Mixing times are hitting times of large sets

Yuval Peres[*]   Perla Sousi[†]

April 21, 2013

Abstract

We consider irreducible reversible discrete time Markov chains on a finite state space. Mixing times and hitting times are fundamental parameters of the chain. We relate them by showing that the mixing time of the lazy chain is equivalent to the maximum over initial states x and large sets A of the hitting time of A starting from x. We also prove that the first time when averaging over two consecutive time steps is close to stationarity is equivalent to the mixing time of the lazy version of the chain.

Keywords and phrases. Markov chain, mixing time, hitting time, stopping time.
MSC 2010 *subject classifications.* Primary 60J10.

1 Introduction

Mixing times and hitting times are among the most fundamental notions associated with a finite Markov chain. A variety of tools have been developed to estimate both these notions; in particular, hitting times are closely related to potential theory and they can be determined by solving a system of linear equations. In this paper we establish a new connection between mixing times and hitting times for reversible Markov chains (Theorem 1.1).

Let $(X_t)_{t \geq 0}$ be an irreducible Markov chain on a finite state space with transition matrix P and stationary distribution π. For x, y in the state space we write

$$P^t(x, y) = \mathbb{P}_x(X_t = y),$$

for the transition probability in t steps.

Let $d(t) = \max_x \|P^t(x, \cdot) - \pi\|$, where $\|\mu - \nu\|$ stands for the total variation distance between the two probability measures μ and ν. Let $\varepsilon > 0$. The total variation mixing is defined as follows:

$$t_{\mathrm{mix}}(\varepsilon) = \min\{t \geq 0 : d(t) \leq \varepsilon\}.$$

We write P_L^t for the transition probability in t steps of the lazy version of the chain, i.e. the chain with transition matrix $\frac{P+I}{2}$. If we now let $d_L(t) = \max_x \|P_L^t(x, \cdot) - \pi\|$, then we can define the mixing time of the lazy chain as follows:

$$t_{\mathrm{L}}(\varepsilon) = \min\{t \geq 0 : d_L(t) \leq \varepsilon\}. \tag{1.1}$$

---

[*]Microsoft Research, Redmond, Washington, USA; peres@microsoft.com
[†]University of Cambridge, Cambridge, UK; p.sousi@statslab.cam.ac.uk

For notational convenience we will simply write t_L and t_{mix} when $\varepsilon = 1/4$.

Before stating our first theorem, we introduce the maximum hitting time of "big" sets. Let $\alpha < 1/2$, then we define

$$t_H(\alpha) = \max_{x, A:\pi(A) \geq \alpha} \mathbb{E}_x[\tau_A],$$

where τ_A stands for the first hitting time of the set A by the Markov chain with transition matrix P.

It is clear (and we prove it later) that if the Markov chain has not hit a big set, then it cannot have mixed. Thus for every $\alpha > 0$, there is a positive constant c'_α so that

$$t_L \geq c'_\alpha t_H(\alpha).$$

In the following theorem, we show that the converse is also true when a chain is reversible.

Theorem 1.1. *Let $\alpha < 1/2$. Then there exist positive constants c'_α and c_α so that for every reversible chain*

$$c'_\alpha t_H(\alpha) \leq t_L \leq c_\alpha t_H(\alpha).$$

Remark 1.2. Aldous in [2] showed that the mixing time, t_{cts}, of a continuous time reversible chain is equivalent to $t_{prod} = \max_{x, A:\pi(A) > 0} \pi(A)\mathbb{E}_x[\tau_A]$. The inequality $t_{prod} \leq c_1 t_{cts}$, for a positive constant c_1, which was the hard part in Aldous' proof, follows from Theorem 1.1 and the equivalence $t_L \asymp t_{cts}$ (see [6, Theorem 20.3]). For the other direction we give a new proof in Section 8.

Remark 1.3. In Section 9 we present an application of Theorem 1.1 to robustness of the mixing time. Namely, we show that for a finite binary tree, assigning bounded conductances to the edges can only change the mixing time of the lazy random walk on the tree by a bounded factor. However, we note that not all graphs are robust to conductance perturbations. A counterexample is given by Ding and Peres in [4].

To avoid periodicity and near-periodicity issues, one often considers the lazy version of a discrete time Markov chain. In the following theorem we show that averaging over two successive times suffices, i.e. $t_L \asymp t_{ave}\left(\frac{1}{4}\right)$ where

$$t_{ave}(\varepsilon) = \min\left\{t \geq 0 : \max_x \left\| \frac{P^t(x, \cdot) + P^{t+1}(x, \cdot)}{2} - \pi \right\| \leq \varepsilon\right\}.$$

For notational convenience we will simply write t_{ave} when $\varepsilon = 1/4$.

Theorem 1.4. *There exist universal positive constants c and c' so that for every reversible Markov chain*

$$ct_L \leq t_{ave} \leq c't_L.$$

The problem of relating t_{ave} to the mixing time t_{cts} of the continuous-time chain was raised in Aldous-Fill [1], Chapter 4, Open Problem 17. Since $t_{cts} \asymp t_L$ (see [6, Theorem 20.3]), Theorem 1.4 gives a partial answer to that problem.

2 Preliminaries and further equivalences

In this section we first introduce some more notions of mixing. We will then state some further equivalences between them mostly in the reversible case and will prove them in later sections. These equivalences will be useful for the proofs of the main results, but are also of independent interest.

The following notion of mixing was first introduced by Aldous in [2] in the continuous time case and later studied in discrete time by Lovász and Winkler in [7, 8]. It is defined as follows:

$$t_{\text{stop}} = \max_x \min\{\mathbb{E}_x[\Lambda_x] : \Lambda_x \text{ is a stopping time s.t. } \mathbb{P}_x(X_{\Lambda_x} \in \cdot) = \pi(\cdot)\}. \qquad (2.1)$$

The definition does not make it clear why stopping times achieving the minimum always exist. We will recall the construction of such a stopping time in Section 3.

The mixing time of the lazy chain and the average mixing are related to t_{stop} in the following way.

Lemma 2.1. *There exists a uniform positive constant c_1 so that for every reversible Markov chain*

$$t_{\text{ave}} \leq c_1 t_{\text{stop}}.$$

Lemma 2.2. *There exists a uniform positive constant c_2 so that for every reversible Markov chain*

$$t_{\text{stop}} \leq c_2 t_{\text{L}}.$$

We will prove Lemma 2.1 in Section 3. Lemma 2.2 was proved by Aldous in [2], but we include the proof in Section 4 for completeness.

In Section 4 we will show that for any chain we have the following:

Lemma 2.3. *For every $\varepsilon \leq 1/4$, there exists a positive constant c_3 so that for every Markov chain we have that*

$$t_{\text{L}}(\varepsilon) \leq c_3 t_{\text{ave}}(\varepsilon).$$

Definition 2.4. We say that two mixing parameters s and r are **equivalent** for a class of Markov chains $\mathcal{M}$ and write $s \asymp r$, if there exist universal positive constants c and c' so that $cs \leq r \leq c's$ for every chain in $\mathcal{M}$. We also write $s \lesssim r$ and $s \gtrsim r$ if there exist universal positive constants c_1 and c_2 such that $s \leq c_1 r$ and $s \geq c_2 r$ respectively.

Proof of Theorem 1.4. Lemmas 2.1, 2.2 and 2.3 give the desired equivalence between t_{ave} and t_{L}. □

Combining the three lemmas above we get the following:

Corollary 2.5. *For every reversible Markov chain t_{L} and t_{stop} are equivalent.*

Remark 2.6. Aldous in [2] was the first to show the equivalence between the mixing time of a continuous time reversible chain and t_{stop}.

We will now define the notion of mixing in a geometric time. The idea of using this notion of mixing to prove Theorem 1.1 was suggested to us by Oded Schramm (private communication June 2008). This notion is also of independent interest, because of its properties that we will prove in this section.

For each t, let Z_t be a Geometric random variable taking values in $\{1, 2, \ldots\}$ of mean t and success probability t^{-1}. We first define

$$d_G(t) = \max_x \|\mathbb{P}_x(X_{Z_t} = \cdot) - \pi\|.$$

The geometric mixing is then defined as follows

$$t_{\text{G}} = t_{\text{G}}(1/4) = \min\{t \geq 0 : d_G(t) \leq 1/4\}.$$

We start by establishing the monotonicity property of $d_G(t)$.

Lemma 2.7. *The total variation distance $d_G(t)$ is decreasing as a function of t.*

Before proving this lemma, we note the following standard fact.

Claim 2.1. *Let T and T' be two independent positive random variables, also independent of the Markov chain. Then for all x*

$$\|\mathbb{P}_x(X_{T+T'} = \cdot) - \pi\| \leq \|\mathbb{P}_x(X_T = \cdot) - \pi\|.$$

Proof of Lemma 2.7. We first describe a coupling between the two Geometric random variables, Z_t and Z_{t+1}. Let $(U_i)_{i \geq 1}$ be a sequence of i.i.d. random variables uniform on $[0,1]$. We now define

$$Z_t = \min\left\{i \geq 1 : U_i \leq \frac{1}{t}\right\} \text{ and}$$

$$Z_{t+1} = \min\left\{i \geq 1 : U_i \leq \frac{1}{t+1}\right\}.$$

It is easy to see that

$$Z_{t+1} - Z_t \quad \text{is independent of} \quad Z_t.$$

Indeed, $\mathbb{P}(Z_{t+1} = Z_t | Z_t) = \frac{t}{t+1}$ and similarly for every $k \geq 1$ we have $\mathbb{P}(Z_{t+1} = Z_t + k | Z_t) = \left(\frac{t}{t+1}\right)^{k-1}\left(\frac{1}{t+1}\right)^2$.

We can thus write $Z_{t+1} = (Z_{t+1} - Z_t) + Z_t$, where the two terms are independent.

Claim 2.1 and the independence of $Z_{t+1} - Z_t$ and Z_t give the desired monotonicity of $d_G(t)$. $\qquad\square$

Lemma 2.8. *For all chains we have that*

$$t_G \leq 4t_{\text{stop}} + 1.$$

The converse of Lemma 2.8 is true for reversible chains in a more general setting. Namely, let N_t be a random variable independent of the Markov chain and of mean t. We define the total variation distance $d_N(t)$ in this setting as follows:

$$d_N(t) = \max_x \|\mathbb{P}_x(X_{N_t} = \cdot) - \pi\|.$$

Defining $t_N = t_N(1/4) = \min\{t \geq 0 : d_N(t) \leq 1/4\}$ we have the following:

Lemma 2.9. *There exists a positive constant c_4 such that for all reversible chains*

$$t_{\text{stop}} \leq c_4 t_N.$$

In particular, $t_{\text{stop}} \leq c_4 t_G$.

We will give the proofs of Lemmas 2.8 and 2.9 in Section 5.

Combining Corollary 2.5 with Lemmas 2.8 and 2.9 we deduce:

Theorem 2.10. *For a reversible Markov chain t_G and t_L are equivalent.*

We end this section by stating and proving a result relating t_{mix} and t_{ave} for any Markov chain. First by the triangle inequality it is clear that always $t_{\text{ave}} \leq t_{\text{mix}}$. For the converse we have the following:

4

Proposition 2.11. *Let $0 < \delta < 1$. There exists a positive constant c_5 so that if P is a transition matrix satisfying $P(x, x) \geq \delta$, for all x, then*

$$t_{\mathrm{mix}} \leq c_5 \left(t_{\mathrm{ave}} \vee \frac{1}{\delta(1-\delta)} \right).$$

Proof. By the triangle inequality we have that for all x

$$\|P^t(x, \cdot) - \pi\| \leq \left\| \tfrac{1}{2} P^t(x, \cdot) + \tfrac{1}{2} P^{t+1}(x, \cdot) - \pi \right\| + \left\| \tfrac{1}{2} P^t(x, \cdot) - \tfrac{1}{2} P^{t+1}(x, \cdot) \right\|.$$

Thus it suffices to show that for all starting points x and all times t there exists a positive constant c_6 such that

$$\|P^t(x, \cdot) - P^{t+1}(x, \cdot)\| \leq \frac{c_6}{\sqrt{t\delta(1-\delta)}}, \tag{2.2}$$

since $t_{\mathrm{mix}}(\varepsilon) \leq c_7 t_{\mathrm{mix}} \left(\tfrac{4}{3}\varepsilon \right)$, for a positive constant c_7 and $\varepsilon \leq \tfrac{1}{4}$.

We will now construct a coupling (X_t, Y_{t+1}) of $P^t(x, \cdot)$ with $P^{t+1}(x, \cdot)$ such that

$$\mathbb{P}(X_t \neq Y_{t+1}) \leq \frac{c_6}{\sqrt{t\delta(1-\delta)}}.$$

Since for all x we have that $P(x, x) \geq \delta$, we can write

$$P = \delta I + (1 - \delta)Q,$$

for a stochastic matrix Q. Let Z be a chain with transition matrix Q that starts from x. Let N_t and N_t' be independent and both distributed according to $\mathrm{Bin}(t, 1 - \delta)$. We are now going to describe the coupling for the two chains, X and Y. Let $(W_s)_{s \geq 1}$ and $(W_s')_{s \geq 1}$ be i.i.d. random variables with $\mathbb{P}(W_1 = 0) = 1 - \mathbb{P}(W_1 = 1) = \delta$. We define $N_t = \sum_{s=1}^{t} W_s$ and define a process (N_t') by setting $N_0 = 0$ and

$$N_t' = \begin{cases} \displaystyle\sum_{s=1}^{t} W_s & \text{if } N_{t-1} \neq N_t, \\ N_{t+1}' & \text{if } N_{t-1} = N_t. \end{cases}$$

It is straightforward to check that N' is a Markov chain with transition matrix

$$A(n, n) = \delta = 1 - A(n, n+1) \quad \text{for all } n \in \mathbb{N}.$$

Hence, if for all t we set $X_t = Z_{N_t'}$ and $Y_t = Z_{N_t}$, then it follows that both X and Y are Markov chains with transition matrix P. We now let $\tau = \min\{t \geq 0 : X_t = Y_{t+1}\}$. If $W_1 = 0$, i.e. $Y_1 = X_0 = x$, then $\tau = 0$. Otherwise, on the event $W_1 = 1$, we can bound τ by

$$\tau \leq \min \left\{ t \geq 0 : N_t' = 1 + \sum_{s=2}^{t+1} W_s \right\}.$$

We thus see that τ is stochastically dominated by the first time that $N_t' - \sum_{s=2}^{t+1} W_s$ hits 1. But $N_t' - \sum_{s=2}^{t+1} W_s$ is a symmetric random walk on the real line with transition probabilities $p(k, k+1) = p(k, k-1) = \delta(1 - \delta)$ for all k. By time t this random walk has moved L number of times, where

$$L \sim \mathrm{Bin}(t, 2\delta(1-\delta)).$$

5

By the Chernoff bound for Binomial random variables we get that

$$\mathbb{P}\left(L < \frac{t\delta(1-\delta)}{2}\right) \le e^{-9t\delta(1-\delta)/16}. \tag{2.3}$$

Therefore we have that

$$\mathbb{P}(\tau > t) \le \mathbb{P}\left(L < \frac{t\delta(1-\delta)}{2}\right) + \mathbb{P}\left(\tau > t, L \ge \frac{t\delta(1-\delta)}{2}\right) \le e^{-9t\delta(1-\delta)/16} + \mathbb{P}_0\left(T_1 > \frac{t\delta(1-\delta)}{2}\right),$$

where T_1 denotes the first hitting time of 1 for a simple random walk on $\mathbb{Z}$. By a classical result for simple random walks on $\mathbb{Z}$ (see for instance [6, Theorem 2.17])

$$\mathbb{P}_0\left(T_1 > \frac{t\delta(1-\delta)}{2}\right) \le \frac{12\sqrt{2}}{\sqrt{t\delta(1-\delta)}}$$

and this concludes the proof. $\qquad\square$

Remark 2.12. We note that the upper bound given in Proposition 2.11 is tight, in the sense that both t_{ave} and $\frac{1}{\delta}$ can be attained. Indeed, for lazy chains t_{mix} and t_{ave} are equivalent. This follows from the observation above that $t_{\text{ave}} \le t_{\text{mix}}$ and [6, Proposition 5.6]. For $\delta \le 1/2$, consider the following transition matrix $\begin{pmatrix} \delta & 1-\delta \\ 1-\delta & \delta \end{pmatrix}$. It is easy to see that in this case the mixing time is of order $\frac{1}{\delta}$.

3   Stopping times and a bound for t_{ave}

In this section we will first give the construction of a stopping time T that achieves stationarity, i.e. for all x, y we have that $\mathbb{P}_x(X_T = y) = \pi(y)$, and also for a fixed x attains the minimum in the definition of t_{stop} in (2.1), i.e.

$$\mathbb{E}_x[T] = \min\{\mathbb{E}_x[\Lambda_x] : \Lambda_x \text{ is a stopping time s.t. } \mathbb{P}_x(X_{\Lambda_x} \in \cdot) = \pi(\cdot)\}. \tag{3.1}$$

The stopping time that we will construct is called the **filling rule** and it was first discussed in [3]. This construction can also be found in [1, Chapter 9], but we include it here for completeness.

First for any stopping time S and any starting distribution μ one can define a sequence of vectors

$$\theta_x(t) = \mathbb{P}_\mu(X_t = x, S \ge t), \quad \sigma_x(t) = \mathbb{P}_\mu(X_t = x, S = t). \tag{3.2}$$

These vectors clearly satisfy

$$0 \le \sigma(t) \le \theta(t), \quad (\theta(t) - \sigma(t))\mathbf{P} = \theta(t+1) \;\forall t; \; \theta(0) = \mu. \tag{3.3}$$

We can also do the converse, namely given vectors $(\theta(t), \sigma(t); t \ge 0)$ satisfying (3.3) we can construct a stopping time S satisfying (3.2). We want to define S so that

$$\mathbb{P}(S = t | S > t - 1, X_t = x, X_{t-1} = x_{t-1}, \ldots, X_0 = x_0) = \frac{\sigma_x(t)}{\theta_x(t)}. \tag{3.4}$$

Formally we define the random variable S as follows: Let $(U_i)_{i\ge 0}$ be a sequence of independent random variables uniform on $[0, 1]$. We now define S via

$$S = \inf\left\{t \ge 0 : U_t \le \frac{\sigma_{X_t}(t)}{\theta_{X_t}(t)}\right\}.$$

From this definition it is clear that (3.4) is satisfied and that S is a stopping time with respect to an enlarged filtration containing also the random variables $(U_i)_{i \geq 0}$, namely $\mathcal{F}_s = \sigma(X_0, U_0, \ldots, X_s, U_s)$. Also, equations (3.2) are satisfied. Indeed, setting $x_t = x$ we have

$$\mathbb{P}_\mu(X_t = x, S \geq t) = \sum_{x_0, x_1, \ldots, x_{t-1}} \mu(x_0) \prod_{k=0}^{t-1} \left(1 - \frac{\sigma_{x_k}(k)}{\theta_{x_k}(k)}\right) P(x_k, x_{k+1}) = \theta_x(t),$$

since $\theta_y(0) = \mu(y)$ for all y and also $\theta(t+1) = (\theta(t) - \sigma(t))\mathbf{P}$ so cancelations happen. Similarly we get the other equality of (3.2).

We are now ready to give the construction of the **filling rule** T. Before defining it formally, we give the **intuition** behind it. Every state x has a quota which is equal to $\pi(x)$. Starting from an initial distribution μ we want to calculate inductively the probability that we have stopped so far at each state. When we reach a new state, we decide to stop there if doing so does not increase the probability of stopping at that state above the quota. Otherwise we stop there with the right probability to exactly fill the quota and we continue with the complementary probability.

We will now give the **rigorous** construction by defining the sequence of vectors $(\theta(t), \sigma(t); t \geq 0)$ for any starting distribution μ. If we start from x, then simply $\mu = \delta_x$. First we set $\theta(0) = \mu$. We now introduce another sequence of vectors $(\Sigma(t); t \geq -1)$. Let $\Sigma_x(-1) = 0$ for all x. We define inductively

$$\sigma_x(t) = \begin{cases} \theta_x(t), & \text{if } \Sigma_x(t-1) + \theta_x(t) \leq \pi(x); \\ \pi(x) - \Sigma_x(t-1), & \text{otherwise.} \end{cases}$$

Then we let $\Sigma_x(t) = \sum_{s \leq t} \sigma_x(s)$ and define $\theta(t+1)$ via (3.3). Then σ will satisfy (3.2) and $\Sigma_x(t) = \mathbb{P}_\mu(X_T = x, T \leq t)$. Also note from the description above it follows that $\Sigma_x(t) \leq \pi(x)$, for all x and all t. Thus we get that

$$\mathbb{P}_\mu(X_T = x) = \lim_{t \to \infty} \Sigma_x(t) \leq \pi(x)$$

and since both $\mathbb{P}_\mu(X_T = \cdot)$ and $\pi(\cdot)$ are probability distributions, we get that they must be equal. Hence the above construction yielded a stationary stopping time. It only remains to prove the mean-optimality (3.1). Before doing so we give a definition.

Definition 3.1. Let S be a stopping time. A state z is called a **halting** state for the stopping time if $S \leq T_z$ a.s. where T_z is the first hitting time of state z.

We will now show that the filling rule has a halting state and then the following theorem gives the mean-optimality.

Theorem 3.2 (Lovász and Winkler). *Let μ and ρ be two distributions. Let S be a stopping time such that $\mathbb{P}_\mu(X_S = x) = \rho(x)$ for all x. Then S is mean optimal in the sense that*

$$\mathbb{E}_\mu[S] = \min\{\mathbb{E}_\mu[U] : U \text{ is a stopping time s.t. } \mathbb{P}_\mu(X_U \in \cdot) = \rho(\cdot)\}$$

if and only if it has a halting state.

Now we will prove that there exists z such that $T \leq T_z$ a.s. For each x we define

$$t_x = \min\{t : \Sigma_x(t) = \pi(x)\} \leq \infty.$$

Take z such that $t_z = \max_x t_x \leq \infty$. We will show that $T \leq T_z$ a.s. If there exists a t such that $\mathbb{P}_\mu(T > t, T_z = t) > 0$, then $\Sigma_x(t) = \pi(x)$, for all x, since the state z is the last one to be filled. So if the above probability is positive, then we get that

$$\mathbb{P}_\mu(T \leq t) = \sum_x \Sigma_x(t) = 1,$$

which is a contradiction. Hence, we obtain that $\mathbb{P}_\mu(T > t, T_z = t) = 0$ and thus by summing over all t we deduce that $\mathbb{P}_\mu(T \leq T_z) = 1$.

Proof of Theorem 3.2. We define the exit frequencies for S via $\nu_x = \mathbb{E}_\mu\left[\sum_{k=0}^{S-1} \mathbf{1}(X_k = x)\right]$, for all x.

Since $\mathbb{P}_\mu(X_S = \cdot) = \rho(\cdot)$, we can write

$$\mathbb{E}_\mu\left[\sum_{k=0}^{S} \mathbf{1}(X_k = x)\right] = \mathbb{E}_\mu\left[\sum_{k=0}^{S-1} \mathbf{1}(X_k = x)\right] + \rho(x) = \nu_x + \rho(x).$$

We also have that

$$\mathbb{E}_\mu\left[\sum_{k=0}^{S} \mathbf{1}(X_k = x)\right] = \mu(x) + \mathbb{E}_\mu\left[\sum_{k=1}^{S} \mathbf{1}(X_k = x)\right].$$

Since S is a stopping time, it is easy to see that

$$\mathbb{E}_\mu\left[\sum_{k=1}^{S} \mathbf{1}(X_k = x)\right] = \sum_y \nu_y P(y, x).$$

Hence we get that

$$\nu_x + \rho(x) = \mu(x) + \sum_y \nu_y P(y, x). \tag{3.5}$$

Let T be another stopping time with $\mathbb{P}_\mu(X_T = \cdot) = \rho(\cdot)$ and let ν_x' be its exit frequencies. Then they would satisfy (3.5), i.e.

$$\nu_x' + \rho(x) = \mu(x) + \sum_y \nu_y' P(y, x).$$

Thus if we set $d = \nu' - \nu$, then d as a vector satisfies

$$d = dP,$$

and hence d must be a multiple of the stationary distribution, i.e. for a constant α we have that $d = \alpha\pi$.

Suppose first that S has a halting state, i.e. there exists a state z such that $\nu_z = 0$. Therefore we get that $\nu_z' = \alpha\pi(z)$, and hence $\alpha \geq 0$. Thus $\nu_x' \geq \nu_x$ for all x and

$$\mathbb{E}_\mu[T] = \sum_x \nu'(x) \geq \sum_x \nu_x = \mathbb{E}_\mu[S],$$

and hence proving mean-optimality.

We will now show the converse, namely that if S is mean-optimal then it should have a halting state. The filling rule was proved to have a halting state and thus is mean-optimal. Hence using the same argument as above we get that S is mean optimal if and only if $\min_x \nu_x = 0$, which is the definition of a halting state. $\qquad\square$

Before giving the proof of Lemma 2.1 we state and prove a preliminary result.

Lemma 3.3. *Let X be a reversible Markov chain on the state space Γ and let L, U be positive constants. Let T be a stopping time that achieves stationarity starting from x, i.e. $\mathbb{P}_x(X_T = y) = \pi(y)$, for all y. For all y and all times u we define $f_y(u) = \frac{1}{2}\mathbb{P}_x(X_u = y, T \leq L) + \frac{1}{2}\mathbb{P}_x(X_{u+1} = y, T \leq L)$. Then there exists $u \leq L + U$ such that*

$$\sum_y \frac{f_y(u)^2}{\pi(y)} \leq 1 + \frac{L}{U}.$$

Proof. In this proof we will write $P_{x,y}(t) = P^t(x, y)$ for notational convenience. We define a measure ν on $\Gamma \times [0, L]$ by

$$\nu(\cdot, \cdot) = \mathbb{P}_x(T \leq L, (X_T, T) \in (\cdot, \cdot)).$$

We define $g_y(u) = \frac{1}{2}\mathbb{P}_x(X_{L+u} = y, T \leq L) + \frac{1}{2}\mathbb{P}_x(X_{L+u+1} = y, T \leq L)$ for $0 \leq u \leq U - 1$. By conditioning on (X_T, T) we get

$$g_y(u) = \frac{1}{2}\sum_{(z,s)}(P_{z,y}(L + u - s) + P_{z,y}(L + u + 1 - s))\nu(z, s),$$

where the sum is over (z, s) in $\Gamma \times [0, L]$. Thus

$$4\sum_y \pi(y)^{-1}g_y(u)^2 = I_1 + I_2 + I_3 + I_4, \tag{3.6}$$

where

$$I_1 = \sum_{\substack{(z_1,s_1)\\(z_2,s_2)}}\sum_y \pi^{-1}(y)P_{z_1,y}(L + u - s_1)P_{z_2,y}(L + u - s_2)\nu(z_1, s_1)\nu(z_2, s_2),$$

$$I_2 = \sum_{\substack{(z_1,s_1)\\(z_2,s_2)}}\sum_y \pi^{-1}(y)P_{z_1,y}(L + u - s_1)P_{z_2,y}(L + u + 1 - s_2)\nu(z_1, s_1)\nu(z_2, s_2),$$

$$I_3 = \sum_{\substack{(z_1,s_1)\\(z_2,s_2)}}\sum_y \pi^{-1}(y)P_{z_1,y}(L + u + 1 - s_1)P_{z_2,y}(L + u - s_2)\nu(z_1, s_1)\nu(z_2, s_2) \text{ and}$$

$$I_4 = \sum_{\substack{(z_1,s_1)\\(z_2,s_2)}}\sum_y \pi^{-1}(y)P_{z_1,y}(L + u + 1 - s_1)P_{z_2,y}(L + u + 1 - s_2)\nu(z_1, s_1)\nu(z_2, s_2).$$

By reversibility we have that

$$I_1 = \sum_{\substack{(z_1,s_1)\\(z_2,s_2)}} \pi(z_2)^{-1}P_{z_1,z_2}(2L + 2u - s_1 - s_2)\nu(z_1, s_1)\nu(z_2, s_2),$$

$$I_2 = \sum_{\substack{(z_1,s_1)\\(z_2,s_2)}} \pi(z_2)^{-1}P_{z_1,z_2}(2L + 2u + 1 - s_1 - s_2)\nu(z_1, s_1)\nu(z_2, s_2),$$

$$I_3 = \sum_{\substack{(z_1,s_1)\\(z_2,s_2)}} \pi(z_2)^{-1}P_{z_1,z_2}(2L + 2u + 1 - s_1 - s_2)\nu(z_1, s_1)\nu(z_2, s_2) \text{ and}$$

$$I_4 = \sum_{\substack{(z_1,s_1)\\(z_2,s_2)}} \pi(z_2)^{-1}P_{z_1,z_2}(2L + 2u + 2 - s_1 - s_2)\nu(z_1, s_1)\nu(z_2, s_2).$$

By considering two cases depending on whether $s_1 + s_2$ is odd or even it is elementary to check that

$$\frac{1}{U}\sum_{u=0}^{U-1}(P_{z_1,z_2}(2L+2u-s_1-s_2) + P_{z_1,z_2}(2L+2u+1-s_1-s_2)) \le \frac{1}{U}\sum_{u=0}^{2L+2U-1} P_{z_1,z_2}(u),$$

since $s_1, s_2 \in [0, L]$. Similarly

$$\frac{1}{U}\sum_{u=0}^{U-1}(P_{z_1,z_2}(2L+2u+2-s_1-s_2) + P_{z_1,z_2}(2L+2u+1-s_1-s_2)) \le \frac{1}{U}\sum_{u=1}^{2L+2U-1} P_{z_1,z_2}(u).$$

In this last average we have no dependence on s_1, s_2. Hence using (3.6), the fact that $\nu(z, [0, L]) \le \pi(z)$ for all z and stationarity of π, we get that

$$\frac{1}{U}\sum_{u=0}^{U-1}\sum_{y}\pi^{-1}(y)g_y(u)^2 \le \frac{1}{4U}\sum_{z_1,z_2}\pi^{-1}(z_2)\left(\sum_{u=0}^{2L+2U-1} P_{z_1,z_2}(u) + \sum_{u=1}^{2L+2U-1} P_{z_1,z_2}(u)\right)\pi(z_1)\pi(z_2)$$

$$= 1 + L/U.$$

This is an upper bound for the average, hence there exists some $u \le U - 1$ such that

$$\sum_{y}\pi^{-1}(y)g_y(u)^2 \le 1 + L/U.$$

$\square$

Remark 3.4. We note that the above lemma uses the same approach as in Aldous [2, Lemma 38]. Aldous' proof is carried out in continuous time. The proof of Lemma 3.3 cannot be done in discrete time for the non lazy version of the chain, since in this case defining $f_y(u) = \mathbb{P}_x(X_u = y, T \le L)$, we would get that $\sum_y \frac{f_y(u)^2}{\pi(y)} \le 2 + \frac{L}{U}$. This is where the averaging plays a crucial role.

We now have all the ingredients needed to give the proof of Lemma 2.1.

Proof of Lemma 2.1. We fix x. Let T be the filling rule as defined at the beginning of this section, which was shown to achieve the minimum appearing in the definition of t_{stop}. Thus, since in the definition of t_{stop} there is a maximum over the starting points, we have that

$$\mathbb{E}_x[T] \le t_{\text{stop}}. \tag{3.7}$$

Let $f_y(u) = \frac{1}{2}\mathbb{P}_x(X_u = y, T \le L) + \frac{1}{2}\mathbb{P}_x(X_{u+1} = y, T \le L)$ as appears in Lemma 3.3, where L and U are two positive constants whose precise value will be determined later in the proof and $u \le L + U$ is such that

$$\sum_{y}\frac{f_y(u)^2}{\pi(y)} \le 1 + \frac{L}{U}. \tag{3.8}$$

We then have

$$\left\|\frac{1}{2}P^u(x, \cdot) + \frac{1}{2}P^{u+1}(x, \cdot) - \pi\right\| = \frac{1}{2}\sum_{y}\left|\frac{1}{2}P^u(x, y) + \frac{1}{2}P^{u+1}(x, y) - \pi(y)\right|$$

$$\le \frac{1}{2}\left(\sum_{y}\left|\frac{1}{2}P^u(x, y) + \frac{1}{2}P^{u+1}(x, y) - f_y(u)\right| + \sum_{y}|f_y(u) - \pi(y)|\right)$$

$$= \frac{1}{2}\left(\mathbb{P}_x(T > L) + \sum_{y}|f_y(u) - \pi(y)|\right),$$

10

since $f_y(u) \leq \frac{1}{2}P^u(x,y) + \frac{1}{2}P^{u+1}(x,y)$ and $\sum_y f_y(u) = \mathbb{P}_x(T \leq L)$.

By the Cauchy–Schwarz inequality we deduce that

$$\left(\sum_y |f_y(u) - \pi(y)|\right)^2 = \left(\sum_y \pi(y)^{1/2} \left|\frac{f_y(u) - \pi(y)}{\pi(y)^{1/2}}\right|\right)^2 \leq \sum_y \pi(y)^{-1}(f_y(u) - \pi(y))^2$$
$$= \sum_y \pi(y)^{-1}f_y(u)^2 - 2\sum_y f_y(u) + 1 = \sum_y \pi(y)^{-1}f_y(u)^2 - 2\mathbb{P}_x(T \leq L) + 1.$$

Using (3.8) we get that this last expression is bounded from above by

$$2\mathbb{P}_x(T > L) + \frac{L}{U}.$$

Since $\left\|\frac{1}{2}P^t(x,\cdot) + \frac{1}{2}P^{t+1}(x,\cdot) - \pi\right\|$ is decreasing in t, we conclude that

$$\left\|\frac{1}{2}P^{L+U}(x,\cdot) + \frac{1}{2}P^{L+U+1}(x,\cdot) - \pi\right\| \leq \frac{1}{2}\left(\mathbb{P}_x(T > L) + \left(2\mathbb{P}_x(T > L) + \frac{L}{U}\right)^{1/2}\right).$$

If we now take $L = 20t_{\text{stop}}$ and $U = 10L$, then by Markov's inequality and (3.7) we get that the total variation distance

$$\left\|\frac{1}{2}P^{L+U}(x,\cdot) + \frac{1}{2}P^{L+U+1}(x,\cdot) - \pi\right\| \leq \frac{1}{4}.$$

Thus we get that $t_{\text{ave}} \leq L + U = 220t_{\text{stop}}$ and this concludes the proof of the lemma. □

4   Proofs of equivalences

In Section 2 we defined the notion of t_{stop}. In order to prove Lemma 2.2 we will first show a preliminary result that compares t_{stop} to t_{stop}^L, where the latter is defined as

$$t_{\text{stop}}^L = \max_x \min\{\mathbb{E}_x[U_x] : U_x \text{ is a stopping time s.t. } \mathbb{P}_x(X_{U_x}^L \in \cdot) = \pi(\cdot)\},$$

where X^L stands for the lazy version of the chain X.

Lemma 4.1. *For every chain we have that*

$$t_{\text{stop}} \leq \frac{1}{2}t_{\text{stop}}^L.$$

Proof. Let X^L denote the lazy version of the chain X. Then X^L can be realized by viewing X at a $\text{Bin}(t, 1/2)$ time, namely let $f(t) \sim \text{Bin}(t, 1/2)$, then $X_t^L = X_{f(t)}$ a.s. We can express $f(t)$ as $f(t) = \sum_{j=0}^t \xi(j)$, where $(\xi(j))_{j\geq 0}$ are i.i.d. fair coin tosses. Let T be a stopping time for the lazy chain X^L. We enlarge the filtration by adding all the coin tosses. In particular for each k we consider the following filtration:

$$\mathcal{F}_k = \sigma(X_0, \ldots, X_k, (\xi_j)_{j\geq 0}).$$

It is obvious that X has the Markov property with respect to the filtration $\mathcal{F}$ too. Also $f(T)$ is a stopping time for that filtration. Indeed,

$$\{f(T) = t\} = \left\{ \sum_{j=0}^{T} \xi_j = t \right\} = \bigcup_{\ell \geq t} \left\{ T = \ell, \sum_{j=0}^{\ell} \xi_j = t \right\}$$

and for each $\ell \geq t$ we have that

$$\left\{ T = \ell, \sum_{j=0}^{\ell} \xi_j = t \right\} \in \sigma(X_0, \ldots, X_t, (\xi_j)_{j \geq 0}),$$

since on the event $f(\ell) = t$ we have that $X_\ell^L = X_{f(\ell)} = X_t$. Hence $f(T)$ is a stopping time for X and it achieves stationarity, since for all x and y

$$\mathbb{P}_x \left(X_{f(T)} = y \right) = \mathbb{P}_x(X_T^L = y) = \pi(y),$$

since T achieves stationarity for the lazy chain. By Wald's identity for stopping times we get that for all x

$$\mathbb{E}_x[f(T)] = \mathbb{E}_x \left[\sum_{j=1}^{T} \xi(j) \right] = \mathbb{E}_x[T]\mathbb{E}_x[\xi] = \frac{1}{2}\mathbb{E}_x[T].$$

Hence using a stopping time of the lazy chain X^L achieving stationarity we defined a stopping time for the base chain X achieving stationarity and with expectation equal to half of the original one. Thus for all x we obtain that

$$\{\mathbb{E}_x[T] : T \text{ stopping time s.t. } \mathbb{P}_x \left(X_T^L = \cdot \right) = \pi\}$$
$$\subset \{2E_x[T'] : T' \text{ stopping time s.t. } \mathbb{P}_x \left(X_{T'} = \cdot \right) = \pi\}.$$

Therefore taking the minimum concludes the proof. $\qquad\square$

Before giving the proof of Lemma 2.2 we introduce some notation and a preliminary result that will also be used in the proof of Lemma 2.9. For any t we let

$$s(t) = \max_{x,y} \left[1 - \frac{P^t(x,y)}{\pi(y)} \right] \quad \text{and} \quad \bar{d}(t) = \max_{x,y} \| P^t(x,\cdot) - P^t(y,\cdot) \|.$$

We will call s the total separation distance from stationarity.

We finally define the separation mixing as follows

$$t_{\text{sep}} = \min\{t \geq 0 : s(t) \leq 3/4\}.$$

Lemma 4.2. *For a reversible Markov chain we have that*

$$d(t) \leq \bar{d}(t) \leq 2d(t) \quad and \quad s(2t) \leq 1 - (1 - \bar{d}(t))^2.$$

Proof. A proof of this result can be found in [1, Chapter 4, Lemma 7] or [6, Lemma 4.11 and Lemma 19.3]. $\qquad\square$

Remark 4.3. Lemma 4.2 above gives that $t_{\text{sep}} \leq 2t_{\text{mix}}$.

Lemma 4.4. *There exists a positive constant c so that for all chains we have*

$$t_{\text{stop}} \leq ct_{\text{sep}}$$

Proof. Fix $t = t_{\text{sep}}$. Then we have that for all x, y

$$P^t(x, y) \geq (1 - 3/4)\pi(y) = \frac{1}{4}\pi(y).$$

Hence, we can write

$$P^t(x, y) = \frac{1}{4}\pi(y) + \frac{3}{4}\nu_x(y),$$

where for a fixed x we have that ν_x is a probability measure. We can now construct a stopping time $S \in \{t, 2t, \ldots\}$ so that for all x

$$\mathbb{P}_x(X_S \in \cdot, S = t) = \frac{1}{4}\pi(\cdot)$$

and by induction on m such that

$$\mathbb{P}_x(X_S \in \cdot, S = mt) = \left(\frac{3}{4}\right)^{m-1}\frac{1}{4}\pi(\cdot).$$

Therefore it is clear that X_S is distributed according to π and $\mathbb{E}_x[S] = 4t$. Hence we get that $t_{\text{stop}} \leq 4t_{\text{sep}}$. $\qquad\square$

Proof of Lemma 2.2. Let t_{sep}^L stand for the separation mixing of the lazy chain. Then Lemma 4.4 gives that

$$t_{\text{stop}}^L \leq ct_{\text{sep}}^L.$$

Finally, Lemma 4.1 and Remark 4.3 conclude the proof. $\qquad\square$

Proof of Lemma 2.3. Fix t. Let T be a random variable taking values t and $t + 1$ each with probability $1/2$, i.e.

$$T = \begin{cases} t, & \text{w.p.}\frac{1}{2} \\ t + 1, & \text{w.p.}\frac{1}{2}. \end{cases}$$

Thus T can be written as $T = Y_1 + t$, where Y_1 is Bernoulli with probability $\frac{1}{2}$. Then we have that for all x and y

$$\mathbb{P}_x(X_T = y) = \frac{1}{2}\mathbb{P}_x(X_t = y) + \frac{1}{2}\mathbb{P}_x(X_{t+1} = y).$$

Let $Z \sim \text{Bin}(3t, \frac{1}{2})$. Then we can write Z as $Z = Y_1 + Z_1$, where Z_1 is distributed according to $\text{Bin}(3t-1, \frac{1}{2})$ and is independent of Y_1. Therefore Z can be expressed as the sum of two independent random variables, $Z = T + (Z_1 - t)$. (With high probability $Z_1 - t \approx (Z_1 - t)_+$.) We fix x. By the triangle inequality for the total variation distance, we obtain

$$\|\mathbb{P}_x(X_Z = \cdot) - \pi\| \leq \|\mathbb{P}_x(X_{T+(Z_1-t)_+} = \cdot) - \pi\| + \|\mathbb{P}_x(X_{T+(Z_1-t)} = \cdot) - \mathbb{P}_x(X_{T+(Z_1-t)_+} = \cdot)\|.$$

Since T and $(Z_1 - t)_+$ are independent and $(Z_1 - t)_+ \geq 0$, by the monotonicity of the total variation distance Claim 2.1, we deduce that

$$\|\mathbb{P}_x(X_{T+(Z_1-t)_+} = \cdot) - \pi\| \leq \|\mathbb{P}_x(X_T = \cdot) - \pi\|. \tag{4.1}$$

It is easy to see that

$$\|\mathbb{P}_x(X_{T+(Z_1-t)} = \cdot) - \mathbb{P}_x(X_{T+(Z_1-t)_+} = \cdot)\| \leq \mathbb{P}_x(Z_1 < t) \leq e^{-ct}, \tag{4.2}$$

for a positive constant c, since Z_1 follows the Binomial distribution. Hence by (4.1) and (4.2) we get that

$$\|\mathbb{P}_x(X_Z = \cdot) - \pi\| \leq \|\mathbb{P}_x(X_T = \cdot) - \pi\| + e^{-ct} \tag{4.3}$$

The mixing time for the lazy chain was defined in (1.1). Equivalently it is given by

$$t_{\mathrm{L}}(\varepsilon) = \min \left\{ t : \max_i \|\mathbb{P}_i(X_{Z'} = \cdot) - \pi\| < \varepsilon \right\},$$

where Z' is distributed according to $\mathrm{Bin}(t, 1/2)$. Thus

$$t_{\mathrm{L}}(\varepsilon) \leq 3 \min \left\{ t : \max_i \|\mathbb{P}_i(X_Z = \cdot) - \pi\| < \varepsilon \right\}.$$

Finally, from (4.3) we get that there exists a constant $c_2 > 0$ such that

$$t_{\mathrm{L}} \left(\tfrac{4}{3}\varepsilon \right) \leq c_2 t_{\mathrm{ave}}(\varepsilon).$$

But $t_{\mathrm{L}} \left(\tfrac{4}{3}\varepsilon \right) \geq c_3 t_{\mathrm{L}}(\varepsilon)$, since $\varepsilon \leq \tfrac{1}{4}$ and this concludes the proof. $\qquad \square$

5   Mixing at a geometric time

Before giving the proof of Lemma 2.8 we state two easy facts about total variation distance.

Claim 5.1. *Let Y be a discrete random variable with values in $\mathbb{N}$ and satisfying*

$$\mathbb{P}(Y = j) \leq c, \text{ for all } j > 0 \quad \text{and} \quad \mathbb{P}(Y = j) \text{ is decreasing in } j,$$

where c is a positive constant. Let Z be an independent random variable with values in $\mathbb{N}$. Then

$$\|\mathbb{P}(Y + Z = \cdot) - \mathbb{P}(Y = \cdot)\| \leq c\mathbb{E}[Z]. \tag{5.1}$$

Proof. Using the definition of total variation distance and the assumption on Y we have for all $k \in \mathbb{N}$

$$\|\mathbb{P}(Y + k = \cdot) - \mathbb{P}(Y = \cdot)\| = \sum_{j : \mathbb{P}(Y=j) \geq \mathbb{P}(Y+k=j)} (\mathbb{P}(Y = j) - \mathbb{P}(Y + k = j)) \leq kc.$$

Finally, since Z is independent of Y, we obtain (5.1). $\qquad \square$

The coupling definition of total variation distance gives the following:

Claim 5.2. *Let X be a Markov chain and W and V be two random variables with values in $\mathbb{N}$. Then*

$$\|\mathbb{P}(X_W = \cdot) - \mathbb{P}(X_V = \cdot)\| \leq \|\mathbb{P}(W = \cdot) - \mathbb{P}(V = \cdot)\|.$$

Proof of Lemma 2.8. We fix x. Let τ be a stationary time, i.e. $\mathbb{P}_x(X_\tau = \cdot) = \pi$. Then $\tau + s$ is also a stationary time for all $s \geq 1$. Hence, if Z_t is a Geometric random variable independent of τ, then $Z_t + \tau$ is also a stationary time, i.e. $\mathbb{P}_x(X_{Z_t + \tau} = \cdot) = \pi$. Since Z_t and τ are independent, and Z_t satisfies the assumptions of Claim 5.1, we get

$$\|\mathbb{P}_x(Z_t + \tau = \cdot) - \mathbb{P}_x(Z_t = \cdot)\| \leq \frac{\mathbb{E}_x[\tau]}{t}. \tag{5.2}$$

From Claim 5.2, we obtain

$$\|\mathbb{P}_x(X_{Z_t + \tau} = \cdot) - \mathbb{P}_x(X_{Z_t} = \cdot)\| \leq \|\mathbb{P}_x(Z_t + \tau = \cdot) - \mathbb{P}_x(Z_t = \cdot)\| \leq \frac{\mathbb{E}_x[\tau]}{t},$$

and since $\mathbb{P}_x(X_{Z_t + \tau} = \cdot) = \pi$, taking $t \geq 4\mathbb{E}_x[\tau]$ concludes the proof. $\qquad\square$

Recall from Section 2 the definition of N_t as a random variable independent of the Markov chain and of mean t. We also defined

$$d_N(t) = \max_x \|\mathbb{P}_x(X_{N_t} = \cdot) - \pi\|.$$

Let $N_t^{(1)}, N_t^{(2)}$ be i.i.d. random variables distributed as N_t and set $V_t = N_t^{(1)} + N_t^{(2)}$. We now define

$$s_N(t) = \max_{x,y} \left[1 - \frac{\mathbb{P}_x(X_{V_t} = y)}{\pi(y)}\right] \quad \text{and} \quad \bar{d}_N(t) = \max_{x,y} \|\mathbb{P}_x(X_{N_t} = \cdot) - \mathbb{P}_y(X_{N_t} = \cdot)\|.$$

When N is a geometric random variable we will write $d_G(t)$ and $\bar{d}_G(t)$ respectively.

Lemma 5.1. *For all t we have that*

$$d_N(t) \leq \bar{d}_N(t) \leq 2d_N(t) \quad \text{and} \quad s_N(t) \leq 1 - (1 - \bar{d}_N(t))^2.$$

Proof. Fix t and consider the chain Y with transition matrix $Q(x,y) = \mathbb{P}_x(X_{N_t} = y)$. Then $Q^2(x,y) = \mathbb{P}_x(X_{V_t} = y)$, where V_t is as defined above. Thus, if we let

$$s_Y(u) = \max_{x,y} \left[1 - \frac{Q^u(x,y)}{\pi(y)}\right] \quad \text{and} \quad \bar{d}_Y(u) = \max_{x,y} \|\mathbb{P}_x(Y_u = \cdot) - \mathbb{P}_y(Y_u = \cdot)\|,$$

then we get that $s_N(t) = s_Y(2)$ and $\bar{d}_N(t) = \bar{d}_Y(1)$. Hence, the lemma follows from Lemma 4.2. $\quad\square$

We now define

$$t_{s,N} = \min\left\{t \geq 0 : s_N(t) \leq \tfrac{3}{4}\right\}.$$

Lemma 5.2. *There exists a positive constant c so that for every chain*

$$t_{\text{stop}} \leq c t_{s,N}.$$

Proof. Fix $t = t_{s,N}$. Consider the chain Y with transition kernel $Q(x,y) = \mathbb{P}_x(X_{V_t} = y)$, where V_t is as defined above.
By the definition of $s_N(t)$ we have that for all x and y

$$Q(x,y) \geq (1 - s_N(t))\pi(y) \geq \frac{1}{4}\pi(y).$$

Thus, in the same way as in the proof of Lemma 4.4, we can construct a stopping time S such that Y_S is distributed according to π and $\mathbb{E}_x[S] = 4$ for all x.

Let $V_t^{(1)}, V_t^{(2)}, \ldots$ be i.i.d. random variables distributed as V_t. Then we can write $Y_u = X_{V_t^{(1)}+\ldots+V_t^{(u)}}$. If we let $T = V_t^{(1)} + \ldots + V_t^{(S)}$, then T is a stopping time for X such that $\mathcal{L}(X_T) = \pi$ and by Wald's identity for stopping times we get that for all x

$$\mathbb{E}_x[T] = \mathbb{E}_x[S]\mathbb{E}[V_t] = 8t.$$

Therefore we proved that

$$t_{\text{stop}} \leq 8t_{s,N}.$$

$\square$

Proof of Lemma 2.9. From Lemma 5.1 we get that

$$t_{s,N} \leq 2t_N.$$

Finally Lemma 5.2 completes the proof. $\square$

Remark 5.3. Let N_t be a uniform random variable in $\{1, \ldots, t\}$ independent of the Markov chain. The mixing time associated to N_t is called Cesàro mixing and it has been analyzed by Lovász and Winkler in [8]. From [6, Theorem 6.15] and the lemmas above we get the equivalence between the Cesàro mixing and the mixing of the lazy chain in the reversible case. In Section 7 we show that the Cesàro mixing time is equivalent to t_G for all chains.

Remark 5.4. From the remark above we see that the mixing at a geometric time and the Cesàro mixing are equivalent for a reversible chain. The mixing at a geometric time though has the advantage that its total variation distance, namely $d_G(t)$, has the monotonicity property Lemma 2.7, which is not true for the corresponding total variation distance for the Cesàro mixing.

Recall that $\bar{d}(t) = \max_{x,y} \|\mathbb{P}_x(X_t = \cdot) - \mathbb{P}_y(X_t = \cdot)\|$ is submultiplicative as a function of t (see for instance [6, Lemma 4.12]). In the following lemma and corollary, which will be used in the proof of Theorem 1.1, we show that $\bar{d}_G$ satisfies some sort of submultiplicativity.

Lemma 5.5. *Let $\beta < 1$ and let t be such that $\bar{d}_G(t) \leq \beta$. Then for all $k \in \mathbb{N}$ we have that*

$$\bar{d}_G(2^k t) \leq \left(\frac{1+\beta}{2}\right)^k \bar{d}_G(t).$$

Proof. As in the proof of Lemma 2.7 we can write $Z_{2t} = (Z_{2t} - Z_t) + Z_t$, where $Z_{2t} - Z_t$ and Z_t are independent. Hence it is easy to show (similar to the case for deterministic times) that

$$\bar{d}_G(2t) \leq \bar{d}_G(t) \max_{x,y} \|\mathbb{P}_x(X_{Z_{2t}-Z_t} = \cdot) - \mathbb{P}_y(X_{Z_{2t}-Z_t} = \cdot)\|. \tag{5.3}$$

By the coupling of Z_{2t} and Z_t it is easy to see that $Z_{2t} - Z_t$ can be expressed as follows:

$$Z_{2t} - Z_t = (1 - \xi) + \xi G_{2t},$$

where ξ is a Bernoulli$(\frac{1}{2})$ random variable and G_{2t} is a Geometric random variable of mean $2t$ independent of ξ. By the triangle inequality we get that

$$\|\mathbb{P}_x(X_{Z_{2t}-Z_t} = \cdot) - \mathbb{P}_y(X_{Z_{2t}-Z_t} = \cdot)\| \leq \frac{1}{2} + \frac{1}{2}\|\mathbb{P}_x(X_{G_{2t}} = \cdot) - \mathbb{P}_y(X_{G_{2t}} = \cdot)\| = \frac{1}{2} + \frac{1}{2}\bar{d}_G(2t),$$

and hence (5.3) becomes

$$\bar{d}_G(2t) \le \bar{d}_G(t) \left(\frac{1}{2} + \frac{1}{2}\bar{d}_G(2t) \right) \le \frac{1}{2}\bar{d}_G(t) \left(1 + \bar{d}_G(t) \right),$$

where for the second inequality we used the monotonicity property of $\bar{d}_G$ (same proof as for $d_G(t)$). Thus, since t satisfies $\bar{d}_G(t) \le \beta$, we get that

$$\bar{d}_G(2t) \le \left(\frac{1+\beta}{2} \right) \bar{d}_G(t),$$

and hence iterating we deduce the desired inequality. □

Combining Lemma 5.5 with Lemma 5.1 we get the following:

Corollary 5.6. *Let $\beta < 1$. If t is such that $d_G(t) \le \beta/2$, then for all k we have that*

$$d_G(2^k t) \le 2 \left(\frac{1+\beta}{2} \right)^k d_G(t).$$

Also if $d_G(t) \le \alpha < 1/2$, then there exists a constant $c = c(\alpha)$ depending only on α, such that $d_G(ct) \le 1/4$.

6 Hitting large sets

In this section we are going to give the proof of Theorem 1.1. We first prove an equivalence that does not require reversibility.

Theorem 6.1. *Let $\alpha < 1/2$. For every chain $t_G \asymp t_H(\alpha)$. (The implied constants depend on α.)*

Proof. We will first show that $t_G \ge ct_H(\alpha)$. By Corollary 5.6 there exists $k = k(\alpha)$ so that $d_G(2^k t_G) \le \frac{\alpha}{2}$. Let $t = 2^k t_G$. Then for any starting point x we have that

$$\mathbb{P}_x(X_{Z_t} \in A) \ge \pi(A) - \alpha/2 \ge \alpha/2.$$

Thus by performing independent experiments, we deduce that τ_A is stochastically dominated by $\sum_{i=1}^N G_i$, where N is a Geometric random variable of success probability $\alpha/2$ and the G_i's are independent Geometric random variables of success probability $\frac{1}{t}$. Therefore for any starting point x we get that

$$\mathbb{E}_x[\tau_A] \le \frac{2}{\alpha}t,$$

and hence this gives that

$$\max_{x, A:\pi(A) \ge \alpha} \mathbb{E}_x[\tau_A] \le \frac{2}{\alpha}2^k t_G.$$

In order to show the other direction, let $t' < t_G$. Then $d_G(t') > 1/4$. For a given $\alpha < 1/2$, we fix $\gamma \in (\alpha, 1/2)$. From Corollary 5.6 we have that there exists a positive constant $c = c(\gamma)$ such that

$$d_G(ct') > \gamma.$$

Set $t = ct'$. Then there exists a set A and a starting point x such that

$$\pi(A) - \mathbb{P}_x(X_{Z_t} \in A) > \gamma,$$

17

and hence $\pi(A) > \gamma$, or equivalently

$$\mathbb{P}_x(X_{Z_t} \in A) < \pi(A) - \gamma.$$

We now define a set B as follows:

$$B = \{y : \mathbb{P}_y(X_{Z_t} \in A) \geq \pi(A) - \alpha\},$$

where c is a constant smaller than α. Since π is a stationary distribution, we have that

$$\pi(A) = \sum_{y \in B} \mathbb{P}_y(X_{Z_t} \in A)\pi(y) + \sum_{y \notin B} \mathbb{P}_y(X_{Z_t} \in A)\pi(y) \leq \pi(B) + \pi(A) - \alpha,$$

and hence rearranging, we get that

$$\pi(B) \geq \alpha.$$

We will now show that for a constant θ to be determined later we have that

$$\max_z \mathbb{E}_z[\tau_B] > \theta t. \tag{6.1}$$

We will show that for a θ to be specified later, assuming

$$\max_z \mathbb{E}_z[\tau_B] \leq \theta t \tag{6.2}$$

will yield a contradiction.

By Markov's inequality, (6.2) implies that

$$\mathbb{P}_x(\tau_B \geq 2\theta t) \leq \frac{1}{2}. \tag{6.3}$$

For any positive integer M we have that

$$\mathbb{P}_x(\tau_B \geq 2M\theta t) = \mathbb{P}_x(\tau_B \geq 2M\theta t | \tau_B \geq 2(M-1)\theta t)\mathbb{P}_x(\tau_B \geq 2(M-1)\theta t),$$

and hence iterating we get that

$$\mathbb{P}_x(\tau_B \geq 2M\theta t) \leq \frac{1}{2^M}. \tag{6.4}$$

By the memoryless property of the Geometric distribution and the strong Markov property applied at the stopping time τ_B, we get that

$$\begin{aligned}
\mathbb{P}_x(X_{Z_t} \in A) &\geq \mathbb{P}_x(\tau_B \leq 2\theta M t, Z_t \geq \tau_B, X_{Z_t} \in A) \\
&\geq \mathbb{P}_x(\tau_B \leq 2\theta M t, Z_t \geq \tau_B)\mathbb{P}_x(X_{Z_t} \in A | \tau_B \leq 2\theta M t, Z_t \geq \tau_B) \\
&\geq \mathbb{P}_x(\tau_B \leq 2\theta M t)\mathbb{P}_x(Z_t \geq \lfloor 2\theta M t \rfloor)\left(\inf_{w \in B} \mathbb{P}_w(X_{Z_t} \in A)\right),
\end{aligned}$$

where in the last inequality we used the independence between Z and τ_B. But since Z_t is a Geometric random variable, we obtain that

$$\mathbb{P}_x(Z_t \geq \lfloor 2\theta M t \rfloor) \geq \left(1 - \frac{1}{t}\right)^{2\theta M t},$$

which for $2\theta M t > 1$ gives that

$$\mathbb{P}_x(Z_t \geq \lfloor 2\theta M t \rfloor) \geq 1 - 2\theta M. \tag{6.5}$$

18

((6.2) implies that $\theta t \geq 1$, so certainly $2\theta Mt > 1$.)

We now set $\theta = \frac{1}{2M2^M}$. Using (6.3) and (6.5) we deduce that

$$\mathbb{P}_x(X_{Z_t} \in A) \geq \left(1 - 2^{-M}\right)^2 (\pi(A) - \alpha).$$

Since $\gamma > \alpha$, we can take M large enough so that $\left(1 - 2^{-M}\right)^2 (\pi(A) - \alpha) > \pi(A) - \gamma$, and we get a contradiction to (6.2).

Thus (6.1) holds; since $\pi(B) \geq \alpha$, this completes the proof. $\qquad \square$

Proof of Theorem 1.1. Combining Theorem 2.10 with Theorem 6.1 gives the result in the reversible case. $\qquad \square$

7 Equivalence between Cesàro mixing and t_{G}

In this section we will show that the notion of mixing at a geometric time defined in Section 2 and the Cesàro mixing used by Lovász and Winkler [8] are equivalent for all chains. First, let us recall the definition of Cesàro mixing. Let U_t be a random variable independent of the chain uniform on $\{1, \ldots, t\}$. We define

$$t_{\mathrm{Ces}} = \min \left\{ t \geq 0 : \max_x \|\mathbb{P}_x(X_{U_t} = \cdot) - \pi\| \leq \frac{1}{4} \right\}.$$

Proposition 7.1. *For all chains $t_{\mathrm{G}} \asymp t_{\mathrm{Ces}}$.*

Proof. For each s, let U_s be a uniform random variable in $\{1, \ldots, s\}$ and Z_s an independent geometric random variable of mean s.

We will first show that there exists a positive constant c_1 such that

$$t_{\mathrm{Ces}} \leq c_1 t_{\mathrm{G}}. \tag{7.1}$$

Let $t = t_{\mathrm{G}}(1/8)$, then for all x

$$\|\mathbb{P}_x(X_{Z_t} = \cdot) - \pi\| \leq \frac{1}{8}. \tag{7.2}$$

From Claims 5.1 and 5.2 we get that

$$\|\mathbb{P}_x(X_{U_{8t}} = \cdot) - \mathbb{P}_x(X_{U_{8t}+Z_t} = \cdot)\| \leq \|\mathbb{P}_x(U_{8t} = \cdot) - \mathbb{P}_x(U_{8t} + Z_t = \cdot)\| \leq \frac{1}{8}.$$

By the triangle inequality for total variation we deduce

$$\|\mathbb{P}_x(X_{U_{8t}} = \cdot) - \pi\| \leq \|\mathbb{P}_x(X_{U_{8t}} = \cdot) - \mathbb{P}_x(X_{U_{8t}+Z_t} = \cdot)\| + \|\mathbb{P}_x(X_{U_{8t}+Z_t} = \cdot) - \pi\|$$

From (7.2) and Claim 2.1 it follows that

$$\|\mathbb{P}_x(X_{U_{8t}+Z_t} = \cdot) - \pi\| \leq \|\mathbb{P}_x(X_{Z_t} = \cdot) - \pi\| \leq \frac{1}{8}.$$

Hence, we conclude

$$\|\mathbb{P}_x(X_{U_{8t}} = \cdot) - \pi\| \leq \frac{1}{4},$$

which gives that $t_{\text{Ces}} \leq 8t$. From Corollary 5.6 we get that there exists a constant c such that $t_{\text{G}}(1/8) \leq ct_{\text{G}}$ and this concludes the proof of (7.1).

We will now show that there exists a positive constant c_2 such that

$$t_{\text{G}} \leq c_2 t_{\text{Ces}}.$$

Let $t = t_{\text{Ces}}$, i.e. for all x

$$\|\mathbb{P}_x(X_{U_t} = \cdot) - \pi\| \leq \frac{1}{4}. \tag{7.3}$$

From Claims 5.1 and 5.2 we get that

$$\|\mathbb{P}_x(X_{Z_{8t}} = \cdot) - \mathbb{P}_x(X_{U_t + Z_{8t}} = \cdot)\| \leq \|\mathbb{P}_x(Z_{8t} = \cdot) - \mathbb{P}_x(Z_{8t} + U_t = \cdot)\| \leq \frac{1}{8}.$$

So, in the same way as in the proof of (7.1) we obtain

$$\|\mathbb{P}_x(X_{Z_{8t}} = \cdot) - \pi\| \leq \frac{3}{8}.$$

Hence, we deduce that $t_{\text{G}}(3/8) \leq 8t$ and from Corollary 5.6 again there exists a positive constant c' such that $t_{\text{G}} \leq c' t_{\text{G}}(3/8)$ and this finishes the proof. $\qquad\square$

8 A new proof of $t_{\text{prod}} \asymp t_{\text{L}}$ for reversible chains

Recall the definition $t_{\text{prod}} = \max_{x,A} \pi(A)\mathbb{E}_x[\tau_A]$ from Remark 1.2. As noted there, Aldous [2] showed the equivalence between the mixing time t_{cts} of a continuous time reversible chain and t_{prod}. Using the equivalence $t_{\text{L}} \asymp t_{\text{cts}}$ (see [6, Theorem 20.3]) it follows that for a reversible chain $t_{\text{prod}} \asymp t_{\text{L}}$. In this section we give a direct proof. Recall that $t_{\text{prod}} \geq ct_{\text{L}}$ for a reversible chain, where c is a positive constant, follows from Theorem 1.1.

We will first state and prove a preliminary lemma, which is a variant of Kac's lemma (see for instance [6, Lemma 21.13]). To that end we define for all k and all sets A

$$\tau_A^+ = \min\{t \geq 1 : X_t \in A\} \quad \text{and} \quad \tau_A^{(k)} = \min\{t \geq k : X_t \in A\}.$$

Lemma 8.1. *We have that*

$$\sum_{x \in A} \pi(x)\mathbb{E}_x[\tau_A^{(k)}] \leq k.$$

Proof. Let $\hat{P}$ be the transition matrix of the reversed chain, i.e.

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

Then for all $t \geq k$ and $x_0, \ldots, x_t$ in the state space S, we have

$$\pi(x_0) \prod_{i=1}^{t} P(x_{i-1}, x_i) = \pi(x_t) \prod_{i=1}^{t} \hat{P}(x_i, x_{i-1}).$$

20

Summing over all $x_0 = x \in A, x_1 \in S, \ldots, x_{k-1} \in S, x_k \notin A, \ldots, x_{t-1} \notin A, x_t = y \in S$ we obtain

$$\sum_{x \in A} \pi(x) \mathbb{P}_x(\tau_A^{(k)} \geq t) \leq \sum_y \pi(y) \hat{\mathbb{P}}_y(\hat{\tau}_A^+ \in (t-k, t]), \tag{8.1}$$

where $\hat{\tau}_A^+$ stands for the first positive entrance time to A for the reversed chain. Summing (8.1) over all t we get that

$$\sum_{x \in A} \pi(x) \mathbb{E}_x[\tau_A^{(k)}] = \sum_t \sum_{x \in A} \pi(x) \mathbb{P}_x(\tau_A^{(k)} \geq t) \leq \sum_t \sum_y \pi(y) \sum_{s=t-k+1}^{t} \hat{\mathbb{P}}_y(\hat{\tau}_A^+ = s)$$

$$= \sum_y \pi(y) \sum_s \sum_{t=s}^{s+k-1} \hat{\mathbb{P}}_y(\hat{\tau}_A^+ = s) = \sum_y \pi(y) \sum_s k \hat{\mathbb{P}}_y(\hat{\tau}_A^+ = s) = k.$$

$\square$

Proof of $t_{\mathrm{prod}} \leq c' t_{\mathrm{L}}$. To simplify notation, let the chain X be lazy and reversible. From Lemma 8.1 and Markov's inequality it follows that for all k and all sets A

$$\mathbb{P}_{\pi|_A}\left(\tau_A^{(k)} \geq \frac{2k}{\pi(A)}\right) \leq \frac{1}{2}, \tag{8.2}$$

where $\pi|_A$ stands for the restriction of the stationary measure π on A.

Take now $k = 2t_{\mathrm{L}}$. Then using submultiplicativity we get that $\bar{d}_L(k) \leq \bar{d}_L(t_{\mathrm{L}})^2 \leq \frac{1}{4}$. Let $X_0 \sim \pi|_A$ and $z \in S$. Then

$$\|P_L^k(X_0, \cdot) - P_L^k(z, \cdot)\| \leq \frac{1}{4}.$$

We can couple the two chains, $X_k, X_{k+1}, \ldots$ with $X_0 \sim \pi|_A$ and $Y_k, Y_{k+1}, \ldots$ with $Y_0 = z$, so that they disagree with probability $\|P_L^k(X_0, \cdot) - P_L^k(z, \cdot)\|$.

Thus we obtain

$$\left| \mathbb{P}_{\pi|_A}\left(\tau_A^{(k)} \geq \frac{2k}{\pi(A)}\right) - \mathbb{P}_z\left(\tau_A^{(k)} \geq \frac{2k}{\pi(A)}\right) \right| \leq \mathbb{P}_{\pi|_A, z}\left(\left\{\tau_A^{(k)}(X) \geq \frac{2k}{\pi(A)}\right\} \triangle \left\{\tau_A^{(k)}(Y) \geq \frac{2k}{\pi(A)}\right\}\right)$$

$$\leq \mathbb{P}(\text{coupling fails}) \leq \frac{1}{4},$$

and hence using (8.2) we get that

$$\mathbb{P}_z\left(\tau_A^{(k)} \geq \frac{2k}{\pi(A)}\right) \leq \frac{3}{4}.$$

Therefore for all z we have that

$$\mathbb{P}_z\left(\tau_A \geq \frac{2k}{\pi(A)}\right) \leq \mathbb{P}_z\left(\tau_A^{(k)} \geq \frac{2k}{\pi(A)}\right) \leq \frac{3}{4}.$$

By performing independent experiments we see that τ_A is stochastically dominated by $\frac{2k}{\pi(A)} \mathrm{Geo}\left(\frac{3}{4}\right)$, where Geo stands for a Geometric random variable, and hence for all z we get that

$$\mathbb{E}_z[\tau_A] \leq \frac{8k}{3\pi(A)} = \frac{16t_{\mathrm{L}}}{3\pi(A)}$$

and this finishes the proof. $\square$

9 Application to robustness of mixing

Theorem 9.1. *Let T be a finite tree on n vertices with unit conductances on the edges. Let $\widetilde{T}$ be a tree on the same set of vertices and edges as T but with conductances on the edges satisfying $c \le c(x,y) \le c'$, for all edges $e = \langle x,y \rangle$, where c and c' are two positive constants. Then the mixing time of the lazy random walk on T and on $\widetilde{T}$ are equivalent, i.e. in our notation, $t_{\mathrm{L}}(T) \asymp t_{\mathrm{L}}(\widetilde{T})$.*

Before proving the theorem, we state and prove two lemmas which will be used in the proof but are also of independent interest.

Lemma 9.2. *Let T be a finite tree with edge conductances. For each subset A of vertices and any vertex v we have*

$$\max_x \mathbb{E}_x[\tau_A] \le t_v \left(1 + \frac{1}{\pi(A)} \right),$$

where τ_A stands for the first hitting time of A by a simple random walk on T and $t_v = \max_x \mathbb{E}_x[\tau_v]$.

Proof. If $v \in A$, then the result is clear, so we assume that $v \notin A$.

For all x we have

$$\mathbb{E}_x[\tau_A] \le \mathbb{E}_x[\tau_v] + \mathbb{E}_v[\tau_A] \le t_v + \mathbb{E}_v[\tau_A].$$

Thus it suffices to show that

$$\mathbb{E}_v[\tau_A] \le \frac{t_v}{\pi(A)}. \tag{9.1}$$

In order to show that, we are going to look at excursions of the random walk from v. Defining Z_A to be the time that the walk spends in A in an excursion from v, i.e., $Z_A = \sum_{t=1}^{\tau_v^+} \mathbf{1}(X_t \in A)$, we can write

$$\mathbb{P}_v(\tau_A < \tau_v^+) = \frac{\mathbb{E}_v[Z_A]}{\mathbb{E}_v[Z_A | Z_A > 0]}.$$

Clearly

$$\mathbb{E}_v[Z_A] = \frac{\pi(A)}{\pi(v)} \quad \text{and} \quad \mathbb{E}_v[Z_A | Z_A > 0] \le t_v.$$

Hence

$$\mathbb{P}_v(\tau_A < \tau_v^+) \ge \frac{\pi(A)}{\pi(v)} \frac{1}{t_v}.$$

Therefore we get

$$\mathbb{E}_v[\tau_A] \le \mathbb{E}_v\left[\sum_{i=1}^N \ell_i \right],$$

where N is a geometric random variable of success probability $\frac{\pi(A)}{\pi(v)} \frac{1}{t_v}$ and ℓ_i is the length of the i-th excursion from v. By Wald's identity we have

$$\mathbb{E}_v[\tau_A] \le \mathbb{E}_v[N] \mathbb{E}_v[\tau_v^+] \le \frac{\pi(v) t_v}{\pi(A)} \frac{1}{\pi(v)} = \frac{t_v}{\pi(A)}$$

and this completes the proof. $\square$

We call a node v in T *central* if each component of $T - \{v\}$ has stationary probability at most $1/2$. It is easy to see that central nodes exist. Indeed, for any node u of the tree denote by $C(u)$ the component of $T - \{u\}$ with the largest stationary probability. Now consider the vertex u^* that achieves $\min_u |\pi(C(u))|$. This is clearly a central node, since if $\pi(C(u^*)) > 1/2$, then the neighbour $w \in C(u^*)$ of u^* would satisfy $\pi(C(w)) < \pi(C(u^*))$, contradicting the choice of u^*.

Lemma 9.3. *Let T be a tree on n vertices with conductances on the edges. Then for any central node v of T*

$$t_{\mathrm{L}} \asymp t_v,$$

where $t_v = \max_x \mathbb{E}_x[\tau_v]$.

Proof. First of all from Lemma 9.2 and Theorem 1.1 we obtain that for any central node v

$$t_{\mathrm{L}} \leq c t_v, \tag{9.2}$$

for an absolute constant c.

To finish the proof of the lemma we have to show that for any central node v

$$t_{\mathrm{L}} \geq c t_v, \tag{9.3}$$

for a positive absolute constant c.

It is easy to see that $\mathbb{E}_x[\tau_v] = \mathbb{E}_x[\tau_B]$, for $x \neq v$, where B is the union of $\{v\}$ and the components of $T - \{v\}$ that do not contain x. The definition of a central node gives that $\pi(B) \geq 1/2$. Hence,

$$t_v \leq t_{\mathrm{H}}(1/2). \tag{9.4}$$

Inequality (9.3) now follows from Theorem 1.1. $\square$

We now recall a formula from [1, Lemma 1, Chapter 5] for the expected hitting time on trees.

Lemma 9.4. *Let T be a finite tree with edge conductances $c(u, v)$, for all edges $\langle u, v \rangle$. Let x and y be two vertices of T and let $\{v_0 = x, v_1, \ldots, v_n = y\}$ be the unique path joining them. Let $T_x(z)$ be the union of $\{z\}$ and the connected component of $T - \{z\}$ containing x. Writing $C_i = \sum_{w, z \in T_x(v_{i+1})} c(w, z)$, we then have*

$$E_x[\tau_y] = \sum_{i=0}^{n-1} \left(\frac{C_i}{c(v_i, v_{i+1})} - 1 \right).$$

Proof of Theorem 9.1. From Lemma 9.4 and the boundedness of the conductances we get that for any two vertices x and v

$$\mathbb{E}_x[\tau_v] \asymp \mathbb{E}_x[\widetilde{\tau}_v],$$

where $\widetilde{\tau}$ denotes hitting times for the random walk on $\widetilde{T}$.

Lemma 9.3 then finishes the proof. $\square$

We end this section with another application of our results on the robustness of mixing when the probability of staying in place changes in a bounded way. The following corollary answers a question suggested to us by K. Burdzy (private communication).

Corollary 9.5. *Let P be an irreducible transition matrix on the state space E and suppose that $(a(x,x))_{x \in E}$ satisfy $c_1 \leq a(x,x) \leq c_2$ for all $x \in E$, where $c_1, c_2 \in (0,1)$. Let Q be the transition matrix of the Markov chain with transitions: when at x it stays at x with probability $a(x,x)$. Otherwise, with probability $1 - a(x,x)$ it jumps to a state $y \in E$ with probability $P(x,y)$. We then have*

$$t_{\mathrm{mix}}(Q) \asymp t_{\mathrm{L}},$$

where $t_{\mathrm{mix}}(Q)$ is the mixing time of the transition matrix Q.

Proof. Since the loop probabilities $a(x,x)$ are bounded from below and above, it follows that if $\widetilde{\pi}$ is the stationary probability of the matrix Q, then $\widetilde{\pi} \asymp \pi$. As we noted in the Introduction, the lower bound of Theorem 1.1 is always true and thus we have

$$\max_{x, A: \widetilde{\pi}(A) \geq 1/4} \mathbb{E}_x[\widetilde{\tau}_A] \lesssim t_{\mathrm{mix}}(Q). \tag{9.5}$$

For every $y \in E$ let $(\xi_i^{(y)})_{i \in \mathbb{N}}$ be i.i.d. geometric random variables of mean $1/a(y,y)$. Then we can write

$$\widetilde{\tau}_A = \sum_{y \in E} \sum_{i=1}^{L_y} \xi_i^{(y)}, \tag{9.6}$$

where L_y is the local time at y up to the first hitting time of A by the chain with transition matrix P. Wald's identity gives

$$\mathbb{E}_x[\widetilde{\tau}_A] = \sum_{y \in E} \frac{\mathbb{E}_x[L_y]}{a(y,y)}.$$

If τ_A' is the hitting of A by the lazy version of the chain, i.e. taking $a(y,y) = 1/2$ for all y, then using the assumption on the boundedness of the probabilities $(a(y,y))$ we get

$$\mathbb{E}_x[\widetilde{\tau}_A] \asymp \mathbb{E}_x[\tau_A'].$$

From (9.6) applying Wald's identity again we deduce

$$\mathbb{E}_x[\tau_A'] = 2\mathbb{E}_x[\tau_A],$$

where τ_A is the first hitting time of the set A by the Markov chain with transition matrix P. Hence using Theorem 1.1 and (9.5) we deduce that

$$t_{\mathrm{mix}}(Q) \gtrsim t_{\mathrm{L}}.$$

It remains to show

$$t_{\mathrm{mix}}(Q) \lesssim t_{\mathrm{L}}. \tag{9.7}$$

Using Proposition 2.11 we get that

$$t_{\mathrm{mix}}(A) \lesssim t_{\mathrm{ave}}.$$

This together with Theorem 1.4 finishes the proof of (9.7). $\qquad\square$

10 Examples and Questions

We start this section with examples that show that the reversibility assumption in Theorem 1.1 and Corollary 2.5 is essential.

Example 10.1. Biased random walk on the cycle.
Let $\mathbb{Z}_n = \{1, 2, \ldots, n\}$ denote the n-cycle and let $P(i, i+1) = \frac{2}{3}$ for all $1 \leq i < n$ and $P(n, 1) = \frac{2}{3}$. Also $P(i, i-1) = \frac{1}{3}$, for all $1 < i \leq n$, and $P(1, n) = \frac{1}{3}$. Then it is easy to see that the mixing time of the lazy random walk is of order n^2, while the maximum hitting time of large sets is of order n. Also, in this case $t_{\text{stop}} = O(n)$, since for any starting point, the stopping time that chooses a random target according to the stationary distribution and waits until it hits it, is stationary and has mean of order n. This example demonstrates that for non-reversible chains, t_{H} and t_{stop} can be much smaller than t_{L}.

Example 10.2. The greasy ladder.
Let $S = \{1, \ldots, n\}$ and $P(i, i+1) = \frac{1}{2} = 1 - P(i, 1)$ for $i = 1, \ldots, n-1$ and $P(n, 1) = 1$. Then it is easy to check that

$$\pi(i) = \frac{2^{-i}}{1 - 2^{-n}}$$

is the stationary distribution and that t_{L} and t_{H} are both of order 1.
This example was presented in Aldous [2], who wrote that t_{stop} is of order n. We give an easy proof here. Essentially the same example is discussed by Lovász and Winkler [8] under the name "the winning streak".
Let τ_π be the first hitting time of a stationary target, i.e. a target chosen according to the stationary distribution. Then starting from 1, this stopping time achieves the minimum in the definition of t_{stop}, i.e.

$$\mathbb{E}_1[\tau_\pi] = \min\{\mathbb{E}_1[\Lambda] : \Lambda \text{ is a stopping time s.t. } \mathbb{P}_1(X_\Lambda \in \cdot) = \pi(\cdot)\}.$$

Indeed, starting from 1 the stopping time τ_π has a halting state, which is n, and hence from Theorem 3.2 we get the mean optimality. By the random target lemma [1] and [6] we get that $\mathbb{E}_i[\tau_\pi] = \mathbb{E}_1[\tau_\pi]$, for all $i \leq n$. Since for all i we have that

$$\mathbb{E}_i[\tau_\pi] \geq \min\{\mathbb{E}_i[\Lambda] : \Lambda \text{ is a stopping time s.t. } \mathbb{P}_i(X_\Lambda \in \cdot) = \pi(\cdot)\},$$

it follows that $t_{\text{stop}} \leq \mathbb{E}_1[\tau_\pi]$. But also $\mathbb{E}_1[\tau_\pi] \leq t_{\text{stop}}$, and hence $t_{\text{stop}} = \mathbb{E}_1[\tau_\pi]$. By straightforward calculations, we get that $\mathbb{E}_1[T_i] = 2^i(1 - 2^{-n})$, for all $i \geq 2$, and hence

$$t_{\text{stop}} = \mathbb{E}_1[\tau_\pi] = \sum_{i=2}^{n} 2^i(1 - 2^{-n}) \frac{2^{-i}}{1 - 2^{-n}} = n - 1.$$

This example shows that for a non-reversible chain t_{stop} can be much bigger than t_{L} or t_{H}.

Question 10.3. The equivalence $t_{\text{H}}(\alpha) \asymp t_{\text{L}}$ in Theorem 1.1 is not valid for $\alpha > \frac{1}{2}$, since for two n-vertex complete graphs with a single edge connecting them, t_{L} is of order n^2 and $t_{\text{H}}(\alpha)$ is at most n for any $\alpha > 1/2$. Does the equivalence $t_{\text{H}}(1/2) \asymp t_{\text{L}}$ hold for all reversible chains?
(After this question was posed in the first version of this paper, it was answered positively by Griffiths et al [5].)

Acknowledgments

References

[1] David Aldous and J. Fill. *Reversible Markov Chains and Random Walks on Graphs.* In preparation, http://www.stat.berkeley.edu/~aldous/RWG/book.html.

[2] David J. Aldous. Some inequalities for reversible Markov chains. *J. London Math. Soc. (2),* 25(3):564–576, 1982.

[3] J. R. Baxter and R. V. Chacon. Stopping times for recurrent Markov processes. *Illinois J. Math.,* 20(3):467–475, 1976.

[4] J. Ding and Y. Peres. Sensitivity of mixing times, 2013. arXiv:1304.0244.

[5] S. Griffiths, R. J. Kang, R. Imbuzeiro Oliveira, and V. Patel. Tight inequalities among set hitting times in Markov chains, 2012. to appear in *Proc. AMS.*

[6] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times.* American Mathematical Society, Providence, RI, 2009. With a chapter by James G. Propp and David B. Wilson.

[7] László Lovász and Peter Winkler. Efficient stopping rules for markov chains. In *Proceedings of the twenty-seventh annual ACM symposium on Theory of computing,* STOC '95, pages 76–82, New York, NY, USA, 1995. ACM.

[8] László Lovász and Peter Winkler. Mixing times. In *Microsurveys in discrete probability (Princeton, NJ, 1997),* volume 41 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.,* pages 85–133. Amer. Math. Soc., Providence, RI, 1998.

[9] Roberto Imbuzeiro Oliveira. Mixing and hitting times for finite Markov chains. *Electron. J. Probab.,* 17:no. 70, 12, 2012.