)\left(Be^{-x^2/2}f(x)\right).$$

Note that $AB/(\sqrt{2\pi}e^{t^2/4})$ is the density of $X + Y$, where X has the standard normal distribution, Y is continuous with density function $Be^{-x^2/2}f(x)$ and X, Y are independent. This implies $AB = 1/\sqrt{2}$ and

$$e^{-u^2} = \mathbb{E}(e^{iu(X+Y)}) = e^{-u^2/2}\mathbb{E}(e^{iuY}), \quad \forall u \in \mathbb{R}.$$

Clearly, Y has the standard normal distribution and, thus, f is a constant a.e. ☐

3.5. **A remark on the spectral gap and mixing time.** In this subsection, we make a comparison of spectral gaps between continuous time chains and δ-lazy discrete time chains. Let (Ω, K, π) be an irreducible and reversible finite Markov chain with spectral gap λ, the smallest non-zero eigenvalue of $I-K$. First, we consider the continuous time case. Since (K,π) is reversible, there is a function f defined on $\{0,1,...,n\}$ such that $Kf=(1-\lambda)f$. This implies

$$d_{\mathrm{TV}}^{(c)}(t)=\frac{1}{2}\|H_t-\pi\|_{\infty\to\infty}\ge\frac{\|(H_t-\pi)f\|_\infty}{2\|f\|_\infty}=\frac{e^{-\lambda t}}{2},$$

where $\|A\|_{\infty\to\infty}:=\sup\{\|Ag\|_\infty:\|g\|_\infty=1\}$. Consequently, we obtain

$$T_{\mathrm{TV}}^{(c)}(\epsilon)\ge\frac{-\log(2\epsilon)}{\lambda}.$$

For the lazy discrete time case, a similar discussion yields

$$d_{\mathrm{TV}}^{(\delta)}(t)\ge\beta_\delta^t/2,\quad T_{\mathrm{TV}}^{(\delta)}(\epsilon)\ge\left\lfloor\frac{\log(2\epsilon)}{\log\beta_\delta}\right\rfloor,$$

where β_δ is the second largest absolute value of all nontrivial eigenvalue values of K_δ. By setting $\delta_0=\inf\{\delta\in(0,1)|\beta_\delta=1-(1-\delta)\lambda\}$, it is easy to see that $\delta_0\le 1/2$ and, for $\delta\in[\delta_0,1)$, $\beta_\delta=1-(1-\delta)\lambda$. As a function of δ, β_δ is decreasing on $(0,\delta_0)$ and increasing on $(\delta_0,1)$. Note that $|1-(1-\delta)\lambda|\le\beta_\delta\le\max\{1-2\delta,1-(1-\delta)\lambda\}$. The first inequality implies $1-\beta_\delta\le(1-\delta)\lambda$. Using the second inequality, if $\beta_\delta>1-2\delta$, then $1-\beta_\delta=(1-\delta)\lambda$. If $\beta_\delta\le 1-2\delta$, then $1-\beta_\delta\ge 2\delta\ge\delta\lambda$, where the last inequality uses the fact $\lambda\le 2$. We summarize the discussion in the following lemma.

Lemma 3.10. *Let K be an irreducible transition matrix on a finite set Ω with stationary distribution π. For $\delta\in(0,1)$, let K_δ be the δ-lazy walk given by (1.1). Suppose (π,K) is reversible, that is, $\pi(x)K(x,y)=\pi(y)K(y,x)$ for all $x,y\in\Omega$ and let λ be the smallest non-zero eigenvalue of $I-K$ and β_δ be the largest absolute value of all nontrivial eigenvalues of K_δ. Then, it holds true that*

$$\min\{1-\delta,\delta\}\lambda\le 1-\beta_\delta\le 1-|1-(1-\delta)\lambda|\le(1-\delta)\lambda,\quad\forall\delta\in(0,1).$$

Furthermore, for $\epsilon\in(0,1/2)$,

$$T_{\mathrm{TV}}^{(c)}(\epsilon)\ge\frac{-\log(2\epsilon)}{\lambda},\quad T_{\mathrm{TV}}^{(\delta)}(\epsilon)\ge\left\lfloor\frac{\log(2\epsilon)}{\log\beta_\delta}\right\rfloor\ge\left\lfloor\frac{-\log(2\epsilon)}{2\max\{1-\delta,\log(2/\delta)\}\lambda}\right\rfloor,$$

where the last inequality assumes $|\Omega|\ge 2/\delta$.

Proof. It remains to prove the second inequality in the lower bound of the mixing time for the δ-lazy chain. Note that if $\lambda\le 1/2$, then

$$-\log\beta_\delta\le-\log(1-(1-\delta)\lambda)\le 2(1-\delta)\lambda,$$

where the last inequality uses the fact $\log(1-x)\ge -2x$ for $x\in(0,1/2)$. For $\lambda\ge 1/2$, let $\theta_1(\delta),...,\theta_{|\Omega|}(\delta)$ be eigenvalues of K_δ. Then, $\theta_i(\delta)=\delta+(1-\delta)\theta_i(0)$ and $\sum_{i=1}^{|\Omega|}\theta_i(0)\ge 0$. See [12] for a reference on the second inequality. This implies

$$1+(|\Omega|-1)\beta_\delta\ge\sum_{i=1}^{|\Omega|}\theta_i(\delta)\ge|\Omega|\delta.$$

Assuming $|\Omega| \geq 2/\delta$, the above inequality yields

$$\beta_\delta \geq \frac{|\Omega|\delta - 1}{|\Omega| - 1} \geq \frac{\delta}{2}, \quad -\log\beta_\delta \leq \left(2\log\frac{2}{\delta}\right)\lambda.$$

□

4. Finite birth and death chains

In this section, we consider the total variation cutoff for birth and death chains. A birth and death chain on $\{0, 1, ..., n\}$ with birth rate p_i, death rate q_i and holding rate r_i is a Markov chain with transition matrix K given by

$$K(i, i+1) = p_i, \quad K(i, i-1) = q_i, \quad K(i, i) = r_i, \quad \forall 0 \leq i \leq n, \tag{4.1}$$

where $p_i + q_i + r_i = 1$ and $p_n = q_0 = 0$. It is obvious that K is irreducible if and only if $p_i q_{i+1} > 0$ for $0 \leq i < n$. Under the assumption of irreducibility, the unique stationary distribution π of K is given by $\pi(i) = c(p_0 \cdots p_{i-1})/(q_1 \cdots q_i)$, where c is a positive constant such that $\sum_{i=0}^n \pi(i) = 1$.

In the next two subsections, we recall some results developed in [9, 10] and make an improvement on them using the result in Section 3. In the third subsection, we go back to the issue of cutoffs and make a comparison of total variation and separation cutoffs.

4.1. The total variation cutoff.

Throughout this subsection, we let

$$\mathcal{F} = \{(\Omega_n, K_n, \pi_n) | n = 1, 2, ...\} \tag{4.2}$$

denote a family of irreducible birth and death chains with $\Omega_n = \{0, 1, ..., n\}$ and transition matrix

$$K_n(i, i+1) = p_{n,i}, \quad K_n(i, i-1) = q_{n,i}, \quad K_n(i, i) = r_{n,i}, \quad \forall 0 \leq i \leq n, \tag{4.3}$$

where $p_{n,i} + q_{n,i} + r_{n,i} = 1$ and $p_{n,n} = q_{n,0} = 0$. Write $\lambda_n = \lambda(K_n)$ as the spectral gap of K_n. As before, $\mathcal{F}_c$ denotes the family of associated continuous semigroups and, for $\delta \in (0, 1)$, $\mathcal{F}_\delta$ denotes the family of δ-lazy chains. Recall one of the main results in [10] as follows.

Theorem 4.1 (Theorems 3-3.1 in [10])**.** *Consider the family in* (4.2)*. For $n \geq 1$, let λ_n be the smallest nonzero eigenvalue of $I - K_n$ and let $\beta_{n,\delta}$ be the second largest absolute value of all nontrivial eigenvalues of $K_{n,\delta}$. Then, $\mathcal{F}_c$ (resp. $\mathcal{F}_\delta$ with $\delta \in (0, 1)$) has a total variation cutoff if and only if $T^{(c)}_{n,\mathrm{TV}}(1/4)\lambda_n \to \infty$ (resp. $T^{(\delta)}_{n,\mathrm{TV}}(1/4)(1 - \beta_{n,\delta}) \to \infty$). Moreover, if $\mathcal{F}_c$ (resp. $\mathcal{F}_\delta$) has a cutoff, then the window has size at most $\sqrt{T^{(c)}_{n,\mathrm{TV}}(1/4)/\lambda_n}$ (resp. $\sqrt{T^{(\delta)}_{n,\mathrm{TV}}(1/4)/(1 - \beta_{n,\delta})}$).*

Remark 4.1. By Lemma 3.10, the total variation cutoff in discrete time case is equivalent to $T^{(\delta)}_{n,\mathrm{TV}}(1/4)\lambda_n \to \infty$. By Theorems 3.1-4.1 and Lemma 3.10, if $\mathcal{F}_c$ or $\mathcal{F}_\delta$ has a cutoff, then the window size is at most $\sqrt{T^{(c)}_{n,\mathrm{TV}}(1/4)/\lambda_n}$ or $\sqrt{T^{(\delta)}_{n,\mathrm{TV}}(1/4)/\lambda_n}$.

Remark 4.2. There are examples with cutoffs, but the order of the optimal window size is less than $\sqrt{T^{(c)}_{n,\mathrm{TV}}(1/4)\lambda_n}$. See Remark 3.1.

The combination of the above theorem and Theorem 3.1 yields

Theorem 4.2. *Referring to Theorem 4.1, the following are equivalent.*

(1) *$\mathcal{F}_c$ has a total variation cutoff.*
(2) *$\mathcal{F}_\delta$ has a total variation cutoff.*
(3) *$\mathcal{F}_c$ has a total variation precutoff.*
(4) *$\mathcal{F}_\delta$ has a total variation precutoff.*
(5) *$T^{(c)}_{n,\mathrm{TV}}(\epsilon)\lambda_n \to \infty$ for some $\epsilon \in (0,1)$.*
(6) *$T^{(\delta)}_{n,\mathrm{TV}}(\epsilon)\lambda_n \to \infty$ for some $\epsilon \in (0,1)$.*

Proof of Theorem 4.2. It remains to show (3)⇒(5) and this is given by the inequality $d^{(c)}_{n,\mathrm{TV}}(t) \geq e^{-\lambda_n t}/2$. □

Theorem 4.3. *Consider the family in (4.2). It holds true that $T^{(c)}_{n,\mathrm{TV}}(\epsilon/2) \asymp T^{(\delta)}_{n,\mathrm{TV}}(\eta/2)$ for all $\epsilon, \eta, \delta \in (0,1)$. Furthermore, if there is $\epsilon_0 \in (0,1)$ such that $T^{(c)}_{n,\mathrm{TV}}(\epsilon_0/2)\lambda_n$ or $T^{(\delta)}_{n,\mathrm{TV}}(\epsilon_0/2)\lambda_n$ is bounded, then $T^{(c)}_{n,\mathrm{TV}}(\epsilon/2) \asymp 1/\lambda_n$ and $T^{(\delta)}_{n,\mathrm{TV}}(\epsilon/2) \asymp 1/\lambda_n$ for all $\epsilon, \delta \in (0,1)$.*

Proof of Theorem 4.3. Assume that there is a subsequence n_k and $\epsilon, \eta \in (0,1/2)$ such that either $T^{(c)}_{n_k,\mathrm{TV}}(\epsilon)/T^{(\delta)}_{n_k,\mathrm{TV}}(\eta) \to \infty$ or $T^{(\delta)}_{n_k,\mathrm{TV}}(\eta)/T^{(c)}_{n_k,\mathrm{TV}}(\epsilon) \to \infty$. By Lemma 3.10, we have $T^{(c)}_{n_k,\mathrm{TV}}(\epsilon)\lambda_{n_k} \to \infty$ or $T^{(\delta)}_{n_k,\mathrm{TV}}(\eta)\lambda_{n_k} \to \infty$. In either case, Theorems 3.1-4.1 imply that the subfamily indexed by $(n_k)_{k=1}^\infty$ has a cutoff in both continuous time and δ-lazy discrete time cases. As a consequence of Theorem 3.1, we obtain $T^{(c)}_{n_k,\mathrm{TV}}(\epsilon) \sim (1-\delta)T^{(\delta)}_{n_k,\mathrm{TV}}(\eta)$, which contradicts with the assumption. □

Concerning the window size, a combination of Theorem 3.4 and Theorem 4.1 yields

Theorem 4.4. *Let $\mathcal{F}, \lambda_n$ be as in Theorem 4.1. Suppose that $\mathcal{F}_c$ or $\mathcal{F}_\delta$ has a total variation cutoff and $\lambda_n \asymp 1$. Then, for any $\epsilon, \eta \in (0,1)$ with $\epsilon \neq \eta$,*

$$\left|T^{(c)}_{n,\mathrm{TV}}(\epsilon) - T^{(c)}_{n,\mathrm{TV}}(\eta)\right| \asymp \sqrt{T^{(c)}_{n,\mathrm{TV}}(\epsilon)} \asymp \left|T^{(\delta)}_{n,\mathrm{TV}}(\epsilon) - T^{(\delta)}_{n,\mathrm{TV}}(\eta)\right|.$$

4.2. **The separation cutoff.** In this subsection, we apply the results obtained in the previous subsection to the separation cutoff. First, we give a definition of the separation in the following. Given an irreducible finite Markov chain K on Ω with initial distribution μ and stationary distribution π, the separation distance at time m is defined by

$$d_{\mathrm{sep}}(\mu, m) := \max_{x \in \Omega} \left\{1 - \frac{\mu K^m(x)}{\pi(x)}\right\}.$$

Aldous and Diaconis [2] introduce the concept of the strong stationary time to identify the separation distance. Set $d_{\mathrm{sep}}(m) = \max_i d_{\mathrm{sep}}(i,m)$. A well-known bound on the separation is achieved by Aldous and Fill in Lemma 7 of [1, Chapter 4], which says

$$\bar{d}(m) \leq d_{\mathrm{sep}}(m), \quad d_{\mathrm{sep}}(2m) \leq 1 - (1 - \bar{d}(m))^2, \tag{4.4}$$

where $\bar{d}(m) := \max_{i,j} \|K^m(i,\cdot) - K^m(j,\cdot)\|_{\mathrm{TV}}$. It is clear from the definitions that $d_{\mathrm{TV}}(m) \leq \bar{d}(m) \leq 2d_{\mathrm{TV}}(m)$. Let $T_{\mathrm{sep}}(\epsilon)$ be the separation mixing time. The above inequalities imply

$$T_{\mathrm{TV}}(\epsilon) \leq T_{\mathrm{sep}}(\epsilon) \leq 2T_{\mathrm{TV}}(\epsilon/4), \quad \forall \epsilon \in (0,1). \tag{4.5}$$

Note that the above discussions are also valid for the continuous time case. As the separation distance is between $(0,1)$, the separation cutoff is similar to the total variation cutoff as in Definition 2.1. By (4.5), we obtain the following lemma.

Lemma 4.5. *Let $\mathcal{F}$ be a family of finite Markov chains in either discrete or continuous time case. Assume that $T_{n,\mathrm{TV}}(\epsilon) \to \infty$ or $T_{n,\mathrm{sep}}(\epsilon) \to \infty$ for some $\epsilon \in (0,1)$ in discrete time case. Then, $\mathcal{F}$ has a total variation precutoff if and only if $\mathcal{F}$ has a separation precutoff.*

For birth and death chains, the application of (4.5) to Theorem 4.3 leads to the following theorem.

Theorem 4.6. *Theorem 4.3 also holds in separation. Furthermore, for $\epsilon, \eta \in (0,1/2)$, $T^{(c)}_{n,\mathrm{TV}}(\epsilon) \asymp T^{(c)}_{n,\mathrm{sep}}(\eta)$.*

Let K be an irreducible birth and death chain on $\{0,1,...,n\}$ with stationary distribution π. The authors in [10] obtain the following fact

$$d^{(c)}_{\mathrm{sep}}(t) = 1 - \frac{H_t(0,n)}{\pi(n)}, \quad d^{(\delta)}_{\mathrm{sep}}(m) = 1 - \frac{K^m_\delta(0,n)}{\pi(n)} \quad \forall \delta \in [1/2,1). \tag{4.6}$$

The authors in [9] provide a criterion on the separation cutoff for continuous time chains and monotone discrete time chains. The result says that a separation cutoff exists if and only if the product of the spectral gap and the separation mixing time tends to infinity. The next theorem is a consequence of this fact and Theorems 4.2 and 4.6, which is also obtained in [10].

Theorem 4.7. *Let $\mathcal{F}$ be a family of birth and death chains given by (4.2). The following are equivalent.*

(1) *$\mathcal{F}_c$ has a cutoff in total variation.*
(2) *For $\delta \in (0,1)$, $\mathcal{F}_\delta$ has a cutoff in total variation.*
(3) *$\mathcal{F}_c$ has a cutoff in separation.*
(4) *For $\delta \in [1/2,1)$, $\mathcal{F}_\delta$ has a cutoff in separation.*

The next theorem is a simple corollary of Theorems 4.2-4.7 and Lemma 4.5.

Theorem 4.8. *Theorem 4.2 also holds in separation distance with $\delta \in [1/2,1)$.*

4.3. **The cutoff time in total variation and separation.** In this subsection, we introduce a spectral representation of the total variation mixing time. Let K be the transition kernel of an irreducible birth and death chain on $\{0,1,...,n\}$. Suppose that K is irreducible with stationary distribution π and let $0 < \lambda_1 < \cdots < \lambda_n$ be the eigenvalues of $I-K$. Consider the continuous time case. Using [9, Theorem 4.1] and [10, Corollary 4.5], we have

$$d^{(c)}_{\mathrm{sep}}(t) = 1 - \frac{H_t(0,n)}{\pi(n)} = 1 - \frac{H_t(n,0)}{\pi(0)} = \mathbb{P}(S > t),$$

where S is a sum of n independent exponential random variables with parameters $\lambda_1, ..., \lambda_n$. By the one-sided Chebyshev inequality, one has

$$\mathbb{E}S - \sqrt{\mathrm{Var}(S)/(1/\epsilon - 1)} \le T^{(c)}_{\mathrm{sep}}(\epsilon) \le \mathbb{E}S + \sqrt{(1/\epsilon - 1)\mathrm{Var}(S)}, \quad \forall \epsilon \in (0,1).$$

Note that

$$\mathbb{E}S = \sum_{i=1}^n \frac{1}{\lambda_i}, \quad \mathrm{Var}(S) = \sum_{i=1}^n \frac{1}{\lambda_i^2} \le (\mathbb{E}S)^2.$$

Clearly, this implies

$$\frac{(\sqrt{1-\epsilon} - \sqrt{\epsilon})\mathbb{E}S}{\sqrt{1-\epsilon}} \le T^{(c)}_{\mathrm{sep}}(\epsilon) \le \frac{(\sqrt{\epsilon} + \sqrt{1-\epsilon})\mathbb{E}S}{\sqrt{\epsilon}}, \quad \forall \epsilon \in (0,1). \tag{4.7}$$

The above equation says that, given $\epsilon \in (0, 1/2)$, the separation mixing time is bounded by $\sum_{i=1}^n \lambda_i^{-1}$ up to universal constants. The above discussion is also valid for discrete time case with the assumption that $K(i, i+1) + K(i+1, i) \leq 1$ for $0 \leq i < n$. See [9] for the details. The next proposition is an application of (4.5) and (4.7).

Proposition 4.9. *Let K be an irreducible birth and death chain on $\{0, 1, ..., n\}$. Let $0, \lambda_1, ..., \lambda_n$ be eigenvalues of K and set $s = \sum_{i=1}^n \lambda_i^{-1}$. Then,*

$$\left(\frac{\sqrt{1-\epsilon}-\sqrt{\epsilon}}{\sqrt{1-\epsilon}}\right) s \leq T_{\mathrm{sep}}^{(c)}(\epsilon) \leq \left(\frac{\sqrt{\epsilon}+\sqrt{1-\epsilon}}{\sqrt{\epsilon}}\right) s, \quad \forall \epsilon \in (0, 1/2),$$

and

$$\frac{1}{2}\left(\frac{\sqrt{1-4\epsilon}-\sqrt{4\epsilon}}{\sqrt{1-4\epsilon}}\right) s \leq T_{\mathrm{TV}}^{(c)}(\epsilon) \leq \left(\frac{\sqrt{\epsilon}+\sqrt{1-\epsilon}}{\sqrt{\epsilon}}\right) s, \quad \forall \epsilon \in (0, 1/8).$$

The above also holds in discrete time case with the assumption that $K(i, i+1) + K(i+1, i) \leq 1$ for $0 \leq i < n$.

Applying Proposition 4.9 to Theorems 4.2-4.7 yields the following theorem, where the result in separation is included in [9] and the result in total variation is implicitly obtained in [10].

Theorem 4.10 (Cutoffs from the spectrum)**.** *Let $\mathcal{F}$ be the family in (4.2). For $n \geq 1$, let $\lambda_{n,1}, ..., \lambda_{n,n}$ be non-zero eigenvalues of $I - K_n$ and set*

$$\lambda_n = \min_{1 \leq i \leq n} \lambda_{n,i}, \quad s_n = \frac{1}{\lambda_{n,1}} + \cdots + \frac{1}{\lambda_{n,n}}.$$

Then, the following are equivalent.

(1) *$\mathcal{F}_c$ has a total variation cutoff.*
(2) *For $\delta \in (0, 1)$, $\mathcal{F}_\delta$ has a total variation cutoff.*
(3) *$\mathcal{F}_c$ has a total variation precutoff.*
(4) *For $\delta \in (0, 1)$, $\mathcal{F}_\delta$ has a total variation precutoff.*
(5) *$s_n \lambda_n \to \infty$.*

The above also holds in separation with $\delta \in [1/2, 1)$. In particular, if (5) holds, then, for $\epsilon \in (0, 1)$,

$$\frac{1}{2} \leq \liminf_{n \to \infty} \frac{T_{n,\mathrm{TV}}^{(c)}(\epsilon)}{s_n} \leq \limsup_{n \to \infty} \frac{T_{n,\mathrm{TV}}^{(c)}(\epsilon)}{s_n} \leq 1.$$

The last result establishes a relation between the mixing time and birth and death rates. Consider an irreducible birth and death chain $(X_m)_{m=0}^\infty$ on $\{0, 1, ..., n\}$ with transition matrix K and stationary distribution π. Let N_t be a Poisson process of parameter 1 that is independent of X_m and set, for $0 \leq i \leq n$,

$$\tau_i := \inf\{t \geq 0 | X_{N_t} = i\}.$$

Brown and Shao discuss the distribution of τ_i in [3] and obtain the following result.

$$\mathbb{P}_0(\tau_n > t) = \sum_{j=1}^n \left(\prod_{k \neq j} \frac{\theta_k}{\theta_k - \theta_j}\right) e^{-\theta_j t},$$

where $\mathbb{P}_i$ is the conditional probability given $X_0 = i$ and $\theta_1, ..., \theta_n$ are eigenvalues of the submatrix of $I - K$ indexed by $\{0, 1, ..., n-1\}$. Let $\mathbb{E}_i$ be the conditional expectation given $X_0 = i$. Clearly, this implies $\mathbb{E}_0\tau_n = \sum_{j=1}^n 1/\theta_j$. Note that $\mathbb{E}_0\tau_n$ can be formulated by the birth and death rates using the strong Markov property. This leads to

$$\mathbb{E}_0\tau_n = \sum_{j=1}^{n} \frac{1}{\theta_j} = \sum_{k=0}^{n-1} \frac{\pi([0,k])}{\pi(k)p_k}, \tag{4.8}$$

where $\pi(A) := \sum_{i\in A} \pi(i)$.

Fix $0 \le i_0 \le n$. By (4.8), we have

$$\mathbb{E}_0\tau_{i_0} = \sum_{i=1}^{i_0} \frac{1}{\lambda'_i}, \quad \mathbb{E}_n\tau_{i_0} = \sum_{i=1}^{n-i_0} \frac{1}{\lambda''_i},$$

where $\lambda'_1, ..., \lambda'_{i_0}$ and $\lambda''_1, ..., \lambda''_{n-i_0}$ are eigenvalues of the submatrices of $I - K$ indexed respectively by $\{0, ..., i_0 - 1\}$ and $\{i_0 + 1, ..., n\}$. Let $\bar\lambda_1 \le \cdots \le \bar\lambda_n$ be a rearrangement of $\lambda'_1, ..., \lambda'_{i_0}, \lambda''_1, ..., \lambda''_{n-i_0}$. Clearly, $\bar\lambda_1, ..., \bar\lambda_n$ are eigenvalues of the submatrix obtained by removing the i_0-th row and the i_0-th column of $I - K$. Let $\lambda_1 < \cdots < \lambda_n$ be nonzero eigenvalues of $I - K$. By Theorem 4.3.8 in [11], we have $\bar\lambda_i \le \lambda_i \le \bar\lambda_{i+1}$ and this leads to

$$\sum_{i=2}^{n} \frac{1}{\bar\lambda_i} \le \sum_{i=1}^{n} \frac{1}{\lambda_i} \le \sum_{i=1}^{n} \frac{1}{\bar\lambda_i} = \sum_{k=0}^{i_0-1} \frac{\pi([0,k])}{\pi(k)p_k} + \sum_{k=i_0+1}^{n} \frac{\pi([k,n])}{\pi(k)q_k},$$

where the first equality uses (4.8). By Proposition 4.9, we obtain, for $\epsilon \in (0,1)$,

$$T^{(c)}_{\mathrm{TV}}(\epsilon) \le T^{(c)}_{\mathrm{sep}}(\epsilon) \le \left(\frac{\sqrt{\epsilon} + \sqrt{1-\epsilon}}{\sqrt{\epsilon}}\right) \min_{0\le i\le n} \left\{ \sum_{k=0}^{i-1} \frac{\pi([0,k])}{\pi(k)p_k} + \sum_{k=i+1}^{n} \frac{\pi([k,n])}{\pi(k)q_k} \right\}.$$

The above discussion also holds in discrete time case with the assumption that $p_i + q_{i+1} \le 1$ for all $0 \le i < n$. This includes the δ-lazy chain for $\delta \in [1/2, 1)$ and we apply it to get the following corollary.

Corollary 4.11. *Let $\mathcal{F} = \{(\Omega_n, K_n, \pi_n)|n = 1, 2, ...\}$ be a family of irreducible birth and death chain in* (4.2) *with birth, death and holding rates $p_{n,i}, q_{n,i}, r_{n,i}$. For $n \ge 1$, set*

$$t_n = \min_{0\le i\le n} \left\{ \sum_{k=0}^{i-1} \frac{\pi_n([0,k])}{\pi_n(k)p_{n,k}} + \sum_{k=i+1}^{n} \frac{\pi_n([k,n])}{\pi_n(k)q_{n,k}} \right\}.$$

If $\mathcal{F}_c$ or $\mathcal{F}_\delta$ has a total variation cutoff, then, for $\epsilon \in (0,1)$ and $\delta \in [1/2, 1)$,

$$\limsup_{n\to\infty} \frac{T^{(c)}_{n,\mathrm{sep}}(\epsilon)}{t_n} \le 1, \quad \limsup_{n\to\infty} \frac{T^{(\delta)}_{n,\mathrm{sep}}(\epsilon)}{t_n} \le \frac{1}{1-\delta},$$

and, for $\epsilon \in (0,1)$,

$$\limsup_{n\to\infty} \frac{T^{(c)}_{n,\mathrm{TV}}(\epsilon)}{t_n} \le 1 \quad \limsup_{n\to\infty} \frac{T^{(\delta)}_{n,\mathrm{TV}}(\epsilon)}{t_n} \le \frac{1}{1-\delta}.$$

Remark 4.3. In [6], the constant t_n in Corollary 4.11 is proved to be of the same order as the constant s_n in Theorem 4.10 and the following term

$$\sum_{k=0}^{i_n-1} \frac{\pi_n([0,k])}{\pi_n(k)p_{n,k}} + \sum_{k=i_n+1}^{n} \frac{\pi_n([k,n])}{\pi_n(k)q_{n,k}},$$

where i_n satisfies $\pi_n([0, i_n]) \ge 1/2$ and $\pi_n([i_n, n]) \ge 1/2$.

Remark 4.4. The bound in Corollary 4.11 is also be obtained implicitly in [10] using a coupling argument.

Acknowledgements

We thank the anonymous referee for her/his very careful reading of the manuscript.

References

[1] D. Aldous and J. A. Fill. Reversible markov chains and random walks on graphs. Monograph at http://www.stat.berkeley.edu/users/aldous/RWG/book.html.

[2] David Aldous and Persi Diaconis. Strong uniform times and finite random walks. *Adv. in Appl. Math.*, 8(1):69–97, 1987.

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

[4] Guan-Yu Chen. *The cutoff phenomenon for finite Markov chains*. PhD thesis, Cornell University, 2006.

[5] Guan-Yu Chen and Laurent Saloff-Coste. The cutoff phenomenon for ergodic markov processes. *Electron. J. Probab.*, 13:26–78, 2008.

[6] Guan-Yu Chen and Laurent Saloff-Coste. On the mixing time and spectral gap for birth and death chains. In preparation, 2012.

[7] Persi Diaconis. *Group representations in probability and statistics*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 11. Institute of Mathematical Statistics, Hayward, CA, 1988.

[8] Persi Diaconis. The cutoff phenomenon in finite Markov chains. *Proc. Nat. Acad. Sci. U.S.A.*, 93(4):1659–1664, 1996.

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

[10] Jian Ding, Eyal Lubetzky, and Yuval Peres. Total variation cutoff in birth-and-death chains. *Probab. Theory Related Fields*, 146(1-2):61–85, 2010.

[11] Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, 1990. Corrected reprint of the 1985 original.

[12] L. Saloff-Coste. Lectures on finite Markov chains. In *Lectures on probability theory and statistics (Saint-Flour, 1996)*, volume 1665 of *Lecture Notes in Math.*, pages 301–413. Springer, Berlin, 1997.

[1]Department of Applied Mathematics, National Chiao Tung University, Hsinchu 300, Taiwan

E-mail address: gychen@math.nctu.edu.tw

[2]Malott Hall, Department of Mathematics, Cornell University, Ithaca, NY 14853-4201

E-mail address: lsc@math.cornell.edu