Mixing times and hitting times: lecture notes

Yuval Peres* Perla Sousi†

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 $d_L(t) = \max_x \|P_L^t(i,\cdot) - \pi\|$, then we can define the mixing time of the lazy chain as follows:

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

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

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

where τ_A stands for the first hitting time of the set A.

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

$$t_{\mathrm{L}} \geq c'_\alpha t_{\mathrm{H}}(\alpha).$$

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

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

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

$$c'_\alpha t_{\mathrm{H}}(\alpha) \le t_{\mathrm{L}} \le c_\alpha t_{\mathrm{H}}(\alpha).$$

Remark 1.2. The proof of Theorem 1.1 is somewhat technical and is given in [7]. Instead in Theorem 6.1 we will prove the equivalence of $t_{\mathrm{H}}(\alpha)$ to t_{G}, mixing at a geometric time, defined in the next section.

Remark 1.3. Aldous in [2] showed that the mixing time, t_{cts}, of a continuous time reversible chain is equivalent to $t_{\mathrm{prod}} = \max_{x,A:\pi(A)>0} \pi(A)\mathbb{E}_x[\tau_A]$. The inequality $t_{\mathrm{prod}} \le c_1 t_{\mathrm{cts}}$, for a positive constant c_1, which was the hard part in Aldous' proof, follows from Theorem 1.1 and the equivalence $t_{\mathrm{L}} \asymp t_{\mathrm{cts}}$ (see [4, Theorem 20.3]).

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 [5, 6]. It is defined as follows:

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

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.

Definition 2.1. 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 $cs \le r \le c's$ for every chain in $\mathcal{M}$.

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

$$t_{\mathrm{G}} = t_{\mathrm{G}}(1/4) = \min\{t \ge 0 : d_G(t) \le 1/4\}.$$

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

Lemma 2.2. *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 2.2. 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 $\mathbb{P}(Z_{t+1} = Z_t + k | Z_t) = \left(\frac{t}{t+1}\right)^{k-1} \frac{1}{t+1}$.

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 $d_G(t)$. □

Lemma 2.3. *For all chains we have that*

$$t_{\mathrm{G}} \leq 4t_{\mathrm{stop}} + 1.$$

The converse of Lemma 2.3 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 2.4. *There exists a positive constant c_4 such that for all reversible chains*

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

In particular, $t_{\mathrm{stop}} \leq c_4 t_{\mathrm{G}}$.

We will give the proofs of Lemmas 2.3 and 2.4 in Section 5.

3 Stationary stopping times

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

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

These vectors clearly satisfy

$$0 \leq \sigma(t) \leq \theta(t), \quad (\theta(t) - \sigma(t))\mathbf{P} = \theta(t+1) \ \forall t; \ \theta_0 = \mu. \tag{3.3}$$

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

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

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

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

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 $\mathcal{F}_s = \sigma(X_0, U_0, \dots, X_s, U_s)$. Also, equations (3.2) are satisfied. Indeed, setting $x_t = x$ we have

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

since $\theta_0(y) = \mu(y)$ 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 $(\theta(t), \sigma(t); t \geq 0)$. Let more generally the starting distribution be 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

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

and next we let $\Sigma_x(t) = \sum_{s \leq t} \sigma_x(s)$. 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 $S \le T_z$ 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 $T \le T_z$ a.s. For each x we define

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

Take z such that $t_z = \max_x t_x \le \infty$. We will show that $T \le T_z$ 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

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

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 $\mathbb{P}_\mu(T \le T_z) = 1$.

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

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

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. □

4 Mixing at a geometric time

Proof of Lemma 2.3. 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 we have that

$$\begin{aligned}
t\pi(y) &= \sum_{s=1}^{\infty} \left(1 - \frac{1}{t}\right)^{s-1} \mathbb{P}_x(X_{\tau+s} = y) = \sum_{s=1}^{\infty} \left(1 - \frac{1}{t}\right)^{s-1} \sum_{\ell=0}^{\infty} \mathbb{P}_x(X_{\ell+s} = y, \tau = \ell) \\
&= \sum_{m=1}^{\infty} \sum_{\ell=0}^{m-1} \left(1 - \frac{1}{t}\right)^{m-\ell-1} \mathbb{P}_x(X_m = y, \tau = \ell) \geq \sum_{m=1}^{\infty} \left(1 - \frac{1}{t}\right)^{m-1} \mathbb{P}_x(X_m = y, \tau < m).
\end{aligned}$$

As in Section 2, let Z_t be a geometric random variable of success probability $\frac{1}{t}$. Then

$$\nu_t(y) := \mathbb{P}_x(X_{Z_t} = y) = \sum_{s=1}^{\infty} \frac{1}{t} \left(1 - \frac{1}{t}\right)^{s-1} \mathbb{P}_x(X_s = y).$$

Thus we obtain

$$\begin{aligned}
t\nu_t(y) - t\pi(y) &= \sum_{s=1}^{\infty} \left(1 - \frac{1}{t}\right)^{s-1} \left(\mathbb{P}_x(X_s = y) - \mathbb{P}_x(X_s = y, \tau < s)\right) \\
&= \sum_{s=1}^{\infty} \left(1 - \frac{1}{t}\right)^{s-1} \mathbb{P}_x(X_s = y, \tau \geq s).
\end{aligned}$$

Summing over all y such that $t\nu_t(y) - t\pi(y) > 0$, we get that

$$t\|\nu_t - \pi\| \leq \sum_{s=1}^{\infty} \left(1 - \frac{1}{t}\right)^{s-1} \mathbb{P}_x(\tau \geq s) \leq \mathbb{E}_x[\tau].$$

Therefore, if we take $t \geq 4\mathbb{E}_x[\tau]$, then we get that

$$\|\nu_t - \pi\| \leq \frac{1}{4},$$

and hence $t_{\mathrm{G}} \leq 4t_{\mathrm{stop}} + 1$. □

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

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

We now introduce some notation and a preliminary result that will be used in the proof of Lemma 2.4. 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.

Lemma 4.1. *For a reversible Markov chain we have that*

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

Proof. A proof of this result can be found in [1, Chapter 4, Lemma 7] or [4, Lemma 4.11 and Lemma 19.3]. □

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 $\bar{d}_G(t)$.

Lemma 4.2. *For all t we have that*

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

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

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

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 5.1. □

We now define

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

Lemma 4.3. *There exists a positive constant c so that for every chain*

$$t_{\mathrm{stop}} \leq c t_{\mathrm{sep}}.$$

Proof. Fix $t = t_{\mathrm{sep}}$. 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

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

Hence, we can write

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

where $\nu_x(\cdot)$ is a probability measure.
We can thus construct a stopping time $S \in \{1, 2, \ldots\}$ such that for all x

$$\mathbb{P}_x(Y_S \in \cdot, S = 1) = (1 - 3/4)\pi(\cdot)$$

and by induction on k such that

$$\mathbb{P}_x(Y_S \in \cdot, S = k) = (3/4)^{k-1}(1/4)\pi(\cdot).$$

Hence, it is clear 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 inequality 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

$$t_{\mathrm{stop}} \leq 8 t_{\mathrm{sep}}.$$

□

Proof of Lemma 2.4. From Lemma 5.2 we get that

$$t_{\mathrm{sep}} \leq 2 t_N.$$

Finally Lemma 5.3 completes the proof. □

Remark 4.4. 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 Cesaro mixing and it has been analyzed by Lovász and Winkler in [6]. From [4, Theorem 6.15] and the lemmas above we get the equivalence between the Cesaro mixing and the mixing of the lazy chain.

Remark 4.5. From the remark above we see that the mixing at a geometric time and the Cesaro 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 2.2, which is not true for the corresponding total variation distance for the Cesaro 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 [4, 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 4.6. *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 2.2 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

$$\bar{d}_G(2t) \leq \bar{d}_G(t) \max_{x,y} \|\mathbb{P}_x(X_{Z_{2t}-Z_t} = \cdot) - \mathbb{P}_y(X_{Z_{2t}-Z_t} = \cdot)\|. \tag{4.1}$$

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 (5.1) becomes

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

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 $\bar{d}_G(t) \leq \beta$, we get that

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

and hence iterating we deduce the desired inequality. □

Combining Lemma 5.6 with Lemma 5.2 we get the following:

Corollary 4.7. *If t is such that $d_G(t) \leq \beta$, then for all k we have that*

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

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

5 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 5.1. *Let $\alpha < 1/2$. For every chain $t_{\mathrm{G}} \asymp t_{\mathrm{H}}(\alpha)$. (The implied constants depend on α.)*

Proof. We will first show that $t_{\mathrm{G}} \geq c t_{\mathrm{H}}(\alpha)$. By Corollary 5.7 there exists $k = k(\alpha)$ so that $d_G(2^k t_{\mathrm{G}}) \leq \frac{\alpha}{2}$. Let $t = 2^k t_{\mathrm{G}}$. Then for any starting point x we have that

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

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

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

and hence this gives that

$$\max_{x, A: \pi(A) \geq \alpha} \mathbb{E}_x[\tau_A] \leq \frac{2}{\alpha} 2^k t_{\mathrm{G}}.$$

In order to show the other direction, let $t' < t_{\mathrm{G}}$. Then $d_G(t') > 1/4$. For a given $\alpha < 1/2$, we fix $\gamma \in (\alpha, 1/2)$. From Corollary 5.7 we have that there exists a 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

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

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

$$\max_z \mathbb{E}_z[\tau_B] > \theta t. \tag{5.1}$$

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

$$\max_z \mathbb{E}_z[\tau_B] \leq \theta t \tag{5.2}$$

will yield a contradiction.
By Markov's inequality, (6.2) implies that

$$\mathbb{P}_x(\tau_B \geq 2\theta t) \leq \frac{1}{2}. \tag{5.3}$$

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

$$\mathbb{P}_x(\tau_B \geq 2M\theta t) \leq \frac{1}{2^M}. \tag{5.4}$$

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

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

But since Z_t is a Geometric random variable, we obtain that

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

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

$$\mathbb{P}_x(Z_t \geq 2\theta M t) \geq 1 - 2\theta M. \tag{5.5}$$

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

6 Examples and Questions

We start this section with examples that show that the reversibility assumption is essential.

Example 6.1. 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 6.2. 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 [6] 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 [4] 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 $t_{\mathrm{stop}} \leq \mathbb{E}_1[\tau_\pi]$. But also $\mathbb{E}_1[\tau_\pi] \leq t_{\mathrm{stop}}$, and hence $t_{\mathrm{stop}} = \mathbb{E}_1[\tau_\pi]$. By straightforward calculations, we get that $\mathbb{E}_1[T_i] = 2^i(1 - 2^{-n})$, for all $i \geq 2$, and hence

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

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

Question 6.3. The equivalence $t_{\mathrm{H}}(\alpha) \asymp t_{\mathrm{L}}$ 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 $t_{\mathrm{H}}(\alpha)$ is at most n for any $\alpha > 1/2$. Does the equivalence $t_{\mathrm{H}}(1/2) \asymp t_{\mathrm{L}}$ hold for all reversible chains?

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] 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.

[5] 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.

[6] 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.

[7] Y. Peres and P. Sousi. Mixing times are hitting times of large sets. preprint, 2011.