ALEA, Lat. Am. J. Probab. Math. Stat. 21, 1727–1767 (2024)
DOI: 10.30757/ALEA.v21-65

Mixing reversible Markov chains in the max-ℓ^2-distance

Guan-Yu Chen

Department of Applied Mathematics, National Yang Ming Chiao Tung University, Hsinchu 300, Taiwan
E-mail address: gychen@math.nctu.edu.tw
URL: https://jupiter.math.nycu.edu.tw/~gychen/index2.htm

Abstract. In this article, we consider the max-ℓ^2-distances of reversible finite Markov chains and study their mixing times and cutoff phenomena. The theoretical developments generalize the frameworks of Chen and Saloff-Coste in 2010 and Chen, Hsu and Sheu in 2017. Based on the precise spectral information of transition matrices, we provide tight bounds on the mixing time, criteria on the existence of cutoffs, and formulas computing cutoff times. For an illustration, birth and death chains with constant rates, random walks on hypercubes and clustering chains are introduced and discussed in detail.

1. Introduction

Consider a discrete time finite Markov chain on $\mathcal{X}$ with irreducible transition matrix K and stationary distribution π. It is well understood that when K is aperiodic, $K^m(x,\cdot)$ converges to π as m tends to infinity for any $x \in \mathcal{X}$. Briefly, we use the triple $(\mathcal{X}, K, \pi)$ to denote such a chain throughout. To quantify the convergence of K^m, we adopt the following measurement

$$d_2(m) = \max_{x\in\mathcal{X}} \|K^m(x,\cdot)/\pi - 1\|_2 = \max_{x\in\mathcal{X}} \left(\sum_{y\in\mathcal{X}} \left| \frac{K^m(x,y)}{\pi(y)} - 1 \right|^2 \pi(y) \right)^{1/2}, \tag{1.1}$$

and call it the max-ℓ^2-distance in this article. As $d_2(\cdot)$ is non-increasing, it is natural to consider the following quantity

$$T_2(\epsilon) = \min\{m \geq 0 | d_2(m) \leq \epsilon\}, \quad \forall \epsilon > 0,$$

where $\min \emptyset := \infty$. Here, we call $T_2(\cdot)$ the max-ℓ^2-mixing time.

Regarding K as an operator defined on $\ell^2(\pi)$ with $Kf(x) = \sum_y K(x,y)f(y)$, we may rewrite the max-ℓ^2-distance as

$$d_2(m) = \|K^m - E_\pi\|_{\ell^2(\pi)\to\ell^\infty(\pi)} := \sup_{\|f\|_2\leq 1} \|K^m f - E_\pi f\|_\infty, \tag{1.2}$$

Received by the editors October 15th, 2023; accepted August 5th, 2024.
2010 *Mathematics Subject Classification.* 60J10,60J27.
Key words and phrases. Markov chain, reversibility, cutoff.
The author has been partially supported by grant MOST 107-2628-M-009-005-MY4, Taiwan.

where E_π is an operator that maps f to a constant function on $\mathcal{X}$ with value $\pi(f) = \sum_y f(y)\pi(y)$, $\|f\|_2^2 = \sum_y f^2(y)\pi(y)$ and $\|f\|_\infty = \max_y |f(y)|$. By the operator theory, one may use the spectral radius ω and the largest singular value μ of $K - E_\pi$ to bound the max-ℓ^2-mixing time as follows,

$$\frac{1}{-\log \omega} \le T_2(1/e) \le \frac{1 + (1/2)\log \pi_*^{-1}}{-\log \mu} + 1, \tag{1.3}$$

where $\pi_* = \min_x \pi(x)$ and $c/0 := \infty$ for any $c > 0$. In general, $\omega \le \mu$ and the equality holds when K is reversible. As $\pi_*^{-1} \ge |\mathcal{X}|$, the upper and lower bounds in (1.3) are not comparable to each other when the state space is large.

Concerning the ℓ^2-distance with a specified initial state, we set $d_2(x, m) = \|K^m(x, \cdot)/\pi - 1\|_2$. When K is reversible with eigenvalues $\beta_0 = 1$, β_1,..., $\beta_{|\mathcal{X}|-1}$ and $\ell^2(\pi)$-orthonormal eigenvectors $\phi_0 = \mathbf{1}_\mathcal{X}$, ϕ_1,..., $\phi_{|\mathcal{X}|-1}$, the ℓ^2-distance can be expressed as

$$d_2^2(x, m) = \sum_{i=1}^{|\mathcal{X}|-1} |\phi_i(x)|^2 |\beta_i|^{2m}. \tag{1.4}$$

See Saloff-Coste (1997, Lemma 1.2.9) for a proof of the above formula. In addition, by arranging those eigenvalues in the order that $|\beta_i| \ge |\beta_{i+1}|$ for $0 \le i < |\mathcal{X}| - 1$, the above identity can be rewritten as

$$d_2^2(x, m) = \int_{(0,\infty)} e^{-m\lambda} dV_x(\lambda) = m \int_{(0,\infty)} V_x(\lambda) e^{-m\lambda} d\lambda, \tag{1.5}$$

where the second equality is a result of the integration by parts and

$$V_x(\lambda) = \sum_{i=1}^{j-1} |\phi_i(x)|^2, \quad \text{for } -2\log|\beta_{j-1}| \le \lambda < -2\log|\beta_j|,\ 1 \le j \le |\mathcal{X}|, \tag{1.6}$$

with $\beta_{|\mathcal{X}|} := 0$ and $-\log 0 := \infty$. The first and second equalities of (1.5) were introduced in Chen and Saloff-Coste (2010) and Chen et al. (2017) for the purpose of bounding the ℓ^2-mixing time from above and below respectively. We refer the reader to Chen et al. (2017, Proposition 2.8) for an illustration and to Proposition 2.1 in this article for its generalization.

Taking the maximum of (1.4) over $\mathcal{X}$ leads to the square of the max-ℓ^2-distance, which can be regarded as a maximum of Laplace transforms from (1.5). Later on, bounds of mixing times for maxima of Laplace transforms will be produced in Section 2 and applied to reversible Markov chains in Section 3. The following lemma, whose proof is addressed in Subsection 3.1, is an illustration of the achieved results in this aspect.

Lemma 1.1. *Let $(\mathcal{X}, K, \pi)$ be an irreducible and reversible finite Markov chain with eigenvalues $\beta_0 = 1$, β_1,..., $\beta_{|\mathcal{X}|-1}$ and $\ell^2(\pi)$-orthonormal eigenvectors $\phi_0 = \mathbf{1}_\mathcal{X}$, ϕ_1,..., $\phi_{|\mathcal{X}|-1}$. Assume that $|\beta_i| \ge |\beta_{i+1}|$ for $i \ge 0$ and set*

$$\tau = \max_{j \ge 1} \frac{\log(1 + \max_x \sum_{i=1}^{j} |\phi_i(x)|^2)}{-2\log|\beta_j|}. \tag{1.7}$$

Then, the max-ℓ^2-mixing time T_2 of $(\mathcal{X}, K, \pi)$ satisfies

$$\tau \le T_2(\epsilon) \le \frac{6\tau}{\epsilon^4} + 1, \quad \forall \epsilon \in (0, 1/\sqrt{2}]. \tag{1.8}$$

Note that, for fixed $\epsilon \in (0, 1/\sqrt{2}]$, the upper and lower bounds of $T_2(\epsilon)$ in (1.8) are comparable up to universal multiple constants. This improves the result in (1.3) to some extent, while similar bounds for the max-ℓ^p-mixing time with $1 < p \le \infty$ will be provided in Lemma 4.1.

It is remarkable that the formula in (1.8) can not be easily seen from past work though the form of τ looks similar to Chen et al. (2017, Equation (3.2)). One may in fact derive the same bounds as (3.5) of this article from Chen et al. (2017, Proposition 3.4), but τ is indeed more suitable to analyzing the cutoff phenomenon rather than $\tau(c)$ in (3.4). Despite the similarity between (1.8) and (3.5), Lemma 1.1 reveals the fact that $\tau(c) = \tau$ for $c \in (0,1)$.

Next, let's consider a family of finite Markov chains, say $\mathcal{F} = (\mathcal{X}_n, K_n, \pi_n)_{n=1}^{\infty}$. For $n \geq 1$, let $d_{n,2}$ and $T_{n,2}$ be the max-ℓ^2-distance and max-ℓ^2-mixing time of $(\mathcal{X}_n, K_n, \pi_n)$. We say that $\mathcal{F}$ has a max-ℓ^2-cutoff if there is a sequence of positive reals, $(t_n)_{n=1}^{\infty}$, such that

$$\lim_{n\to\infty} d_{n,2}(\lceil ct_n \rceil) = 0, \quad \forall c > 1; \quad \lim_{n\to\infty} d_{n,2}(\lfloor ct_n \rfloor) = \infty, \quad \forall 0 < c < 1. \tag{1.9}$$

Here, $(t_n)_{n=1}^{\infty}$ is called a max-ℓ^2-cutoff time for $\mathcal{F}$. When $T_{n,2}(\epsilon) \to \infty$ for some $\epsilon > 0$, (1.9) is equivalent to

$$\lim_{n\to\infty} \frac{T_{n,2}(\epsilon)}{T_{n,2}(\delta)} = 1, \quad \forall \epsilon, \delta \in (0,\infty). \tag{1.10}$$

Note that the choice of $(t_n)_{n=1}^{\infty}$ is not unique but, from (1.10), $(T_{n,2}(\epsilon))_{n=1}^{\infty}$ is a natural candidate for any $\epsilon > 0$.

From (1.8), one can see that τ in (1.7) is a suitable replacement of the mixing time in (1.14). Such an observation leads to a new criterion on the max-ℓ^2-cutoff and its proof, which is different from that in Chen and Saloff-Coste (2008), guides us to the aim of identifying τ as a max-ℓ^2-cutoff time. The theoretical development in this aspect is made in a more general setting in Section 2, while its application to reversible finite Markov chains are demonstrated in Section 3. We summarize part of Theorem 3.3 in the following.

Theorem 1.2. *Consider a family, $\mathcal{F} = (\mathcal{X}_n, K_n, \pi_n)_{n=1}^{\infty}$, of irreducible and reversible finite Markov chains. For $n \geq 1$, let μ_n be an eigenvalue of K_n with second largest absolute value, $\lambda_n = -\log|\mu_n|$ and τ_n be the quantity in (1.7) corresponding to $(\mathcal{X}_n, K_n, \pi_n)$. Assume that $T_{n,2}(\epsilon_0) \to \infty$ for some $\epsilon_0 > 0$ or $\tau_n \to \infty$. Then, the following statements are equivalent.*

1. *(1) $\mathcal{F}$ has a max-ℓ^2-cutoff.*
2. *(2) $T_{n,2}(\epsilon)\lambda_n \to \infty$ for some $\epsilon > 0$.*
3. *(3) $\tau_n\lambda_n \to \infty$.*

Furthermore, if $\mathcal{F}$ presents a max-ℓ^2-cutoff, then $(\tau_n)_{n=1}^{\infty}$ is a max-ℓ^2-cutoff time and, for $\epsilon, \delta \in (0,\infty)$,

$$|T_{n,2}(\epsilon) - T_{n,2}(\delta)| = O(\lambda_n^{-1} \vee 1), \quad |T_{n,2}(\epsilon) - \tau_n| = O\left([\lambda_n^{-1}\log(\tau_n\lambda_n)] \vee 1\right),$$

where $x \vee y := \max\{x, y\}$ and $a_n = O(b_n)$ means that $\{|a_n/b_n| : n \geq 1\}$ is bounded.

In Theorem 1.2, if $T_{n,2}(\epsilon)$ and τ_n are respectively selected as max-ℓ^2-cutoff times, then $\lambda_n^{-1} \vee 1$ and $[\lambda_n^{-1}\log(\tau_n\lambda_n)] \vee 1$ are their corresponding cutoff windows, which are important concepts to be introduced in Definition 3.1(2). By Theorem 1.2(3), the former cutoff window is of order less, which means better, than the latter one in general. Concerning the optimization of cutoff windows, this means that τ_n might not be the best choice of a max-ℓ^2-cutoff time. We refer the reader to Corollary 3.6 for a possible refinement of Theorem 1.2 in this aspect.

An immediate application of Theorem 1.2 is the identification of the max-ℓ^p-cutoff. In Chen and Saloff-Coste (2008, Propositions 5.3-5.4), it is proved that, for reversible Markov chains, the max-ℓ^p-cutoff is equivalent to the max-ℓ^2-cutoff for $1 < p \leq \infty$. In addition with Theorem 1.2, one may achieve a spectral criterion on the max-ℓ^p-cutoff. See Corollary 4.2 for details.

To illustrate Theorem 1.2, we consider the following random walks. For $n \geq 1$, let $(\mathcal{X}_n, K_n, \pi_n)$ be a Markov chain on $\{0, 1, ..., n\}$ with transition probabilities

$$\begin{cases} K_n(x, x+1) = p_n, \ K_n(x+1, x) = q_n, \quad \forall 0 \leq x < n, \\ K_n(x, x) = r_n, \quad \forall 0 < x < n, \\ K_n(0, 0) = q_n + r_n, \quad K_n(n, n) = p_n + r_n, \end{cases} \tag{1.11}$$

where $p_n + q_n + r_n = 1$. Clearly, K_n is reversible and it's a straightforward exercise to check that K_n has eigenvalues

$$\beta_{n,0} = 1, \quad \beta_{n,i} = r_n + 2\sqrt{p_n q_n} \cos \frac{i\pi}{n+1}, \quad \forall 1 \le i \le n, \tag{1.12}$$

and $\ell^2(\pi_n)$-orthonormal eigenvectors $\phi_{n,0}(x) = 1$ and

$$\phi_{n,i}(x) = C_{n,i} \left\{ \left(\frac{q_n}{p_n}\right)^{(x+1)/2} \sin \frac{i(x+1)\pi}{n+1} - \left(\frac{q_n}{p_n}\right)^{(x+2)/2} \sin \frac{ix\pi}{n+1} \right\},$$

for $1 \le i \le n$ and $x \in \mathcal{X}_n$, where $C_{n,i}^{-2} = (n+1)q_n\pi_n(0)(1-\beta_{n,i})/(2p_n^2)$. The cutoff phenomenon for the family $\mathcal{F} = (\mathcal{X}_n, K_n, \pi_n)_{n=1}^\infty$ with specified initial states was early discussed in Diaconis and Saloff-Coste (2006); Chen and Saloff-Coste (2010) respectively in separation and in the ℓ^2-distance. In separation, the eigenvalues of K_n are sufficient to identify the cutoff, while the eigenvectors of K_n are also required to determine the ℓ^2-cutoff.

In the max-ℓ^2-distance, let τ_n and λ_n be the constants in Theorem 1.2 associated with the transition matrix in (1.11). From (1.12), it is easy to see that $\lambda_n = -\log \beta_{n,1}$ but the computation of τ_n can be challenging since $\phi_{n,i}$ has a complicated form and the absolute values of $\beta_{n,1}, ..., \beta_{n,n}$ are not arranged in the desired order. In Proposition 5.1, the exact order of τ_n is provided and leads to the following criterion on the max-ℓ^2-cutoff, whose proof is presented in Subsection 5.1.

Theorem 1.3. *The family $\mathcal{F} = (\mathcal{X}_n, K_n, \pi_n)_{n=1}^\infty$, where K_n is given by (1.11), has a max-ℓ^2-cutoff if and only if $n|p_n - q_n|/(p_n+q_n) \to \infty$. Furthermore, if $\mathcal{F}$ has a max-ℓ^2-cutoff, then the max-ℓ^2-mixing time $T_{n,2}$ of $(\mathcal{X}_n, K_n, \pi_n)$ satisfies*

$$\left| T_{n,2}(\epsilon) - \frac{n|\log(p_n/q_n)}{-2\log(r_n + 2\sqrt{p_n q_n})} \right| = O\left(\frac{\log|n \log(p_n/q_n)|}{-\log(r_n + 2\sqrt{p_n q_n})} \vee 1 \right), \quad \forall \epsilon > 0. \tag{1.13}$$

Remark 1.4. For the family in Theorem 1.3, the equivalent condition of the max-ℓ^2-cutoff is the same as that of the ℓ^2-cutoff when the initial state has the smallest stationary probability. We refer the reader to Chen and Saloff-Coste (2010, Theorem 7.2) for details.

To end the introduction, we give a brief description of related works. The concept of cutoffs was initiated by Aldous and Diaconis in early 1980s for the purpose of catching up a phase transition of convergence of Markov chains. Since the first example of cutoffs was displayed in Diaconis and Shahshahani (1981), there have been many techniques introduced to analyze this phenomenon. One of the most heuristic conjectures was proposed by Peres in 2004, which said that $\mathcal{F}$ has a cutoff if and only if the following product condition holds,

$$\text{Mixing time} \times \text{Spectral gap} \to \infty. \tag{1.14}$$

In the maximum total variation, this conjecture was confirmed for finite birth and death chains in Ding et al. (2010), while counterexamples in the general case were given by Aldous and Pak respectively. See e.g. Chen and Saloff-Coste (2008, Section 6) for an illustration. In the max-ℓ^p-distance with $1 < p \le \infty$, Peres's conjecture holds for reversible Markov chains, which was proved in Chen and Saloff-Coste (2008), but can fail for nonreversible ones, where counterexamples were provided in Chen (2006). In separation and the relative entropy, the cutoff was considered over some specific classes of Markov chains in Diaconis and Saloff-Coste (2006); Hermon and Peres (2018a,b). It is notable from those literatures that the cutoff phenomenon and the order of mixing times can be sensitive to the distance measuring the convergence of Markov chains.

Referring to (1.14), one can see that, instead of exact values, the orders of mixing times and spectral gaps are sufficient to determine the existence of a cutoff. However, when a cutoff is claimed, it is important to know when the cutoff happens (the cutoff time) and how long the cutoff takes (the cutoff window). Generally, estimating a cutoff time is more challenging than determining the existence of a cutoff, let alone finding a cutoff window. In maximum separation, Diaconis and Saloff-Coste (2006) expressed the cutoff time of finite birth and death chains as a simple formula of nontrivial eigenvalues of transition matrices. In the maximum total variation, Basu et al. (2017) determined the cutoff time of reversible Markov chains using the tail probability of hitting times to large sets. Independently, Chen and Saloff-Coste (2015) proved that the maximum total variation cutoff time for birth and death chains can be the maximum expected hitting time to the median of the stationary distribution from boundary states. Through a comparison of the total variation and the Hellinger distance, Chen and Kumagai (2018) identified cutoffs in both measurements and characterized cutoff times of product chains. For the total variation with specified initial states, Bordenave et al. (2019) considered random walks on sparse random directed graphs and identified the cutoff time with the entropic time, which is a constant determined by the cardinality of the state space and the average row entropy of the transition matrix. In the ℓ^2-distance, Chen and Saloff-Coste (2010) obtained a spectral formula of the cutoff time when the initial distribution is specified. Later, Chen et al. (2017) refined the cutoff criterion introduced in Chen and Saloff-Coste (2010) and applied it to derive a comparison of cutoffs between discrete time and continuous time Markov chains. It is remarkable that Hermon and Peres (2018a) provided a probabilistic interpretation of mixing times and cutoffs in the max-ℓ^2-distance and the maximum relative entropy.

Contrary to the study of cutoff times, there are very few works on cutoff windows except in the max-ℓ^p-distance and the maximum total variation and separation. For the max-ℓ^p-distance of reversible Markov chains with $1 < p \leq \infty$, Chen and Saloff-Coste (2008) proved that, when the mixing time is selected as a cutoff time, the order of a cutoff window is no bigger than the order of the relaxation time, the reciprocal of the spectral gap. For the maximum total variation of lazy random walks, which can be nonreversible, a cutoff window is proved to be of order at least the square root of a cutoff time in Chen and Saloff-Coste (2013). For finite birth and death chains, Diaconis and Saloff-Coste (2006) provided a spectral formula of the optimal cutoff window in maximum separation, which can be of order larger than the relaxation time, while Chen and Saloff-Coste (2015) proved that the maximum standard deviation of the hitting time to the medium of the stationary distribution from boundary states is a choice of cutoff window. We refer the reader to Barrera and Ycart (2014); Chen and Saloff-Coste (2010) for more discussions of cutoff windows in the ℓ^2-distance and to Chen and Saloff-Coste (2008) for other miscellaneous results.

The remaining sections of this article are organized as follows. In Section 2, the theoretical framework of mixing times and cutoffs for maxima of Laplace transforms is developed and, particular, formulas on cutoff times and cutoff windows are established. Through the connection between the max-ℓ^2-distance and the maximum of Laplace transforms, the results produced in Section 2 are used to study the max-ℓ^2-cutoff for families of reversible finite Markov chains in Section 3. In Section 4, two theoretical applications are considered, including spectral bounds on the max-ℓ^p-mixing time and comparisons of cutoffs between discrete time and continuous time Markov chains. For illustration, birth and death chains, random walks on hypercubes and their clustering chains are introduced and discussed in Section 5.

Here are some notations to be used later. Let $x, y \in \mathbb{R}$ and a_n, b_n be sequences of positive reals. We use $x \wedge y$ to denote the minimum of x and y and write $a_n \asymp b_n$ if $a_n = O(b_n)$ and $b_n = O(a_n)$. When $a_n/b_n \to 1$, we write $a_n \sim b_n$. When $a_n/b_n \to 0$, we write $a_n = o(b_n)$. When writing $O(a_n)$ and $o(b_n)$ as single terms, we mean sequences, c_n and d_n, satisfying $|c_n/a_n| = O(1)$ and $|d_n/b_n| = o(1)$. Besides, we also use the convention of $\sup \emptyset := 0$, $\max \emptyset := 0$ and $\inf \emptyset := \infty$.

2. Maxima of Laplace transforms and their cutoffs

The study of cutoffs for Laplace transforms started in Chen and Saloff-Coste (2010) and continued in Chen et al. (2017). In this section, we consider families of maxima of Laplace transforms in a more general setting and derive criteria to examine their cutoffs. The whole work here extends the framework in Chen et al. (2017); Chen and Saloff-Coste (2010).

In the following, we give a blueprint of the theoretical construction but leave lengthy technical proofs to Appendix A. For the consistency between past and current works, we adopt those notations introduced in Chen et al. (2017). Let $\mathcal{V}$ be the class of all non-decreasing and right-continuous functions V on $(0,\infty)$ satisfying

$$\lim_{\lambda\to 0^+} V(\lambda)=0, \quad 0<\lim_{\lambda\to\infty} V(\lambda)<\infty.$$

For $V\in\mathcal{V}$, $\mathcal{W}\subset\mathcal{V}$ and $\epsilon>0$, define

$$\mathcal{L}_V(t):=\int_{(0,\infty)} e^{-t\lambda}dV(\lambda), \quad T_V(\epsilon)=\inf\{t\ge 0|\mathcal{L}_V(t)\le\epsilon\},$$

and

$$\mathcal{L}_{\mathcal{W}}(t):=\sup_{V\in\mathcal{W}}\mathcal{L}_V(t), \quad T_{\mathcal{W}}(\epsilon):=\inf\{t\ge 0|\mathcal{L}_{\mathcal{W}}(t)\le\epsilon\}.$$

Clearly, $\mathcal{L}_V$ is the Laplace transform of the measure induced by function V. As $\mathcal{L}_V$ and $\mathcal{L}_{\mathcal{W}}$ are related to the ℓ^2-distance and the max-ℓ^2-distance of Markov chains, we simply name T_V and $T_{\mathcal{W}}$ mixing times.

To state the results, we need some additional setting. Let $c>0$. For any non-decreasing and non-negative function V on $(0,\infty)$, define

$$\lambda_V(c):=\inf\{\lambda|V(\lambda)>c\}, \quad \tau_V(c):=\sup_{\lambda\ge\lambda_V(c)}\frac{\log(1+V(\lambda))}{\lambda}. \tag{2.1}$$

For any collection $\mathcal{W}$ of non-decreasing and non-negative functions on $(0,\infty)$, define

$$\lambda_{\mathcal{W}}(c):=\inf_{V\in\mathcal{W}}\lambda_V(c), \quad \tau_{\mathcal{W}}(c):=\sup_{V\in\mathcal{W}}\tau_V(c). \tag{2.2}$$

It is easy to see that the values of V at its discontinuous points play no role in (2.1)–(2.2) and, hence, V can be replaced by its right-continuous variant in $\mathcal{V}$. In this article, those notations in (2.1)–(2.2) are mainly used for the circumstance of $V\in\mathcal{V}$ and $\mathcal{W}\subset\mathcal{V}$, while the general setting makes some statements clearer, e.g. (A.6) in Lemma A.5.

Our first result provides comparable upper and lower bounds of mixing times.

Proposition 2.1. *Let $\mathcal{W}\subset\mathcal{V}$ and ϵ,c,c_1,c_2 be constants in $(0,\mathcal{L}_{\mathcal{W}}(0))$. Set $\alpha=\sqrt{\tau_{\mathcal{W}}(c)\lambda_{\mathcal{W}}(c)}$ and assume $\mathcal{L}_{\mathcal{W}}(0)<\infty$.*

(1) *If $\tau_{\mathcal{W}}(c)<\infty$, then*

$$\left(\frac{\alpha}{\alpha+A}\right)T_{\mathcal{W}}\left(c+\frac{A+\alpha}{Ae^{A\alpha}}\right)\le\tau_{\mathcal{W}}(c)\le T_{\mathcal{W}}\left(\frac{c}{1+c}\right), \quad \forall A>0. \tag{2.3}$$

(2) *If $T_{\mathcal{W}}(\epsilon)<\infty$ and $\lambda_{\mathcal{W}}(c_2)>0$, then*

$$T_{\mathcal{W}}\left(c_1+c_2e^{-T_{\mathcal{W}}(\epsilon)\lambda_{\mathcal{W}}(c_1)}+2\epsilon e^{-B}\right)\le T_{\mathcal{W}}(\epsilon)+\frac{2B}{\lambda_{\mathcal{W}}(c_2)}, \quad \forall B>0. \tag{2.4}$$

In particular, for $0<c<(\mathcal{L}_{\mathcal{W}}(0)\wedge 1)/2$, if $\tau_{\mathcal{W}}(c/2)<\infty$, then

$$\tau_{\mathcal{W}}(2c)\le T_{\mathcal{W}}(c)\le\frac{6}{c^2}\tau_{\mathcal{W}}(c/2). \tag{2.5}$$

From (2.3)–(2.4), one can see that the ratio between the upper and lower bounds of $\tau_{\mathcal{W}}(c)$ is close to 1 when α is large. Such an observation is crucial in identifying cutoffs and characterizing cutoff times and cutoff windows, which will be defined below immediately. The readers are referred to Remark A.4 for possible refinement of cutoff windows if any.

We now introduce the various cutoffs as follows.

Definition 2.2. Let $\mathcal{W}_n \subset \mathcal{V}$ and assume

$$M := \limsup_{n\to\infty} \mathcal{L}_{\mathcal{W}_n}(0) > 0.$$

The sequence $(\mathcal{L}_{\mathcal{W}_n})_{n=1}^{\infty}$ is said to present

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

$$\lim_{n\to\infty} \mathcal{L}_{\mathcal{W}_n}(Bt_n) = 0, \quad \liminf_{n\to\infty} \mathcal{L}_{\mathcal{W}_n}(At_n) > 0.$$

(2) a cutoff if there is a sequence $t_n > 0$, which is called a cutoff time, such that

$$\lim_{n\to\infty} \mathcal{L}_{\mathcal{W}_n}(at_n) = \begin{cases} 0 & \text{for } a > 1, \\ M & \text{for } 0 < a < 1. \end{cases}$$

(3) a (t_n, b_n) cutoff if $t_n > 0$, $b_n > 0$, $b_n = o(t_n)$ and

$$\lim_{a\to\infty} \limsup_{n\to\infty} \mathcal{L}_{\mathcal{W}_n}(t_n + ab_n) = 0, \quad \lim_{a\to\infty} \liminf_{n\to\infty} \mathcal{L}_{\mathcal{W}_n}(t_n - ab_n) = M.$$

Here, b_n is called a cutoff window associated with t_n.

Remark 2.3. In Definition 2.2, it is easy to see that (3)⇒(2)⇒(1). Note that $(\mathcal{L}_{\mathcal{W}_n})_{n=1}^{\infty}$ has a cutoff if and only if $T_{\mathcal{W}_n}(\epsilon) \sim T_{\mathcal{W}_n}(\delta)$ for all $0 < \delta < \epsilon < M$. More precisely, $(\mathcal{L}_{\mathcal{W}_n})_{n=1}^{\infty}$ has a (t_n, b_n) cutoff if and only if $|T_{\mathcal{W}_n}(\epsilon) - t_n| = O(b_n)$ for any $0 < \epsilon < M$. In particular, when a cutoff exists, $T_{\mathcal{W}_n}(\epsilon)$ can be selected as a cutoff time for any $\epsilon > 0$.

The next proposition identifies precutoffs with cutoffs, which can fail in general.

Proposition 2.4. *Let $\mathcal{W}_n \subset \mathcal{V}$ and assume that $\limsup_{n\to\infty} \mathcal{L}_{\mathcal{W}_n}(0) > 0$. Then, $(\mathcal{L}_{\mathcal{W}_n})_{n=1}^{\infty}$ has a precutoff if and only if $(\mathcal{L}_{\mathcal{W}_n})_{n=1}^{\infty}$ has a cutoff.*

Remark 2.5. When $|\mathcal{W}_n| = 1$ for all $n \geq 1$, Proposition 2.4 is exactly Theorem 2.2 in Chen et al. (2017). In the proof of Proposition 2.4, it is revealed that, in Definition 2.2, $M = \infty$ is necessary for the existence of a cutoff.

The following theorem is the main result of the framework, which generalizes Chen et al. (2017, Theorem 2.4) and Chen and Saloff-Coste (2010, Theorem 3.8).

Theorem 2.6. *For $n \geq 1$, let $\mathcal{W}_n \subset \mathcal{V}$. Assume that $\mathcal{L}_{\mathcal{W}_n}(0) < \infty$, $\lambda_{\mathcal{W}_n}(c) > 0$ for $0 < c < \mathcal{L}_{\mathcal{W}_n}(0)$, and $\mathcal{L}_{\mathcal{W}_n}(0) \to \infty$. Then, the following statements are equivalent.*

(1) *$(\mathcal{L}_{\mathcal{W}_n})_{n=1}^{\infty}$ has a cutoff.*
(2) *For all $\epsilon > 0$ and $c > 0$, $T_{\mathcal{W}_n}(\epsilon)\lambda_{\mathcal{W}_n}(c) \to \infty$.*
(3) *There is $\epsilon > 0$ such that $T_{\mathcal{W}_n}(\epsilon)\lambda_{\mathcal{W}_n}(c) \to \infty$ for all $c > 0$.*
(4) *For all $c > 0$, $\tau_{\mathcal{W}_n}(c)\lambda_{\mathcal{W}_n}(c) \to \infty$.*
(5) *For all $\widetilde{c} > 0$ and $c > 0$, $\tau_{\mathcal{W}_n}(\widetilde{c})\lambda_{\mathcal{W}_n}(c) \to \infty$.*
(6) *There is $\widetilde{c} > 0$ such that $\tau_{\mathcal{W}_n}(\widetilde{c})\lambda_{\mathcal{W}_n}(c) \to \infty$ for all $c > 0$.*

In particular, if $(\mathcal{L}_{\mathcal{W}_n})_{n=1}^{\infty}$ has a cutoff, then $\tau_{\mathcal{W}_n}(c)$ is a cutoff time for any $c > 0$. Moreover, for any positive constants δ, ϵ, c,

$$|T_{\mathcal{W}_n}(\epsilon) - T_{\mathcal{W}_n}(\delta)| = O\left(\frac{1}{\lambda_{\mathcal{W}_n}(c)}\right), \tag{2.6}$$

and

$$|T_{\mathcal{W}_n}(\epsilon) - \tau_{\mathcal{W}_n}(c)| = O\left(\frac{\log[\tau_{\mathcal{W}_n}(c)\lambda_{\mathcal{W}_n}(c)]}{\lambda_{\mathcal{W}_n}(c)}\right). \tag{2.7}$$

As a cutoff window bounds the duration of a cutoff from above, its order reflects the accuracy of a cutoff time. From (2.6) and (2.7), one can see that the orders of cutoff windows with respect to $T_{\mathcal{W}_n}(\epsilon)$ and $\tau_{\mathcal{W}_n}(c)$ are respectively $\lambda_{\mathcal{W}_n}(c)^{-1}$ and $\lambda_{\mathcal{W}_n}(c)^{-1}\log[\tau_{\mathcal{W}_n}(c)\lambda_{\mathcal{W}_n}(c)]$, where the former is tighter than the latter. It is unknown whether $\lambda_{\mathcal{W}_n}(c)^{-1}$ can be a cutoff window associated with $\tau_{\mathcal{W}_n}(c)$. The following corollary, whose proof is clear from Remark 2.3 and (2.6), compensates Theorem 2.6 in this aspect.

Corollary 2.7. *Refer to the setting in Theorem 2.6 and assume that $(\mathcal{L}_{\mathcal{W}_n})_{n=1}^\infty$ has a cutoff. If $(t_n)_{n=1}^\infty$ is a sequence satisfying*

$$0 < \liminf_{n\to\infty} \mathcal{L}_{\mathcal{W}_n}(t_n) \le \limsup_{n\to\infty} \mathcal{L}_{\mathcal{W}_n}(t_n) < \infty, \tag{2.8}$$

then $(\mathcal{L}_{\mathcal{W}_n})_{n=1}^\infty$ has a $(t_n, 1/\lambda_{\mathcal{W}_n}(c))$ cutoff for any $c > 0$.

3. The max-ℓ^2-cutoff for reversible Markov chains

In this section, we apply those results in Section 2 to reversible finite Markov chains and generate criteria on their max-ℓ^2-cutoffs.

3.1. *Discrete time cases.* Let $(\mathcal{X}, K, \pi)$ be an irreducible and reversible finite Markov chain and d_2, T_2 be its max-ℓ^2-distance and corresponding mixing time. To produce bounds on T_2 from Proposition 2.1, we need a continuous time variant of d_2. Refer to (1.6) and set

$$\mathcal{L}(t) = \max_{x\in\mathcal{X}} \int_{(0,\infty)} e^{-t\lambda} dV_x(\lambda), \quad \forall t \ge 0, \quad T_{\mathcal{L}}(\epsilon) = \inf\{t \ge 0 | \mathcal{L}(t) \le \epsilon\}. \tag{3.1}$$

Clearly, $\mathcal{L}(m) = d_2^2(m)$ for $m \in \mathbb{N}$ and

$$T_{\mathcal{L}}(\epsilon) \le T_2(\sqrt{\epsilon}) = \lceil T_{\mathcal{L}}(\epsilon)\rceil. \tag{3.2}$$

By (2.5), one has

$$\tau(2\epsilon) \le T_{\mathcal{L}}(\epsilon) \le \frac{6}{\epsilon^2}\tau(\epsilon/2), \quad \forall \epsilon \in (0, 1/2), \tag{3.3}$$

where, for any $c > 0$,

$$\tau(c) = \max_{x\in\mathcal{X}} \max_{j\ge j(x,c)} \frac{\log(1 + \sum_{i=1}^j |\phi_i(x)|^2)}{-2\log|\beta_j|}, \tag{3.4}$$

and

$$j(x, c) = \min\left\{ j \ge 1 \,\middle|\, \sum_{i=1}^j |\phi_i(x)|^2 > c \right\}.$$

As a result, this implies

$$\tau(2\epsilon) \le T_2(\sqrt{\epsilon}) \le \left\lceil \frac{6}{\epsilon^2}\tau(\epsilon/2) \right\rceil, \quad \forall \epsilon \in (0, 1/2). \tag{3.5}$$

Thank to Lemma A.5, $\tau(c)$ can be rewritten as

$$\tau(c) = \max_{j\ge j(c)} \frac{\log(1 + \max_x \sum_{i=1}^j |\phi_i(x)|^2)}{-2\log|\beta_j|}, \tag{3.6}$$

where

$$j(c) := \min_{x\in\mathcal{X}} j(x, c) = \min\left\{ j \ge 1 \,\middle|\, \max_{x\in\mathcal{X}} \sum_{i=1}^j |\phi_i(x)|^2 > c \right\}.$$

Through (3.6), the bounds in (3.5) could be computed in a more direct way rather than using (3.4). Further, by the following fact,

$$\max_{x\in\mathcal{X}}\sum_{i=1}^{j}|\phi_i(x)|^2 \ge \sum_{x\in\mathcal{X}}\sum_{i=1}^{j}|\phi_i(x)|^2\pi(x)=j,$$

one has $j(c)=1$ and $\tau(c)=\tau$ for $c\in(0,1)$, where τ is the constant in (1.7). Immediately, this leads to Lemma 1.1, where the case of $\epsilon=1/\sqrt{2}$ follows from the continuity of $T_{\mathcal{L}}$.

Next, we consider the cutoff phenomenon. In the introduction, a family of Markov chains is said to present a max-ℓ^2-cutoff if (1.9) holds. In the following, we classify the cutoff in more detail.

Definition 3.1. Let $\mathcal{F}=(\mathcal{X}_n,K_n,\pi_n)_{n=1}^\infty$ be a family of irreducible finite Markov chains and $d_{n,2}$ be the max-ℓ^2-distance of the nth chain in $\mathcal{F}$. The family $\mathcal{F}$ is said to present

(1) a max-ℓ^2-precutoff if there are $B>A>0$ and $t_n>0$ such that

$$\lim_{n\to\infty} d_{n,2}(\lceil Bt_n\rceil)=0,\quad \liminf_{n\to\infty} d_{n,2}(\lfloor At_n\rfloor)>0.$$

(2) a (t_n,b_n) max-ℓ^2-cutoff if $t_n>0$, $b_n>0$, $b_n=o(t_n)$ and

$$\lim_{c\to\infty}\limsup_{n\to\infty} d_{n,2}(\lceil t_n+cb_n\rceil)=0,\quad \lim_{c\to\infty}\liminf_{n\to\infty} d_{n,2}(\lfloor t_n-cb_n\rfloor)=\infty.$$

Here, b_n is called a max-ℓ^2-cutoff window with respect to t_n.

Clearly, a max-ℓ^2-precutoff is necessary for the existence of a max-ℓ^2-cutoff, while a (t_n,b_n) max-ℓ^2-cutoff reveals more intrinsic information of the convergence of Markov chains. Let $T_{n,2}$ be the max-ℓ^2-mixing time of the nth chain in $\mathcal{F}$ and assume that $T_{n,2}(\epsilon)\to\infty$ for some $\epsilon>0$. It has been displayed in the introduction that the max-ℓ^2-cutoff for $\mathcal{F}$ is equivalent to (1.10). Similarly, one can show that $\mathcal{F}$ has a max-ℓ^2-precutoff if and only if

$$\lim_{\epsilon\to 0}\limsup_{n\to\infty}\frac{T_{n,2}(\epsilon)}{T_{n,2}(\delta)}<\infty,\qquad \text{for some } \delta>0. \tag{3.7}$$

In addition with the assumption of $\liminf_n b_n>0$, the (t_n,b_n) max-ℓ^2-cutoff for $\mathcal{F}$ is in fact equivalent to

$$|T_{n,2}(\epsilon)-t_n|=O(b_n),\quad \forall\epsilon>0. \tag{3.8}$$

Those equivalences can be directly derived from the various definitions of cutoffs. We refer the reader to Chen and Saloff-Coste (2008, Proposition 2.4) for their proofs in a more general setting.

Thereafter, we focus on the specific case that all Markov chains in $\mathcal{F}$ are reversible. By (3.1)–(3.2), if $T_{n,2}(\epsilon)\to\infty$ for some $\epsilon>0$ and $\liminf_n b_n>0$, then all variants of the max-ℓ^2-cutoff for $\mathcal{F}$ are equivalent to cutoffs for $(d_{n,2}^2)_{n=1}^\infty$ in the sense of Definition 2.2. Through this connection, all results in Section 2 are applicable to reversible finite Markov chains. First, by Proposition 2.4, cutoffs and precutoffs are identified as follows.

Lemma 3.2. *Let $\mathcal{F}$ be a family of irreducible and reversible fintie Markov chains and $T_{n,2}$ be the max-ℓ^2-mixing time of the nth chain in $\mathcal{F}$. Assume that $T_{n,2}(\epsilon)\to\infty$ for some $\epsilon>0$. Then, $\mathcal{F}$ has a max-ℓ^2-cutoff if and only if $\mathcal{F}$ has a max-ℓ^2-precutoff.*

To identify the max-ℓ^2-cutoff and characterize its cutoff time, we need the following setting. Let $\beta_{n,0}=1,\dots,\beta_{n,|\mathcal{X}_n|-1}$ be eigenvalues of K_n and $\phi_{n,0}=\mathbf{1}_{\mathcal{X}_n},\dots,\phi_{n,|\mathcal{X}_n|-1}$ be corresponding $\ell^2(\pi_n)$-orthonormal eigenvectors. Assume that $|\beta_{n,i}|\ge|\beta_{n,i+1}|$ for $0\le i<|\mathcal{X}_n|-1$. Set

$$\lambda_n(c)=-\log|\beta_{n,j_n(c)}|,\quad \tau_n(c)=\max_{j\ge j_n(c)}\frac{\log(1+\max_x\sum_{i=1}^{j}|\phi_{n,i}(x)|^2)}{-2\log|\beta_{n,j}|}, \tag{3.9}$$

where

$$j_n(c)=\min\left\{j\ge 1:\max_{x\in\mathcal{X}_n}\sum_{i=1}^{j}|\phi_{n,i}(x)|^2>c\right\},$$

and

$$\lambda_n=-\log|\beta_{n,1}|,\quad \tau_n=\max_{j\ge 1}\frac{\log(1+\max_x\sum_{i=1}^{j}|\phi_{n,i}(x)|^2)}{-2\log|\beta_{n,j}|}. \tag{3.10}$$

By Theorem 2.6, we have the following theorem.

Theorem 3.3. *Let $\mathcal{F}=(\mathcal{X}_n,K_n,\pi_n)_{n=1}^{\infty}$ be a family of irreducible and reversible finite Markov chains. For $n\ge 1$, let $T_{n,2}$ be the max-ℓ^2-mixing time of $(\mathcal{X}_n,K_n,\pi_n)$ and refer to the setting in (3.9)–(3.10). Assume that $\min_x\pi_n(x)\to 0$ and that $T_{n,2}(\epsilon)\to\infty$ for some $\epsilon>0$ or $\tau_n(c)\to\infty$ for some $c>0$. Then, the following are equivalent.*

1. *$\mathcal{F}$ has a max-ℓ^2-cutoff.*
2. *$T_{n,2}(\epsilon)\lambda_n\to\infty$ for any $\epsilon>0$.*
3. *$T_{n,2}(\epsilon)\lambda_n\to\infty$ for some $\epsilon>0$.*
4. *$\tau_n\lambda_n\to\infty$.*
5. *$\tau_n(c)\lambda_n\to\infty$ for any $c>0$.*
6. *$\tau_n(c)\lambda_n\to\infty$ for some $c>0$.*

Furthermore, if $\mathcal{F}$ has a max-ℓ^2-cutoff, then $(\tau_n(c))_{n=1}^{\infty}$ is a max-ℓ^2-cutoff time for any $c>0$ and

$$|T_{n,2}(\epsilon)-T_{n,2}(\delta)|=O\left(\lambda_n^{-1}\vee 1\right),\quad |T_{n,2}(\epsilon)-\tau_n(c)|=O\left(\frac{\log(\tau_n\lambda_n)}{\lambda_n}\vee 1\right), \tag{3.11}$$

for all $\delta,\epsilon,c\in(0,\infty)$.

Remark 3.4. The assumption of $\min_x\pi_n(x)\to 0$ is consistent with the requirement of $\mathcal{L}_{\mathcal{W}_n}(0)\to\infty$ in Theorem 2.6. As the mixing time is integer-valued in the discrete time case, (2.6) and (2.7) turn into the form of (3.11).

Remark 3.5. It is easy to see from Theorem 3.3 that (3), (4) and (6) are useful in proving a max-ℓ^2-cutoff, while (2), (4) and (5) make disproving a max-ℓ^2-cutoff easier. Referring to Definition 3.1(2), one may conclude from (3.11) that $\mathcal{F}$ has a (t_n,b_n) max-ℓ^2-cutoff with $(t_n,b_n)=(T_{n,2}(\epsilon),\lambda_n^{-1}\vee 1)$ or $(t_n,b_n)=(\tau_n,[\lambda_n^{-1}\log(\tau_n\lambda_n)]\vee 1)$.

From (3.11), one can see that the order of the max-ℓ^2-cutoff window can be larger than $\lambda_n^{-1}\vee 1$ when $\tau_n(c)$ is selected as the max-ℓ^2-cutoff time. By Corollary 2.7, we provide in the following a sufficient condition for which $\lambda_n^{-1}\vee 1$ can be a max-ℓ^2-cutoff window.

Corollary 3.6. *Refer to the setting in Theorem 3.3 and assume that $\mathcal{F}$ has a max-ℓ^2-cutoff. If $(t_n)_{n=1}^{\infty}$ is a sequence such that*

$$0<\liminf_{n\to\infty}d_{n,2}(\lceil t_n\rceil)\le\limsup_{n\to\infty}d_{n,2}(\lceil t_n\rceil)<\infty, \tag{3.12}$$

then $\mathcal{F}$ has a $(t_n,\lambda_n^{-1}\vee 1)$ max-ℓ^2-cutoff.

3.2. *Continuous time cases.* In this subsection, we write $(\mathcal{X},Q)$ for a continuous time Markov chain on the finite set $\mathcal{X}$ with generating matrix Q, that is, Q is a matrix indexed by $\mathcal{X}$ and satisfying $Q(x,y)\ge 0$ for $x\ne y$ and $\sum_y Q(x,y)=0$ for all $x\in\mathcal{X}$. Here, Q is called irreducible if $e^{tQ}(x,y)>0$ for all $x,y\in\mathcal{X}$ and $t>0$. In this case, there is a unique probability π on $\mathcal{X}$ such that $\pi Q=\mathbf{0}$, where $\mathbf{0}$ is the constant vector with value 0, and we call π the stationary distribution of Q. Further, it is well-known that

$$\lim_{t\to\infty}e^{tQ}(x,y)=\pi(y),\quad\forall x,y\in\mathcal{X}.$$

From now on, we focus on irreducible continuous time finite Markov chains and simply write $(\mathcal{X}, Q, \pi)$ for them.

In a way similar to the discrete time case, we define the max-ℓ^2-distance of $(\mathcal{X}, Q, \pi)$ and its corresponding mixing time as

$$d_2(t) = \max_{x\in\mathcal{X}} \left(\sum_{y\in\mathcal{X}} \left| \frac{e^{tQ}(x,y)}{\pi(y)} - 1 \right|^2 \pi(y) \right)^{1/2}, \quad T_2(\epsilon) := \min\{t \geq 0 | d_2(t) \leq \epsilon\}.$$

Just like (1.3), one may derive

$$\frac{1}{\zeta} \leq T_2(1/e) \leq \frac{1 + (1/2)\log \pi_*^{-1}}{\lambda}, \tag{3.13}$$

where ζ is the smallest real part of all nonzero eigenvalues of $-Q$ and λ is the smallest nonzero eigenvalue of $-(Q + Q^*)/2$. See Saloff-Coste (1997, Chapter 2) for a proof of the above inequalities. In general, one has $\lambda \leq \zeta$ and, particularly, the equality holds when Q is reversible, i.e. $\pi(x)Q(x,y) = \pi(y)Q(y,x)$ for all $x, y \in \mathcal{X}$. Similar to (1.3), if the cardinality of the state space is large, then the upper and lower bounds of (3.13) are not comparable to each other.

Suppose that Q is reversible and let $\alpha_0 = 0 < \alpha_1 \leq \cdots \leq \alpha_{|\mathcal{X}|-1}$ be eigenvalues of $-Q$ with $\ell^2(\pi)$-orthonormal eigenvectors $\phi_0 = \mathbf{1}_{\mathcal{X}}, \phi_1, ..., \phi_{|\mathcal{X}|-1}$. With the spectral information of Q, we may write the max-ℓ^2-distance as

$$d_2^2(t) = \max_{x\in\mathcal{X}} \sum_{i=1}^{|\mathcal{X}|-1} |\phi_i(x)|^2 e^{-2\alpha_i t} = \max_{x\in\mathcal{X}} \int_{(0,\infty)} e^{-\lambda t} dV_x(\lambda), \tag{3.14}$$

where

$$V_x(\lambda) = \sum_{i=1}^{j-1} |\phi_i(x)|^2, \quad \forall 2\alpha_{j-1} \leq \lambda < 2\alpha_j,\ 1 \leq j \leq |\mathcal{X}|,$$

and $\alpha_{|\mathcal{X}|} := \infty$. Being different from the discrete time case, the square of the max-ℓ^2-distance of a reversible continuous time finite Markov chain is exactly a maximum of Laplace transforms. As an immediate consequence of (2.5) in Proposition 2.1, the max-ℓ^2-mixing time of $(\mathcal{X}, Q, \pi)$ satisfies

$$\tau \leq T_2(\epsilon) \leq \frac{6\tau}{\epsilon^4}, \quad \forall \epsilon \in (0, 1/\sqrt{2}), \tag{3.15}$$

where

$$\tau = \max_{j\geq 1} \frac{\log(1 + \max_x \sum_{i=1}^{j} |\phi_i(x)|^2)}{2\alpha_j}. \tag{3.16}$$

For the cutoff in the continuous time case, all variants are defined in the same way as (1.9) and Definition 3.1 except the removal of $\lceil\cdot\rceil$ and $\lfloor\cdot\rfloor$. As the time index ranges over nonnegative reals, the equivalence of various cutoffs to (1.10), (3.7) and (3.8) also applies for $\mathcal{F}$ without the assumptions of $T_{n,2}(\epsilon) \to \infty$ for some $\epsilon > 0$ and $\liminf_n b_n > 0$. By Proposition 2.4, Theorem 2.6 and Corollary 2.7, we may achieve continuous time versions of Lemma 3.2, Theorem 3.3 and Corollary 3.6 as follows.

Theorem 3.7. *Let $\mathcal{F} = (\mathcal{X}_n, Q_n, \pi_n)_{n=1}^\infty$ be a family of irreducible and reversible continuous time finite Markov chains. For $n \geq 1$, let $\alpha_{n,0}, ..., \alpha_{n,|\mathcal{X}_n|-1}$ be eigenvalues of $-Q_n$ with $\ell^2(\pi_n)$-orthonormal eigenvectors $\phi_{n,0}, ..., \phi_{n,|\mathcal{X}_n|-1}$. Assume that*

$$\alpha_{n,i} \leq \alpha_{n,i+1}, \quad \forall 0 \leq i < |\mathcal{X}_n| - 1, \tag{3.17}$$

and let $\tau_n(c)$, τ_n and λ_n be those in (3.9)–(3.10) under the replacement of $-\log|\beta_{n,i}|$ with $\alpha_{n,i}$. Then, Lemma 3.2, Theorem 3.3 and Corollary 3.6 hold for $\mathcal{F}$ without the assumption that $T_{n,2}(\epsilon) \to \infty$ for some $\epsilon > 0$ or $\tau_n(c) \to \infty$ for some $c > 0$ and with the removal of $\lceil\cdot\rceil$ and $\cdot \vee 1$.

3.3. *Normal Markov chains.* It has been revealed in the introduction and Subsections 3.1 and 3.2 that (1.4) and (3.14) are bridges connecting the max-ℓ^2-distance of finite Markov chains and the maximum of Laplace transforms. In fact, (1.4) and (3.14) hold in a more general realm than the reversibility, while the normality is one typical aspect.

Let $(\mathcal{X}, K, \pi)$ be an irreducible finite Markov chain. Regarding K as an operator on $\ell^2(\pi)$ with kernel $K(x,y)$, we write K^* for the adjoint operator of K. If $K^*(x,y)$ is the corresponding kernel of K^*, then $K^*(x,y) = \pi(y)K(y,x)/\pi(x)$ for all $x, y \in \mathcal{X}$. It is easy to see from the above setting that the transition matrix K is reversible if and only if the operator K is self-adjoint. Following the notion of linear algebra, we say that $(\mathcal{X}, K, \pi)$ is normal if $KK^* = K^*K$. As normal operators are unitarily similar to diagonal matrices, one can prove without difficulty that (1.4) also holds for normal Markov chains, where $|z|$ turns out the modulus of complex number z in this formula. For the continuous time case, we call $(\mathcal{X}, Q, \pi)$ normal if $QQ^* = Q^*Q$, where Q^* is the adjoint operator of Q in $\ell^2(\pi)$. Being similar to the discrete time case, if Q is normal, then (3.14) becomes

$$d_2^2(t) = \max_{x\in\mathcal{X}} \sum_{i=1}^{|\mathcal{X}|-1} |\phi_i(x)|^2 e^{-2t\Re\alpha_i},$$

where $\Re z$ denotes the real part of complex number z. Immediately, we may generalize theorems in Subsections 3.1 and 3.2 as follows.

Corollary 3.8. *For discrete time cases, Lemma 3.2, Theorem 3.3 and Corollary 3.6 hold for normal Markov chains. For continuous time cases, when $\alpha_{n,i}$ is replaced by $\Re\alpha_{n,i}$, Theorem 3.7 holds for normal Markov chains.*

4. Two applications

In this section, we apply the work in Section 3 to push the theoretical development of mixing times forward in some aspects. In the first subsection, we introduce some measurements related to the max-ℓ^2-distance and derive criteria on their cutoffs. In the second subsection, a comparison of max-ℓ^2-cutoffs is made between discrete time and continuous time Markov chains.

4.1. *The mixing times and cutoffs in the max-ℓ^p-distance.* Let $(\mathcal{X}, K\,\pi)$ be an irreducible finite Markov chain. For $p \in [1,\infty]$, the max-ℓ^p-distance of $(\mathcal{X}, K, \pi)$ is defined by

$$d_p(m) := \max_{x\in\mathcal{X}} \left(\sum_{y\in\mathcal{X}} \left| \frac{K^m(x,y)}{\pi(y)} - 1 \right|^p \right)^{1/p}, \quad \forall p \in [1,\infty),$$

and

$$d_\infty(m) := \max_{x,y\in\mathcal{X}} \left| \frac{K^m(x,y)}{\pi(y)} - 1 \right|.$$

Correspondingly, let $T_p(\epsilon)$ be the max-ℓ^p-mixing time of $(\mathcal{X}, K, \pi)$. It was proved in Chen and Saloff-Coste (2008, Proposition 5.1) that, when K is reversible,

$$T_p(\epsilon) \le T_q(\epsilon) \le m_{p,q} T_p(\epsilon^{1/m_{p,q}}), \quad \forall 1 < p \le q \le \infty,\ \epsilon > 0,$$

where $m_{p,q} = \lceil p'/q' \rceil$ and $1/p + 1/p' = 1/q + 1/q' = 1$. This implies that, for all $\epsilon > 0$,

$$T_2(\epsilon) \le T_p(\epsilon) \le 2T_2(\sqrt{\epsilon}), \quad \forall 2 \le p \le \infty, \tag{4.1}$$

and

$$C_p^{-1} T_2(\epsilon^{C_p}) \le T_p(\epsilon) \le T_2(\epsilon), \quad \forall 1 < p \le 2, \tag{4.2}$$

where $C_p = \lceil p/[2(p-1)] \rceil$.

Assume that K is reversible and let $\mathcal{L}$, $T_{\mathcal{L}}$, $\tau(c)$ and τ be as in (3.1), (3.6) and (1.7). By the fact of $\tau(c) = \tau$ for $c \in (0,1)$, one may conclude from the second inequality of (3.3) that $T_{\mathcal{L}}(1/2) \le 24\tau$.

Note that $d_2^2(m) = \mathcal{L}(m)$, $\mathcal{L}(t+s) \le \mathcal{L}(t)|\beta_1|^{2s}$ for $s \ge 0$ and $|\beta_1|^2 \le 2^{-1/\tau}$. This implies $\mathcal{L}(s+t) \le \epsilon$ for $t = T_{\mathcal{L}}(1/2)$ and $s = -\tau \log_2(2\epsilon)$ with $\epsilon \in (0, 1/2]$. As a result, we obtain

$$T_{\mathcal{L}}(\epsilon) \le T_{\mathcal{L}}(1/2) - \tau \log_2(2\epsilon) \le \tau(24 - \log_2(2\epsilon)) = \tau(23 - \log_2 \epsilon),$$

for all $\epsilon \in (0, 1/2]$. In addition with the equality of (3.2) and the first inequality of (1.8), the max-ℓ^2-mixing time can be bounded as follows.

$$\tau \le T_2(\epsilon) = \lceil T_{\mathcal{L}}(\epsilon^2) \rceil \le \lceil \tau(23 - 2\log_2 \epsilon) \rceil, \quad \forall \epsilon \in (0, 1/\sqrt{2}]. \tag{4.3}$$

Note that, based on (3.15), the above discussion also holds in the continuous time case. The following lemma is an immediate consequence of (4.1)–(4.3), while the proof is clear and omitted.

Lemma 4.1. *For $1 < p \le \infty$, set $C_p = \lceil p/[2(p-1)] \rceil$ and let T_p be the max-ℓ^p-mixing time of an irreducible and reversible finite Markov chain.*

(1) *For the discrete time case, one has*

$$\tau \le T_p(\epsilon) \le 2\lceil \tau(23 - 2\log_2 \epsilon) \rceil, \quad \forall 2 \le p \le \infty,$$

and

$$C_p^{-1}\tau \le T_p(\epsilon) \le \lceil \tau(23 - 2\log_2 \epsilon) \rceil, \quad \forall 1 < p \le 2,$$

for all $\epsilon \in (0, 1/\sqrt{2}]$, where τ is the constant in (1.7).

(2) *For the continuous time case, the above inequalities hold under the removal of $\lceil \cdot \rceil$ and the replacement of τ by (3.16).*

Next, let's consider cutoffs in the max-ℓ^p-distance. For a family of irreducible finite Markov chains, its max-ℓ^p-cutoffs, for $p \in (1, \infty]$, are defined by replacing $d_{n,2}$ with $d_{n,p}$ in (1.9) and Definition 3.1 in the discrete time case, while $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ are removed in the continuous time case. In Chen and Saloff-Coste (2008, Theorems 5.3-5.4), the max-ℓ^p-cutoff is proved to be equivalent, for reversible Markov chains, to the max-ℓ^2-cutoff when $1 < p \le \infty$. Along with Theorems 3.3 and 3.7, we achieve the following spectral criterion on the max-ℓ^p-cutoff.

Corollary 4.2. *Let $\mathcal{F}$ be a family of irreducible and reversible finite Markov chains. Refer to the setting and assumptions in Theorem 3.3 for the discrete time case and in Theorem 3.7 for the continuous time case. For $p \in (1, \infty]$, $\mathcal{F}$ has a max-ℓ^p-cutoff if and only if $\tau_n \lambda_n \to \infty$.*

4.2. *Comparisons of mixing times and cutoffs in the max-ℓ^2-distance.* When the state space is finite, a continuous time Markov chain can be generated from a discrete time one. In detail, let $(\mathcal{X}, Q)$ be a continuous time finite Markov chain and write $Q = q(K - I)$, where $q = \max_x |Q(x, x)|$ and I is the identity matrix. Clearly, K is the transition matrix of some finite Markov chain on $\mathcal{X}$. If $(X_n)_{n=0}^\infty$ is a realization of $(\mathcal{X}, K)$ and $(N_t)_{t \ge 0}$ is a Poisson process independent of $(X_n)_{n=0}^\infty$, then $(X_{N_{qt}})_{t \ge 0}$ is a realization of $(\mathcal{X}, Q)$. Note that the only difference of distributions between $(\mathcal{X}, Q)$ and $(\mathcal{X}, K - I)$ is the time-scaling factor q. As a result, the irreducibility of K is consistent with that of Q and both Q and K have the same stationary distribution. Regardless of the factor q, when one considers a continuous time Markov chain, it is natural to start with a discrete time one, say $(\mathcal{X}, K)$, and focus on its continuous time variant $(\mathcal{X}, K - I)$.

The comparison of distributions between $(\mathcal{X}, K)$ and $(\mathcal{X}, K - I)$ provides a way to understand one from the other. Note that the distribution of $(\mathcal{X}, K - I)$ can be expressed by that of $(\mathcal{X}, K)$ through

$$e^{t(K-I)} = \sum_{m=0}^{\infty} \left(e^{-t} \frac{t^m}{m!} \right) K^m. \tag{4.4}$$

From (4.4), one may interpret $e^{t(K-I)}$ as a Poisson(t) randomization of $(K^m)_{m=0}^\infty$, but this only results in one-way comparison in general. Correspondingly, it would be more natural to consider $(\mathcal{X}, K_\delta)$, where $K_\delta = \delta I + (1-\delta)K$ and $\delta \in (0, 1)$, rather than $(\mathcal{X}, K)$ since $(K_\delta)^m = \sum_{i=0}^m \binom{m}{i} \delta^i (1-$

$\delta)^{m-i}K^{m-i}$ and the right hand side is a binomial(m,δ) randomization of $(I,K,K^2,...,K^m)$. Afterward, $(\mathcal{X},K_\delta)$ will be called the δ-lazy variant of $(\mathcal{X},K)$. In Chen and Saloff-Coste (2013), a precise comparison of cutoffs is made between $e^{t(K-I)}$ and $(K_\delta)^m$ in the maximum total variation.

To compare the max-ℓ^2-mixing times of $(\mathcal{X},K_\delta)$ and $(\mathcal{X},K-I)$, let's start with the analysis of their spectra. First, fix an irreducible and reversible finite Markov chain $(\mathcal{X},K,\pi)$. Let $\beta_0=1$, β_1, ..., $\beta_{|\mathcal{X}|-1}$ be eigenvalues of K and $\phi_0=\mathbf{1}_{\mathcal{X}}$, ϕ_1, ..., $\phi_{|\mathcal{X}|-1}$ be corresponding $\ell^2(\pi)$-orthonormal eigenvectors. For $0\le i<|\mathcal{X}|$, set

$$\alpha_i=1-\beta_i,\quad \beta_i^{(\delta)}=\delta+(1-\delta)\beta_i. \tag{4.5}$$

Clearly, α_i and $\beta_i^{(\delta)}$ are eigenvalues of $I-K$ and K_δ and ϕ_i is their corresponding eigenvector. Let $T_2^{(c)}$ and $T_2^{(\delta)}$ be the max-ℓ^2-mixing times of $(\mathcal{X},K-I,\pi)$ and $(\mathcal{X},K_\delta,\pi)$. Assuming that $\beta_i\ge\beta_{i+1}$ for $1\le i<|\mathcal{X}|-1$ yields

$$0<\alpha_i\le\alpha_{i+1},\quad 0<2\delta-1\le\beta_{i+1}^{(\delta)}\le\beta_i^{(\delta)}<1,$$

for $1\le i<|\mathcal{X}|-1$ and $\delta\in(1/2,1)$. By Lemma 4.1, this leads to

$$\tau^{(c)}\le T_2^{(c)}(\epsilon)\le\tau^{(c)}(23-2\log_2\epsilon),\quad \tau^{(\delta)}\le T_2^{(\delta)}(\epsilon)\le\lceil\tau^{(\delta)}(23-2\log_2\epsilon)\rceil, \tag{4.6}$$

for $\delta\in(1/2,1)$ and $\epsilon\in(0,1/\sqrt{2}]$, where

$$\tau^{(c)}=\max_{j\ge1}\frac{\log(1+\max_x\sum_{i=1}^j|\phi_i(x)|^2)}{2\alpha_j}, \tag{4.7}$$

and

$$\tau^{(\delta)}=\max_{j\ge1}\frac{\log(1+\max_x\sum_{i=1}^j|\phi_i(x)|^2)}{-2\log\beta_j^{(\delta)}}. \tag{4.8}$$

Note that

$$\frac{(1-t)(\log a)}{1-a}\le\log t\le(t-1),\quad \forall 0<a\le t\le 1. \tag{4.9}$$

When $t=\beta_j^{(\delta)}$ and $a=2\delta-1$, (4.9) becomes

$$(1-\delta)\alpha_j\le-\log\beta_j^{(\delta)}\le\frac{-\log(2\delta-1)}{2}\alpha_j,\quad \forall\delta\in(1/2,1),\ j\ge1. \tag{4.10}$$

As a consequence, we obtain

$$\frac{2\tau^{(c)}}{-\log(2\delta-1)}\le\tau^{(\delta)}\le\frac{\tau^{(c)}}{1-\delta},\quad \forall\delta\in(1/2,1). \tag{4.11}$$

In addition with the fact of $\log t\ge(t-1)/t$ for $t>0$ and the observation of $\tau^{(c)}\ge(\log 2)/4>1/6$, one may achieve the following lemma from (4.6) and (4.11).

Lemma 4.3. *Consider an irreducible and reversible finite Markov chain $(\mathcal{X},K,\pi)$. Let $T_2^{(c)}$ and $T_2^{(\delta)}$ be the max-ℓ^2-mixing times of $(\mathcal{X},K-I,\pi)$ and $(\mathcal{X},K_\delta,\pi)$. Then, for $\epsilon\in(1/\sqrt{2}]$ and $\delta\in(1/2,1)$,*

$$\frac{2\delta-1}{(1-\delta)(23-2\log_2\epsilon)}T_2^{(c)}(\epsilon)\le T_2^{(\delta)}(\epsilon)\le\frac{26-2\log_2\epsilon}{1-\delta}T_2^{(c)}(\epsilon).$$

Concerning different families of Markov chains, we introduce the following notations for convenience. Let $\mathcal{F}=(\mathcal{X}_n,K_n,\pi_n)_{n=1}^\infty$ be a family of irreducible finite Markov chains and $\Delta=(\delta_n)_{n=1}^\infty$ be a sequence in $(0,1)$. Write

$$\mathcal{F}_c=(\mathcal{X}_n,K_n-I_n,\pi_n)_{n=1}^\infty,\quad \mathcal{F}_\Delta=(\mathcal{X}_n,(K_n)_{\delta_n},\pi_n)_{n=1}^\infty, \tag{4.12}$$

for the family of continuous time Markov chains associated with $\mathcal{F}$, where I_n denotes the identity matrix indexed by $\mathcal{X}_n$, and the family of lazy discrete time Markov chains associated with $\mathcal{F}$ and

Δ. If Δ is a constant sequence with value δ, we simply write $\mathcal{F}_\delta$ for $\mathcal{F}_\Delta$. The next theorem provides an equivalence of max-ℓ^2-cutoffs between continuous time and lazy discrete time Markov chains.

Theorem 4.4. *Consider a family, $\mathcal{F} = (\mathcal{X}_n, K_n, \pi_n)_{n=1}^\infty$, of irreducible and reversible finite Markov chains and a sequence, $\Delta = (\delta_n)_{n=1}^\infty$, in $(1/2, 1)$. Let $\mathcal{F}_c$ and $\mathcal{F}_\Delta$ be the family in (4.12). For $n \ge 1$, let $T_{n,2}^{(\Delta)}$ and $T_{n,2}^{(c)}$ be the max-ℓ^2-mixing times of the nth chains in $\mathcal{F}_\Delta$ and $\mathcal{F}_c$. Assume that $\min_x \pi_n(x) \to 0$, $\inf_n \delta_n > 1/2$ and that there is $\epsilon > 0$ such that $T_{n,2}^{(\Delta)}(\epsilon) \to \infty$ or $(1-\delta_n)^{-1}T_{n,2}^{(c)}(\epsilon) \to \infty$. Then, the following are equivalent.*

(1) *$\mathcal{F}_c$ has a max-ℓ^2-cutoff.*
(2) *$\mathcal{F}_\Delta$ has a max-ℓ^2-cutoff.*

Furthermore, if $\mathcal{F}_c$ and $\mathcal{F}_\Delta$ have max-ℓ^2-cutoffs, then

$$\limsup_{n\to\infty} \frac{2T_{n,2}^{(c)}(\epsilon)}{[-\log(2\delta_n - 1)]T_{n,2}^{(\Delta)}(\epsilon)} \le 1 \le \liminf_{n\to\infty} \frac{T_{n,2}^{(c)}(\epsilon)}{(1-\delta_n)T_{n,2}^{(\Delta)}(\epsilon)}, \quad \forall \epsilon > 0. \tag{4.13}$$

Particularly, if $\delta_n \to 1$, then

$$\lim_{n\to\infty} \frac{T_{n,2}^{(c)}(\epsilon)}{(1-\delta_n)T_{n,2}^{(\Delta)}(\epsilon)} = 1, \quad \forall \epsilon > 0. \tag{4.14}$$

If $\delta_n \to \delta \in (1/2, 1)$ and $T_{n,2}^{(\delta)}$ is the max-ℓ^2-mixing time of the nth chain in $\mathcal{F}_\delta$, then

$$\lim_{n\to\infty} \frac{T_{n,2}^{(\delta)}(\epsilon)}{T_{n,2}^{(\Delta)}(\epsilon)} = 1, \quad \forall \epsilon > 0. \tag{4.15}$$

Proof: The equivalence of (1) and (2) comes immediately from (4.10), (4.11) and Theorems 3.3 and 3.7. For $n \ge 1$, let $\tau_n^{(c)}$ and $\tau_n^{(\Delta)}$ be the constants in (4.7) and (4.8) corresponding to the nth chains in $\mathcal{F}_c$ and $\mathcal{F}_\Delta$. By Theorems 3.3 and 3.7, if $\mathcal{F}_c$ and $\mathcal{F}_\Delta$ have max-ℓ^2-cutoffs, then $T_{n,2}^{(c)}(\epsilon) \sim \tau_n^{(c)}$ and $T_{n,2}^{(\Delta)}(\epsilon) \sim \tau_n^{(\Delta)}$ for all $\epsilon > 0$. Immediately, (4.13) follows from (4.11).

Next, assume that $\mathcal{F}_\Delta$ has a max-ℓ^2-cutoff with $\delta_n \to \delta$. When $\delta = 1$, (4.14) can be easily seen from (4.13). For the case of $\delta \in (1/2, 1)$, let $\tau_n^{(\delta)}$ be the constant in (4.8) corresponding to the nth chain in $\mathcal{F}_\delta$. From the equivalence of (1) and (2), $\mathcal{F}_\delta$ also has a max-ℓ^2-cutoff. By Theorems 3.3 and 3.7, to prove (4.15), it suffices to show that $\tau_n^{(\Delta)} \sim \tau_n^{(\delta)}$. To see this, we need a comparison similar to (4.11). Let $1/2 < \eta < \theta < 1$ and $\tau^{(\eta)}, \tau^{(\theta)}$ be the constant in (4.8). Note that

$$\tau^{(\eta)} = \max_{j\ge 1} \frac{\log(1 + \max_x \sum_{i=1}^j |\phi_i(x)|^2)}{-2\log[\eta + (1-\eta)\beta_i]}.$$

Fix $\beta \in [-1, 1)$ and set

$$f(t) = \frac{1}{-\log[t + (1-t)\beta]}, \quad \forall 1/2 < t < 1.$$

In some computations, one has

$$0 < f'(t) = \frac{1-\beta}{t + (1-t)\beta} f^2(t) \le \frac{1}{(2t-1)(1-t)} f(t),$$

where the last inequality is a result of the second inequality of (4.9). By the mean value theorem of calculus, this implies

$$0 < f(\theta) - f(\eta) \le \frac{\theta - \eta}{(2\eta - 1)(1-\theta)} f(\theta),$$

which leads to

$$\left(1-\frac{\theta-\eta}{(2\eta-1)(1-\theta)}\right)\tau^{(\theta)}\le\tau^{(\eta)}\le\tau^{(\theta)}. \tag{4.16}$$

Back to the proof of (4.15), set $\eta_n=\delta_n\wedge\delta$ and $\theta_n=\delta_n\vee\delta$. Then, replacing (η,θ) with (η_n,θ_n) in (4.16) yields

$$\frac{|\tau_n^{(\Delta)}-\tau_n^{(\delta)}|}{\tau_n^{(\Delta)}\vee\tau_n^{(\delta)}}=\frac{\tau_n^{(\Delta)}\vee\tau_n^{(\delta)}-\tau_n^{(\Delta)}\wedge\tau_n^{(\delta)}}{\tau_n^{(\Delta)}\vee\tau_n^{(\delta)}}\le\frac{|\delta_n-\delta|}{(2\eta_n-1)(1-\theta_n)}\to 0,\quad\text{as } n\to\infty.$$

Consequently, we obtain $\tau_n^{(\Delta)}\sim\tau_n^{(\delta)}$ as desired. □

Remark 4.5. In Theorem 4.4, if Δ is a constant sequence with value $\delta\in(1/2,1)$, then (4.13) turns out

$$1-\delta\le\liminf_{n\to\infty}\frac{T_{n,2}^{(c)}(\epsilon)}{T_{n,2}^{(\delta)}(\epsilon)}\le\limsup_{n\to\infty}\frac{T_{n,2}^{(c)}(\epsilon)}{T_{n,2}^{(\delta)}(\epsilon)}\le\frac{-\log(2\delta-1)}{2}, \tag{4.17}$$

for all $\epsilon>0$. Here, we illustrate with examples that neither the upper bound nor the lower bound is the exact asymptotic ratio of $T_{n,2}^{(c)}$ to $T_{n,2}^{(\delta)}$. Let π_n be a probability on $\mathcal{X}_n=\{1,2,...,n\}$, $\Pi_n(x,y)=\pi(y)$ for $x,y\in\mathcal{X}_n$ and $K_n=\theta I_n+(1-\theta)\Pi_n$, where I_n is the identity matrix indexed by $\mathcal{X}_n$ and $\theta\in(0,1)$. Assuming $\pi_n(x)>0$ for all $x\in\mathcal{X}_n$, one can see that K_n is irreducible and reversible with eigenvalues $1,\theta,\theta,...,\theta$. Note that if $\phi_{n,1},...,\phi_{n,n}$ are corresponding $\ell^2(\pi_n)$-orthonormal eigenvectors of K_n, then $\sum_{i=1}^n\phi_{n,i}^2(x)=1/\pi_n(x)$. As a consequence, if $\tau_n^{(c)},\lambda_n^{(c)}$ and $\tau_n^{(\delta)},\lambda_n^{(\delta)}$ are constants τ_n,λ_n in Theorem 3.7 and (3.10) with respect to $(\mathcal{X}_n,K_n-I_n,\pi_n)$ and $(\mathcal{X}_n,(K_n)_\delta,\pi_n)$, then

$$\lambda_n^{(c)}=1-\theta,\quad\lambda_n^{(\delta)}=-\log(\delta+(1-\delta)\theta),\quad\tau_n^{(c)}=\frac{\log\pi_{n,*}^{-1}}{2\lambda_n^{(c)}},\quad\tau_n^{(\delta)}=\frac{\log\pi_{n,*}^{-1}}{2\lambda_n^{(\delta)}},$$

for $\delta\in(1/2,1)$, where $\pi_{n,*}=\min_x\pi_n(x)$. Clearly, $\pi_{n,*}\to 0$ and this implies $\tau_n^{(c)}\lambda_n^{(c)}\to\infty$ and $\tau_n^{(c)}\to\infty$. By Theorems 3.7 and 4.4, $\mathcal{F}_c$ and $\mathcal{F}_\delta$, with $\delta\in(1/2,1)$, have the max-ℓ^2-cutoff and

$$\lim_{n\to\infty}\frac{T_{n,2}^{(c)}(\epsilon)}{T_{n,2}^{(\delta)}(\epsilon)}=\lim_{n\to\infty}\frac{\tau_n^{(c)}}{\tau_n^{(\delta)}}=\frac{-\log(\delta+(1-\delta)\theta)}{1-\theta}.$$

Applying (4.9) with $(t,a)=(\delta+(1-\delta)\theta,\delta)$ and $(t,a)=(\delta,2\delta-1)$ yields

$$1-\delta<\frac{-\log(\delta+(1-\delta)\theta)}{1-\theta}<-\log\delta<\frac{-\log(2\delta-1)}{2}.$$

Remark 4.6. In this remark, we consider the case that Δ, in Theorem 4.4, is a constant sequence with value 1/2. Let α_i and $\beta_i^{(1/2)}$ be constants in (4.5) and $\beta_i\ge\beta_{i+1}$ for $1\le i<|\mathcal{X}|-1$. Clearly, the first inequality of (4.10) holds with the convention of $-\log 0:=\infty$. This implies the second inequality of (4.11), i.e. $\tau^{(1/2)}\le 2\tau^{(c)}$. Note that, when $0<\beta_1<1$, the replacement of $(t,a)=(\beta_1^{(1/2)},1/2)$ in (4.9) yields $-\log\beta_1^{(1/2)}\le\alpha_1$. When $-1\le\beta_1\le 0$, it is obvious that $\alpha_1\ge 1$. Both cases combine to the fact of $\alpha_1\ge 1\wedge(-\log\beta_1^{(1/2)})$ and, then,

$$\tau^{(c)}\alpha_1\ge\frac{\tau^{(1/2)}\wedge[\tau^{(1/2)}(-\log\beta_1^{(1/2)})]}{2}.$$

Referring to the setting in Theorem 4.4, one may apply the above observation and Theorems 3.3 and 3.7 to conclude that

$$\mathcal{F}_{1/2}\text{ has a max-}\ell^2\text{-cutoff}\quad\Rightarrow\quad\mathcal{F}_c\text{ has a max-}\ell^2\text{-cutoff}. \tag{4.18}$$

Furthermore, if $\mathcal{F}_{1/2}$ has a max-ℓ^2-cutoff, then

$$\liminf_{n\to\infty} \frac{T_{n,2}^{(c)}(\epsilon)}{T_{n,2}^{(1/2)}(\epsilon)} \geq \frac{1}{2}, \quad \forall \epsilon > 0,$$

which is exactly the first inequality in (4.17) with $\delta = 1/2$. However, it is not clear whether the converse of (4.18) is correct.

It is notable from (4.14) that the max-ℓ^2-cutoff times of $\mathcal{F}_c$ and $\mathcal{F}_\Delta$ are closely related when δ_n converges to 1. Based on this observation, we provide a way to determine the max-ℓ^2-cutoff time of $\mathcal{F}_c$ using the max-ℓ^2-cutoff time of $\mathcal{F}_\delta$ in the following corollary.

Corollary 4.7. *Consider a family, $\mathcal{F} = (\mathcal{X}_n, K_n, \pi_n)_{n=1}^\infty$, of irreducible and reversible finite Markov chains. Let $(\delta_k)_{k=1}^\infty$ be a sequence in $(1/2, 1)$ converging to 1 and $(t_n)_{n=1}^\infty$ be a sequence of positive reals tending to infinity. Assume that, for $k \geq 1$, $\mathcal{F}_{\delta_k}$ has a max-ℓ^2-cutoff with cutoff time $(C_k t_n)_{n=1}^\infty$. Then, the following limit exists*

$$C = \lim_{k\to\infty} \frac{C_k}{1-\delta_k},$$

and $\mathcal{F}_c$ has a max-ℓ^2-cutoff with cutoff time $(Ct_n)_{n=1}^\infty$.

Proof: By Theorem 4.4, $\mathcal{F}_c$ has a max-ℓ^2-cutoff with cutoff time, say $(s_n)_{n=1}^\infty$. It follows immediately from (4.17) that

$$1-\delta_k \leq C_k \liminf_{n\to\infty} \frac{t_n}{s_n} \leq C_k \limsup_{n\to\infty} \frac{t_n}{s_n} \leq \frac{-\log(2\delta_k - 1)}{2}, \quad \forall k \geq 1.$$

This implies

$$1 \leq \liminf_{k\to\infty} \frac{C_k}{1-\delta_k} \liminf_{n\to\infty} \frac{t_n}{s_n} \leq \limsup_{k\to\infty} \frac{C_k}{1-\delta_k} \limsup_{n\to\infty} \frac{t_n}{s_n} \leq 1.$$

Consequently, the limits of $C_k/(1-\delta_k)$ and s_n/t_n exist and equal to each other. □

5. Examples

In this section, we introduce classical models to illustrate the work in Sections 3 and 4. For convenience, when a discrete time Markov chain with transition matrix K is considered, we refer its continuous time variant to a continuous time Markov chain with infinitesimal generator $K - I$.

5.1. *Birth and death chains with constant rates.* A birth and death chain on $\mathcal{X} = \{0, 1, ..., N\}$ is a discrete time Markov chain with transition matrix

$$K(x,y) = \begin{cases} p_x & \text{for } y = x+1, \\ q_x & \text{for } y = x-1, \\ r_x & \text{for } y = x, \end{cases}$$

where $p_x + q_x + r_x = 1$ for $x \in \mathcal{X}$ and $p_N = q_0 = 0$. It is clear that K is irreducible if and only if $p_x q_{x+1} > 0$ for $0 \leq x \leq N-1$. With the irreducibility, K has stationary distribution $\pi(x) = \pi(0)p_0 p_1 \cdots p_{x-1}/(q_1 q_2 \cdots q_x)$ for $1 \leq x \leq N$ and $(\mathcal{X}, K, \pi)$ is reversible. Particularly, we call $(\mathcal{X}, K, \pi)$ a birth and death chain on $\{0, 1, ..., N\}$ with constant rates (p, q, r) if $p_x = p$ and $q_{x+1} = q$ for $0 \leq x \leq N-1$ and $r_x = r$ for $1 \leq x \leq N-1$. In this case, $\pi(x) = \pi(0)(p/q)^x$ for $0 \leq x \leq N$, where

$$\pi(0) = \frac{(p/q) - 1}{(p/q)^{N+1} - 1} \quad \text{for } p \neq q, \quad \pi(0) = \frac{1}{N+1} \quad \text{for } p = q.$$

Moreover, the eigenvalues of K are given by

$$\beta_0 = 1, \quad \beta_i = r + 2\sqrt{pq}\cos\frac{i\pi}{N+1}, \quad \forall 1 \le i \le N, \tag{5.1}$$

with $\ell^2(\pi)$-orthonormal eigenvectors

$$\phi_0(x) = 1, \quad \phi_i(x) = \sqrt{\frac{2}{\pi(x)(N+1)}}\sin\left(\frac{ix\pi}{N+1} + \theta_i\right), \quad \forall 1 \le i \le N,\ x \in \mathcal{X}, \tag{5.2}$$

where θ_i is the constant in $(0,\pi)$ such that

$$\sin\theta_i = \sqrt{\frac{p}{1-\beta_i}}\sin\frac{i\pi}{N+1}, \quad \cos\theta_i = \sqrt{\frac{p}{1-\beta_i}}\cos\frac{i\pi}{N+1} - \sqrt{\frac{q}{1-\beta_i}}.$$

Thereafter, we focus on birth and death chains with constant rates.

In Lemma 4.1, one may bound the max-ℓ^p-mixing time with τ. Refer to the setting in (5.1) and (5.2). In the discrete time case, τ turns out

$$\tau = \max_{1 \le j \le N} \frac{\log(1 + \max_x \sum_{i=1}^j |\varphi_i(x)|^2)}{-2\log|\xi_j|}, \tag{5.3}$$

where $(\xi_i)_{i=1}^N$ is a rearrangement of $(\beta_i)_{i=1}^N$ with decreasing absolute values in i and $(\varphi_i)_{i=1}^N$ is the corresponding rearrangement of $(\phi_i)_{i=1}^N$. In the continuous time case, we use $\tau^{(c)}$ to denote τ in avoidance of confusion and has

$$\tau^{(c)} = \max_{1 \le j \le N} \frac{\log(1 + \max_x \sum_{i=1}^j |\phi_i(x)|^2)}{2\alpha_j}, \tag{5.4}$$

where $\alpha_j = 1 - \beta_j$. In the following, we provide bounds on τ and $\tau^{(c)}$ but leave the lengthy proof to the appendix.

Proposition 5.1. *Let τ and $\tau^{(c)}$ be as in (5.3)–(5.4). Set $C = N(p-q)/(p+q)$ and assume $p \ge q$.*

(1) *When $C \le N/3$, one has*

$$\frac{(N+1)^2}{16(C+3)^2(p+q)} \le \tau \le \frac{(C+1)(N+1)^2}{2(p+q)},$$

where the lower bound of τ holds for $N \ge 3$, and

$$\frac{(N+1)^2}{8(C+3)^2(p+q)} \le \tau^{(c)} \le \frac{(C+1)(N+1)^2}{p+q}.$$

(2) *When $C \ge 2$, one has*

$$\frac{N\log(p/q)}{-2\log(r+2\sqrt{pq})}\left(1 - \frac{222}{C^2} - \frac{3\log C + 5}{C}\right) \le \tau \le \frac{N\log(p/q)}{-2\log(r+2\sqrt{pq})}\left(1 + \frac{2}{C}\right),$$

and

$$\frac{N\log(p/q)}{2(\sqrt{p}-\sqrt{q})^2}\left(1 - \frac{42}{C^2} - \frac{3\log C + 5}{C}\right) \le \tau^{(c)} \le \frac{N\log(p/q)}{2(\sqrt{p}-\sqrt{q})^2}\left(1 + \frac{2}{C}\right).$$

Clearly, when C is bounded, τ and $\tau^{(c)}$ are of the same order as $N^2/(p+q)$. When C tends to infinity, tight bounds of τ and $\tau^{(c)}$ can be derived respectively and, further, one may use the following inequalities,

$$-\log(r + 2\sqrt{pq}) \ge (\sqrt{p} - \sqrt{q})^2, \quad \frac{\sqrt{p}}{\sqrt{q}} > \frac{1}{r + 2\sqrt{pq}} > 1,$$

to conclude that, asymptotically, $\tau^{(c)}$ is bigger than τ, while τ is bounded below by N.

Next, we consider the max-ℓ^2-cutoff of birth and death chains. For $n \geq 1$, let $(\mathcal{X}_n, K_n, \pi_n)$ be a birth and death chain on $\{0, 1, ..., n\}$ with constant rates (p_n, q_n, r_n) and τ_n be the associated constant in (5.3). By Proposition 5.1, one has

$$\tau_n \begin{cases} \asymp n^2/(p_n + q_n) & \text{if } c_n = O(1), \\ \sim t_n & \text{if } c_n \to \infty, \end{cases} \tag{5.5}$$

where

$$c_n = \frac{n|p_n - q_n|}{p_n + q_n}, \quad t_n = \frac{n|\log(p_n/q_n)|}{-2\log(r_n + 2\sqrt{p_n q_n})}. \tag{5.6}$$

Let λ_n be the constant in Theorem 3.3. Refer to (5.1), one has $\lambda_n = -\log \beta_{n,1}$, where $\beta_{n,1} = r_n + 2\sqrt{p_n q_n}\cos(\pi/(n+1))$. Note that

$$\begin{aligned} 1 - \beta_{n,1} &= \frac{(p_n - q_n)^2}{(\sqrt{p_n} + \sqrt{q_n})^2} + 4\sqrt{p_n q_n}\sin^2\frac{\pi}{2(n+1)} \\ &\leq \frac{(p_n - q_n)^2}{p_n + q_n} + 2(p_n + q_n) \cdot \frac{\pi^2}{4n^2} = \frac{(p_n + q_n)(c_n^2 + \pi^2/2)}{n^2}, \end{aligned}$$

and

$$n\left|\log\frac{p_n}{q_n}\right| > c_n, \quad \lambda_n \geq -\log(r_n + 2\sqrt{p_n q_n}).$$

Immediately, the former implies $1 - \beta_{n,1} \to 0$ when $c_n = O(1)$, while the latter yields $t_n\lambda_n > c_n/2$. In addition with (5.5), we obtain

$$\tau_n\lambda_n \begin{cases} = O(c_n^2 + 1) = O(1) & \text{if } c_n = O(1), \\ \sim t_n\lambda_n \geq c_n/2 \to \infty & \text{if } c_n \to \infty. \end{cases}$$

Further, thanks to (B.23) and (B.4) of Lemma B.1, if $c_n \to \infty$, then

$$\lambda_n \sim -\log(r_n + 2\sqrt{p_n q_n}), \quad |\tau_n - t_n| = O\left(\frac{\log c_n}{\lambda_n}\right) = O\left(\frac{\log(t_n\lambda_n)}{\lambda_n}\right).$$

Consequently, we achieve Theorem 1.3.

Remark 5.2. Assume that the constant c_n in (5.6) tends to infinity. By Proposition 5.1(ii) and Theorem 1.2, one has $|\tau_n - t_n| = O(b_n)$ and $|T_{n,2}(\epsilon) - \tau_n| = O(d_n)$, where $b_n = (\log c_n)t_n/c_n$ and $d_n = \max\{\lambda_n^{-1}\log(t_n\lambda_n) \vee 1\}$. This implies $|T_{n,2}(\epsilon) - t_n| = O(b_n \vee d_n)$, while (1.13) says $|T_{n,2}(\epsilon) - t_n| = O(d_n)$. It is notable that the former could be worse than the latter since, in general, $d_n = \lambda_n^{-1}\log(t_n\lambda_n)$ and then

$$b_n \vee d_n = t_n \max\left\{\frac{\log c_n}{c_n}, \frac{\log(t_n\lambda_n)}{t_n\lambda_n}\right\} \asymp b_n,$$

where the last step is a result of the fact $t_n\lambda_n > c_n/2$ and the monotonicity of $\log s/s$ over $[e, \infty)$.

For the continuous time case, let $\tau_n^{(c)}$ be the constant in (5.4) associated with $(\mathcal{X}_n, K_n - I_n, \pi_n)$, c_n be as in (5.6) and set $\lambda_n = p_n + q_n - 2\sqrt{p_n q_n}\cos(\pi/(n+1))$. By Proposition 5.1, one has

$$\tau_n^{(c)} \asymp \frac{n^2}{p_n + q_n}, \quad \text{if } c_n = O(1); \quad \tau_n^{(c)} \sim \frac{n|\log(p_n/q_n)|}{2(\sqrt{p_n} - \sqrt{q_n})^2}, \quad \text{if } c_n \to \infty.$$

Note that, by writing

$$\lambda_n = (\sqrt{p_n} - \sqrt{q_n})^2 + 4\sqrt{p_n q_n}\sin^2\frac{\pi}{2(n+1)} \asymp \frac{(p_n + q_n)c_n^2 + \sqrt{p_n q_n}}{n^2},$$

we have

$$\lambda_n = O((p_n + q_n)/n^2), \quad \text{if } c_n = O(1); \quad \lambda_n \asymp (\sqrt{p_n} - \sqrt{q_n})^2, \quad \text{if } c_n \to \infty.$$

As a result, this implies that $\tau_n^{(c)}\lambda_n = O(1)$ when $c_n = O(1)$, and $\tau_n^{(c)}\lambda_n \to \infty$ when $c_n \to \infty$. Consequently, we have the following theorem from Theorem 3.7 and Lemma B.1.

Theorem 5.3. *Let $\mathcal{F}_c = (\mathcal{X}_n, K_n - I_n, \pi_n)_{n=1}^\infty$, where K_n is the transition matrix of a birth and death chain on $\{0,1,...,n\}$ with constant rates (p_n, q_n, r_n). Then, $\mathcal{F}_c$ has a max-ℓ^2-cutoff if and only if $n|p_n - q_n|/(p_n+q_n) \to \infty$. Furthermore, if $\mathcal{F}_c$ has a max-ℓ^2-cutoff, then the max-ℓ^2-mixing time $T_{n,2}^{(c)}$ of the nth chain in $\mathcal{F}_c$ satisfies*

$$\left| T_{n,2}^{(c)}(\epsilon) - \frac{n|\log(p_n/q_n)|}{2(\sqrt{p_n}-\sqrt{q_n})^2} \right| = O\left(\frac{\log|n\log(p_n/q_n)|}{(\sqrt{p_n}-\sqrt{q_n})^2} \right), \quad \forall \epsilon > 0.$$

5.2. *Random walks on hypercubes.* For $n \in \mathbb{N}$, let $\mathcal{X}_n = (\mathbb{Z}_2)^n$ be the group of binary n-vectors and P_n be a probability on $\mathcal{X}_n$. A random walk on $\mathcal{X}_n$ driven by P_n is a discrete time Markov chain on $\mathcal{X}_n$ with transition matrix K_n given by

$$K_n(x,y) = P_n(y-x), \quad \forall x, y \in \mathcal{X}_n. \tag{5.7}$$

Clearly, K_n is symmetric and has the uniform distribution π_n on $\mathcal{X}_n$ as a stationary distribution. By setting

$$\beta_{n,x} = \sum_{y \in \mathbb{Z}_2^n} P_n(y)(-1)^{\langle x,y\rangle}, \quad \phi_{n,x}(y) = (-1)^{\langle x,y\rangle},$$

where $\langle x,y\rangle$ denotes the inner product of x and y, one can directly check that, for $x \in \mathcal{X}_n$, $\beta_{n,x}$ is an eigenvalue of K_n and $\phi_{n,x}$ is a corresponding eigenvector. As $\{\phi_{n,x}|x \in \mathcal{X}_n\}$ forms an $\ell^2(\pi_n)$-orthonormal basis of $\mathbb{R}^{\mathcal{X}_n}$, the ℓ^2-distance of $(\mathcal{X}_n, K_n, \pi_n)$ can be expressed as

$$\left\| \frac{K_n^m(x,\cdot)}{\pi_n} - 1 \right\|_2 = \left(\sum_{y \neq \mathbf{0}} |\beta_{n,y}|^{2m} \right)^{1/2},$$

which is independent of the initial state.

In the following, we focus on the specific case that

$$P_n(x) = P_n(y), \quad \forall x, y \in \mathcal{X}_n,\ H(x,\mathbf{0}) = H(y,\mathbf{0}), \tag{5.8}$$

where $\mathbf{0}$ is the zero vector in $\mathcal{X}_n$ and $H(x,y)$ is the Hamming distance denoting the number of coordinates where x and y disagree. Let S_n be the symmetric group and consider the group action $\sigma x = (x_{\sigma(1)}, ..., x_{\sigma(n)})$ for $\sigma \in S_n$ and $x = (x_1, ..., x_n) \in \mathcal{X}_n$. Under this action, the orbits of $\mathcal{X}_n$ consist of $C_i := \{x \in \mathcal{X}_n | H(x,\mathbf{0}) = i\}$ for $0 \le i \le n$. Obviously, $\sigma(C_i) = C_i$ and $P_n(\sigma x) = P_n(x)$ for $\sigma \in S_n$, $0 \le i \le n$ and $x \in \mathcal{X}_n$. This implies

$$K_n(\sigma x, \sigma y) = K_n(x,y), \quad \forall x, y \in \mathcal{X}_n,\ \sigma \in S_n \tag{5.9}$$

and

$$\sum_{y \in C_i} K_n(x,y) = \sum_{y \in C_i} K_n(\sigma x, y), \quad \forall x \in \mathcal{X}_n,\ 0 \le i \le n,\ \sigma \in S_n. \tag{5.10}$$

Based on these observations, a Markov chain on the clustering sets $\{C_0, ..., C_n\}$ arises naturally. In detail, set $\Pi(x) = i$ for $x \in C_i$ and $0 \le i \le n$. Let $(X_m)_{m=0}^\infty$ be a realization of $(\mathcal{X}_n, K_n, \pi_n)$ and $Y_m = \Pi(X_m)$. It is easy to check that $(Y_m)_{m=0}^\infty$ is also a Markov chain and, by (5.10),

$$\mathbb{P}(Y_{m+1} = j | Y_m = i) = L_n(i,j) := \sum_{y \in C_j} K_n(x,y), \tag{5.11}$$

for all $x \in C_i$ and $i, j \in \mathcal{Y}_n := \{0, 1, ..., n\}$. By setting $\nu_n(i) = \pi_n(C_i)$ for $0 \le i \le n$, one may derive

$$\nu_n(i)L_n(i,j) = \sum_{x\in C_i} \pi_n(x) \sum_{y\in C_j} K_n(x,y) = \sum_{y\in C_j} \pi_n(y) \sum_{x\in C_i} K_n(y,x) = \nu_n(j)L_n(j,i), \quad \forall i,j \in \mathcal{Y}_n,$$

which implies that ν_n is a stationary distribution of L_n and L_n is self-adjoint in $\ell^2(\nu_n)$.

Next, we explore the spectral information of L_n. Observe that, for $\sigma \in S_n$ and $y \in \mathcal{X}_n$,

$$\sum_{x\in C_i} \phi_{n,x}(y) = \sum_{x\in C_i} (-1)^{\langle x,y\rangle} = \sum_{z\in C_i} (-1)^{\langle z,\sigma y\rangle} = \sum_{z\in C_i} \phi_{n,z}(\sigma y), \tag{5.12}$$

and

$$\beta_{n,y} = \sum_{k=0}^n \frac{P_n(C_k)}{|C_k|} \sum_{x\in C_k} (-1)^{\langle x,y\rangle} = \sum_{k=0}^n \frac{P_n(C_k)}{|C_k|} \sum_{z\in C_k} (-1)^{\langle z,\sigma y\rangle} = \beta_{n,\sigma y}. \tag{5.13}$$

For $i, k \in \mathcal{Y}_n$, define $\psi_{n,i}(k) = \sum_{x\in C_i} \phi_{n,x}(y)$ and $\beta_{n,k} = \beta_{n,y}$, where $y \in C_k$. Then, one has

$$\sum_{k=0}^n \psi_{n,i}(k)\psi_{n,j}(k)\nu_n(k) = \sum_{x\in C_i, y\in C_j} \sum_{z\in\mathcal{X}_n} \phi_{n,x}(z)\phi_{n,y}(z)\pi_n(z) = |C_i|\delta_i(j), \tag{5.14}$$

and, for $x \in C_i$,

$$\begin{aligned}\sum_{k=0}^n L_n(i,k)\psi_{n,j}(k) &= \sum_{k=0}^n \sum_{z\in C_k} K_n(x,z) \sum_{y\in C_j} \phi_{n,y}(z) \\ &= \sum_{y\in C_j} \beta_{n,y}\phi_{n,y}(x) = \beta_{n,j}\psi_{n,j}(i).\end{aligned} \tag{5.15}$$

These computations show that $\beta_{n,0}, ..., \beta_{n,n}$ are eigenvalues of L_n with $\ell^2(\nu_n)$-orthogonal eigenvectors $\psi_{n,0}, ..., \psi_{n,n}$. Note that $\beta_{n,0} = 1$, $\psi_{n,0}(k) = 0$ for all $0 \le k \le n$ and $|C_i| = \binom{n}{i}$. As a result, the ℓ^2-distance of $(\mathcal{Y}_n, L_n, \nu_n)$ is given by

$$\left\| \frac{L_n^m(i,\cdot)}{\nu_n} - 1 \right\|_2 = \left(\sum_{j=1}^n \binom{n}{j}^{-1} |\psi_{n,j}(i)|^2 \cdot |\beta_{n,j}|^{2m} \right)^{1/2}. \tag{5.16}$$

Different from K_n, L_n has an ℓ^2-distance sensitive to the initial state. It is worthwhile to note that $(\psi_{n,j})_{j=0}^n$ are related to the well-known Krawtchouk polynomials and independent of P_n. From the following formulas,

$$\beta_{n,j} = \sum_{k=0}^n P_n(C_k) \sum_{\ell=0}^k (-1)^\ell \frac{\binom{j}{\ell}\binom{n-j}{k-\ell}}{\binom{n}{k}}, \quad \psi_{n,j}(i) = \sum_{k=0}^j (-1)^k \binom{i}{k}\binom{n-i}{j-k}, \tag{5.17}$$

one can see that the computation of τ in (1.7) can be challenging.

In the following, we make up a connection of ℓ^p-distances between $(\mathcal{X}_n, K_n, \pi_n)$ and $(\mathcal{Y}_n, L_n, \nu_n)$. From (5.9), one may derive $K_n^m(\sigma x, \sigma y) = K_n^m(x,y)$ for $x, y \in \mathcal{X}_n$, $\sigma \in S_n$ and $m \ge 0$. By (5.11), this implies

$$\begin{aligned}\sum_{y\in C_i} K_n^{m+1}(x,y) &= \sum_{j=0}^n \sum_{z\in C_j} \sum_{y\in C_i} K_n^m(x,z)K_n(z,y) = \sum_{j=0}^n \sum_{z\in C_j} K_n^m(x,z)L_n(j,i) \\ &= \sum_{j=0}^n \sum_{w\in C_j} K_n^m(\sigma x, w)L_n(j,i) = \sum_{y\in C_i} K_n^{m+1}(\sigma x, y),\end{aligned}$$

which generalizes (5.10). In particular, an inductive argument on the second equality of the above computation leads to

$$\sum_{y\in C_j} K_n^m(x,y) = L_n^m(i,j), \quad \forall x\in C_i,\ i,j\in\mathcal{Y}_n.$$

Let μ_n be a probability on $\mathcal{X}_n$, which is uniform on C_i for $0\le i\le n$, and ρ_n be a probability on $\mathcal{Y}_n$ satisfying $\rho_n(i)=\mu_n(C_i)$ for $0\le i\le n$. Then, $\mu_n K_n^m$ is also uniform on C_i for $0\le i\le n$ and $m\ge 1$, since

$$\mu_n K_n^m(y) = \sum_{i=0}^n\sum_{x\in C_i}\mu_n(x)K_n^m(x,y) = \sum_{i=0}^n\sum_{z\in C_i}\mu_n(z)K_n^m(z,\sigma y) = \mu_n K_n^m(\sigma y),$$

for all $y\in\mathcal{X}_n$ and $\sigma\in S_n$. As a consequence, we obtain that, for $y\in C_j$,

$$\frac{\mu_n K_n^m(y)}{\pi_n(y)} = \frac{\mu_n K_n^m(C_j)}{\nu_n(j)} = \frac{1}{\nu_n(j)}\sum_{i=0}^n\sum_{x\in C_i}\mu_n(x)L_n^m(i,j) = \frac{\rho_n L_n^m(j)}{\nu_n(j)}, \tag{5.18}$$

which leads to the following lemma.

Lemma 5.4. *Fix $n\in\mathbb{N}$ and let P_n be a probability on $\mathcal{X}_n := (\mathbb{Z}_2)^n$ satisfying (5.8). Refer to the setting in (5.7) and (5.11) and set $\pi_n(x)=2^{-n}$ for $x\in\mathcal{X}_n$ and $\nu_n(i)=\binom{n}{i}2^{-n}$ for $0\le i\le n$. Then, for $1\le p\le\infty$ and $i_0\in\{0,n\}$,*

$$\max_{x\in\mathcal{X}_n}\left\|\frac{K_n^m(x,\cdot)}{\pi_n}-1\right\|_p = \max_{0\le i\le n}\left\|\frac{L_n^m(i,\cdot)}{\nu_n}-1\right\|_p = \left\|\frac{L_n^m(i_0,\cdot)}{\nu_n}-1\right\|_p, \quad \forall m\ge 0.$$

In particular, when $p=2$,

$$\max_{0\le i\le n}\left\|\frac{L_n^m(i,\cdot)}{\nu_n}-1\right\|_2 = \left(\sum_{j=1}^n\binom{n}{j}|\beta_{n,j}|^{2m}\right)^{1/2},$$

where $\beta_{n,j}$ is the constant in (5.17).

Proof: We prove this lemma with $1\le p<\infty$, while the case $p=\infty$ can be dealt with in a similar way and is omitted. Fix $p\in[1,\infty)$. First, observe that

$$\left\|\frac{K_n^m(x,\cdot)}{\pi_n}-1\right\|_p^p = \sum_{y\in\mathcal{X}_n}\left|P_n^{(m)}(y-x)2^n-1\right|^p 2^{-n} = \sum_{y\in\mathcal{X}_n}\left|P_n^{(m)}(y)2^n-1\right|^p 2^{-n},$$

where $P_n^{(m+1)}(y) := \sum_{z\in\mathcal{X}_n}P_n^{(m)}(z)P_n(y-z)$. This implies that $\|K_n^m(x,\cdot)/\pi_n-1\|_p$ is constant in x. Second, by the Minkowski inequality, one has

$$\left\|\frac{\mu_n K_n^m}{\pi_n}-1\right\|_p \le \sum_{x\in\mathcal{X}_n}\mu_n(x)\left\|\frac{K_n^m(x,\cdot)}{\pi_n}-1\right\|_p = \left\|\frac{K_n^m(\mathbf{0},\cdot)}{\pi_n}-1\right\|_p,$$

for any probability μ_n on $\mathcal{X}_n$. Besides, the replacement of μ_n and ρ_n with $|C_i|^{-1}\mathbf{1}_{C_i}$ and $\mathbf{1}_{\{i\}}$ in (5.18) yields

$$\left\|\frac{L_n^m(i,\cdot)}{\nu_n}-1\right\|_p^p = \sum_{j=0}^n\sum_{y\in C_j}\left|\frac{L_n^m(i,j)}{\nu_n(j)}-1\right|^p\pi_n(y) = \left\|\frac{\mu_n K_n^m}{\pi_n}-1\right\|_p^p.$$

As a result, we obtain that, for $(x_0, i_0) \in \{(\mathbf{0}, 0), (\mathbf{1}, n)\}$,

$$\left\| \frac{K_n^m(x_0, \cdot)}{\pi_n} - 1 \right\|_p = \left\| \frac{L_n^m(i_0, \cdot)}{\nu_n} - 1 \right\|_p \leq \max_{0 \leq i \leq n} \left\| \frac{L_n^m(i, \cdot)}{\nu_n} - 1 \right\|_p$$
$$\leq \left\| \frac{K_n^m(\mathbf{0}, \cdot)}{\pi_n} - 1 \right\|_p = \left\| \frac{K_n^m(x_0, \cdot)}{\pi_n} - 1 \right\|_p,$$

as desired. □

Remark 5.5. By setting $P_n(C_0) = 1/(2n+1)$ and $P_n(C_1) = 2n/(2n+1)$, one may derive from Lemma 5.4 and (5.16) that $|\psi_{n,1}(i)| \leq |\psi_{n,1}(0)|$ for all $0 < i \leq n$. In fact, such an observation also holds for $\psi_{n,j}$ with $j > 1$, since

$$|\psi_{n,j}(i)| \leq \sum_{k=0}^{j} \binom{i}{k} \binom{n-i}{j-k} = \binom{n}{j} = \psi_{n,j}(0).$$

Now, we consider the case of $P_n(C_1) = 1$ for an illustration. First, one can see that

$$L_n(i, i+1) = 1 - \frac{i}{n}, \quad L_n(i+1, i) = \frac{i+1}{n}, \quad \forall 0 \leq i < n, \tag{5.19}$$

which is the transition matrix of the well-known Ehrenfest chain and has eigenvalues $\beta_{n,j} = 1-2j/n$ for $0 \leq j \leq n$. To remove the periodicity of L_n, we consider its δ-lazy variant with $\delta \in (0,1)$ and write it as $L_{n,\delta}$. Note that $L_{n,\delta}$ is associated with probability $P_{n,\delta}$ on $\mathcal{X}_n$, where $P_{n,\delta}(C_0) = \delta$ and $P_{n,\delta}(C_1) = 1 - \delta$, and has eigenvalues

$$\beta_{n,j}^{(\delta)} = \delta + (1-\delta)\beta_{n,j} = 1 - \frac{2(1-\delta)j}{n}, \quad \forall 0 \leq j \leq n.$$

Clearly, one has

$$\beta_{n,j}^{(\delta)} \geq \beta_{n,j+1}^{(\delta)}, \quad \forall 0 \leq j < n, \quad 1 - \beta_{n,j}^{(\delta)} < \beta_{n,n-j}^{(\delta)} - (-1), \quad \forall 1 \leq j < n/2.$$

This implies $|\beta_{n,n-j}^{(\delta)}| \leq |\beta_{n,j}^{(\delta)}| = \beta_{n,j}^{(\delta)}$ for $1 \leq j < n/2$. Let $d_{n,2}^{(\delta)}$ be the max-ℓ^2-distance of $(\mathcal{X}_n, L_{n,\delta}, \nu_n)$. By Lemma 5.4, we have

$$D_{n,\delta}(m) \leq \left[d_{n,2}^{(\delta)}(m)\right]^2 = \sum_{j=1}^{n} \binom{n}{j} \left|\beta_{n,j}^{(\delta)}\right|^{2m} \leq 2D_{n,\delta}(m), \quad \forall m \geq 0, \tag{5.20}$$

where

$$D_{n,\delta}(m) = |1-2\delta|^{2m} + \sum_{1 \leq j \leq n/2} \binom{n}{j} \left[\beta_{n,j}^{(\delta)}\right]^{2m}. \tag{5.21}$$

Set $T(D_{n,\delta}, \epsilon) = \min\{m \geq 0 | D_{n,\delta}(m) \leq \epsilon\}$. As a consequence, if $T_{n,2}^{(\delta)}$ is the max-ℓ^2-mixing time of $(\mathcal{Y}_n, L_{n,\delta}, \nu_n)$, then

$$T(D_{n,\delta}, \epsilon^2) \leq T_{n,2}^{(\delta)}(\epsilon) \leq T(D_{n,\delta}, \epsilon^2/2), \quad \forall \epsilon > 0.$$

To estimate $T(D_{n,\delta}, \epsilon)$, let $N = \lfloor n/2 \rfloor + 1$ and write $D_{n,\delta}(m) = \sum_{i=1}^{N} a_i b_i^{2m}$, where $b_1, ..., b_N$ be a rearrangement of $\{|1-2\delta|, \beta_{n,j}^{(\delta)}, 1 \leq j \leq n/2\}$ in a decreasing order and $a_1, ..., a_N$ be the corresponding rearrangement of $\{1, \binom{n}{j}, 1 \leq j \leq n/2\}$. By Lemma 1.1, we have

$$\tau_{n,\delta} \leq T(D_{n,\delta}, \epsilon) \leq \frac{6\tau_{n,\delta}}{\epsilon^2} + 1, \quad \forall \epsilon \in (0, 1/2),$$

where

$$\tau_{n,\delta} = \max_{1 \leq i \leq N} \frac{\log(a_1 + \cdots + a_i)}{-2\log b_i}. \tag{5.22}$$

Suppose $b_{i_0} = |1-2\delta|$. Note that $\sum_{i=1}^{j}\binom{n}{i} \le n^j$ for $1 \le j \le n$. This implies

$$a_1 + \cdots + a_i \le \begin{cases} n^i & \forall 1 \le i < i_0, \\ n^{i-1}+1 & \forall i_0 \le i \le N, \end{cases}$$

and then, by the fact of $\log(1+t) \le \log t + 1/t$ for $t > 0$,

$$\frac{\log(1+a_1+\cdots+a_i)}{-2\log b_i} \le \frac{n(\log n + 1)}{4(1-\delta)}, \quad \forall 1 \le i \le N,\ i \ne i_0.$$

For $i = i_0$, if $i_0 > 1$, then

$$\frac{\log(1+a_1+\cdots+a_{i_0})}{-2\log b_{i_0}} \le \frac{\log(n^{i_0-1}+2)}{-2\log \beta^{(\delta)}_{n,i_0-1}} \le \frac{n(\log n + 2)}{4(1-\delta)}.$$

If $i_0 = 1$, then $[\log(1+a_1)]/[-2\log b_1] = (\log 2)/(-2\log|1-2\delta|)$. For the lower bound of $\tau_{n,\delta}$, observe that $\tau_{n,\delta} \ge [\log(1+a_1)]/[-2\log b_1]$. By treating the cases of $i_0 = 1$ and $i_0 > 1$ respectively, one may achieve $\tau_{n,\delta} \ge (\log 2)/(-2\log|1-2\delta|)$ and

$$\tau_{n,\delta} \ge \frac{\log(n+1)}{-2\log \beta^{(\delta)}_{n,1}} \ge \frac{n\log n}{4(1-\delta)}\left(1 - \frac{2(1-\delta)}{n}\right) \ge \frac{(n-2)\log n}{4(1-\delta)},$$

where the last inequality uses the fact of $\log t \ge (t-1)/t$ for $t > 0$. We summarize the above discussions in the following lemma.

Lemma 5.6. *For $n \ge 1$ and $\delta \in (0,1)$, let L_n be the transition matrix in (5.19) and $L_{n,\delta}$ be its δ-lazy variant. Let $T^{(\delta)}_{n,2}$ be the max-ℓ^2-mixing time of the Markov chain with transition matrix $L_{n,\delta}$. Then, $A_{n,\delta} \le T^{(\delta)}_{n,2}(\epsilon) \le 24\epsilon^{-4}B_{n,\delta} + 1$ for $\epsilon \in (0, 1/\sqrt{2})$, where*

$$A_{n,\delta} = \frac{(n-2)\log n}{4(1-\delta)} \vee \frac{\log 2}{-2\log|1-2\delta|}, \quad B_{n,\delta} = \frac{n(\log n+2)}{4(1-\delta)} \vee \frac{\log 2}{-2\log|1-2\delta|}.$$

Remark 5.7. In fact, what has been proved above is $A_{n,\delta} \le \tau_{n,\delta} \le B_{n,\delta}$. Immediately, this implies

$$\left|\tau_{n,\delta} - \frac{n\log n}{4(1-\delta)} \vee \frac{\log 2}{-2\log|1-2\delta|}\right| \le \frac{n}{2(1-\delta)}.$$

Next, we explore the max-ℓ^2-cutoff of $\mathcal{F} := (\mathcal{Y}_n, L_{n,\delta_n}, \nu_n)_{n=1}^\infty$, where $(\delta_n)_{n=1}^\infty$ is a sequence in $(0,1)$. By (5.20), the max-ℓ^2-cutoff of $\mathcal{F}$ is equivalent to the cutoff of $(D_{n,\delta_n})_{n=1}^\infty$. Furthermore, when a cutoff time is specified, both families share the same cutoff window. Set $\lambda_n = \min\{-\log|1-2\delta_n|, -\log \beta^{(\delta_n)}_{n,1}\}$. Clearly, one has

$$\lambda_n \asymp \ell_n := \min\{-\log|1-2\delta_n|, (1-\delta_n)/n\} \tag{5.23}$$

and, by Remark 5.7,

$$\begin{aligned}\tau_{n,\delta_n}\lambda_n &\asymp \max\left\{\frac{n\log n}{1-\delta_n}, \frac{1}{-\log|1-2\delta_n|}\right\}\min\left\{-\log|1-2\delta_n|, \frac{1-\delta_n}{n}\right\} \\ &= \left(\frac{(n\log n)(-\log|1-2\delta_n|)}{1-\delta_n} \vee 1\right) \wedge \left(\log n \vee \frac{1-\delta_n}{n(-\log|1-2\delta_n|)}\right).\end{aligned}$$

This implies

$$\tau_{n,\delta_n}\lambda_n \to \infty \quad \Leftrightarrow \quad C_n := \frac{(n\log n)(-\log|1-2\delta_n|)}{1-\delta_n} \to \infty.$$

Note that $-\log|1-2\delta_n| \ge 1 - |1-2\delta_n|$. If $\delta_n \in [1/2, 1)$, then $C_n \ge 2n\log n \to \infty$. If $\delta_n \in (0, 1/2)$, then $2\delta_n n\log n \le C_n \le (2\delta_n n\log n)/(1-2\delta_n)$. In particular, when $\liminf_n \delta_n > 0$, one has $C_n \to \infty$.

When $\delta_n \to 0$, one obtains that $C_n \to \infty$ if and only if $\delta_n n \log n \to \infty$. As a result, we may conclude from Theorem 2.6 and the above discussions that

$$(D_{n,\delta_n})_{n=1}^{\infty} \text{ has a cutoff} \quad \Leftrightarrow \quad \delta_n n \log n \to \infty.$$

To see a cutoff time, assume that $\delta_n n \log n \to \infty$ and set $t_n = (n \log n)/[4(1-\delta_n)]$. As $C_n \to \infty$, $|1-2\delta_n|^{2t_n} \to 0$. In addition with the following computations,

$$|\beta_{n,1}^{(\delta_n)}|^{2t_n} = \exp\left\{\frac{2n\log n}{4(1-\delta_n)} \cdot \frac{-2(1-\delta_n)}{n}\left(1+O\left(\frac{1}{n}\right)\right)\right\} \sim \frac{1}{n},$$

and

$$|\beta_{n,j}^{(\delta_n)}|^{2t_n} \le \exp\left\{\frac{2n\log n}{4(1-\delta_n)} \cdot \frac{-2j(1-\delta_n)}{n}\right\} = \frac{1}{n^j}, \quad \forall 1 \le j \le \frac{n}{2}.$$

we obtain

$$\liminf_{n\to\infty} D_{n,\delta_n}(\lceil t_n \rceil) \ge \liminf_{n\to\infty} n|\beta_{n,1}^{(\delta_n)}|^{2(t_n+1)} = 1,$$

and

$$\limsup_{n\to\infty} D_{n,\delta_n}(\lceil t_n \rceil) \le \limsup_{n\to\infty} D_{n,\delta_n}(t_n) \le \lim_{n\to\infty} \sum_{j=1}^{\lfloor n/2 \rfloor} \binom{n}{j} n^{-j} \le e-1.$$

By Corollary 2.7, $(D_{n,\delta_n})_{n=1}^{\infty}$ has a $(t_n, \ell_n^{-1} \vee 1)$ cutoff, where ℓ_n is the constant in (5.23). Note that

$$\ell_n = \frac{1-\delta_n}{n}, \quad \text{when } \delta_n \ge \frac{1}{2n}; \quad \delta_n \le \ell_n \le \frac{2\delta_n}{1-2\delta_n}, \quad \text{when } 0 < \delta_n < \frac{1}{2n}.$$

Both cases combine to $\ell_n \asymp \max\{(1-\delta_n)/n, \delta_n\} \le 1$. Consequently, we achieve the following theorem.

Theorem 5.8. *Let $\mathcal{F} = (\mathcal{Y}_n, L_{n,\delta_n}, \nu_n)_{n=1}^{\infty}$, where $\mathcal{Y}_n = \{0,1,...,n\}$, $\nu_n(i) = \binom{n}{i}2^{-n}$ for $i \in \mathcal{Y}_n$ and L_{n,δ_n} is the δ_n-lazy variant of L_n in (5.19) with $\delta_n \in (0,1)$. Then, $\mathcal{F}$ has a max-ℓ^2-cutoff if and only if $\delta_n n \log n$ tends to infinity. Moreover, if $\delta_n n \log n \to \infty$, then $\mathcal{F}$ has a (t_n, b_n) max-ℓ^2-cutoff, where*

$$t_n = \frac{n\log n}{4(1-\delta_n)}, \quad b_n = \max\left\{\frac{n}{1-\delta_n}, \frac{1}{\delta_n}\right\}.$$

5.3. *A Metropolis chain on the hypercube.* A Metropolis algorithm is a widely used procedure for sampling a specified probability on a finite set when a direct simulation of i.i.d. samples is not feasible. Such an algorithm was introduced by Metropolis et al. (1953) and is implemented with a targeted probability π on a finite set $\mathcal{X}$ and an irreducible Markov chain $(\mathcal{X}, K)$, which is called a base chain, satisfying $K(x,y) > 0$ if and only if $K(y,x) > 0$. In detail, a Metropolis chain associated with (K,π) is a Markov chain with transition matrix M given by

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

where $A(x,y) := \pi(y)K(y,x)/[\pi(x)K(x,y)]$ and is called the acceptance ratio. In general, π need not be a stationary distribution of K. It is easy to see that $(\mathcal{X}, M, \pi)$ is irreducible, reversible and mostly aperiodic. Thus, $M^m(x,y)$ converges to $\pi(y)$ for all $x, y \in \mathcal{X}$, as desired. We refer the reader to Diaconis and Saloff-Coste (1998) for more discussions on this subject.

Here, we consider a Metropolis chain introduced by Diaconis and Hanlon (1992). Let K_n be the matrix in (5.7) with $P_n(e_{n,i}) = 1/n$ for $1 \le i \le n$, where $e_{n,i} \in \mathcal{X}_n = (\mathbb{Z}_2)^n$ has entries 0 except the ith one which is 1. The targeted probability is

$$\pi_{n,\theta}(x) = \frac{\theta^{H(\mathbf{0},x)}}{(1+\theta)^n}, \quad \forall x \in \mathcal{X}_n,$$

where $\theta \in (0,1)$ and H is the Hamming distance as before. Then, the Metropolis chain associated with $(K_n, \pi_{n,\theta})$ has the following transition matrix.

$$M_{n,\theta}(x,y) = \begin{cases} 1/n & \text{if } H(x,y)=1,\ H(\mathbf{0},y) < H(\mathbf{0},x), \\ \theta/n & \text{if } H(x,y)=1,\ H(\mathbf{0},y) > H(\mathbf{0},x), \\ (1-\theta)[1-H(\mathbf{0},x)/n] & \text{if } H(x,y)=0, \\ 0 & \text{otherwise.} \end{cases} \tag{5.24}$$

Clearly, $M_{n,\theta}$ is aperiodic. Note that one may write

$$M_{n,\theta} = \frac{1}{n}\sum_{i=1}^{n} I_1 \otimes \cdots \otimes I_{i-1} \otimes K_\theta \otimes I_{i+1} \otimes \cdots \otimes I_n,$$

where I_j is the 2×2 identity matrix for $1 \le j \le n$, K_θ is a transition matrix on $\{0,1\}$ with

$$K_\theta(1,0) = 1, \quad K_\theta(0,1) = \theta,$$

and $A \times B$ denotes the tensor product of matrices A and B. A straightforward computation says that $\pi_\theta := (1/(1+\theta), \theta/(1+\theta))$ is the stationary distribution of K_θ and $1, -\theta$ are eigenvalues of K_θ with $\ell^2(\pi_\theta)$-orthonormal eigenvectors $(1,1)$ and $(\theta^{1/2}, -\theta^{-1/2})$. As a result of this observation, the eigenvalues and $\ell^2(\pi_{n,\theta})$-orthonormal eigenvectors of $M_{n,\theta}$ are given by

$$\beta_{n,\theta,x} = 1 - \frac{(1+\theta)H(\mathbf{0},x)}{n}, \quad \phi_{n,\theta,x}(y) = (-1)^{\langle x,y\rangle}\theta^{\frac{1}{2}H(\mathbf{0},x)-\langle x,y\rangle},$$

for all $x, y \in \mathcal{X}_n$. Let $\mathbf{1}$ be the n-vector with entries 1. Obviously, one can see that

$$\phi^2_{n,\theta,x}(y) = \theta^{H(\mathbf{0},x)-2\langle x,y\rangle} \le \theta^{H(\mathbf{0},x)-2\langle x,\mathbf{1}\rangle} = \phi^2_{n,\theta,x}(\mathbf{1}), \quad \forall x, y \in \mathcal{X}_n,$$

which leads to

$$\max_{y\in\mathcal{X}_n}\left\|\frac{M^m_{n,\theta}(y,\cdot)}{\pi_{n,\theta}} - 1\right\|_2^2 = \left\|\frac{M^m_{n,\theta}(\mathbf{1},\cdot)}{\pi_{n,\theta}} - 1\right\|_2^2 = \sum_{i=1}^{n}\binom{n}{i}\theta^{-i}\left|1 - \frac{(1+\theta)i}{n}\right|^{2m}.$$

As before, set $C_i = \{x \in \mathcal{X}_n | H(\mathbf{0},x) = i\}$ for $0 \le i \le n$ and $\mathcal{Y}_n = \{0,1,...,n\}$. From (5.24), one may check that the properties of (5.9) and (5.10) also hold for $M_{n,\theta}$. This implies that the Markov chain on the clustering sets $C_0, ..., C_n$, which are briefly written as $0, ..., n$, associated with $M_{n,\theta}$ has transition matrix

$$N_{n,\theta}(i,j) := \sum_{y\in C_j} M_{n,\theta}(x,y) = \begin{cases} \frac{i}{n} & \text{if } j = i-1, \\ (1-\frac{i}{n})\theta & \text{if } j = i+1, \\ (1-\frac{i}{n})(1-\theta) & \text{if } j = i, \end{cases} \tag{5.25}$$

for all $i, j \in \mathcal{Y}_n$ and $x \in C_i$, and stationary distribution

$$\nu_{n,\theta}(i) = \pi_{n,\theta}(C_i) = \binom{n}{i}\theta^i(1+\theta)^{-n}, \quad \forall 0 \le i \le n. \tag{5.26}$$

Moreover, (5.18) also holds under the replacement of K_n, L_n, π_n, ν_n with $M_{n,\theta}$, $N_{n,\theta}$, $\pi_{n,\theta}$, $\nu_{n,\theta}$, which yields

$$\left\|\frac{\mu_n M^m_{n,\theta}}{\pi_{n,\theta}} - 1\right\|_2 = \left\|\frac{\rho_n N^m_{n,\theta}}{\nu_{n,\theta}} - 1\right\|_2,$$

where μ_n is a probability on $\mathcal{X}_n$ uniform over each clustering set and $\rho_n(i) = \mu_n(C_i)$ for $0 \le i \le n$. As a result, we obtain

$$\max_{0\le j\le n}\left\|\frac{N^m_{n,\theta}(j,\cdot)}{\nu_{n,\theta}} - 1\right\|_2^2 = \left\|\frac{N^m_{n,\theta}(n,\cdot)}{\nu_{n,\theta}} - 1\right\|_2^2 = \sum_{i=1}^{n}\binom{n}{i}\theta^{-i}\left|1 - \frac{(1+\theta)i}{n}\right|^{2m}.$$

Remark 5.9. Here is another proof of the first equality in the above. Note that the conclusions of (5.12) and (5.13) also hold for $\phi_{n,\theta,x}(y)$ and $\beta_{n,\theta,y}$. This implies that the following terms,

$$\psi_{n,\theta,i}(k) := \sum_{x\in C_i} \phi_{n,\theta,x}(y) = \sum_{j=0}^{i}(-1)^j \binom{k}{j}\binom{n-k}{i-j}\theta^{i/2-j}, \quad \beta_{n,\theta,k} := \beta_{n,\theta,y},$$

where $y \in C_k$, are well-defined. Furthermore, the computations in (5.14) and (5.15) also hold under the replacement of L_n, $\psi_{n,i}$, $\nu_n, \beta_{n,j}$, $\phi_{n,x}$ with $N_{n,\theta}$, $\psi_{n,\theta,i}$, $\nu_{n,\theta}$, $\beta_{n,\theta,j}$, $\phi_{n,\theta,x}$. As before, we obtain

$$\left\|\frac{N^m_{n,\theta}(j,\cdot)}{\nu_{n,\theta}} - 1\right\|_2^2 = \sum_{i=1}^{n}\binom{n}{i}^{-1}|\psi_{n,\theta,i}(j)|^2|\beta_{n,\theta,i}|^{2m}.$$

The desired equality is then given by the following inequality,

$$|\psi_{n,\theta,i}(k)| \le \sum_{j=0}^{i}\binom{k}{j}\binom{n-k}{i-j} = \binom{n}{i}\theta^{-i/2} = |\psi_{n,\theta,i}(n)|.$$

Now, let's examine the max-ℓ^2-cutoff for $\mathcal{F} = (\mathcal{Y}_n, L_{n,\theta_n}, \nu_{n,\theta_n})_{n=1}^\infty$, where θ_n is a sequence in $(0,1)$. Let $d_{n,2}$ be the max-ℓ^2-distance of $(\mathcal{Y}_n, L_{n,\theta_n}, \nu_{n,\theta_n})$. Similar to (5.20), one may derive

$$D_n(m) \le d^2_{n,2}(m) \le 2D_n(m), \quad \forall m \ge 0,$$

where

$$D_n(m) = \theta_n^{2m-n} + \sum_{i=1}^{\lfloor n/2\rfloor}\binom{n}{i}\theta_n^{-i}\left|1 - \frac{(1+\theta_n)i}{n}\right|^{2m}.$$

This implies that the max-ℓ^2-cutoff for $\mathcal{F}$ is equivalent to the cutoff for $(D_n)_{n=1}^\infty$ and both families share the same cutoff time and cutoff window if any. First, we consider the case of $n(1-\theta_n) = O(1)$. Set $C = \sup_n n(1-\theta_n)$ and let $D_{n,\delta}$ be the function defined in (5.21). Note that

$$\theta_n^{-1}D_{n,(1-\theta_n)/2}(m) \le D_n(m) \le \theta_n^{-n}D_{n,(1-\theta_n)/2}(m) \le e^{C/\theta_n}D_{n,(1-\theta_n)/2}(m).$$

Since θ_n converges to 1, the cutoffs for $(D_n)_{n=1}^\infty$ and $(D_{n,(1-\theta_n)/2})_{n=1}^\infty$ are equivalent. By Theorem 5.8, when $n(1-\theta_n) = O(1)$, $\mathcal{F}$ has a max-ℓ^2-cutoff if and only if $(1-\theta_n)n\log n$ tends to infinity. Furthermore, if $(1-\theta_n)n\log n \to \infty$, then $\mathcal{F}$ has a $(t_n, (1-\theta_n)^{-1})$ max-ℓ^2-cutoff with $t_n = (n\log n)/[2(1+\theta_n)]$.

Next, we consider the case of $n(1-\theta_n) \to \infty$. For convenience, set

$$c_{n,i} = 1 - (1+\theta_n)i/n, \quad \forall 1 \le i \le n/2, \quad \widetilde{D}_n(m) = \sum_{i=1}^{\lfloor n/2\rfloor}\binom{n}{i}\theta_n^{-i}c_{n,i}^{2m}.$$

Then, $D_n(m) = \widetilde{D}_n(m) + \theta_n^{2m-n}$. To see the cutoff for $(\widetilde{D}_n)_{n=1}^\infty$, set

$$\tau_n = \max_{1\le j\le n/2}\frac{\log\left(1 + \sum_{i=1}^{j}\binom{n}{i}\theta_n^{-i}\right)}{-2\log c_{n,j}}, \quad \lambda_n = -\log c_{n,1}.$$

Note that, for $1 \le j \le n/2$,

$$\sum_{i=1}^{j}\binom{n}{i}\theta_n^{-i} \le \sum_{i=1}^{j}\frac{1}{i!}\left(\frac{n}{\theta_n}\right)^i \le \left(\frac{n}{\theta_n}\right)^j\sum_{i=1}^{j}\frac{1}{i!n^{j-i}} \le \left(\frac{n}{\theta_n}\right)^j.$$

This implies

$$\log\left(1 + \sum_{i=1}^{j}\binom{n}{i}\theta_n^{-i}\right) \le \log\left(1 + \left(\frac{n}{\theta_n}\right)^j\right) \le j\log\left(1 + \frac{n}{\theta_n}\right).$$

Moreover, using the fact that $0 < 1 - a - b \le (1-a)(1-b)$ for $a, b \in (0,1)$ and $a + b < 1$, one may derive $c_{n,j} \le c_{n,1}^j$, which leads to

$$\tau_n = \max_{1\le j\le n/2} \frac{\log\left(1 + \sum_{i=1}^j \binom{n}{i}\theta_n^{-i}\right)}{-2\log c_{n,j}} \le \frac{\log(1+n/\theta_n)}{-2\log c_{n,1}} \le \tau_n.$$

Consequently, we obtain

$$\lambda_n \sim \frac{1+\theta_n}{n} \asymp \frac{1}{n}, \quad \tau_n\lambda_n \sim \frac{1}{2}\log\frac{n}{\theta_n} \to \infty, \quad \tau_n \sim \frac{n\log(n/\theta_n)}{2(1+\theta_n)} \to \infty.$$

By Theorem 2.6, $(\widetilde{D}_n)_{n=1}^\infty$ has a cutoff. In addition with the following computations,

$$\lim_{n\to\infty} \theta_n^{2\lceil c\tau_n\rceil - n} \le \lim_{n\to\infty} \exp\left\{-n(1-\theta_n)\left(\frac{c\log(1+n/\theta_n)}{n\lambda_n} - 1\right)\right\} = 0, \quad \forall c > 0,$$

and

$$\limsup_{n\to\infty} \widetilde{D}_n(\lceil\tau_n\rceil) \le \limsup_{n\to\infty} \sum_{i=1}^{\lfloor n/2\rfloor} \frac{(n/\theta_n)^i c_{n,1}^{2i\tau_n}}{i!} \le e - 1,$$

and

$$\liminf_{n\to\infty} \widetilde{D}_n(\lceil\tau_n\rceil) \ge \liminf_{n\to\infty} (n/\theta_n) c_{n,1}^{2(\tau_n+1)} = 1,$$

one may conclude from Corollary 2.7 that $(D_n)_{n=1}^\infty$ has a (τ_n, n) cutoff. We summarize the above discussion in the following theorem.

Theorem 5.10. *Refer to* (5.25) *and* (5.26). *The family* $\mathcal{F} = (\mathcal{Y}_n, N_{n,\theta_n}, \nu_{n,\theta_n})_{n=1}^\infty$ *has a max-*ℓ^2*-cutoff if and only if* $(1-\theta_n)n\log n \to \infty$. *Furthermore, if* $(1-\theta_n)n\log n \to \infty$, *then* $\mathcal{F}$ *has a* (s_n, b_n) *max-*ℓ^2*-cutoff with*

$$s_n = \frac{\log(n/\theta_n)}{-2\log(1-(1+\theta_n)/n)}, \quad b_n = n \vee \frac{1}{1-\theta_n}.$$

Proof: The equivalent condition of the max-ℓ^2-cutoff has been proved before. Moreover, assuming $(1-\theta_n)n\log n \to \infty$, $\mathcal{F}$ has a $(t_n, (1-\theta_n)^{-1})$ max-ℓ^2-cutoff when $n(1-\theta_n) = O(1)$ and has a (τ_n, n) max-ℓ^2-cutoff when $n(1-\theta_n) \to \infty$, where

$$t_n = \frac{n\log n}{2(1+\theta_n)}, \quad \tau_n = \frac{\log(1+n/\theta_n)}{-2\log(1-(1+\theta_n)/n)}.$$

The desired cutoff time is then given by the fact of

$$s_n = \begin{cases} \tau_n + O(1) & \text{when } n(1-\theta_n) \to \infty, \\ t_n + O(\log n) & \text{when } n(1-\theta_n) = O(1). \end{cases}$$

□

Remark 5.11. Note that, in Theorem 5.10, $s_n \sim r_n := [n\log(n/\theta_n)]/[2(1+\theta_n)]$. It seems that r_n looks simpler than s_n, but taking r_n to be the cutoff time could enlarge the corresponding cutoff window since $r_n - s_n \sim (1/4)\log(n/\theta_n)$ and the latter can be of order bigger than n, e.g. $\theta_n = \exp\{-n^2\}$.

Appendix A. Proofs for Section 2

This section is dedicated to proving the theoretical framework created in Section 2. We shall derive the proofs of Propositions 2.1 and 2.4 and Theorem 2.6 in order.

A.1. *Proofs for Proposition 2.1.* We first give some basic properties of $\mathcal{L}_{\mathcal{W}}$ and $T_{\mathcal{W}}$.

Lemma A.1. *Let $\mathcal{W} \subset \mathcal{V}$.*

(1) $T_{\mathcal{W}}(\epsilon) = \sup\{T_V(\epsilon) | V \in \mathcal{W}\}$ *for* $\epsilon > 0$.
(2) *If* $T_{\mathcal{W}}(\epsilon_0) < \infty$, *then* $T_{\mathcal{W}}(\epsilon_0) = \min\{t \geq 0 | \mathcal{L}_{\mathcal{W}}(t) \leq \epsilon_0\}$.
(3) *If* $\mathcal{L}_{\mathcal{W}}(t_0) < \infty$, *then* $\mathcal{L}_{\mathcal{W}}$ *is continuous on* $[t_0, \infty)$.

Proof: For (1), it is clear that $T_{\mathcal{W}}(\epsilon) \geq \sup\{T_V(\epsilon)|V \in \mathcal{W}\}$. To see the equality, we consider the following two cases. When $T_{\mathcal{W}}(\epsilon) = \infty$, $\mathcal{L}_{\mathcal{W}}(t) > \epsilon$ for all $t > 0$. To this case, one may choose, for each $n \in \mathbb{N}$, some $V_n \in \mathcal{W}$ such that $\mathcal{L}_{\mathcal{W}}(n) \geq \mathcal{L}_{V_n}(n) > \epsilon$. This implies $T_{V_n}(\epsilon) \geq n$ and, hence, $\sup\{T_V(\epsilon)|V \in \mathcal{W}\} = \infty$. When $T_{\mathcal{W}}(\epsilon) < \infty$, we may select, for each $\eta \in (0, T_{\mathcal{W}}(\epsilon))$, some $V_\eta \in \mathcal{W}$ such that $\mathcal{L}_{\mathcal{W}}(T_{\mathcal{W}}(\epsilon) - \eta) \geq \mathcal{L}_{V_\eta}(T_{\mathcal{W}}(\epsilon) - \eta) > \epsilon$, since $\mathcal{L}_{\mathcal{W}}(T_{\mathcal{W}}(\epsilon) - \eta) > \epsilon$. As a consequence, this implies $T_{\mathcal{W}}(\epsilon) - \eta \leq T_{V_\eta}(\epsilon) \leq \sup\{T_V(\epsilon)|V \in \mathcal{W}\}$, which leads to the desired equality.

For (2), set $s = T_{\mathcal{W}}(\epsilon_0)$. Following the definition of $T_{\mathcal{W}}$, one has $\mathcal{L}_{\mathcal{W}}(t) \leq \epsilon_0$ for $t > s$. As $\mathcal{L}_V$ is continuous, we have

$$\mathcal{L}_V(s) = \lim_{t>s, t\to s} \mathcal{L}_V(t) \leq \lim_{t>s, t\to s} \mathcal{L}_{\mathcal{W}}(t) \leq \epsilon_0, \quad \forall V \in \mathcal{W}.$$

This implies $\mathcal{L}_{\mathcal{W}}(s) \leq \epsilon_0$.

To prove the continuity of $\mathcal{L}_{\mathcal{W}}$, we recall the following lemma.

Lemma A.2. *(Chen and Saloff-Coste, 2010, Corollary 3.3) Let $V_n \in \mathcal{V}$ and assume* $\sup_n \mathcal{L}_{V_n}(0) < \infty$. *For any sequence $t_n > 0$, the following functions*

$$\overline{F}(a) := \limsup_{n\to\infty} \mathcal{L}_{V_n}(at_n), \quad \underline{F}(a) := \liminf_{n\to\infty} \mathcal{L}_{V_n}(at_n).$$

are continuous on $(0, \infty)$. *Further, if* $\overline{F}(a) = 0$ *(resp.* $\underline{F}(a) = 0$*) for some* $a > 0$, *then* $\overline{F}(a) = 0$ *(resp.* $\underline{F}(a) = 0$*) for all* $a > 0$.

By setting $\mathcal{W}' = \{V' \in \mathcal{V} | dV'(\lambda) = e^{-t_0\lambda} dV(\lambda),\, V \in \mathcal{W}\}$, one has $\mathcal{L}_{\mathcal{W}}(t) = \mathcal{L}_{\mathcal{W}'}(t - t_0)$ for $t \geq t_0$. Based on this observation, we may assume without loss of generality that $t_0 = 0$. Let $(r_n)_{n=1}^\infty$ be a sequence consisting of all positive rational numbers. For $n, m \in \mathbb{N}$, select $V_{n,m} \in \mathcal{W}$ such that

$$\mathcal{L}_{\mathcal{W}}(r_n) \leq \mathcal{L}_{V_{n,m}}(r_n) + \frac{1}{m}.$$

Let $\ell_k = k(k+1)/2$ and define

$$V_i = V_{i-\ell_k, \ell_{k+1}-i+1}, \quad \forall \ell_k < i \leq \ell_{k+1},\ k \geq 0.$$

It is easy to check from the above setting that

$$\mathcal{L}_{\mathcal{W}}(t) = \overline{F}(t) := \limsup_{i\to\infty} \mathcal{L}_{V_i}(t), \quad \forall t \in \mathbb{Q} \cap (0, \infty).$$

By Lemma A.2, $\overline{F}$ is continuous on $(0, \infty)$ and, by the monotonicity of $\mathcal{L}_{\mathcal{W}}$, we obtain $\mathcal{L}_{\mathcal{W}} = \overline{F}$. To see that $\mathcal{L}_{\mathcal{W}}$ is continuous at 0, let $\epsilon > 0$ and choose $V \in \mathcal{W}$ such that $\mathcal{L}_{\mathcal{W}}(0) < \mathcal{L}_V(0) + \epsilon$. Note that $\mathcal{L}_{\mathcal{W}}(0) \geq \mathcal{L}_{\mathcal{W}}(t) \geq \mathcal{L}_V(t)$. This implies

$$\mathcal{L}_{\mathcal{W}}(0) \geq \limsup_{t\to 0} \mathcal{L}_{\mathcal{W}}(t) \geq \liminf_{t\to 0} \mathcal{L}_{\mathcal{W}}(t) \geq \mathcal{L}_V(0) \geq \mathcal{L}_{\mathcal{W}}(0) - \epsilon.$$

Letting $\epsilon \to 0$ gives the desired continuity. □

The next two lemmas and their following remarks provide more information of $\lambda_{\mathcal{W}}$ and $\tau_{\mathcal{W}}$, which are crucial to the proofs of Proposition 2.1 and Theorem 2.6.

Lemma A.3. *Let $\mathcal{W} \subset \mathcal{V}$ and ϵ, c, c_1, c_2 be constants in $(0, \mathcal{L}_{\mathcal{W}}(0))$. Suppose $\mathcal{L}_{\mathcal{W}}(0) < \infty$.*

(1) *If* $\tau_{\mathcal{W}}(c) < \infty$, *then* $\mathcal{L}_{\mathcal{W}}(\tau_{\mathcal{W}}(c)) \geq c/(1+c)$ *and, for* $s > 0$,

$$\mathcal{L}_{\mathcal{W}}(\tau_{\mathcal{W}}(c) + s) \leq c + \frac{\tau_{\mathcal{W}}(c) + s}{s e^{s\lambda_{\mathcal{W}}(c)}}. \tag{A.1}$$

(2) *If* $T_{\mathcal{W}}(\epsilon) < \infty$*, then* $\mathcal{L}_{\mathcal{W}}(T_{\mathcal{W}}(\epsilon)) = \epsilon$ *and, for* $r \geq 0$*,* $s > 0$ *and* $c_1 < c_2$*,*

$$\mathcal{L}_{\mathcal{W}}(T_{\mathcal{W}}(\epsilon) + r + s) \leq c_1 + c_2 e^{-(T_{\mathcal{W}}(\epsilon)+r+s)\lambda_{\mathcal{W}}(c_1)} + \frac{\epsilon(T_{\mathcal{W}}(\epsilon) + r + s)}{(T_{\mathcal{W}}(\epsilon) + s)e^{r\lambda_{\mathcal{W}}(c_2)}}. \tag{A.2}$$

Remark A.4. By Lemma A.3, one may derive another lower bound of $\tau_{\mathcal{W}}(c)$ other than (2.3). Set $\beta = \tau_{\mathcal{W}}(c)\lambda_{\mathcal{W}}(c)/\log[1+\tau_{\mathcal{W}}(c)\lambda_{\mathcal{W}}(c)]$ and assume that $\mathcal{L}_{\mathcal{W}}(0) < \infty$, $c \in (0, \mathcal{L}_{\mathcal{W}}(0))$ and $\tau_{\mathcal{W}}(c) < \infty$. By replacing s with $\tau_{\mathcal{W}}(c)/\beta$ in (A.1), we obtain

$$\mathcal{L}_{\mathcal{W}}(c)(\tau_{\mathcal{W}}(c) + s) \leq c + \frac{1+\beta}{1 + \tau_{\mathcal{W}}(c)\lambda_{\mathcal{W}}(c)} \leq c + \frac{2}{\log[1 + \tau_{\mathcal{W}}(c)\lambda_{\mathcal{W}}(c)]}.$$

This implies

$$\tau_{\mathcal{W}}(c) \geq \left(\frac{\beta}{\beta+1}\right) T_{\mathcal{W}}\left(c + \frac{2}{\log[1 + \tau_{\mathcal{W}}(c)\lambda_{\mathcal{W}}(c)]}\right). \tag{A.3}$$

Concerning lower bounds in (2.3) and (A.3), let's assume that $\tau_{\mathcal{W}}(c)\lambda_{\mathcal{W}}(c)$ is large, which happens at the presence of cutoffs. Clearly, α is much smaller than β in this case. By (2.6), the mixing time $T_{\mathcal{W}}$ in (2.3) differs from that in (A.3) by at most some multiple of $1/\lambda_{\mathcal{W}}(c)$. Consequently, when a cutoff exists, (A.3) performs better than the lower bound in (2.3) and finally results in a refined cutoff window. The reader is referred to the proof of Theorem 2.6 for more details of the above discussion.

Proof of Lemma A.3: When $\mathcal{W} = \{V\}$, Lemma A.3 is exactly Chen et al. (2017, Lemma 2.7). For such a special case, (A.1) and (A.2) in fact hold without the restriction of $\epsilon, c, c_1, c_2 \in (0, \mathcal{L}_V(0))$ and their proofs can be generalized from that of Chen et al. (2017, Lemma 2.7), which are skipped here. We shall assume this fact throughout this proof.

By Lemma A.1, $\mathcal{L}_{\mathcal{W}}$ is continuous on $[0, \infty)$. Since $c, \epsilon \in (0, \mathcal{L}_{\mathcal{W}}(0))$, it is clear that $\tau_{\mathcal{W}}(c) > 0$ and $T_{\mathcal{W}}(\epsilon) > 0$. For (1), assume that $\tau_{\mathcal{W}}(c) < \infty$. Let $0 < \delta < \tau_{\mathcal{W}}(c)$ and select $V \in \mathcal{W}$ such that $\tau_{\mathcal{W}}(c) < \tau_V(c) + \delta$. As $\tau_V(c) > \tau_{\mathcal{W}}(c) - \delta > 0$, c must be less than $\mathcal{L}_V(0)$. This implies

$$\mathcal{L}_{\mathcal{W}}(\tau_{\mathcal{W}}(c) - \delta) \geq \mathcal{L}_V(\tau_V(c) - [\tau_V(c) + \delta - \tau_{\mathcal{W}}(c)]) > \mathcal{L}_V(\tau_V(c)) \geq c/(1+c),$$

where the last inequality is given by Chen et al. (2017, Lemma 2.7). By the continuity of $\mathcal{L}_{\mathcal{W}}$ on $[0, \infty)$, letting $\delta \to 0$ leads to the first inequality. To see the second inequality, we fix $s > 0$ and choose, for any $\delta' > 0$, $V' \in \mathcal{W}$ such that $\mathcal{L}_{\mathcal{W}}(\tau_{\mathcal{W}}(c) + s) \leq \mathcal{L}_{V'}(\tau_{\mathcal{W}}(c) + s) + \delta'$. By Chen et al. (2017, Lemma 2.7), one has

$$\mathcal{L}_{V'}(\tau_{\mathcal{W}}(c) + s) \leq \mathcal{L}_{V'}(\tau_{V'}(c) + s) \leq c + \frac{\tau_{V'}(c) + s}{se^{s\lambda_{V'}(c)}} \leq c + \frac{\tau_{\mathcal{W}}(c) + s}{se^{s\lambda_{\mathcal{W}}(c)}}.$$

Letting $\delta' \to 0$ gives (A.1).

For (2), we assume $T_{\mathcal{W}}(\epsilon) < \infty$. Clearly, the first equality is given by the continuity of $\mathcal{L}_{\mathcal{W}}$. For (A.2), let $\delta'' > 0$ and choose $V'' \in \mathcal{W}$ such that $\mathcal{L}_{\mathcal{W}}(T_{\mathcal{W}}(\epsilon) + r + s) \leq \mathcal{L}_{V''}(T_{\mathcal{W}}(\epsilon) + r + s) + \delta''$. Set $u = s + T_{\mathcal{W}}(\epsilon) - T_{V''}(\epsilon) > 0$. Again, by Chen et al. (2017, Lemma 2.7), we have

$$\begin{aligned}\mathcal{L}_{V''}(T_{\mathcal{W}}(\epsilon) + r + s) &= \mathcal{L}_{V''}(T_{V''}(\epsilon) + r + u)\\ &\leq c_1 + c_2 e^{-(T_{V''}(\epsilon)+r+u)\lambda_{V''}(c_1)} + \frac{\epsilon(T_{V''}(\epsilon) + r + u)}{(T_{V''}(\epsilon) + u)e^{r\lambda_{V''}(c_2)}}\\ &= c_1 + c_2 e^{-(T_{\mathcal{W}}(\epsilon)+r+s)\lambda_{V''}(c_1)} + \frac{\epsilon(T_{\mathcal{W}}(\epsilon) + r + s)}{(T_{\mathcal{W}}(\epsilon) + s)e^{r\lambda_{V''}(c_2)}}\\ &\leq c_1 + c_2 e^{-(T_{\mathcal{W}}(\epsilon)+r+s)\lambda_{\mathcal{W}}(c_1)} + \frac{\epsilon(T_{\mathcal{W}}(\epsilon) + r + s)}{(T_{\mathcal{W}}(\epsilon) + s)e^{r\lambda_{\mathcal{W}}(c_2)}}.\end{aligned}$$

Consequently, (A.2) is obtained by letting $\delta'' \to 0$. □

Lemma A.5. *Let* $\mathcal{W} \subset \mathcal{V}$ *and set* $W(\lambda) = \sup\{V(\lambda) | V \in \mathcal{W}\}$*. Assume that* $\mathcal{L}_{\mathcal{W}}(0) < \infty$*.*

(1) *For $c \in (0, \mathcal{L}_{\mathcal{W}}(0))$, one has $\lambda_{\mathcal{W}}(c) < \infty$ and $\tau_{\mathcal{W}}(c) > 0$ and*

$$\lambda_{\mathcal{W}}(c) > 0 \quad \Leftrightarrow \quad \tau_{\mathcal{W}}(c) < \infty. \tag{A.4}$$

(2) *If $\lambda_{\mathcal{W}}(c) > 0$ for all $c \in (0, \mathcal{L}_{\mathcal{W}}(0))$, then $\mathcal{L}_{\mathcal{W}}(t) \to 0$ as $t \to \infty$.*

(3) *W is non-decreasing and*

$$\lim_{\lambda\to\infty} W(\lambda) = \mathcal{L}_{\mathcal{W}}(0), \qquad \lim_{\lambda\to 0^+} W(\lambda) = \inf\{c > 0 | \lambda_{\mathcal{W}}(c) > 0\}. \tag{A.5}$$

and

$$\lambda_{\mathcal{W}}(c) = \lambda_W(c), \quad \tau_{\mathcal{W}}(c) = \tau_W(c), \quad \forall c \in (0, \mathcal{L}_{\mathcal{W}}(0)). \tag{A.6}$$

(4) *(A.5) and (A.6) still hold when W is replaced by $\overline{W}$, where*

$$\overline{W}(\lambda) := \lim_{\delta\to\lambda^+} W(\delta), \quad \forall \lambda > 0.$$

In particular, if $\lambda_{\mathcal{W}}(c) > 0$ for all $c \in (0, \mathcal{L}_{\mathcal{W}}(0))$, then $\overline{W} \in \mathcal{V}$.

Remark A.6. From Lemma A.5(2), if $\mathcal{L}_{\mathcal{W}}(0) < \infty$ and $\lambda_{\mathcal{W}}(c) > 0$ for all $c \in (0, \mathcal{L}_{\mathcal{W}}(0))$, then $\tau_{\mathcal{W}}(c) < \infty$ and $T_{\mathcal{W}}(c) < \infty$ for all $c > 0$. By Lemma A.5(3), $\lambda_{\mathcal{W}}$ and $\tau_{\mathcal{W}}$ are characterized in a more direct way than their definitions in (2.2).

Proof of Lemma A.5: For (1), the first two inequalities follow from the definitions of $\lambda_{\mathcal{W}}$ and $\tau_{\mathcal{W}}$. To see (A.4) and (2), let $V \in \mathcal{V}$. For $0 < c < \mathcal{L}_V(0)$, one may derive from (2.1) that

$$\frac{\log(1+c)}{\lambda_V(c)} \le \tau_V(c) \le \frac{\log(1+\mathcal{L}_V(0))}{\lambda_V(c)}, \quad \mathcal{L}_V(t) \le c + e^{-\lambda_V(c)t}\mathcal{L}_V(0).$$

Note that the above inequalities also hold for $c \ge \mathcal{L}_V(0)$. Taking the supremum over $\mathcal{W}$, we obtain

$$\frac{\log(1+c)}{\lambda_{\mathcal{W}}(c)} \le \tau_{\mathcal{W}}(c) \le \frac{\log(1+\mathcal{L}_{\mathcal{W}}(0))}{\lambda_{\mathcal{W}}(c)}, \quad \mathcal{L}_{\mathcal{W}}(t) \le c + e^{-\lambda_{\mathcal{W}}(c)t}\mathcal{L}_{\mathcal{W}}(0).$$

Immediately, the first two inequalities imply (A.4), while the last inequality yields (2).

For (3), the monotonicity of W is easy to see from its definition. Note that the first limit of (A.5) is given by

$$\begin{aligned}\mathcal{L}_{\mathcal{W}}(0) &= \sup_{V\in\mathcal{W}} \mathcal{L}_V(0) = \sup_{V\in\mathcal{W}} \lim_{\lambda\to\infty} V(\lambda) \le \lim_{\lambda\to\infty} W(\lambda) \\ &= \lim_{\lambda\to\infty} \sup_{V\in\mathcal{W}} V(\lambda) \le \sup_{V\in\mathcal{W}} \mathcal{L}_V(0) = \mathcal{L}_{\mathcal{W}}(0).\end{aligned}$$

As the proof of the second limit of (A.5) requires (A.6), we proceed to show (A.6) next. Fix $c \in (0, \mathcal{L}_{\mathcal{W}}(0))$. To show the first equality of (A.6), set $\alpha = \lambda_W(c)$ and let $\epsilon > 0$. One may check directly from the definition of W that $\alpha < \infty$. Since $W(\alpha+\epsilon) > c$, we may select $V' \in \mathcal{W}$ such that $c < V'(\alpha+\epsilon) \le W(\alpha+\epsilon)$. This implies $\lambda_{\mathcal{W}}(c) \le \lambda'_V(c) \le \alpha+\epsilon$. Letting $\epsilon \to 0$ yields $\lambda_{\mathcal{W}}(c) \le \alpha$. Immediately, if $\alpha = 0$, then $\alpha = \lambda_{\mathcal{W}}(c)$. For the case $\alpha > 0$, let $\epsilon \in (0, \alpha)$. As $W(\epsilon) \le c$, one has $V(\epsilon) \le c$ for all $V \in \mathcal{W}$. Consequently, this implies $\lambda_V(c) \ge \epsilon$ for $V \in \mathcal{W}$ and, hence, $\lambda_{\mathcal{W}}(c) \ge \epsilon$. Letting $\epsilon \to \alpha$ gives $\lambda_{\mathcal{W}}(c) \ge \alpha$, as desired.

To prove the second equality of (A.6), we set $\beta = \tau_W(c)$. As $\lambda_{\mathcal{W}}(c) < \infty$ and $\lambda_{\mathcal{W}}(c) = \lambda_W(c)$, one has

$$\tau_V(c) = \sup_{\lambda\ge\lambda_V(c)} \frac{\log(1+V(\lambda))}{\lambda} \le \sup_{\lambda\ge\lambda_W(c)} \frac{\log(1+W(\lambda))}{\lambda} = \beta, \quad \forall V \in \mathcal{W},$$

which implies $\tau_{\mathcal{W}}(c) \le \beta$, and

$$\beta = \sup_{\lambda\ge\lambda_W(c)} \sup_{V\in\mathcal{W}} \frac{\log(1+V(\lambda))}{\lambda} = \sup_{V\in\mathcal{W}} \sup_{\lambda\ge\lambda_{\mathcal{W}}(c)} \frac{\log(1+V(\lambda))}{\lambda}. \tag{A.7}$$

By (A.4), if $\lambda_{\mathcal{W}}(c) = 0$, then $\tau_{\mathcal{W}}(c) = \infty$, which leads to $\beta = \infty$. If $\lambda_{\mathcal{W}}(c) > 0$, then $\tau_{\mathcal{W}}(c) < \infty$. For this case, one may apply the following computation

$$\begin{aligned}\sup_{\lambda \geq \lambda_{\mathcal{W}}(c)} \frac{\log(1+V(\lambda))}{\lambda} &= \tau_V(c) \vee \left(\sup_{\lambda_{\mathcal{W}}(c) \leq \lambda < \lambda_V(c)} \frac{\log(1+V(\lambda))}{\lambda} \right) \\ &\leq \tau_V(c) \vee \frac{\log(1+c)}{\lambda_{\mathcal{W}}(c)}, \quad \forall V \in \mathcal{W},\end{aligned}$$

and (A.7) to derive $\beta \leq \tau_{\mathcal{W}}(c) \vee [\lambda_{\mathcal{W}}(c)^{-1} \log(1+c)]$. To finish the proof of this part, it remains to show $\tau_{\mathcal{W}}(c)\lambda_{\mathcal{W}}(c) \geq \log(1+c)$ when $\lambda_{\mathcal{W}}(c) > 0$. For $\delta > 0$, choose $V'' \in \mathcal{W}$ such that $\lambda_{\mathcal{W}}(c) \geq \lambda_{V''}(c) - \delta$. Since $\tau_{V''}(c) \leq \tau_{\mathcal{W}}(c) < \infty$, we have $\lambda_{V''}(c) > 0$ and

$$\begin{aligned}\tau_{\mathcal{W}}(c)\lambda_{\mathcal{W}}(c) &\geq \tau_{\mathcal{W}}(c)(\lambda_{V''}(c) - \delta) \geq \tau_{V''}(c)\lambda_{V''}(c) - \delta\tau_{\mathcal{W}}(c) \\ &\geq \log(1+c) - \delta\tau_{\mathcal{W}}(c),\end{aligned} \tag{A.8}$$

where the last inequality comes immediately from the definition of $\tau_{V''}$. As a result, letting $\delta \to 0$ gives the desired identity.

Now, we proceed to prove of the second limit in (A.5). Set $\theta = \inf\{c > 0 | \lambda_{\mathcal{W}}(c) > 0\}$ and write $W(0^+)$ for the right limit of W at 0. Note that $\lambda_{\mathcal{W}}(\eta) = 0$ for $0 < \eta < \theta$ and $\lambda_{\mathcal{W}}(\eta) > 0$ for $\eta > \theta$. By the first equality of (A.6), this implies that $W(\lambda) > \eta$ for $\lambda > 0$ and $0 < \eta < \theta$; $W(\lambda) \leq \eta$ for $0 < \lambda < \lambda_{\mathcal{W}}(\eta)$ and $\eta > \theta$. As a consequence, we obtain $W(0^+) \geq \eta$ for $0 < \eta < \theta$ and $W(0^+) \leq \eta$ for $\eta > \theta$, as desired.

For (4), since $\overline{W}$ differs from W at most over a countable set, it is clear that (A.5) and (A.6) also hold under the replacement of W with $\overline{W}$. For the specific case, as $\overline{W}$ is non-decreasing and right-continuous, the conclusion of $\overline{W} \in \mathcal{V}$ is an immediate result of the second limit in (A.5). □

Remark A.7. It is proved by (A.8) that $\tau_{\mathcal{W}}(c)\lambda_{\mathcal{W}}(c) \geq \log(1+c)$ for $c \in (0, \mathcal{L}_{\mathcal{W}}(0))$ such that $\lambda_{\mathcal{W}}(c) > 0$ or equivalently $\tau_{\mathcal{W}}(c) < \infty$.

Proof of Proposition 2.1: The proof is based on Lemma A.3 and Remark A.7 and the detail is exactly the same as that of Chen et al. (2017, Proposition 2.8), which is skipped here. □

A.2. *Proofs for Proposition 2.4 and Theorem 2.6.*

Proof of Proposition 2.4: It is obvious from Definition 2.2 that a cutoff implies a precutoff. Now, assume that $(\mathcal{L}_{\mathcal{W}_n})_{n=1}^{\infty}$ has a precutoff and let A, B, t_n be the constants in Definition 2.2. Set

$$\alpha = \liminf_{n\to\infty} \mathcal{L}_{\mathcal{W}_n}(At_n), \quad s_n = T_{\mathcal{W}_n}(\beta), \tag{A.9}$$

where β is a positive constant less than α. By the monotonicity of $\mathcal{L}_{\mathcal{W}_n}(t)$ in t, there is $N > 0$ such that

$$At_n \leq s_n \leq Bt_n < \infty, \quad \forall n \geq N. \tag{A.10}$$

In what follows, we will prove that $(\mathcal{L}_{\mathcal{W}_n})_{n=1}^{\infty}$ has a cutoff with cutoff time s_n.

First, let $\epsilon > 0$ and choose $V_n \in \mathcal{W}_n$ such that

$$\mathcal{L}_{\mathcal{W}_n}((1+\epsilon)s_n) \leq \mathcal{L}_{V_n}((1+\epsilon)s_n) + 1/n. \tag{A.11}$$

Write $\mathcal{L}_{V_n}(t + s_n) = \mathcal{L}_{U_n}(t)$, where $U_n(\lambda) = \int_{(0,\lambda]} e^{-s_n\eta} dV_n(\eta)$. Clearly, $U_n \in \mathcal{V}$. By Lemma A.1(3) and (A.10), one has that, for $n \geq N$,

$$\mathcal{L}_{U_n}(0) = \mathcal{L}_{V_n}(s_n) \leq \mathcal{L}_{\mathcal{W}_n}(s_n) \leq \beta,$$

and

$$\mathcal{L}_{U_n}((B/A - 1)s_n) = \mathcal{L}_{V_n}((B/A)s_n) \leq \mathcal{L}_{V_n}(Bt_n) \leq \mathcal{L}_{\mathcal{W}_n}(Bt_n).$$

Letting $n \to \infty$ implies

$$\limsup_{n\to\infty} \mathcal{L}_{U_n}(0) \le \beta, \quad \limsup_{n\to\infty} \mathcal{L}_{U_n}((B/A-1)s_n) = 0.$$

As a result of (A.11) and Lemma A.2, we obtain that, for $\epsilon > 0$,

$$\limsup_{n\to\infty} \mathcal{L}_{\mathcal{W}_n}((1+\epsilon)s_n) \le \limsup_{n\to\infty} \mathcal{L}_{V_n}((1+\epsilon)s_n) = \limsup_{n\to\infty} \mathcal{L}_{U_n}(\epsilon s_n) = 0.$$

Next, let $\epsilon \in (0,1)$ and choose $V_n' \in \mathcal{W}_n$ such that

$$\mathcal{L}_{V_n'}((1-\epsilon)s_n) \ge \mathcal{L}_{\mathcal{W}_n}((1-\epsilon)s_n) - 1/n.$$

Clearly, for $n \ge (2/\beta) \vee N$, $\mathcal{L}_{V_n'}((1-\epsilon)s_n) > \beta - \beta/2 = \beta/2$, which implies $T_{V_n'}(\beta/2) \ge (1-\epsilon)s_n$. In addition with the following fact,

$$\limsup_{n\to\infty} \mathcal{L}_{V_n'}((1+\epsilon)s_n) \le \limsup_{n\to\infty} \mathcal{L}_{\mathcal{W}_n}((1+\epsilon)s_n) = 0,$$

one can see that $(\mathcal{L}_{V_n'})_{n=1}^\infty$ presents a precutoff. By Chen et al. (2017, Theorem 2.2 and Remark 2.3), $(\mathcal{L}_{V_n'})_{n=1}^\infty$ has a cutoff with cutoff time $T_{V_n'}(\beta/2)$ and, hence,

$$\liminf_{n\to\infty} \mathcal{L}_{\mathcal{W}_n}((1-\epsilon)^2 s_n) \ge \liminf_{n\to\infty} \mathcal{L}_{V_n'}\left((1-\epsilon)T_{V_n'}(\beta/2)\right) = \infty.$$

□

The following lemma highlights the importance of small c in Theorem 2.6(4), which will be used to prove Theorem 2.6.

Lemma A.8. *Let $\mathcal{W}_n \subset \mathcal{V}$ and assume that $\mathcal{L}_{\mathcal{W}_n}(0) < \infty$ and $\mathcal{L}_{\mathcal{W}_n}(0) \to \infty$ and $\lambda_{\mathcal{W}_n}(c) > 0$ for $0 < c < \mathcal{L}_{\mathcal{W}_n}(0)$. If there is $c' > 0$ such that $\tau_{\mathcal{W}_n}(c')\lambda_{\mathcal{W}_n}(c') \to \infty$, then $\tau_{\mathcal{W}_n}(c'')\lambda_{\mathcal{W}_n}(c') \to \infty$ for all $c'' > 0$. In particular, for $c_2 > c_1 > 0$,*

$$\tau_{\mathcal{W}_n}(c_1)\lambda_{\mathcal{W}_n}(c_1) \to \infty \quad \Rightarrow \quad \tau_{\mathcal{W}_n}(c_2)\lambda_{\mathcal{W}_n}(c_2) \to \infty.$$

Proof of Lemma A.8: The second part is given by the first part and the fact of $\lambda_{\mathcal{W}_n}(c_1) \le \lambda_{\mathcal{W}_n}(c_2)$ for $c_1 < c_2$. For the first part, we need the following lemma.

Lemma A.9. *(Chen et al., 2017, Lemma 2.6) For $V \in \mathcal{V}$ and $c \in (0, \mathcal{L}_V(0))$, there is $\gamma \ge \lambda_V(c)$ such that $\tau_V(c) = \gamma^{-1}\log(1+V(\gamma))$.*

Suppose $\tau_{\mathcal{W}_n}(c')\lambda_{\mathcal{W}_n}(c') \to \infty$. By the definition in (2.1), it is obvious that $\tau_{\mathcal{W}_n}(c'') \ge \tau_{\mathcal{W}_n}(c')$ for $0 < c'' < c'$. Immediately, this implies $\tau_{\mathcal{W}_n}(c'')\lambda_{\mathcal{W}_n}(c') \ge \tau_{\mathcal{W}_n}(c')\lambda_{\mathcal{W}_n}(c') \to \infty$. In the following, we assume $c'' > c'$. Since $\mathcal{L}_{\mathcal{W}_n}(0) \to \infty$, one may choose $N > 0$ such that $\mathcal{L}_{\mathcal{W}_n}(0) > c''$ for $n \ge N$. By (A.4), as $\lambda_{\mathcal{W}_n}(c) > 0$ for all $c \in (0, \mathcal{L}_{\mathcal{W}_n}(0))$, $\tau_{\mathcal{W}_n}(c')$ and $\lambda_{\mathcal{W}_n}(c')$ are positive and finite for $n \ge N$. Further, for $n \ge N$, one may select $V_n \in \mathcal{W}_n$ such that $\tau_{\mathcal{W}_n}(c') \le 2\tau_{V_n}(c')$. By Lemma A.9, there is $\gamma_n \ge \lambda_{V_n}(c')$ such that $\tau_{V_n}(c') = \gamma_n^{-1}\log(1+V_n(\gamma_n))$. As a result, we obtain

$$\log(1+V_n(\gamma_n)) = \tau_{V_n}(c')\gamma_n \ge \frac{\tau_{\mathcal{W}_n}(c')\lambda_{V_n}(c')}{2} \ge \frac{\tau_{\mathcal{W}_n}(c')\lambda_{\mathcal{W}_n}(c')}{2} \to \infty.$$

By choosing $N' > N$ such that $V_n(\gamma_n) \ge c''$ for $n \ge N'$, one has $\tau_{V_n}(c') = \tau_{V_n}(c'')$ for $n \ge N'$. Consequently, this leads to

$$\tau_{\mathcal{W}_n}(c'')\lambda_{\mathcal{W}_n}(c') \ge \tau_{V_n}(c'')\lambda_{\mathcal{W}_n}(c') \ge \frac{\tau_{\mathcal{W}_n}(c')\lambda_{\mathcal{W}_n}(c')}{2} \to \infty.$$

□

Proof of Theorem 2.6: By Remark A.6, the assumption of

$$\mathcal{L}_{\mathcal{W}_n}(0) < \infty, \quad \mathcal{L}_{\mathcal{W}_n}(0) \to \infty, \quad \lambda_{\mathcal{W}_n}(c) > 0, \quad \forall c \in (0, \mathcal{L}_{\mathcal{W}_n}(0)),\ n \ge 1,$$

implies that, for any $c > 0$, $\tau_{\mathcal{W}_n}(c)$ and $T_{\mathcal{W}_n}(c)$ are positive and finite for n large enough.

We first prove that (1)–(3) are equivalent. Assume that $(\mathcal{L}_{\mathcal{W}_n})_{n=1}^\infty$ has a cutoff. By Definition 2.2, $T_{\mathcal{W}_n}(\epsilon)$ is a cutoff time and this implies $\mathcal{L}_{\mathcal{W}_n}(2T_{\mathcal{W}_n}(\epsilon))\to 0$ for all $\epsilon>0$. Note that, for $V\in\mathcal{V}$, $c>0$ and $t>0$,

$$\mathcal{L}_V(t)\ge\int_{(0,\lambda_V(c)]}e^{-t\lambda}dV(\lambda)\ge ce^{-t\lambda_V(c)}.$$

Taking the supremum over $\mathcal{W}_n$ yields $\mathcal{L}_{\mathcal{W}_n}(t)\ge ce^{-t\lambda_{\mathcal{W}_n}(c)}$ and, then, replacing t with $2T_{\mathcal{W}_n}(\epsilon)$ leads to (2). (2)⇒(3) is obvious. For (3)⇒(1), assume that there exists $\epsilon>0$ such that $T_{\mathcal{W}_n}(\epsilon)\lambda_{\mathcal{W}_n}(c)\to\infty$ for all $c>0$. Recall the following identity in Chen et al. (2017, Lemma 2.1),

$$\mathcal{L}_V(t)=t\int_{(0,\infty)}V(\lambda)e^{-t\lambda}d\lambda,\quad \forall t>0,\,V\in\mathcal{V}.$$

For $0<s<t$ and $c>0$, one has

$$\begin{aligned}\mathcal{L}_V(t)&\le c+t\int_{[\lambda_V(c),\infty)}V(\lambda)e^{-t\lambda}d\lambda\\&\le c+te^{(s-t)\lambda_V(c)}\int_{[\lambda_V(c),\infty)}V(\lambda)e^{-s\lambda}d\lambda\\&\le c+(t/s)e^{(s-t)\lambda_V(c)}\mathcal{L}_V(s).\end{aligned}$$

Taking the supremum over $\mathcal{W}_n$ yields

$$\mathcal{L}_{\mathcal{W}_n}(t)\le c+(t/s)e^{(s-t)\lambda_{\mathcal{W}_n}(c)}\mathcal{L}_{\mathcal{W}_n}(s),\quad\forall t>s>0,\,c>0. \tag{A.12}$$

By Lemmas A.1(3) and A.5(2), we may select $N\in\mathbb{N}$ such that $T_{\mathcal{W}_n}(\epsilon)\in(0,\infty)$ for $n\ge N$. The replacement of t,s in (A.12) with $aT_{\mathcal{W}_n}(\epsilon),T_{\mathcal{W}_n}(\epsilon)$, where $a>1$, and with $T_{\mathcal{W}_n}(\epsilon),aT_{\mathcal{W}_n}(\epsilon)$, where $a\in(0,1)$, yields

$$\mathcal{L}_{\mathcal{W}_n}(aT_{\mathcal{W}_n}(\epsilon))\le c+a\epsilon e^{(1-a)T_{\mathcal{W}_n}(\epsilon)\lambda_{\mathcal{W}_n}(c)},\quad\forall a>1,\,c>0,$$

and

$$\mathcal{L}_{\mathcal{W}_n}(aT_{\mathcal{W}_n}(\epsilon))\ge(\epsilon-c)ae^{(1-a)T_{\mathcal{W}_n}(\epsilon)\lambda_{\mathcal{W}_n}(c)},\quad\forall a\in(0,1),\,c>0,$$

for $n\ge N$. Letting $n\to\infty$ and then $c\to 0$ implies the cutoff of $(\mathcal{L}_{\mathcal{W}_n})_{n=1}^\infty$.

Next, we show (2)⇒(4)⇒(5) ⇒(6)⇒(3). First, (5)⇒(6) is obvious. Secondly, (4)⇒(5) is an immediate result of Lemma A.8, while (6)⇒(3) is given by the second inequality in (2.3). For (2)⇒(4), let $c>0$ and choose $N'>0$ such that $\mathcal{L}_{\mathcal{W}_n}(0)>c$ for $n\ge N'$. The replacement of $\mathcal{W},A$ with $\mathcal{W}_n$, $\sqrt{\tau_{\mathcal{W}_n}(c)\lambda_{\mathcal{W}_n}(c)}$ in (2.3) yields $2\tau_{\mathcal{W}_n}(c)\ge T_{\mathcal{W}_n}(c+2)$ for $n\ge N'$. By the assumption of (2), we obtain (4).

To see $\tau_{\mathcal{W}_n}(c)$ is a cutoff time, assume that $(\mathcal{L}_{\mathcal{W}_n})_{n=1}^\infty$ has a cutoff. Note that $T_{\mathcal{W}_n}(\epsilon)\sim T_{\mathcal{W}_n}(\delta)$ for $\epsilon,\delta>0$. Let $c>0$ and $\alpha_n(c)=\sqrt{\tau_{\mathcal{W}_n}(c)\lambda_{\mathcal{W}_n}(c)}$. By applying (2.3) with $A=1$, we obtain

$$\left(1-\frac{1}{\alpha_n(c)}\right)T_{\mathcal{W}_n}(c+1)\le\tau_{\mathcal{W}_n}(c)\le T_{\mathcal{W}_n}\left(\frac{c}{1+c}\right), \tag{A.13}$$

for n large enough. In addition with (4), this implies $\tau_{\mathcal{W}_n}(c)\sim T_{\mathcal{W}_n}(c/(1+c))$, which proves that $\tau_{\mathcal{W}_n}(c)$ is a cutoff time.

For (2.6), let $\epsilon>\delta>0$ and $c>0$. Clearly, $T_{\mathcal{W}_n}(\epsilon)\le T_{\mathcal{W}_n}(\delta)$. By (2), one may choose $N''>0$ such that

$$c\exp\{-T_{\mathcal{W}_n}(\epsilon)\lambda_{\mathcal{W}_n}(\delta/3)\}<\delta/3,\quad\mathcal{L}_{\mathcal{W}_n}(0)>c\vee\epsilon,\quad\forall n\ge N''.$$

The desired bound is then given by applying (2.4) to $\mathcal{W}_n$ and replacing c_1,c_2,B with $\delta/3,c,\log(6\epsilon/\delta)$. For (2.7), let $\epsilon,c>0$ and $\beta_n(c)=\tau_{\mathcal{W}_n}(c)\lambda_{\mathcal{W}_n}(c)/\log[1+\tau_{\mathcal{W}_n}(c)\lambda_{\mathcal{W}_n}(c)]$. Note that the second inequality of (2.3) and (A.3) combine to

$$\frac{\beta_n(c)}{\beta_n(c)+1}T_{\mathcal{W}_n}(c+1)\le\tau_{\mathcal{W}_n}(c)\le T_{\mathcal{W}_n}\left(\frac{c}{1+c}\right),$$

for n large enough. This implies

$$\begin{aligned}\left|\tau_{\mathcal{W}_n}(c) - T_{\mathcal{W}_n}\left(\frac{c}{1+c}\right)\right| &\le \left|T_{\mathcal{W}_n}\left(\frac{c}{1+c}\right) - \frac{\beta_n(c)}{\beta_n(c)+1}T_{\mathcal{W}_n}(c+1)\right| \\ &\le \left|T_{\mathcal{W}_n}\left(\frac{c}{1+c}\right) - T_{\mathcal{W}_n}(c+1)\right| + \frac{T_{\mathcal{W}_n}(c+1)}{\beta_n(c)}.\end{aligned}$$

As a consequence, (2.7) is achieved by (2.6) and the fact that $T_{\mathcal{W}_n}(c+1) \sim \tau_{\mathcal{W}_n}(c)$ and $\beta_n(c) \sim \tau_{\mathcal{W}_n}(c)\lambda_{\mathcal{W}_n}(c)/\log[\tau_{\mathcal{W}_n}(c)\lambda_{\mathcal{W}_n}(c)]$. □

Appendix B. Other technical proofs

This section is dedicated to proving Proposition 5.1 and the proof is based on the following lemma.

Lemma B.1. *Let τ and $\tau^{(c)}$ be as in (5.3)–(5.4). Assume that $p \ge q > 0$ and set*

$$i^* = \frac{(p-q)(N+1)}{p+q}, \quad j^* = \frac{(p-q)(N+1)}{3(pq)^{1/4}\sqrt{p+q}} + 1. \tag{B.1}$$

Then, one has

$$\frac{(N+1)^2\log(1+\lfloor j^*\wedge(N/3)\rfloor)}{80\sqrt{pq}\lfloor j^*\rfloor^2} \le \frac{\tau}{r+2\sqrt{pq}} \le \frac{(N+1)^2[N\log(p/q)+3]}{16\sqrt{pq}} \tag{B.2}$$

and

$$\frac{(N+1)^2\log(1+\lfloor j^*\wedge N\rfloor)}{40\sqrt{pq}\lfloor j^*\rfloor^2} \le \tau^{(c)} \le \frac{(N+1)^2[N\log(p/q)+2]}{8\sqrt{pq}}. \tag{B.3}$$

If it is assumed further that $p > q$, then

$$\frac{N\log(p/q) - 222/i^* - 3\log\lceil i^*\rceil - 2}{-2\log(r+2\sqrt{pq})} \le \tau \le \frac{N\log(p/q)+2}{-2\log(r+2\sqrt{pq})} \tag{B.4}$$

and

$$\frac{N\log(p/q) - 42/i^* - 3\log\lceil i^*\rceil - 2}{2(\sqrt{p}-\sqrt{q})^2} \le \tau^{(c)} \le \frac{N\log(p/q)+2}{2(\sqrt{p}-\sqrt{q})^2}. \tag{B.5}$$

Proof of Lemma B.1(Bounds for $\tau^{(c)}$): Note that it suffices to show (B.3) and (B.5) with the assumption of $p+q=1$. Otherwise, one may set $p' = p/(p+q)$ and $q' = q/(p+q)$ and let τ' be the constant in (5.4) under the replacement of p, q with p', q'. Then, the desired bounds for $\tau^{(c)}$ are given by the bounds for τ' in (B.3)–(B.5) and the fact of $\tau' = (p+q)\tau^{(c)}$.

In the following, we shall restrict ourselves to the case of $p+q=1$ with $p \ge q$. To see the upper bounds of $\tau^{(c)}$, as $\max_x \sum_{i=1}^j |\phi_i(x)|^2 \ge j$, one may apply the inequality of $\log(1+t) \le 1/t + \log t$ for $t > 0$ to derive

$$\log\left(1 + \max_{x\in\mathcal{X}}\sum_{i=1}^j |\phi_i(x)|^2\right) \le 1 + \log\max_{x\in\mathcal{X}}\sum_{i=1}^j |\phi_i(x)|^2 \le 1 + \log\frac{2j}{\pi(0)(N+1)}.$$

Furthermore, from the trigonometric identities, it is easy to see that

$$\alpha_j = (\sqrt{p}-\sqrt{q})^2 + 4\sqrt{pq}\sin^2\frac{j\pi}{2(N+1)} \ge (\sqrt{p}-\sqrt{q})^2 + 4\sqrt{pq}\frac{j^2}{(N+1)^2}, \tag{B.6}$$

where the inequality is a result of the fact that $\sin t \ge 2t/\pi$ for $t \in [0, \pi/2]$. As $p \ge q$, we have

$$\log\frac{1}{\pi(0)(N+1)} = \log\frac{1+(p/q)+\cdots+(p/q)^N}{N+1} \le N\log\frac{p}{q}. \tag{B.7}$$

By (B.6), this implies

$$\frac{\tau^{(c)}}{(N+1)^2} \le \max_{1\le j\le N} \frac{N\log(p/q)+1+\log(2j)}{8\sqrt{pq}j^2} \le \frac{N\log(p/q)+2}{8\sqrt{pq}},$$

where the last inequality comes from the fact that the function $t^{-2}\log(2t)$ is decreasing on $[1,\infty)$. This proves the upper bound of $\tau^{(c)}$ in (B.3).

For the case of $p>q$, let i^* be the constant in (B.1). Note that

$$\log\frac{1}{\pi(0)(N+1)} = \log\frac{p(p/q)^N-q}{i^*} \le N\log\frac{p}{q}+\log\frac{1}{i^*}. \tag{B.8}$$

Since $\alpha_i<\alpha_{i+1}$ for $i\ge 1$, one has

$$\tau^{(c)} \le \frac{1+N\log(p/q)}{2\alpha_1} + \max_{1\le j\le N}\frac{\log(2j/i^*)}{2\alpha_j}.$$

By (B.6) and the fact of $\sqrt{p}+\sqrt{q}\ge 1$, we have

$$\alpha_j \ge (\sqrt{p}-\sqrt{q})^2\left(1+4\sqrt{pq}\frac{j^2}{(i^*)^2}\right) \ge (\sqrt{p}-\sqrt{q})^2.$$

As a consequence, this implies

$$\tau^{(c)} \le \frac{1+N\log(p/q)+R}{2(\sqrt{p}-\sqrt{q})^2}, \quad R=\max_{1\le j\le N}\frac{\log(2j/i^*)}{1+4\sqrt{pq}(j/i^*)^2}.$$

When $p>9/10$, it is clear that $p-q>4/5$ and then

$$R \le \log\frac{2N}{i^*} \le \log\frac{2}{(p-q)} < \log\frac{5}{2} < 1. \tag{B.9}$$

When $1/2<p\le 9/10$, it is easy to see that $4\sqrt{pq}\ge 6/5>1$, which leads to

$$R \le \max_{1\le j\le N}\log\frac{\log(2j/i^*)}{1+(j/i^*)^2} \le \sup_{t>0}\frac{\log(2t)}{1+t^2} \le \sup_{0<t\le 5/4}\frac{\log(2t)}{1+t^2} \le \log\frac{5}{2} < 1, \tag{B.10}$$

where the third inequality is immediate from the fact that the function $(1+t^2)^{-1}\log(2t)$ is decreasing on $[5/4,\infty)$. This proves the upper bound of $\tau^{(c)}$ in (B.5).

Next, we prove the lower bounds of $\tau^{(c)}$. For the case of $p\ge q$, let j^* be the constant in (B.1) and set $\kappa=\lfloor j^*\wedge N\rfloor$. By the equality in (B.6) and the fact of $\sin t\le t$ for $t\ge 0$, one may derive

$$\begin{aligned}\alpha_j &\le (\sqrt{p}-\sqrt{q})^2+\frac{10\sqrt{pq}j^2}{(N+1)^2} \le (p-q)^2+\frac{10\sqrt{pq}j^2}{(N+1)^2}\\ &= \frac{10\sqrt{pq}\left((j^*-1)^2+j^2\right)}{(N+1)^2} \le \frac{10\sqrt{pq}\left(\lfloor j^*\rfloor^2+j^2\right)}{(N+1)^2}.\end{aligned} \tag{B.11}$$

As $\max_x\sum_{i=1}^j|\phi_i(x)|^2\ge j$, this implies

$$\tau^{(c)} \ge \frac{\log(1+\kappa)}{2\alpha_\kappa} \ge \frac{(N+1)^2\log(1+\kappa)}{40\sqrt{pq}\lfloor j^*\rfloor^2}.$$

For the case of $p>q$, note that $\tau^{(c)}\ge[\log|\phi_1(0)|^2]/(2\alpha_1)$. Write

$$|\phi_1(0)|^2 = \frac{(p/q)^N-q/p}{i^*}\times\frac{2p^2}{\alpha_1}\sin^2\frac{\pi}{N+1}. \tag{B.12}$$

It is easy to see from the fact of $\log(t-1) \geq \log t - 1/(t-1)$ and $\log t \geq (t-1)/t$ for $t > 1$ that

$$\begin{aligned}\log\left(\left(\frac{p}{q}\right)^N - \frac{q}{p}\right) &\geq N\log\frac{p}{q} - \frac{1}{(p/q)^N - 1} \\ &\geq N\log\frac{p}{q} - \frac{1}{e^{i^*/2}-1} \geq N\log\frac{p}{q} - \frac{2}{i^*},\end{aligned} \tag{B.13}$$

where the last inequality is a result of $e^t \geq t+1$. Using the first two inequalities of (B.11), one may derive respectively

$$\frac{\alpha_1}{(\sqrt{p}-\sqrt{q})^2} \leq 1 + \frac{10\sqrt{pq}(\sqrt{p}+\sqrt{q})^2}{(i^*)^2} \leq 1 + \frac{20\sqrt{pq}}{(i^*)^2} \tag{B.14}$$

and

$$(N+1)^2\alpha_1 \leq (i^*)^2 + 10\sqrt{pq} \leq \lceil i^*\rceil^2 + 5 \leq 6\lceil i^*\rceil^2. \tag{B.15}$$

Applying (B.13), (B.15) and the fact of $\sin(\pi/(N+1)) \geq 2/(N+1)$ and $2p \geq 1$ to (B.12), we obtain

$$\begin{aligned}\log|\phi_1(0)|^2 &\geq N\log\frac{p}{q} - \frac{2}{i^*} + \log\frac{1}{i^*} + \log\frac{1}{3\lceil i^*\rceil^2} \\ &\geq N\log\frac{p}{q} - \frac{2}{i^*} - 3\log\lceil i^*\rceil - 2.\end{aligned} \tag{B.16}$$

By the fact of $1/(1+t) \geq 1-t$ for $t > -1$, (B.14) implies

$$\frac{1}{2\alpha_1} \geq \frac{1}{2(\sqrt{p}-\sqrt{q})^2}\left(1 - \frac{20\sqrt{pq}}{(i^*)^2}\right).$$

In addition with the observation of $\alpha_1 \geq (\sqrt{p}-\sqrt{q})^2$, the above inequalities lead to

$$\tau^{(c)} \geq \frac{[N\log(p/q)][1-20\sqrt{pq}/(i^*)^2] - 2/i^* - 3\log\lceil i^*\rceil - 2}{2(\sqrt{p}-\sqrt{q})^2}. \tag{B.17}$$

The desired lower bound of $\tau^{(c)}$ is then given by

$$N\log\frac{p}{q} = 2N\log\frac{\sqrt{p}}{\sqrt{q}} \leq \frac{2N(\sqrt{p}-\sqrt{q})}{\sqrt{q}} \leq \frac{2N(p-q)}{\sqrt{q}} \leq \frac{2i^*}{\sqrt{q}}.$$

□

Proof of Lemma B.1(Bounds for τ): We first prove the upper bounds of τ. It is easy to see from (5.1) and (5.2) that

$$|\beta_{N+1-i}| \leq |\beta_i| = \beta_i < \beta_{i-1},\ \forall 1 \leq i \leq \frac{N+1}{2}, \quad |\varphi_i(x)|^2 \leq \frac{2}{\pi(0)(N+1)},\ \forall i, x.$$

The former implies $|\xi_{2i}| \leq |\xi_{2i-1}| \leq \beta_i$ for $1 \leq i \leq (N+1)/2$. Using the fact of $\log(1+t) \leq 1 + \log t$ for $t \geq 1$, one may derive

$$\tau \leq \max_{1\leq j\leq(N+1)/2} \frac{1+\log(4j) - \log[\pi(0)(N+1)]}{-2\log\beta_j}. \tag{B.18}$$

In addition with (B.7), we obtain

$$\tau \leq \max_{1\leq j\leq(N+1)/2} \frac{N\log(p/q) + 1 + \log(4j)}{-2\log\beta_j}. \tag{B.19}$$

Observe that, for $1 \le j \le (N+1)/2$,

$$\begin{aligned}\log \beta_j &= \log(r+2\sqrt{pq}) + \log\left(1 - \frac{4\sqrt{pq}}{r+2\sqrt{pq}} \sin^2 \frac{j\pi}{2(N+1)}\right) \\ &\le \log(r+2\sqrt{pq}) - \frac{8\sqrt{pq}j^2}{(r+2\sqrt{pq})(N+1)^2},\end{aligned} \tag{B.20}$$

where the inequality is a result of $\sin t \ge (2\sqrt{2}/\pi)t$ for $t \in [0, \pi/4]$. As $r + 2\sqrt{pq} \le 1$, one may derive from (B.19) and (B.20) that

$$\frac{\tau}{(r+2\sqrt{pq})(N+1)^2} \le \frac{1}{16\sqrt{pq}}\left(N\log(p/q) + 1 + \max_{1\le j\le (N+1)/2} \frac{\log(4j)}{j^2}\right).$$

The upper bound of τ in (B.2) is then given by the monotonicity of $t^{-2}\log(4t)$ on $[1,\infty)$.

In the remaining proof, we set $p' = p/(p+q)$ and $q' = q/(p+q)$. For the upper bound in (B.4), one may derive from (B.20) that

$$\log \beta_j \le [\log(r+2\sqrt{pq})]\left(1 + A\left(\frac{j}{i^*}\right)^2\right), \quad \forall 1 \le j \le \frac{N+1}{2},$$

where

$$A = \frac{8\sqrt{p'q'}(\sqrt{p'}+\sqrt{q'})^2(\sqrt{p}-\sqrt{q})^2}{-(r+2\sqrt{pq})\log(r+2\sqrt{pq})}.$$

By the fact of $r + 2\sqrt{pq} = 1 - (\sqrt{p}-\sqrt{q})^2$ and $\log t \ge (t-1)/t$ for $t > 0$, it is easy to see that $A \ge 8\sqrt{p'q'}(\sqrt{p'}+\sqrt{q'})^2 \ge 8\sqrt{p'q'}$. Note that the inequality of (B.8) also holds here. As a result of (B.18), we achieve

$$\tau \le \max_{1\le j\le (N+1)/2} \frac{N\log p/q + 1 + \log(4j/i^*)}{-2[\log(r+2\sqrt{pq})][1+8\sqrt{p'q'}(j/i^*)^2]} \le \frac{N\log(p/q)+1+R'}{-2\log(r+2\sqrt{pq})},$$

where

$$R' = \max_{1\le j\le (N+1)/2} \frac{\log(4j/i^*)}{1+8\sqrt{p'q'}(j/i^*)^2}.$$

Consequently, one may use estimations similar to (B.9)–(B.10) with the replacement of p, q by p', q' and the fact of $(1+2t^2)^{-1}\log(4t) \le 1$ for $t > 0$ to conclude that $R' < 1$, which leads to the desired upper bound in (B.4).

To see the lower bound of τ, suppose $\xi_{i_k} = \beta_k$ for $1 \le k \le (N+1)/2$. Clearly, $i_k < i_{k+1}$ and $\varphi_{i_k} = \phi_k$. This implies

$$\sum_{j=1}^{i_k} |\varphi_j(x)|^2 \ge \sum_{\ell=1}^{k} |\varphi_{i_\ell}(x)|^2 = \sum_{\ell=1}^{k} |\phi_\ell(x)|^2,$$

and then

$$\tau \ge \max_{1\le j\le (N+1)/2} \frac{\log(1+\max_x \sum_{i=1}^{j} |\phi_i(x)|^2)}{-2\log\beta_j}.$$

Following the equality of (B.20) and applying the fact of $\sin t \le t$ and $\log t \ge (t-1)/t$ for $t > 0$, one may derive that, for $j \le (N+1)/3$,

$$\begin{aligned}(r+2\sqrt{pq})\log\beta_j &\ge (r+2\sqrt{pq})\log(r+2\sqrt{pq}) - \frac{r+2\sqrt{pq}}{r+\sqrt{pq}} \times \frac{\sqrt{pq}j^2\pi^2}{(N+1)^2} \\ &\ge -(\sqrt{p}-\sqrt{q})^2 - \frac{20\sqrt{pq}j^2}{(N+1)^2} \\ &= -(p+q)\left[\left(\sqrt{p'}-\sqrt{q'}\right)^2 + \frac{20\sqrt{p'q'}j^2}{(N+1)^2}\right].\end{aligned} \tag{B.21}$$

Applying the second and third inequalities in (B.11) with the replacement of p,q by p',q', we obtain

$$(r+2\sqrt{pq})\log\beta_j \geq -(p+q)\frac{20\sqrt{p'q'}(\lfloor j^*\rfloor^2+j^2)}{(N+1)^2} = -\frac{20\sqrt{pq}(\lfloor j^*\rfloor^2+j^2)}{(N+1)^2}.$$

The lower bound in (B.2) then follows from the above estimation and the fact of $\tau \geq [\log(1+\kappa')]/(-2\log\beta_{\kappa'})$ with $\kappa' = \lfloor j^* \wedge (N/3)\rfloor$.

For the case of $p>q$, note that $\tau \geq [\log|\phi_1(0)|^2]/(-2\log\beta_1)$ and that the conclusion in (B.16) also holds here. With the first inequality of (B.21), we have

$$\frac{\log\beta_1}{\log(r+2\sqrt{pq})} \leq 1 + \frac{20\sqrt{pq}}{-(r+2\sqrt{pq})[\log(r+2\sqrt{pq})](N+1)^2}. \tag{B.22}$$

When $r+2\sqrt{pq} > 1/e$, one may apply the fact of $\log(r+2\sqrt{pq}) \leq -(\sqrt{p}-\sqrt{q})^2$ to derive

$$\frac{\log\beta_1}{\log(r+2\sqrt{pq})} \leq 1 + \frac{55\sqrt{p'q'}}{(\sqrt{p'}-\sqrt{q'})^2} \leq 1 + \frac{110\sqrt{p'q'}}{(i^*)^2}. \tag{B.23}$$

Similarly to (B.17), this implies

$$\begin{aligned}\tau &\geq \frac{[N\log(p/q)][1-110\sqrt{p'q'}/(i^*)^2] - 2/i^* - 3\log\lceil i^*\rceil - 2}{-2\log(r+2\sqrt{pq})}\\ &\geq \frac{N\log(p/q) - 222/i^* - 3\log\lceil i^*\rceil - 2}{-2\log(r+2\sqrt{pq})},\end{aligned}$$

where the second inequality is a result of $N\log(p/q) = N\log(p'/q') \leq 2i^*/\sqrt{q'}$.

When $r+2\sqrt{pq} \leq 1/e$, one has $\log(4pq) \leq -1/2$. Observe that, by the monotonicity of $t\log t$ on $(0,1/e)$, (B.22) yields

$$\frac{\log\beta_1}{\log(r+2\sqrt{pq})} \leq 1 + \frac{20}{[-\log(4pq)](N+1)^2}.$$

As before, this implies

$$\tau \geq \frac{N\log(p/q) - 20I - 2/i^* - 3\log\lceil i^*\rceil - 2}{-2\log(r+2\sqrt{pq})},$$

where

$$I = \frac{N\log(p/q)}{[-\log(4pq)](N+1)^2} \leq \frac{\log(4p^2)-\log(4pq)}{[-\log(4pq)]i^*} \leq \frac{4}{i^*}.$$

Immediately, this leads to the lower bound in (B.4). □

Proof of Proposition 5.1: Let i^* and j^* be the constants in (B.1) and set $p' = p/(p+q)$ and $q' = q/(p+q)$. First, assume $C \leq N/3$. As $p' = 1-q'$ and $C = N(p'-q')$, one has

$$q' = \frac{1}{2}\left(1 - \frac{C}{N}\right) \geq \frac{1}{3}, \quad \sqrt{p'q'} \geq \frac{\sqrt{2}}{3},$$

which imply

$$j^* \leq \frac{C+1}{3(\sqrt{2}/3)^{1/2}} + 1 \leq \frac{C+3}{2}.$$

As a consequence of (B.3), we obtain

$$\tau^{(c)} \leq \frac{[N(p'-q')/q'+2](N+1)^2}{(8\sqrt{2}/3)(p+q)} \leq \frac{(3C+2)(N+1)^2}{3(p+q)} \leq \frac{(C+1)(N+1)^2}{p+q},$$

and, in addition with the inequality of $\sqrt{p'q'} \leq 1/2$,

$$\tau^{(c)} \geq \frac{(\log 2)(N+1)^2}{5(C+3)^2(p+q)} \geq \frac{(N+1)^2}{8(C+3)^2(p+q)}.$$

In a similar way, the bounds of τ can be derived from (B.2) with the additional fact of

$$\frac{2\sqrt{2}}{3} \le 2\sqrt{p'q'} \le r + 2\sqrt{pq} = 1 + (p+q)\left(2\sqrt{p'q'} - 1\right) \le 1,$$

while the restriction of $N \ge 3$ is used for the requirement of $\lfloor j^* \wedge (N/3) \rfloor \ge 1$.

Next, let's consider the case of $C \ge 2$. Note that

$$i^* > C \ge 2, \quad N \log \frac{p}{q} \ge \frac{N(p'/q' - 1)}{p'/q'} > C, \quad \lceil i^* \rceil \le i^* + 1 \le C + 2 < 2C.$$

By (B.5), the above inequalities yield

$$\tau^{(c)} \le \frac{N \log(p/q)}{2(\sqrt{p} - \sqrt{q})^2} \left(1 + \frac{2}{C}\right)$$

and

$$\tau^{(c)} \ge \frac{N \log(p/q)}{2(\sqrt{p} - \sqrt{q})^2} \left(1 - \frac{42/C + 3\log(2C) + 2}{C}\right).$$

The desired lower bound of $\tau^{(c)}$ is then given by the fact of $3 \log 2 < 3$. For τ, the upper bound is clear from the above discussion, while the lower bound can be proved in a similar way and the details are skipped. $\square$

Acknowledgements

The author would like to thank Takashi Kumagai for his great help of figuring out the intrinsic mechanism of the theoretical framework, especially the formula of the max-ℓ^2-cutoff time, and the innominate referee for carefully reading the early version.

References

Barrera, J. and Ycart, B. Bounds for left and right window cutoffs. *ALEA Lat. Am. J. Probab. Math. Stat.*, **11** (1), 445–458 (2014). MR3265085.

Basu, R., Hermon, J., and Peres, Y. Characterization of cutoff for reversible Markov chains. *Ann. Probab.*, **45** (3), 1448–1487 (2017). MR3650406.

Bordenave, C., Caputo, P., and Salez, J. Cutoff at the "entropic time" for sparse Markov chains. *Probab. Theory Related Fields*, **173** (1-2), 261–292 (2019). MR3916108.

Chen, G.-Y. *The cutoff phenomenon for finite Markov chains*. Ph.D. thesis, Cornell University (2006). MR2708835.

Chen, G.-Y., Hsu, J.-M., and Sheu, Y.-C. The L^2-cutoffs for reversible Markov chains. *Ann. Appl. Probab.*, **27** (4), 2305–2341 (2017). MR3693527.

Chen, G.-Y. and Kumagai, T. Cutoffs for product chains. *Stochastic Process. Appl.*, **128** (11), 3840–3879 (2018). MR3860012.

Chen, G.-Y. and Saloff-Coste, L. The cutoff phenomenon for ergodic Markov processes. *Electron. J. Probab.*, **13**, no. 3, 26–78 (2008). MR2375599.

Chen, G.-Y. and Saloff-Coste, L. The L^2-cutoff for reversible Markov processes. *J. Funct. Anal.*, **258** (7), 2246–2315 (2010). MR2584746.

Chen, G.-Y. and Saloff-Coste, L. Comparison of cutoffs between lazy walks and Markovian semigroups. *J. Appl. Probab.*, **50** (4), 943–959 (2013). MR3161366.

Chen, G.-Y. and Saloff-Coste, L. Computing cutoff times of birth and death chains. *Electron. J. Probab.*, **20**, no. 76, 47 (2015). MR3371435.

Diaconis, P. and Hanlon, P. Eigen-analysis for some examples of the Metropolis algorithm. In *Hypergeometric functions on domains of positivity, Jack polynomials, and applications (Tampa,* *FL, 1991)*, volume 138 of *Contemp. Math.*, pp. 99–117. Amer. Math. Soc., Providence, RI (1992). MR1199122.

Diaconis, P. and Saloff-Coste, L. What do we know about the Metropolis algorithm? *J. Comput. System Sci.*, **57** (1), 20–36 (1998). MR1649805.

Diaconis, P. and Saloff-Coste, L. Separation cut-offs for birth and death chains. *Ann. Appl. Probab.*, **16** (4), 2098–2122 (2006). MR2288715.

Diaconis, P. and Shahshahani, M. Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete*, **57** (2), 159–179 (1981). MR626813.

Ding, J., Lubetzky, E., and Peres, Y. Total variation cutoff in birth-and-death chains. *Probab. Theory Related Fields*, **146** (1-2), 61–85 (2010). MR2550359.

Hermon, J. and Peres, Y. A characterization of L_2 mixing and hypercontractivity via hitting times and maximal inequalities. *Probab. Theory Related Fields*, **170** (3-4), 769–800 (2018a). MR3773799.

Hermon, J. and Peres, Y. On sensitivity of mixing times and cutoff. *Electron. J. Probab.*, **23**, Paper No. 25, 34 (2018b). MR3779818.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.*, **21** (6), 1087–1092 (1953). DOI: 10.1063/1.1699114.

Saloff-Coste, L. Lectures on finite Markov chains. In *Lectures on probability theory and statistics (Saint-Flour, 1996)*, volume 1665 of *Lecture Notes in Math.*, pp. 301–413. Springer, Berlin (1997). MR1490046.