# A characterization of $L_2$ mixing and hypercontractivity via hitting times and maximal inequalities

Jonathan Hermon *    Yuval Peres †

**Abstract**

There are several works characterizing the total-variation mixing time of a reversible Markov chain in term of natural probabilistic concepts such as stopping times and hitting times. In contrast, there is no known analog for the $L_2$ mixing time, $\tau_2$ (while there are sophisticated analytic tools to bound $\tau_2$, in general they do not determine $\tau_2$ up to a constant factor and they lack a probabilistic interpretation). In this work we show that $\tau_2$ can be characterized up to a constant factor using hitting times distributions. We also derive a new extremal characterization of the Log-Sobolev constant, $c_{\mathrm{LS}}$, as a weighted version of the spectral gap. This characterization yields a probabilistic interpretation of $c_{\mathrm{LS}}$ in terms of a hitting time version of hypercontractivity. As applications of our results, we show that (1) for every reversible Markov chain, $\tau_2$ is robust under addition of self-loops with bounded weights, and (2) for weighted nearest neighbor random walks on trees, $\tau_2$ is robust under bounded perturbations of the edge weights.

**Keywords:** Mixing-time, finite reversible Markov chains, maximal inequalities, hitting times, hypercontractivity, Log-Sobolev inequalities, relative entropy, robustness of mixing times.

*Department of Statistics, UC Berkeley, USA. E-mail: jonathan.hermon@stat.berkeley.edu.
†Microsoft Research, Redmond, Washington, USA. E-mail: peres@microsoft.com.

# 1 Introduction

There are numerous essentially equivalent characterizations of mixing in $L_1$ (e.g. [1, Theorem 4.6] and [17]) of a finite reversible Markov chain. Some involve natural probabilistic concepts such as couplings, stopping times and hitting times (see § 3.5). In contrast, (paraphrasing Aldous and Fill [1] last sentence of page 155, which mentions that there is no $L_2$ counterpart to [1, Theorem 4.6]) while there are several sophisticated analytic and geometric tools for bounding the $L_2$ mixing time, $\tau_2$, none of them has a probabilistic interpretation, and none of them determines $\tau_2$ up to a constant factor.

In this work we provide probabilistic characterizations in terms of hitting times distributions for the $L_2$ mixing time and also for the mixing time in relative entropy, $\tau_{\mathrm{Ent}}$ (see (3.2) and (3.5) for definitions), of a reversible Markov chain (Theorem 1.1).

While the spectral gap is a natural and simple parameter, the *Log-Sobolev constant* (see (3.11)), $c_{\mathrm{LS}}$, is a more involved quantity. When one first encounters $c_{\mathrm{LS}}$, it may seem like an artificial parameter that "magically" gives good bounds on $\tau_2$. We give a new extremal characterization of the Log-Sobolev constant as a weighted version of the spectral gap. This characterization gives a direct link between $c_{\mathrm{LS}}$ and $\tau_2$ (answering a question asked by James Lee, see Remark 1.2) and can be interpreted probabilistically as a hitting-time version of hypercontractivity (see the discussion following Fact 3.3). We note that recently, Cattiaux and Guillin [5] established a different connection between hitting times distributions and the existence of a Log-Sobolev inequality for diffusions.

We present our main results in the continuous-time setup. All of our results can be extended to the setup of discrete-time chains (see § 2.3). We note that most of our results can be extended to the general setup of ergodic Markov chains. However, working in such generality leads to many technical difficulties which we chose to avoid for the sake of clarity of presentation.

## 1.1 Characterizations of $\tau_2$ and $\tau_{\mathrm{Ent}}$ using hitting times

We now describe the aforementioned characterizations of $\tau_2$ and $\tau_{\mathrm{Ent}}$. More refined versions will be given later on in Theorems 5.1 and 5.2. Recall that for a Markov chain $(X_t)_{t\geq 0}$ with state space $\Omega$, the *hitting-time* of a set $A \subset \Omega$ is $T_A := \inf\{t : X_t \in A\}$. We say that $A$ is connected if $\mathrm{P}_a[T_b < T_{A^c}] > 0$, for all $a, b \in A$. We denote by $\mathrm{Con}_\delta$ the collection of all connected sets $A$ satisfying $\pi(A) \leq \delta$, where throughout, $\pi$ shall denote the stationary distribution of the chain. Denote

$$\rho := \max_{x\in\Omega} \rho_x \quad \text{and} \quad \rho_{\mathrm{Ent}} := \max_{x\in\Omega} \rho_{\mathrm{Ent},x}, \quad \text{where} \tag{1.1}$$

$$\begin{aligned} \rho_x &:= \min\{t : \mathrm{P}_x[T_{A^c} > t] \leq \pi(A) + \frac{1}{2}\sqrt{\pi(A)\pi(A^c)} \text{ for all } A \in \mathrm{Con}_{1/2}\}, \\ \rho_{\mathrm{Ent},x} &:= \min\{t : \mathrm{P}_x[T_{A^c} > t] \leq \min\left(\frac{C_{\mathrm{Ent}}}{|\log \pi(A)|}, \frac{99}{100}\right), \text{ for all } A \in \mathrm{Con}_{1/2}\}, \end{aligned} \tag{1.2}$$

for some absolute constant $C_{\mathrm{Ent}} > 0$ to be determined later (in § 4.1 in the discussion following (4.1)). Note that allowing $A$ above to range over all $A \subset \Omega$ such that $\pi(A) \leq 1/2$ does not change the values of $\rho_x$ and $\rho_{\mathrm{Ent},x}$.

**Theorem 1.1.** *There exist absolute constants $C_1, C_2, C_3$ such that for every irreducible reversible Markov chain on a finite state space*

$$\rho \le \tau_2 \le \rho + C_1/c_{\mathrm{LS}} \le C_2\rho. \tag{1.3}$$

$$\rho_{\mathrm{Ent}} \le \tau_{\mathrm{Ent}} \le C_3\rho_{\mathrm{Ent}}. \tag{1.4}$$

Note that in the definitions of $\rho$ and $\rho_{\mathrm{Ent}}$, the smaller $A$ is, the smaller we require the chance of not escaping it by time $\rho$ or $\rho_{\mathrm{Ent}}$, respectively, to be. In other words, the smaller $A$ is, the higher the "penalty" we assign to the case the chain did not escape from it. As we explain in § 4.1, the first inequalities in (1.3)-(1.4) are easy and even somewhat "naive".

### 1.2 A new extremal characterization of the Log-Sobolev time.

A lot of attention has been focused on inequalities that interpolate between the Log-Sobolev inequality and the Poincaré (spectral gap) inequality (e.g. [3, 15]). Using similar ideas as described above we prove a new extremal characterization (up to a constant factor) of the *Log-Sobolev constant* (Theorems 1.2), $c_{\mathrm{LS}}$ (see (3.11) for a definition). The *Log-Sobolev time* is defined as $t_{\mathrm{LS}} := 1/c_{\mathrm{LS}}$.

The aforementioned characterization has a relatively simple form which does not involve any entropy. Instead, it describes the Log-Sobolev constant as a weighted version of the spectral gap. This characterization provides some insights regarding the hierarchy of the aforementioned inequalities. Before presenting it, we first need a few definitions.

The *time-reversal* of $P$ is defined as $P^*(x,y) := \pi(y)P(y,x)/\pi(x)$. This is the dual operator of $P$ w.r.t. $L_2(\Omega,\pi)$. We say that $P$ is *reversible* if $P = P^*$. Denote $Q := (P+P^*)/2$. Note that $Q = Q^*$. The *spectral gap* of $P$, denoted by $\lambda$, is defined as the smallest non-zero eigenvalue of $I - Q$. The relaxation-time is defined as $t_{\mathrm{rel}} := 1/\lambda$. Let $A \subsetneq \Omega$. Let $Q_A$ (resp. $P_A$) be the restriction of $Q$ (resp. $P$) to $A$. Note that $Q_A$ and $P_A$ are substochastic. The spectral gap of $P_A$, denoted by $\lambda(A)$, is defined as the minimal eigenvalue of $I - Q_A$. Denote $t_{\mathrm{rel}}(A) := 1/\lambda(A)$. Denote

$$\kappa := 1/\alpha, \quad \alpha := \min_{A \in \mathrm{Con}_{1/2}} \alpha(A), \quad \text{where} \quad \alpha(A) := \lambda(A)/|\log \pi(A)|. \tag{1.5}$$

As mentioned earlier, $\alpha$ is a weighted version of $\lambda$ since ([1, Lemma 4.39] and [9, (1.4)])

$$\lambda/2 \le \min_{A \in \mathrm{Con}_{1/2}} \lambda(A) \le \lambda, \quad \text{and so} \quad t_{\mathrm{rel}} \log 2 \le \kappa. \tag{1.6}$$

**Theorem 1.2.** *For every irreducible Markov chain on a finite state space*

$$\kappa \le t_{\mathrm{LS}} \le 2(\kappa + t_{\mathrm{rel}}(1+\log 49)) \le 2(1 + (1+\log 49)/\log 2)\kappa < 17\kappa. \tag{1.7}$$

**Remark 1.1.** *The inequality $\kappa \le t_{\mathrm{LS}}$ is easy. See Lemma 4.2 in [9] for a stronger inequality. The harder and more interesting direction is $t_{\mathrm{LS}} \le C\kappa$, which is an improvement over the well-known inequality $t_{\mathrm{LS}} \le t_{\mathrm{rel}}\frac{\log[1/\pi_* - 1]}{1-2\pi_*}$, where $\pi_* := \min_{x\in\Omega}\pi(x)$ [6, Corollary A.4].*

**Remark 1.2.** *Despite the fact that $t_{\mathrm{LS}}$ is a geometric quantity, Logarithmic Sobolev inequalities have a strong analytic flavor and little probabilistic interpretation. For instance, the proof of the inequality $t_{\mathrm{LS}} \le 2\tau_2(1/e)$ [6, Corollary 3.11] (where $\tau_2(\varepsilon)$ is the $L_2$ mixing time defined in* (3.2)*) relies on Stein's interpolation Theorem for a family of analytic operators. Our analysis yields a probabilistic proof of the fact that $t_{\mathrm{LS}} \le C\tau_2$ for reversible chains. The problem of finding such a proof was posed by James Lee at the Simons institute in 2015. Indeed by Theorem 1.2 and* (3.17)*, $t_{\mathrm{LS}}/17 \le \kappa \le 3\rho \le 3\tau_2$. The second inequality is relatively easy, and is obtained by analyzing hitting times, rather than by analytic tools. As we show in § 4.1, the inequality $\rho \le \tau_2$ also has a probabilistic interpretation.*

We note that while some effort was made to make most constants explicit in order to demonstrate that they are not large, we did not attempt to optimize constants. We use the convention that $C, C', C_1, \dots$ (resp. $c, c', c_1, \dots$) denote positive absolute constants which are sufficiently large (resp. small). Different appearances of the same constant at different places may refer to different numeric values.

### 1.3 Organization of this work

In § 2 we present some applications of our main results, two of which concern robustness of mixing times and one (§ 2.3) concerns a comparison result between the ($L_2$ and relative entropy) mixing times of the discrete-time averaged chain (defined in § 2.3) and of the continuous-time chain. In § 3 we provide some background about mixing-times, the spectral gap and the Log-Sobolev constant and present some auxiliary results about maximal inequalities and hitting times. In § 4.1 we prove the lower bounds on $\tau_2$ and $\tau_{\mathrm{Ent}}$ from (1.3) and (1.4) and in § 4.2 we present a sketch of the proof of the upper bound on $\tau_2$ from (1.3). In § 5 we prove our main results (we prove Theorem 1.2 and also slightly more refined versions of the upper bounds from (1.3) and (1.4), resp.). The necessary adaptations for the discrete-time setup are given at § 5.4. In § 6 we prove the two applications from § 2 concerning robustness of mixing times (Corollary 2.1 and Theorem 2.1). We conclude with some open problems in § 7.

## 2 Applications

### 2.1 Robustness of $\tau_2$ under addition of self-loops of bounded weights.

**Corollary 2.1.** *Let $(X_t)$ be a reversible irreducible continuous-time Markov chain on a finite state space $\Omega$ with generator $G$. Let $(\tilde{X}_t)$ be a chain with generator $\tilde{G}$ obtained by multiplying for all $x \in \Omega$ the $x$th row of $G$ by some $r_x \in (1/M, M)$ (for some $M \ge 1$). Then for some absolute constant $C$ the corresponding $L_2$ and relative-entropy mixing times satisfy*

$$\tilde{\tau}_2/(CM\log M) \le \tau_2 \le (CM\log M)\tilde{\tau}_2. \tag{2.1}$$

$$\tilde{\tau}_{\mathrm{Ent}}/(CM\log M) \le \tau_{\mathrm{Ent}} \le (CM\log M)\tilde{\tau}_{\mathrm{Ent}}. \tag{2.2}$$

This corollary, proved in § 6.1, is an analog of [17, Corollary 9.5], which gives the corresponding statement for $\tau_1$. While the statement is extremely intuitive, surprisingly, it was recently shown that it may fail for simple random walk on an Eulerian digraph [4, Theorem 1.5].

Observe that the generator $G$ of a reversible chain on a finite state space $\Omega$, can be written as $r(P-I)$, where $P$ is the transition matrix of some nearest neighbor weighted random walk on a network which may contain some weighted self-loops. The operation of multiplying the $x$th row of $G$ by some $r_x \in (1/M, M)$ for all $x \in \Omega$ is the same as changing $r$ above by some constant factor and changing the weights of the self-loops by a constant factor.

**Remark 2.2.** *Similarly, one can show that under reversibility the $L_2$ mixing time in the discrete-time lazy setup is robust under changes of the holding probabilities. More precisely, for every $\delta \in (0, 1/2]$ if we consider a chain that for all $x \in \Omega$, when at state $x$ it stays put w.p. $\delta \leq a(x) \leq 1-\delta$ and otherwise moves to state $y$ w.p. $P(x,y)$ (where $P$ is reversible), then its $L_2$ mixing time can differ from the $L_2$ mixing time of the chain with $a(x) = 1/2$ for all $x$ only by a factor of $C\delta^{-1}|\log \delta|$.*

## 2.2 Robustness of $\tau_\infty$ for trees.

Recall that for reversible chains the $L_2$ mixing time, $\tau_2$, determines the $L_p$-mixing time up to a factor $c_p$ for all $1 < p \leq \infty$ (see (3.3)). Denote the $L_p$ mixing time of simple random walk on a finite connected simple graph $G$ by $\tau_p(G)$. Kozma [12] made the following conjecture:

**Conjecture 2.3** ([12])**.** *Let $G$ and $H$ be two finite $K$-roughly isometric graphs of maximal degree $\leq d$. Then*

$$\tau_\infty(G) \leq C(K,d)\tau_\infty(H). \tag{2.3}$$

It is well-known that (2.3) is true if one replaces $\tau_\infty$ with $t_{\mathrm{LS}}$ (e.g. [6, Lemma 3.4]). Ding and Peres [7] showed that (2.3) is false if one replaces $\tau_\infty$ with $\tau_1$. In part, their analysis relied on the fact that the total variation mixing time can be related to hitting times, which may be sensitive to small changes in the geometry. Hence it is natural to expect that a description of $\tau_\infty$ in terms of hitting times might shed some light on Conjecture 2.3. Indeed this was one of the main motivations for this work. In [10] the first author constructed a counterexample to Conjecture 2.3, where also there the key is sensitivity of hitting times.

Peres and Sousi [17, Theorem 9.1] showed that for weighted nearest neighbor random walks on trees (see § 6.2 for a definition), $\tau_1$ can change only by a constant factor, as a result of a bounded perturbation of the edge weights. As an application of Theorem 1.1 we extend their result to the case of $\tau_2$.

**Theorem 2.1.** *There exists an absolute constant $C$ such that for every finite tree $\mathcal{T} = (V, E)$ with some edge weights $(w_e)_{e \in E}$, the corresponding random walk satisfies that*

$$\max(\tau_1, t_{\mathrm{LS}}/4) \leq \tau_2 \leq \tau_1 + C\max(t_{\mathrm{LS}}, \sqrt{t_{\mathrm{LS}}\tau_1}), \tag{2.4}$$

*Consequently, if $(w'_e)_{e \in E}, (w_e)_{e \in E}$ are two edge weights such that $1/M \leq w_e/w'_e \leq M$ for all $e \in E$, then there exists a constant $C_M$ (depending only on $M$) such that the corresponding $L_\infty$ mixing times, $\tau_\infty$ and $\tau'_\infty$, satisfy*

$$\tau'_\infty/C_M \leq \tau_\infty \leq C_M\tau'_\infty. \tag{2.5}$$

**Remark 2.4.** *Since $t_{\mathrm{LS}}$ is robust under a bounded perturbation of the edge weights (e.g. [6, Lemma 3.3]), indeed* (2.5) *follows from* (2.4) *in conjunction with the aforementioned $L_1$ robustness of trees (and the fact that $\tau_2 \le \tau_\infty \le 2\tau_2$, see* (3.3)*).*

## 2.3 Comparison of continuous-time, discrete-time and averaged chains

Let $(X_k)$ be a finite irreducible reversible discrete-time Markov chain with transition matrix $P$. Since reversible Markov chains can only have period 2, one may wonder whether it suffices to average over two consecutive times (i.e. to make a single lazy step) in order to avoid near-periodicity issues. This motivates considering the following Markov chain. For any $t \ge 1$, denote $A_t := (P^t + P^{t-1})/2$. The ***averaged chain***, $(X_t^{\mathrm{ave}})_{t=0}^{\infty}$, with initial state $x$, is a Markov chain, whose distribution at time $t \ge 1$ is $A_t(x,\cdot)$, where $A_t(x,y) := (P^t(x,y) + P^{t-1}(x,y))/2$. Equivalently, $(X_t^{\mathrm{ave}})_{t=1}^{\infty} := (X_{t-\xi})_{t=1}^{\infty}$, where $\xi$ is a Bernoulli(1/2) random variable, independent of $(X_t)_{t=0}^{\infty}$.

We may consider the $L_p$-mixing times of the discrete-time and averaged chains $\tau_p^{\mathrm{discete}}(\cdot)$ and $\tau_p^{\mathrm{ave}}(\cdot)$, resp., defined in an analogous manner as $\tau_p(\cdot)$, obtained by replacing $h_t(x,y) = H_t(x,y)/\pi(y)$ with $k_t(x,y) := P^t(x,y)/\pi(y)$ and $a_t(x,y) := A_t(x,y)/\pi(y)$, resp. (see § 3.1). Similarly, we may consider the relative-entropy mixing times of the discrete-time and averaged chains $\tau_{\mathrm{Ent}}^{\mathrm{discrete}}(\cdot)$ and $\tau_{\mathrm{Ent}}^{\mathrm{ave}}(\cdot)$, resp.. We define $\rho_{\mathrm{discete}}$ and $\rho_{\mathrm{Ent}}^{\mathrm{discete}}$ in an analogous manner to $\rho$ and $\rho_{\mathrm{Ent}}$, where now the hitting times are defined w.r.t. the discrete-time chain. Denote the eigenvalues of $P$ by $1 = \lambda_1 > \lambda_2 \ge \cdots \ge \lambda_{|\Omega|} \ge -1$. Define $t_{\mathrm{rel}}^{\mathrm{absolute}} := \max\{|\log|\lambda_2||^{-1}, |\log|\lambda_{|\Omega|}||^{-1}\}$. Let $\beta(A) := 1 - \lambda(A)$ and

$$\kappa_{\mathrm{discrete}} := \max_{A \in \mathrm{Con}_{1/2}} \log_{\frac{1}{\beta(A)}} \left( \frac{1}{\pi(A)} \right) \le \kappa.$$

**Theorem 2.5.** *There exist positive absolute constants $c, C_1, C_2, C_3$ such that for every irreducible reversible Markov chain on a finite state space*

$$\rho_{\mathrm{discete}} \le \tau_2^{\mathrm{ave}} \le \rho_{\mathrm{discete}} + C_1 \kappa_{\mathrm{discrete}} \le C_2 \rho_{\mathrm{discete}}. \tag{2.6}$$

$$\rho_{\mathrm{Ent}}^{\mathrm{discete}} \le \tau_{\mathrm{Ent}}^{\mathrm{ave}} \le C_3 \rho_{\mathrm{Ent}}^{\mathrm{discete}}. \tag{2.7}$$

$$\max(\rho_{\mathrm{discete}}, t_{\mathrm{rel}}^{\mathrm{absolute}} \log 2) \le \tau_2^{\mathrm{discete}} \le \rho_{\mathrm{discete}} + C_1(\kappa_{\mathrm{discrete}} + t_{\mathrm{rel}}^{\mathrm{absolute}}) \le C_2(\rho_{\mathrm{discete}} + t_{\mathrm{rel}}^{\mathrm{absolute}}). \tag{2.8}$$

$$\max(\rho_{\mathrm{Ent}}^{\mathrm{discete}}, c t_{\mathrm{rel}}^{\mathrm{absolute}}) \le \tau_{\mathrm{Ent}}^{\mathrm{discrete}} \le C_3(\rho_{\mathrm{Ent}}^{\mathrm{discete}} + t_{\mathrm{rel}}^{\mathrm{absolute}}). \tag{2.9}$$

*Moreover, there exists an absolute constant $M \ge 1/2$ such that*

$$\tau_2^{\mathrm{discete}}(M) \le \rho_{\mathrm{discete}} + C_1 \kappa_{\mathrm{discrete}}. \tag{2.10}$$

In conjunction with Theorem 1.1 and Lemma 3.9, which asserts that $\rho_{\mathrm{discete}} \le \bar{C}\rho$ and $\rho_{\mathrm{Ent}}^{\mathrm{discete}} \le \bar{C}' \rho_{\mathrm{Ent}}$, we get the following corollary.

**Corollary 2.6.** *There exists an absolute constant $C > 0$ such that for every irreducible reversible Markov chain on a finite state space*

$$\tau_2^{\mathrm{ave}} \le C\tau_2 \quad \textit{and} \quad \tau_{\mathrm{Ent}}^{\mathrm{ave}} \le C\tau_{\mathrm{Ent}}.$$

**Remark 2.7.** *To see that the reverse inequalities are false consider simple random walk (SRW) on the $n$-clique, for which $\tau_2^{\mathrm{ave}} \le 2$ while $\tau_2 = \Theta(\log n)$ and $\tau_{\mathrm{Ent}} = \Theta(\log\log n)$. Indeed, it is possible that $\rho^{\mathrm{discrete}} \ll \rho$. Loosely speaking, this type of behavior is possible when $\max(|\lambda_{|\Omega|}|, |\lambda_2|) \ll 1$ (e.g. for SRW on the $n$-clique $\lambda_2 = \lambda_n = -\frac{1}{n-1}$). More generally, consider an arbitrary distribution $\pi$ and the transition matrix $\Pi$ whose rows are all equal to $\pi$. Then $\tau_\infty^{\mathrm{discrete}} = 1$ while $\tau_\infty \approx |\log(2\min_x \pi(x))|$. Note that $\lambda_i = 1_{i=1}$.*

**Remark 2.8.** *The fact that under reversibility $\tau_1^{\mathrm{ave}} \le C\tau_1$ is due to Peres and Sousi [17]. In fact, in [11] the authors confirmed a conjecture by Aldous and Fill [1, Open Problem 4.17] by showing that under reversibility, for all $t, M \ge e$ and $x \in \Omega$*

$$\|H_{t+M\sqrt{t}}(x,\cdot)-\pi(\cdot)\|_{\mathrm{TV}}-e^{-cM^2} \le \|A_t(x,\cdot)-\pi(\cdot)\|_{\mathrm{TV}} \le \|H_{t-(M\log M)\sqrt{t}}(x,\cdot)-\pi(\cdot)\|_{\mathrm{TV}}+C/M,$$

*where $\|\mu-\nu\|_{\mathrm{TV}} = \frac{1}{2}\|\mu-\nu\|_{1,\pi} = \frac{1}{2}\sum_x |\mu(x)-\nu(x)|$ and $H_s := e^{-s(I-P)}$.*

The following proposition refines the inequality $\tau_2^{\mathrm{ave}} \le C\tau_2$. The argument is borrowed from [6, Corollary 2.2].

**Proposition 2.9.** *For every finite irreducible reversible Markov chain, for every $k \ge 2$*

$$\forall k' \ge 1, \quad \|A_{k+k'}(x,\cdot)-\pi(\cdot)\|_{2,\pi}^2 \le \left(\frac{1}{2ek'}\right)^2 (\|\mathrm{P}_x^{k-2}-\pi\|_{2,\pi}^2+1) + (1-\lambda)^{2k'+2}\|\mathrm{P}_x^{k-2}-\pi\|_{2,\pi}^2.$$

# 3 Propaedeutics

## 3.1 Different notions of distance and mixing times and their relations

Generically, we shall denote the state space of a Markov chain $(X_t)$ by $\Omega$ and its stationary distribution by $\pi$. We denote such a chain by $(\Omega, P, \pi)$. We say that the chain is finite, whenever $\Omega$ is finite. The continuous-time version of a chain is a continuous-time Markov chain whose distribution at time $t$ is given by the heat kernel $H_t := e^{-t(I-P)}$. We denote $h_t(x,y) := H_t(x,y)/\pi(y)$.

We denote by $\mathrm{P}_x^t$ (resp. $\mathrm{P}_x$) the distribution of $X_t$ (resp. $(X_t)_{t\ge 0}$), given that the initial state is $x$. The $L_p$ norm of a function $f \in \mathbb{R}^\Omega$ is $\|f\|_p := (\mathbb{E}_\pi[|f|^p])^{1/p}$ for $1 \le p < \infty$ (where $\mathbb{E}_\pi[h] := \sum_x \pi(x)h(x)$) and $\|f\|_\infty := \max_x |f(x)|$. The $L_p$ norm of a signed measure $\sigma$ is

$$\|\sigma\|_{p,\pi} := \|\sigma/\pi\|_p, \quad \text{where} \quad (\sigma/\pi)(x) = \sigma(x)/\pi(x).$$

We denote the worst case $L_p$ distance at time $t$ by $d_p(t) := \max_x d_{p,x}(t)$, where $d_{p,x}(t) := \|\mathrm{P}_x^t - \pi\|_{p,\pi}$. Under reversibility for all $x \in \Omega$ and $t \ge 0$ (e.g. (2.2) in [9])

$$d_{2,x}^2(t) = h_{2t}(x,x) - 1, \quad d_\infty(t) = \max_y h_t(y,y) - 1. \tag{3.1}$$

The $\varepsilon$-$L_p$**-mixing-time** of the chain (resp. for a fixed starting state $x$) is defined as

$$\tau_p(\varepsilon) := \max_x \tau_{p,x}(\varepsilon), \qquad \text{where } \tau_{p,x}(\varepsilon) := \min\{t : d_{p,x}(t) \le \varepsilon\}. \tag{3.2}$$

When $\varepsilon = 1/2$ we omit it from the above notation. Let $m_p := 1 + \lceil (2-p)/(2(p-1)) \rceil$. It follows from (3.1), Jensen's inequality and the Reisz-Thorin interpolation Theorem that for reversible chains, the $L_p$ mixing times can be compared as follows (e.g. [18, Lemma 2.4.6]):

$$\begin{aligned} \tau_2(a) \le & \tau_p(a) \le 2\tau_2(\sqrt{a}) = \tau_\infty(a) \quad \text{for all } p \in (2,\infty] \text{ and } a > 0, \\ \frac{1}{m_p}\tau_2(a^{m_p}) & \le \tau_p(a) \le \tau_2(a) \qquad \text{for all } p \in (1,2) \text{ and } a > 0, \end{aligned} \tag{3.3}$$

Hence for all $1 < p \le \infty$ the $L_p$ convergence profile is determined by that of $L_2$.

The ***relative entropy*** of a distribution $\mu$ w.r.t. $\pi$ is defined as

$$D(\mu||\pi) := \sum_x \mu(x) \log(\mu(x)/\pi(x)) = \mathrm{Ent}_\pi(\mu/\pi), \quad \text{where for } f \in \mathbb{R}_+^\Omega \tag{3.4}$$

$$\mathrm{Ent}_\pi(f) := \mathbb{E}_\pi[f \log f] - \mathbb{E}_\pi[f] \log \mathbb{E}_\pi[f] = \mathbb{E}_\pi[f \log(f/\mathbb{E}_\pi[f])].$$

The mixing time in relative entropy is defined as

$$\tau_{\mathrm{Ent},x} := \inf\{t : D(\mathrm{P}_x^t||\pi) \le 1/2\} \quad \text{and} \quad \tau_{\mathrm{Ent}} = \max_x \tau_{\mathrm{Ent},x}. \tag{3.5}$$

The relative entropy distance can be compared with the $L_1$ and $L_2$ distances as follows: [14, p. 110-112]

$$2D(\mu||\pi) \ge \|\mu - \pi\|_{1,\pi}^2 + (M^{-1}\|\mu-\pi\|_{1,\pi})^M \quad \text{for some constant} \quad M \ge 3, \tag{3.6}$$

and ([8, Theorem 5])

$$D(\mu||\pi) \le \log(1 + \|\mu - \pi\|_{2,\pi}^2). \tag{3.7}$$

### 3.2 Background on the spectral-gap and the Log-Sobolev constant

The following fact (often referred to as the Poincaré inequality) is standard. It can be proved by elementary linear-algebra using the spectral decomposition (e.g. [1, Lemma 3.26]).

**Fact 3.1.** *Let $(\Omega, P, \pi)$ be a finite irreducible Markov chain. Let $x \in \Omega$ and $s, t \ge 0$. Then*

$$\|\mathrm{P}_x^{t+s} - \pi\|_{2,\pi} \le e^{-s/t_{\mathrm{rel}}} \|\mathrm{P}_x^t - \pi\|_{2,\pi}. \tag{3.8}$$

*In particular, for all $x \in \Omega$ and $M \ge 1$,*

$$\tau_{2,x} \le \tau_{2,x}(M/2) + t_{\mathrm{rel}} \log M.$$

The relaxation-time provides a lower bound on $L_1$ mixing time as follows. Let $\beta \ne 1$ be an eigenvalue of $P$. Then under reversibility ([13, Theorem 12.4 and Lemma 20.11])

$$\|P^k(x,\cdot) - \pi(\cdot)\|_{1,\pi} \ge |\beta|^k \quad \text{and} \quad \|\mathrm{P}_x^t - \pi\|_{1,\pi} \ge e^{-(1-\beta)t}.$$

Consequently, under reversibility, for all $\delta \in (0,1]$ we have that

$$\tau_1^{\mathrm{discrete}}(\delta) \ge t_{\mathrm{rel}}^{\mathrm{absolute}} \log(1/\delta) \quad \text{and} \quad \tau_1(\delta) \ge t_{\mathrm{rel}} \log(1/\delta). \tag{3.9}$$

It follows from (3.6) that there exists some absolute constant $c>0$ such that

$$\tau_{\mathrm{Ent}}^{\mathrm{discrete}} \ge c t_{\mathrm{rel}}^{\mathrm{absolute}}. \tag{3.10}$$

With the convention $0\log 0=0$, for all non-zero $f,g\in \mathbb{R}_{+}^{\Omega}$ we define $\langle f,g\rangle_{\pi}:=\mathbb{E}_{\pi}[fg]$, $\mathcal{E}(f,g):=\langle (I-Q)f,g\rangle_{\pi}$ and $\mathcal{E}(f):=\mathcal{E}(f,f)$. The *Log-Sobolev constant* of the chain is

$$c_{\mathrm{LS}}:=\inf\{\mathcal{E}(f)/\mathrm{Ent}_{\pi}(f^2): f \text{ is non-constant}\}. \tag{3.11}$$

Recall that $t_{\mathrm{LS}}:=1/c_{\mathrm{LS}}$. It is always the case that $t_{\mathrm{LS}}\ge 2t_{\mathrm{rel}}$ (e.g. [6, Lemma 3.1]).

There are numerous works aiming towards general geometric upper bounds on $\tau_{\infty}$. Among the most advanced techniques are the spectral profile [9] and Logarithmic Sobolev inequalities (see [6] for a survey on the topic). Let $\pi_{*}:=\min_{x\in\Omega}\pi(x)$. It is classical (e.g. [6, Corollary 3.11]) that for reversible chains

$$t_{\mathrm{LS}}/2\le \tau_2(1/e)\le t_{\mathrm{LS}}(1+\frac{1}{4}\log\log(1/\pi_{*})). \tag{3.12}$$

There are examples demonstrating that each of these bounds can be attained up to a constant factor.

Let $1\le p_1,p_2\le\infty$. The $p_1\to p_2$ norms of a linear operator $\mathbf{A}$ are given by

$$\|\mathbf{A}\|_{p_1\to p_2}:=\max\{\|\mathbf{A}f\|_{p_2}:\|f\|_{p_1}=1\}.$$

If $\|\mathbf{A}\|_{p_1\to p_2}\le 1$ for some $1\le p_1<p_2\le\infty$ we say that $\mathbf{A}$ is a hypercontraction. For all $p_1,p_2$, $\|H_t\|_{p_1\to p_2}$ is non-increasing in $t$. It is a classic result (e.g. [6, Theorem 3.5] and [1, Theorem 8.24]) that the Log-Sobolev time can be characterized in terms of hypercontrativity.

**Fact 3.2.** *Let $(\Omega,P,\pi)$ be a finite reversible chain. Let $s_q:=\inf\{t:\|H_t\|_{2\to q}\le 1\}$. Then $t_{\mathrm{LS}}=4\sup_{q:2<q<\infty}s_q/\log(q-1)$.*

The following result ([6, Theorem 3.10]) will allow us to bound $t_{\mathrm{LS}}$ from above.

**Fact 3.3.** *Let $(\Omega,P,\pi)$ be a finite reversible chain. Fix $2<q<\infty$. Assume that $r_q$ and $M_q$ satisfy that $\|H_{r_q}\|_{2\to q}\le M_q$. Then*

$$t_{\mathrm{LS}}\le \frac{2q}{q-2}r_q+2t_{\mathrm{rel}}(1+\frac{q}{q-2}\log M_q). \tag{3.13}$$

Fix some $0<\varepsilon<1/2$ and $A\in \mathrm{Con}_{2^{-1/\varepsilon}}$. Assume that $\mathrm{P}_{\pi}[T_{A^c}>t]\ge 2\pi(A)^{1+\varepsilon}$. Let $\pi_A$ denote $\pi$ conditioned on $A$ (i.e. $\pi_A(a)=\frac{\pi(a)1_{a\in A}}{\pi(A)}$). Then $\mathrm{P}_{\pi_A}[T_{A^c}>t]\ge 2\pi(A)^{\varepsilon}$ and so

$$B=\{a\in A:\mathrm{P}_a[T_{A^c}>t]\ge \pi(A)^{\varepsilon}\}$$

satisfies $\pi_A(B)\ge\pi(A)^{\varepsilon}$ (i.e. $\pi(B)\ge\pi(A)^{1+\varepsilon}$). Consequently, for $q>\frac{2(1+\varepsilon)}{1-2\varepsilon}$

$$\|H_t 1_A\|_q\ge [\sum_{b\in B}\pi(b)H_t(b,A)^q]^{1/q}\ge \pi(B)^{1/q}\pi(A)^{\varepsilon}\ge \pi(A)^{\varepsilon+(1+\varepsilon)/q}>\sqrt{\pi(A)}=\|1_A\|_2.$$

Thus a natural hitting time version of hypercontractivity is

$$t_{\mathrm{ht}}:=\min\{t:\mathrm{P}_{\pi}[T_{A^c}>t]\le \pi(A)^{5/4} \text{ for all } A\in \mathrm{Con}_{1/2}\}.$$

**Question.** *Is there an absolute constant $C$ such that for every finite irreducible reversible Markov chain $t_{\mathrm{ht}}/C \le t_{\mathrm{LS}} \le Ct_{\mathrm{ht}}$.*

Trivially, $t_{\mathrm{ht}} = \min\{t : \mathrm{P}_{\pi_A}[T_{A^c} > t] \le \pi(A)^{1/4}$ for all $A \in \mathrm{Con}_{1/2}\}$. Note that if we replace $\pi_A$ by the quasi-stationary distribution of $A$, denoted by $\mu_A$, then by (3.16) we get precisely $\kappa/4$. This explains why also $\kappa$ can be interpreted as a hitting time version of hypercontractivity. We note that the above question resembles Open problem 4.38 in [1], which asks whether for reversible chains $t_{\mathrm{rel}} \le C \max_{A \in \mathrm{Con}_{1/2}} \mathbb{E}_{\pi_A}[T_{A^c}]$, where indeed [1, Lemma 4.39] $t_{\mathrm{rel}} \le \max_{A \in \mathrm{Con}_{1/2}} \mathbb{E}_{\mu_A}[T_{A^c}]$ (the formulation in [1] is slightly different, but it is equivalent to our formulation).

### 3.3 Starr maximal inequality and a useful lemma

In this section we prove a maximal inequality which shall be central in what comes. Denote $S_t := e^{-(I-Q)t} = \sum_{k=0}^{\infty} \frac{e^{-t}t^k}{k!} Q^t$. When considering $Q$ instead of $P$ we write $\mathbb{P}_x^t$, $\mathbb{P}_x$ and $Y_t$ instead of $\mathrm{P}_x^t$, $\mathrm{P}_x$ and $X_t$, respectively.

**Theorem 3.4** (Starr's Maximal inequality [19])**.** *Let $(\Omega, P, \pi)$ be an irreducible Markov chain. Let $f \in \mathbb{R}^{\Omega}$. Its corresponding* maximal function *$f^* \in \mathbb{R}^{\Omega}$ is defined as*

$$f^*(x) := \sup_{0 \le t < \infty} |S_t(f)(x)| = \sup_{0 \le t < \infty} |\mathbb{E}_x[f(Y_t)]|.$$

*Then for every $1 < p < \infty$*

$$\|f^*\|_p \le p^* \|f\|_p, \quad \textit{where} \quad p^* := p/(p-1) \textit{ is the conjugate exponent of } p. \tag{3.14}$$

*Moreover, under reversibility, for $f_{*,\mathrm{even}}(x) := \sup_{k \in \mathbb{Z}_+} |P^{2k} f(x)|$ we have that (3.14) holds also with $f_{*,\mathrm{even}}$ in the role of $f^*$ and hence $f_*(x) := \sup_{k \in \mathbb{Z}_+} |P^k f(x)|$ satisfies for $1 < p < \infty$*

$$\|f_*\|_p^p \le \|f_{*,\mathrm{even}}\|_p^p + \|(Pf)_{*,\mathrm{even}}\|_p^p \le (p^*)^p (\|f\|_p^p + \|Pf\|_p^p) \le 2(p^*)^p \|f\|_p^p. \tag{3.15}$$

The following Lemma is essentially due to Norris, Peres and Zhai [16].

**Lemma 3.5.** *Let $(\Omega, P, \pi)$ be a finite irreducible Markov chain. Let $f_A(x) := 1_{x \in A}/\pi(A)$.*

$$\forall A \subset \Omega, \quad \max(\frac{1}{2}\|(f_A)_*\|_1, \|f_A^*\|_1) \le e \max(1, |\log \pi(A)|).$$

*Proof.* We first show that $\|f_A^*\|_1 \le e \max(1, |\log \pi(A)|)$. By (3.14) for all $1 < p < \infty$

$$\|f_A^*\|_1 \le \|f_A^*\|_p \le p^* \|f_A\|_p = p^* [\pi(A)]^{-1/p^*}$$

Taking $p^* := \max(1 + \varepsilon, |\log \pi(A)|)$ and sending $\varepsilon$ to 0 (noting that the r.h.s. is continuous w.r.t. $p_*$) concludes the proof. The same calculation shows that (last inequality)

$$\|(f_A)_*\|_1 \le \|(f_A)_{*,\mathrm{even}}\|_1 + \|(Pf_A)_{*,\mathrm{even}}\|_1 \le 2\|(f_A)_{*,\mathrm{even}}\|_1 \le 2e \max(1, |\log \pi(A)|).$$

□

We note that by [19, Theorem 2] $(1 - e^{-1})\|f_A^*\|_1 - 1 \le \|f_A \log[\max(1, |f_A|)]\|_1 = |\log \pi(A)|$.

### 3.4 Bounding escape probabilities using $\kappa$

Recall that $P_A$ and $Q_A$ are the restriction to $A$ of $P$ and $Q$, resp.. Denote

$$H_t^A(x,y) := e^{-t(I-P_A)}(x,y) = \mathrm{P}_x(X_t = y, T_{A^c} > t) \quad \text{and similarly} \quad S_t^A := e^{-t(I-Q_A)}.$$

Recall that $\lambda(A)$ is the smallest eigenvalue of $I - Q_A$. By the Perron-Frobenius Theorem there exists a distribution $\mu_A$ on $A$, known as the ***quasi-stationary distribution*** of $A$, satisfying that the escape time from $A$ w.r.t. $Q$, starting from $\mu_A$, has an Exponential (resp. Geometric in discrete-time) distribution with mean $t_{\mathrm{rel}}(A) = 1/\lambda(A)$. Equivalently, for all $t \geq 0$

$$\mu_A Q_A = (1 - \lambda(A))\mu_A \text{ and } \mu_A S_t^A = e^{-\lambda(A)t}\mu_A.$$

Throughout we use $\mu_A$ to denote the quasi-stationary distribution of $A$. Recall that we denote $\pi$ conditioned on $A$ by $\pi_A$.

Using the spectral decomposition of $Q_A$ (e.g. [2, Lemma 3.8] or [1, (3.87)]) it follows that

$$\begin{aligned} &\forall A \subsetneq \Omega,\, s \geq 0, \quad \mathbb{P}_{\pi_A}[T_{A^c} > s] \leq \mathbb{P}_{\mu_A}[T_{A^c} > s] = \mu_A S_t^A 1_A = e^{-\lambda(A)s}\mu_A(A) = e^{-\lambda(A)s}. \\ &\forall A \subsetneq \Omega,\, k \geq 0, \quad \pi_A Q_A^k 1_A \leq \mu_A Q_A^k 1_A = (1-\lambda(A))^k \mu_A(A) = (1-\lambda(A))^k. \end{aligned} \tag{3.16}$$

**Proposition 3.6.** *For reversible chains*

$$\kappa \leq 3\rho \quad \text{and} \quad \kappa_{\mathrm{discrete}} \leq 3\rho_{\mathrm{discrete}}. \tag{3.17}$$

*Proof:* We first show that $\kappa \leq 3\rho$. Let $A \in \mathrm{Con}_{1/2}$ be such that $\kappa = t_{\mathrm{rel}}(A)|\log \pi(A)|$. By (3.16) $\mathrm{P}_{\mu_A}[T_{A^c} > \kappa/3] = \pi(A)^{1/3}$. Since $a^{1/3} > a + \frac{1}{2}\sqrt{a(1-a)}$, for all $0 \leq a \leq 1/2$, we have that

$$\max_{x \in A} \mathrm{P}_x[T_{A^c} > \kappa/3] \geq \mathrm{P}_{\mu_A}[T_{A^c} > \kappa/3] = \pi(A)^{1/3} > \pi(A) + \frac{1}{2}\sqrt{\pi(A)\pi(A^c)}.$$

We now show that $\kappa_{\mathrm{discrete}} \leq 3\rho_{\mathrm{discrete}}$. Let $B \in \mathrm{Con}_{1/2}$ be such that $\kappa_{\mathrm{discrete}} = |\log_{\frac{1}{1-\lambda(A)}} \pi(B)|$. Denote the hitting time of $B^c$ w.r.t. the discrete time chain as $T_{B^c}^{\mathrm{discrete}}$. By (3.16)

$$\max_{x \in B} \mathrm{P}_x[T_{B^c}^{\mathrm{discrete}} > \kappa_{\mathrm{discrete}}/3] \geq \mu_B P_B^{\kappa_{\mathrm{discrete}}/3} 1_B = \pi(B)^{1/3} > \pi(B) + \frac{1}{2}\sqrt{\pi(B)\pi(B^c)}. \quad \square$$

**Definition 3.7.** $\bar{\rho} := \max_x \bar{\rho}_x$ *and* $\bar{\rho}_{\mathrm{Ent}} := \max_x \rho_{\mathrm{Ent},x}$, *where*

$$\begin{aligned} &\bar{\rho}_x := \min\{t : \mathrm{P}_x[T_{A^c} > t] \leq \pi(A)^3 \text{ for all } A \in \mathrm{Con}_{1/2}\}. \\ &\bar{\rho}_{\mathrm{Ent},x} := \min\{t : \mathrm{P}_x[T_{A^c} > t] \leq \frac{1}{16e^2[\log(e^{3/2}/\pi(A))]^3} \text{ for all } A \in \mathrm{Con}_{1/2}\}. \end{aligned} \tag{3.18}$$

Note that by the Markov property, $\max_x \mathrm{P}_x[T_{A^c} > mt] \leq (\max_y \mathrm{P}_y[T_{A^c} > t])^m$ and so

$$\rho \leq \bar{\rho} \leq 9\rho \quad \text{and} \quad c'\rho_{\mathrm{Ent}} \leq \bar{\rho}_{\mathrm{Ent}} \leq C'\rho_{\mathrm{Ent}}, \tag{3.19}$$

for some absolute constants $c', C' > 0$ (for $\bar{\rho} \leq 9\rho$, use the inequality $a^{1/3} \geq a + \frac{1}{2}\sqrt{a(1-a)}$, valid for all $0 \leq a \leq 1/2$). The following proposition refines the inequality $\bar{\rho} \leq 9\rho$.

**Proposition 3.8.** *For every reversible chain,*

$$\forall x \in \Omega, \quad \bar{\rho}_x \leq \rho_x + s, \quad \textit{where} \quad s := 8\kappa + 2t_{\mathrm{rel}} \log 8. \tag{3.20}$$

*Proof:* Let $x \in \Omega$ and $A \in \mathrm{Con}_{1/2}$. By (1.6) $2t_{\mathrm{rel}} \geq \max_{B \in \mathrm{Con}_{1/2}} t_{\mathrm{rel}}(B)$ and so by (3.16)

$$\mathrm{P}_{\pi_A}[T_{A^c} > s] \leq e^{-\lambda(A)[t_{\mathrm{rel}}(A)(8|\log \pi(A)|+\log 8)]} = \pi(A)^8/8.$$

Thus the set

$$B = B(A) := \{y : \mathrm{P}_y[T_{A^c} > s] > \pi(A)^3/2\}$$

satisfies

$$\pi(B)/\pi(A) = \pi_A(B) < \mathrm{P}_{\pi_A}[T_{A^c} > s]/(\pi(A)^3/2) \leq \pi(A)^5/4,$$

and so by the definition of $\rho_x$, $\mathrm{P}_x[T_{B^c} > \rho_x] \leq \pi(B) + \frac{1}{2}\sqrt{\pi(B)\pi(B^c)} \leq \sqrt{\pi(B)} \leq \frac{1}{2}\pi(A)^3$ (where we used $\pi(B) < 2^{-8}$). Finally, by the definition of $B$ and the Markov property

$$\mathrm{P}_x[T_{A^c} > \rho_x + s] \leq \mathrm{P}_x[T_{B^c} > \rho_x] + \max_{b \notin B} \mathrm{P}_b[T_{A^c} > s] \leq \frac{1}{2}\pi(A)^3 + \frac{1}{2}\pi(A)^3 = \pi(A)^3. \qquad \square$$

**Lemma 3.9.** *For every finite irreducible Markov chain we have that*

$$\rho_{\mathrm{discete}} \leq C\rho,$$
$$\rho_{\mathrm{Ent}}^{\mathrm{discete}} \leq C'\rho_{\mathrm{Ent}}.$$

*Proof:* Let $A \in \mathrm{Con}_{1/2}$ and $x \in \Omega$. To avoid ambiguity we denote the distributions of the discrete and the continuous-time chains started at $x$ by $\mathrm{P}_x$ and $\mathrm{H}_x$, resp.. Since for all $M \in \mathbb{N}$ we have that $\mathrm{H}_x[T_A > Mt] \leq (\max_y \mathrm{H}_y[T_A > t])^M$ it suffices to show that for all $t \in \mathbb{N}$ we have that $\mathrm{P}_x[T_A > 4t] \leq 4\mathrm{H}_x[T_A > t]$. Indeed, if $N_t \sim \mathrm{Pois}(t)$ then

$$\mathrm{H}_x[T_A > t] = \sum_k \mathbb{P}[N_t = k]\mathrm{P}_x[T_A > k] \geq \mathbb{P}[N_t \leq 4t]\mathrm{P}_x[T_A > 4t] \geq \frac{1}{4}\mathrm{P}_x[T_A > 4t]. \quad \square$$

### 3.5 Related work

Let

$$\mathrm{hit}(\varepsilon) := \max_x \mathrm{hit}_x(\varepsilon), \quad \mathrm{hit}_x(\varepsilon) := \min\{t : \mathrm{P}_x[T_A > t] \leq \varepsilon, \forall A \in \mathrm{Con}_{1/2}\}.$$

Let $t_{\mathrm{mix}}(\varepsilon) := \tau_1(2\varepsilon)$ be the total-variation $\varepsilon$-mixing-time. In [2] it was shown that for finite irreducible reversible chains, for all $\varepsilon \in (0,1), \delta \in (0, \frac{1}{2}\min(\varepsilon, 1-\varepsilon))$ we have that

$$\mathrm{hit}(\varepsilon + \delta) - 4t_{\mathrm{rel}}|\log \delta| \leq t_{\mathrm{mix}}(\varepsilon) \leq \mathrm{hit}(\varepsilon - \delta) + 4t_{\mathrm{rel}}|\log \delta|, \tag{3.21}$$

Generally, $t_{\mathrm{rel}}|\log(2\varepsilon)| \leq t_{\mathrm{mix}}(\varepsilon)$ for all $0 < \varepsilon \leq 1/2$, however often $t_{\mathrm{rel}} \ll \tau_1$. In particular, this is the case for a sequence of reversible chains which exhibits cutoff (i.e. abrupt convergence) in total variation ([13, Lemma 18.4]). In [2, Theorem 3] (3.21) is exploited in order to obtain a characterization of the cutoff phenomenon for reversible Markov chains, in terms of concentration of hitting times of "worst" (in some sense) sets.

The main tool in the proof of (3.21) is Starr's $L_p$ maximal inequality (Theorem 3.4). In other words, an $L_p$ maximal inequality is used to characterize convergence in $L_1$. A look into the proof of (3.21) reveals that it does not require the full strength of Starr's inequality. It is thus natural to try applying Starr's $L_p$ maximal inequality to study stronger notions of convergence. Indeed Theorem 1.1 can be seen as the $p > 1$ counterpart of (3.21). Also in our analysis the main tool is Starr's inequality.

# 4 An overview of our approach

## 4.1 Lower bounding mixing times using hitting times

We start with an illustrating example: if $\mathrm{P}_x[T_{A^c} > t] > 3\pi(A)/2$ for some set $A$, then

$$[H_t(x,A) - \pi(A)]/\pi(A) \geq [\mathrm{P}_x[T_{A^c} > t] - \pi(A)]/\pi(A) > 1/2.$$

Denote $\pi$ conditioned on $A$ by $\pi_A(a) := 1_{a \in A}\pi(a)/\pi(A)$. Finally, note that

$$d_{\infty,x}(t) \geq \max_{a \in A} h_t(x,a) - 1 \geq \sum \pi_A(a)(h_t(x,a) - 1) = [H_t(x,A) - \pi(A)]/\pi(A) > 1/2.$$

Hence $\tau_{\infty,x} \geq \min\{t : \mathrm{P}_x[T_{A^c} > t] \leq 3\pi(A)/2, \text{ for all } A\}$.

This generalizes as follows. Let $\mathscr{P}(\Omega)$ be the collection of all distributions on $\Omega$. Let $A \subsetneq \Omega$, $x \in \Omega$, $t > 0$ and $\delta \in (0,1)$. Let

$$\mathscr{P}_{A,\delta} := \{\mu \in \mathscr{P}(\Omega) : \mu(A) \geq \pi(A) + \delta\pi(A^c)\}.$$

Clearly, if $\mathrm{P}_x[T_{A^c} > t] \geq \pi(A) + \delta\pi(A^c)$, then $\mathrm{P}_x^t \in \mathscr{P}_{A,\delta}$. Note that

$$\nu_{A,\delta} := \delta\pi_A + (1-\delta)\pi \in \mathscr{P}_{A,\delta}.$$

Moreover, $\min\{\delta' : \nu_{A,\delta'} \in \mathscr{P}_{A,\delta}\} = \delta$. It is thus intuitive that for a convex distance function between distributions, $\nu_{A,\delta}$ is the closest distribution to $\pi$ in $\mathscr{P}_{A,\delta}$.

**Proposition 4.1.** *Let $(\Omega, P, \pi)$ be some finite irreducible Markov chain. Let $A \subsetneq \Omega$. Denote $\nu_{A,\delta} := \delta\pi_A + (1-\delta)\pi$. Then for all $\delta \in (0,1)$,*

$$\begin{aligned} \min_{\mu \in \mathscr{P}_{A,\delta}} \|\mu - \pi\|_{2,\pi} &= \|\nu_{A,\delta} - \pi\|_{2,\pi} = \delta\sqrt{\pi(A^c)/\pi(A)}. \\ \min_{\mu \in \mathscr{P}_{A,\delta}} D(\mu\|\pi) &= D(\nu_{A,\delta}\|\pi) = u(\pi(A), \delta), \end{aligned} \tag{4.1}$$

*where $u(x,y) := [y + x(1-y)]\log(1 + \frac{y(1-x)}{x}) + (1-y)(1-x)\log(1-y)$.*

*Proof.* The first equality in both lines can be verified using Lagrange multipliers. The second equality in both lines is straightforward. □

Proposition 4.1 motivates the definitions in (1.2). We argue that (4.1), implies the first inequalities in both (1.3)-(1.4) by making suitable substitutes for $\delta$ in (4.1). For (1.3) substitute $\delta = \frac{1}{2}\sqrt{\frac{\pi(A)}{\pi(A^c)}}$ in the first line of (4.1). For every $x \in \Omega$ and $t < \rho_x$ there is some $A \in \mathrm{Con}_{1/2}$ such that

$$\mathrm{P}_x[T_{A^c} > t] > \pi(A) + \frac{1}{2}\sqrt{\pi(A)\pi(A^c)} = \pi(A) + \delta\pi(A^c),$$

where the equality follows by our choice of $\delta$. As mentioned above, this implies that $\mathrm{P}_x^t \in \mathscr{P}_{A,\delta'}$ for some $\delta' > \delta$ and so by (4.1) and the choice of $\delta$, we have that $\|\mathrm{P}_x^t - \pi\|_{2,\pi} > 1/2$. For (1.4), it is not hard to verify that for some $C', C_{\mathrm{Ent}} > 0$, we have that $u(x, \min(\frac{C'}{|\log x|}, \frac{\frac{99}{100} - x)}{1-x}) \geq 1/2$ and $x + \frac{C'}{|\log x|}(1-x) \leq \frac{C_{\mathrm{Ent}}}{|\log x|}$ for all $x \in (0, 1/2]$. Substituting $\delta = \min(\frac{C'}{|\log x|}, \frac{\frac{99}{100} - x)}{1-x})$ in the second line of (4.1) implies the first inequality of (1.4) in a similar manner to the above derivation of the first inequality of (1.3).

### 4.2 Upper bounding mixing times using hitting times

We now explain the idea behind the proof of the upper bound on $\tau_2$ from (1.3). Let $x \in \Omega$. Denote $t := \rho_x + 8\kappa + 6t_{\mathrm{rel}} \log 2$. By Theorem 1.2 it suffices to bound $d_{2,x}(t)$.

Step 1: Show that (Proposition 3.8)

$$\forall B \in \mathrm{Con}_{1/2}, \quad \mathrm{P}_x[T_{B^c} > t] \le \pi(B)^3.$$

Step 2: Show that (Lemma 5.1) for $A_s := \{y : h_t(x,y) \ge (s+1)\}$

$$\forall M \ge 1 \quad \|\mathrm{P}_x^t - \pi\|_{2,\pi}^2 \le M^2 + \int_M^\infty 2s\pi(A_s)ds.$$

$\Longrightarrow$ By Poincaré ineq. (3.1) it suffices that $s\pi(A_s) \le 2s^{-3/2}$ for $s \ge M$ (for some $M$).

Step 3: For $B_s = \{y : \sup_k H_k(y, A_s) > \frac{s}{2}\pi(A_s)\}$ by step 1 and the Markov property,

$$\begin{aligned} s\pi(A_s) \le H_t(x, A_s) &= \mathrm{P}_x[T_{B_s^c} > t, X_t \in A_s] + \mathrm{P}_x[T_{B_s^c} \le t, X_t \in A_s] \\ &\le \mathrm{P}_x[T_{B_s^c} > t] + \sup_{y \notin B_s, k \ge 0} H_k(y, A_s) \le \pi(B_s)^3 + \frac{s}{2}\pi(A_s). \end{aligned} \tag{4.2}$$

Step 4: If $\pi(B_s) \le s^{-1/2}$, then we are done. Unfortunately, we do not know how to prove this estimate. Hence we have to define the set $B_s$ in a slightly different manner: $B_s := \{y : \sup_k H_k(y, A_s) > e\sqrt{s}|\log \pi(A_s)|\pi(A_s)\}$. By Lemma 3.5 indeed $\pi(B_s) \le s^{-1/2}$. Since $e\sqrt{s}|\log \pi(A_s)|\pi(A_s) \le s\pi(A_s)/2$, unless $\pi(A_s) \le Ce^{-\sqrt{s}}$, repeating the reasoning in (4.2) with the new choice of $B_s$ concludes the proof. □

The proof of Theorem 1.2 is similar. The general scheme is as follows. Define a relevant family of sets $A_s$. Define $B_s$ to be of the following form $\{y : \sup |g_s(y)| > a_s\}$ with appropriate choices of $g_s$ and $a_s \in \mathbb{R}_+$ so that the desired inequality we wish to establish for $A_s$ holds with some room to spare given that $T_{B_s^c} \le t$ (for an appropriate choice of $t$). Finally, control the error term $\mathrm{P}[T_{B_s^c} > t]$ (using the choice of $t$) by controlling $\pi(B_s)$ using an appropriate maximal inequality.

## 5 Proofs of the main results

### 5.1 An upper bound on $\tau_2$

In this section we prove the following theorem, which refines (1.3) from Theorem 1.1.

**Theorem 5.1.** *For every finite irreducible reversible Markov chain $(\Omega, P, \pi)$ we have that*

$$\forall x, \quad \rho_x \le \tau_{2,x} \le \bar{\rho}_x + 5t_{\mathrm{rel}} \le \rho_x + 8\kappa + (5 + 6\log 2)t_{\mathrm{rel}}. \tag{5.1}$$

*The same holds when $x$ is omitted from all of the terms above. Consequently,*

$$\rho \le \tau_2 \le (9 + 15/\log 2)\rho. \tag{5.2}$$

**Lemma 5.1.** *Let $A_{x,t}(s) := \{y : h_t(x,y) \geq s+1\}$. For every finite irreducible reversible chain, for all $x \in \Omega$ and $\ell \geq 1$*

$$\forall t \geq 0, \quad \|\mathrm{P}_x^t - \pi\|_{2,\pi}^2 \leq \ell^2 + \int_\ell^\infty 2s\pi(A_{x,t}(s))ds.$$

*Proof:* Fix some $x \in \Omega$, $t \geq 0$ and $\ell \geq 1$. Let $f(y) := |h_t(x,y) - 1|$. Then $\|\mathrm{P}_x^t - \pi\|_{2,\pi}^2 = \|f\|_2^2 = \mathbb{E}_\pi[f^2]$. Note that for all $s > 1$, $\{f \geq s\} = A_{x,t}(s)$. Observe that

$$\mathbb{E}_\pi[f^2 1_{f>\ell}] = \int_0^\infty 2s\pi(\{f 1_{f>\ell} > s\})ds \leq \pi(f > \ell)\ell^2 + \int_\ell^\infty 2s\pi(A_{x,t}(s))ds.$$

Finally, since $f^2 \leq f^2 1_{f>\ell} + 1_{f \leq \ell}\ell^2$, we get that

$$\mathbb{E}_\pi[f^2] \leq \pi(f \leq \ell)\ell^2 + \mathbb{E}_\pi[f^2 1_{f>\ell}] \leq \ell^2 + \int_\ell^\infty 2s\pi(A_{x,t}(s))ds. \qquad \square$$

*Proof of Theorem 5.1:* We first note that (5.2) follows from (5.1) in conjunction with (3.17) and (1.6). We now prove (5.1). Let $x \in \Omega$. The inequality $\rho_x \leq \tau_{2,x}$ follows from (4.1) and the discussion following it. Set $t := \bar{\rho}_x$. As above, denote $A_s := \{y : h_t(x,y) \geq s+1\}$. By the Poincaré inequality (Fact 3.1) and Lemma 5.1 it suffices to show that

$$\int_{e^4}^\infty 2s\pi(A_s)ds \leq e^7 \leq e^{10}/4 - e^8.$$

Let $g_s(y) := \sup_k H_k(y, A_s)/\pi(A_s)$. By Lemma 3.5 $\|g_s\|_1 \leq e|\log \pi(A_s)|$. Let

$$B_s := \{y : g_s(y) > 2e^{-1}\sqrt{s+1}|\log \pi(A_s)|\} = \{y : \sup_k H_k(y, A_s) \geq 2e^{-1}\sqrt{s+1}\pi(A_s)|\log \pi(A_s)|\}.$$

Let $s \geq e^4$. By Markov inequality $\pi(B_s) \leq e^2/(2\sqrt{s+1}) \leq \frac{1}{2}$ and so by the definition of $\bar{\rho}_x$

$$\mathrm{P}_x[T_{B_s^c} > t, X_t \in A_s] \leq \mathrm{P}_x[T_{B_s^c} > t] \leq \frac{e^6}{8(s+1)^{3/2}}.$$

Also, by the definition of $B_s$ we clearly have that

$$\mathrm{P}_x[T_{B_s^c} \leq t, X_t \in A_s] \leq \sup_{b \notin B_s, k \geq 0} H_k(b, A_s) \leq 2e^{-1}\sqrt{s+1}\pi(A_s)|\log \pi(A_s)|.$$

Since by the definition of $A_s$ (first inequality)

$$(s+1)\pi(A_s) \leq H_t(x, A_s) = \mathrm{P}_x[T_{B_s^c} > t, X_t \in A_s] + \mathrm{P}_x[T_{B_s^c} \leq t, X_t \in A_s],$$

we get that if $\mathrm{P}_x[T_{B_s^c} > t, X_t \in A_s] \leq \mathrm{P}_x[T_{B_s^c} \leq t, X_t \in A_s]$, then

$$(s+1)\pi(A_s) \leq 4e^{-1}\sqrt{s+1}\pi(A_s)|\log \pi(A_s)|,$$

which simplifies as follows

$$2s\pi(A_s) \leq 2se^{-e\sqrt{s+1}/4}.$$

while if $\mathrm{P}_x[T_{B_s^c} > t, X_t \in A_s] > \mathrm{P}_x[T_{B_s^c} \le t, X_t \in A_s]$, then we have that

$$2s\pi(A_s) < 4\mathrm{P}_x[T_{B_s^c} > t, X_t \in A_s] \le \frac{e^6}{2(s+1)^{3/2}}.$$

Let $f(s) = 15(s+1)^{3/2}e^{-e\sqrt{s+1}/4}$. Then $f(e^4) < e^6$ and for $s \ge e^4$ we have that $\frac{d}{ds}(-f(s)) \ge (\frac{15e}{8} - \frac{45}{2e^2})(s+1)e^{-e\sqrt{s+1}/4} \ge 2se^{-e\sqrt{s+1}/4}$. Hence indeed

$$\int_{e^4}^{\infty} 2s\pi(A_s)ds \le \int_{e^4}^{\infty} \max(2se^{-\sqrt{s+1}/(2e)}, \frac{e^6}{2(s+1)^{3/2}})ds \le f(e^4) + \frac{e^6}{(e^4+1)^{\frac{1}{2}}} \le e^7. \quad \square$$

## 5.2 A hitting times characterization of mixing in relative entropy

Recall the definitions of $\rho_{\mathrm{Ent}}, \bar{\rho}_{\mathrm{Ent}}, \rho_{\mathrm{Ent},x}$ and $\bar{\rho}_{\mathrm{Ent},x}$ from (1.1) and (3.18). Recall that by (3.19), $c\rho_{\mathrm{Ent}} \le \bar{\rho}_{\mathrm{Ent}} \le C\rho_{\mathrm{Ent}}$. The following theorem refines (1.4) from Theorem 1.1.

**Theorem 5.2.** *Let $(\Omega, P, \pi)$ be a finite irreducible reversible Markov chain. Then*

$$\forall x, \quad \rho_{x,\mathrm{Ent}} \le \tau_{\mathrm{Ent},x} \le \bar{\rho}_{x,\mathrm{Ent}} + 14t_{\mathrm{rel}}. \tag{5.3}$$

*The same holds when $x$ is omitted from all of the terms above. Consequently*

$$\rho_{\mathrm{Ent}} \le \tau_{\mathrm{Ent}} \le C_1\rho_{\mathrm{Ent}}. \tag{5.4}$$

*Proof of Theorem 5.2:* Let $x \in \Omega$. The inequality $\rho_{x,\mathrm{Ent}} \le \tau_{\mathrm{Ent},x}$ follows from (4.1) and the discussion following (4.1). The inequality $\tau_{\mathrm{Ent}} \le C_1\rho_{\mathrm{Ent}}$ follows from (5.3) and (3.19), in conjunction with the fact that (under reversibility) $ct_{\mathrm{rel}} \le \rho_{\mathrm{Ent}}$ for some absolute constant $c > 0$ (c.f. [2, (3.19)] for the fact that there exist some $A \in \mathrm{Con}_{1/2}$ and $a \in A$ so that $\mathrm{P}_a[T_{A^c} > \varepsilon t_{\mathrm{rel}}] \ge e^{-\varepsilon} \ge 1 - \varepsilon$, for all $\varepsilon \ge 0$). We now prove that $\tau_{\mathrm{Ent},x} \le \bar{\rho}_{x,\mathrm{Ent}} + 14t_{\mathrm{rel}}$. Denote $r := \bar{\rho}_{x,\mathrm{Ent}}$, $r' := 14t_{\mathrm{rel}}$. Let

$$D := \{y : h_r(x,y) > e^{10}\}.$$

Denote $\delta := H_r(x,D) - e^{10}\pi(D)$,

$$\mu(y) := \delta^{-1}1_{y\in D}[H_r(x,y) - e^{10}\pi(y)],$$

$$\nu(y) := (1-\delta)^{-1}[1_{y\notin D}H_r(x,y) + 1_{y\in D}e^{10}\pi(y)].$$

Denote $\mu_\ell := \mu H_\ell$ and $\nu_\ell := \nu H_\ell$. Then $\mathrm{P}_x^{r+r'} = \delta\mu_{r'} + (1-\delta)\nu_{r'}$ and so by convexity (which holds for $D(\cdot\|\pi)$ by Jensen's inequality applied to each $y$ separately) and (3.7)

$$D(\mathrm{P}_x^{r+r'}||\pi) \le \delta D(\mu_{r'}||\pi) + (1-\delta)D(\nu_{r'}||\pi) \le \delta D(\mu_{r'}||\pi) + (1-\delta)\log(1 + \|\nu_{r'} - \pi\|_{2,\pi}^2). \tag{5.5}$$

By (3.8)

$$\|\nu_{r'} - \pi\|_{2,\pi} \le \|\nu - \pi\|_{2,\pi}e^{-14} \le \|\nu - \pi\|_{\infty,\pi}e^{-14} \le (1-\delta)^{-1}e^{-4}.$$

Using $\sqrt{1+a} \le 1 + \sqrt{a}$ and $\log(1+a) \le a$ we get that

$$(1-\delta)\log(1 + \|\nu_{r'} - \pi\|_{2,\pi}^2) \le 2(1-\delta)\log\left(1 + \|\nu_{r'} - \pi\|_{2,\pi}\right) \le 2e^{-4}.$$

By (5.5) to conclude the proof it is left to show that $\delta D(\mu_{r'}||\pi) \le 1/2 - 2e^{-4}$. Denote

$$a_y := 1_{y\in D}[H_r(x,y) - e^{10}\pi(y)], \quad g(y) = a_y/\pi(y).$$

$$\delta D(\mu||\pi) = \sum a_y \log(g(y)/\delta) = \delta|\log\delta| + \mathbb{E}_\pi[g\log g].$$

Since $\delta|\log\delta| \le 1/e$, for all $\delta\in[0,1]$, in order to show that $\delta D(\mu_{r'}||\pi) \le 1/2 - 2e^{-4}$

$$\text{it suffices to show that} \quad \mathbb{E}_\pi[g\log g] \le 1/10 < 1/2 - 1/e - 2e^{-4}. \tag{5.6}$$

Similarly to the proof of Theorem 5.1, let

$$A_s = \{y : g(y) \ge s\} \text{ and } B_s := \{y : \sup_\ell H_\ell(y, A_s) > \sqrt{s+e^{10}}\pi(A_s)|\log\pi(A_s)|\}.$$

Then

$$\mathbb{E}_\pi[g\log g] \le \int_0^\infty \pi(\{y : g(y)\log g(y) > s\})ds = \int_1^\infty (1+\log s)\pi(A_s)ds. \tag{5.7}$$

Note that $(e^{10}+s)\pi(y) \le H_r(x,y)$ for every $y\in A_s$. Hence as in the proof of Theorem 5.1

$$(e^{10}+s)\pi(A_s) \le H_r(x,A_s) \le \mathrm{P}_x[T_{B_s^c} > r] + \mathbb{E}_x[X_r\in A_s \mid T_{B_s^c} \le r]. \tag{5.8}$$

By the definition of $B_s$ and the Markov property,

$$\mathbb{E}_x[X_r\in A_s \mid T_{B_s^c} \le r] \le \sup_{y\notin B_s, \ell\ge 0} H_\ell(y,A_s) \le \sqrt{s+e^{10}}\pi(A_s)|\log\pi(A_s)|. \tag{5.9}$$

By Lemma 3.5 $\pi(B_s) \le e/\sqrt{s+e^{10}} \le 1/2$ and hence by the definition of $r$,

$$\mathrm{P}_x[T_{B_s^c} > r] \le \frac{1}{16e^2(\frac{1}{2}(\log(s+e^{10})+1))^3} = \frac{1}{2e^2(1+\log(s+e^{10}))^3}.$$

As in the proof of Theorem 5.1, it follows that for all $s\ge 1$, $(s+e^{10})\pi(A_s) \le \frac{2}{2e^2(1+\log(s+e^{10}))^3}$, as otherwise by (5.8) $(s+e^{10})\pi(A_s) < 2\mathbb{E}_x[X_r\in A_s \mid T_{B_s^c}\le r]$, which by (5.9) implies that

$$\pi(A_s) \le \exp(-\frac{1}{2}\sqrt{s+e^{10}}) \le \exp(-\sqrt{s/8} - \sqrt{e^{10}/8}) < e^{-50-\sqrt{s/8}} < \frac{(s+e^{10})^{-1}}{e^2(1+\log(s+e^{10}))^3},$$

a contradiction. Thus for all $s\ge 1$,

$$(1+\log s)\pi(A_s) \le \frac{1}{e^2(s+e^{10})(1+\log(s+e^{10}))^2},$$

which yields that $\int_1^\infty (1+\log s)\pi(A_s)ds \le \int_{1+e^{10}}^\infty \frac{e^{-2}ds}{s(1+\log s)^2} = \frac{e^{-2}}{1+\log(1+e^{10})} < e^{-2}/11$. This concludes the proof using (5.6) and (5.7). ☐

### 5.3 Proof of Theorem 1.2

*Proof of Theorem 1.2:* As mentioned in the introduction, it is known that $\kappa \le t_{\mathrm{LS}}$. Denote $r := \frac{1}{2}\kappa$. Note that $P$ and $Q = (P + P^*)/2$ have the same $t_{\mathrm{rel}}$ and $t_{\mathrm{LS}}$. Thus we may work with $S_t = e^{-t(I-Q)}$ instead of $H_t$. By (3.13) it suffices to show that $\|S_r\|_{2\to 4} \le 7$. Fix some $f \in \mathbb{R}^\Omega$ such that $\|f\|_2 = 1$. Our goal is to show that $\|S_r f\|_4 \le 7$. By considering $|f|$ instead of $f$ we may assume that $f \ge 0$. Let

$$A_s := \{x : S_r f(x) \ge s\}.$$

Then $\|S_r f\|_4^4 = \int_0^\infty 4s^3 \pi(A_s) ds \le 6^4 + \int_6^\infty 4s^3 \pi(A_s) ds$. Hence to conclude the proof

$$\text{it suffices to show that} \quad \int_6^\infty 4s^3 \pi(A_s) ds \le 256 \le 7^4 - 6^4. \tag{5.10}$$

Recall that $S_t f(x) = \mathbb{E}_x[f(Y_t)]$ and that for all $A \subset \Omega$, $S_t^A f(a) = \mathbb{E}_a[f(Y_t) 1_{T_{A^c} > t}]$. Let

$$B_s := \{x : \sup_t S_t f(x) > s/2\} = \{f^* > s/2\}, \quad \text{where} \quad f^*(x) = \sup_t S_t f(x)$$

$$D_s := \{x \in B_s : \mathbb{E}_x[f(Y_r) 1_{T_{B_s^c} > r}] \ge s/2\}, \quad F_s := \{x \in B_s : \mathbb{E}_x[f^2(Y_r) 1_{T_{B_s^c} > r}] \ge s^2/4\}.$$

By the Markov property (first inclusion), $A_s \subset D_s \subset F_s$ (the second inclusion follows by the Cauchy-Schwarz inequality). Thus $\pi(A_s) \le \pi(F_s)$. Hence, by (5.10) in order to conclude the proof it suffices to show that $\int_6^\infty 4s^3 \pi(F_s) ds \le 256$. By Starr's maximal inequality (3.14) we know that $\int_0^\infty 16 s \pi(B_s) ds = 64 \|f^*\|_2^2 \le 256 \|f\|_2^2 = 256$. Thus in order to show that $\int_6^\infty 4s^3 \pi(F_s) ds \le 256$, and conclude the proof, it suffices to show that for all $s \ge 6$ we have that $\pi(F_s) \le 4 s^{-2} \pi(B_s)$.

Fix some $s \ge 6$. Note that since $\|f^*\|_2^2 \le 4$, by Markov inequality we have that $\pi(B_s) \le 16/s^2 < 1/2$. Using the spectral decomposition of the restriction of $f$ to $B_s$ (c.f. [2, Lemma 3.8]) and the choice of $r$

$$\mathbb{E}_{\pi_{B_s}}[f^2(Y_r) 1_{T_{B_s^c} > r}] \le \mathbb{E}_{\pi_{B_s}}[f^2(Y_0)] e^{-2\lambda(B_s) r} \le (\|f\|_2^2 / \pi(B_s)) e^{-2\lambda(B_s) r} = (1/\pi(B_s)) \times \pi(B_s) = 1.$$

Thus by the def. of $F_s$, $\frac{1}{4} s^2 \pi_{B_s}(F_s) \le \sum_{y \in F_s} \pi_{B_s}(y) \mathbb{E}_y[f^2(Y_r) 1_{T_{B_s^c} > r}] \le \mathbb{E}_{\pi_{B_s}}[f^2(Y_r) 1_{T_{B_s^c} > r}] \le 1$ and so indeed $\pi(F_s) \le 4 s^{-2} \pi(B_s)$. ☐

### 5.4 The necessary adaptations in the proofs of the results concerning the discrete-time and averaged chains

The proofs of the lower bounds $\rho_{\mathrm{discete}} \le \min(\tau_2, \tau_2^{\mathrm{ave}})$ and $\rho_{\mathrm{Ent}}^{\mathrm{discete}} \le \min(\tau_{\mathrm{Ent}}, \tau_{\mathrm{Ent}}^{\mathrm{ave}})$ in Theorem 2.5 are identical to those from Theorem 1.1 (namely, these are "naive" bounds that can be proven using the same argument as in § 4.1). The inequalities $t_{\mathrm{rel}}^{\mathrm{absolute}} \log 2 \le \tau_2^{\mathrm{discete}}$ and $c t_{\mathrm{rel}}^{\mathrm{absolute}} \le \tau_{\mathrm{Ent}}^{\mathrm{discrete}}$ from (2.8) and (2.9) follow from (3.9)-(3.10), resp.. The proofs of the upper bounds require the following minor adaptations:

(i) In the definition of the sets $A_s$ (both in the proof of the $L_2$ case and of the relative-entropy case) one needs to replace $h_t(x,y)$ with $k_t(x,y) = P^t(x,y)/\pi(y)$.

(ii) In the applications of Starr's inequality one has to work with the discrete-time version, and thus pick up a multiplicative factor of 2 (which is a non-issue). Namely, when applying Lemma 3.5 one has to use the estimate $\|(f_A)_*\|_1 \le 2e\max(1,|\log \pi(A)|)$, instead of $\|f_A^*\|_1 \le e\max(1,|\log\pi(A)|)$.

(iii) One has to replace the Poincaré inequality with the discrete and averaged analogs, to be described shortly (Lemma 5.2).

For the sake of completeness we prove (2.10) in full details, leaving (2.7) and (2.9) as exercises. We first note that the inequalities

$$\tau_2^{\text{ave}} \le \rho_{\text{discete}} + C_1/\alpha_{\text{discrete}} \quad \text{and} \quad \tau_2^{\text{discete}} \le \rho_{\text{discete}} + C_1(1/\alpha_{\text{discrete}} + t_{\text{rel}}^{\text{absolute}})$$

are simple consequences of (2.10). Indeed, this follows from the following extension of the Poincaré inequality.

**Lemma 5.2.** *Assuming reversibility, for all $\mu \in \mathscr{P}(\Omega)$, $M \ge 1$ and $k \in \mathbb{Z}_+$ we have*

$$\|\mu P^k - \pi\|_{2,\pi} \le \|\mu-\pi\|_{2,\pi} e^{-\frac{k}{t_{\text{rel}}^{\text{absolute}}}}, \quad \textit{thus } \tau_2^{\text{discete}} \le \tau_2^{\text{discete}}(M/2) + \lceil t_{\text{rel}}^{\text{absolute}} \log M \rceil. \quad (5.11)$$

$$\|\mu A_k - \pi\|_{2,\pi} \le \|\mu-\pi\|_{2,\pi} \max(|\lambda_2|^k, \frac{1}{2}|\lambda_{|\Omega|}|^k(1+\lambda_{|\Omega|})) \le \|\mu-\pi\|_{2,\pi}\max(e^{-k/t_{\text{rel}}}, \frac{1}{2ek}), \quad (5.12)$$

$$\textit{thus} \quad \tau_2^{\text{ave}} \le \tau_2^{\text{discete}}(M/2) + \lceil \max(t_{\text{rel}}\log M, M) \rceil.$$

*Proof:* We first prove (5.12). The second inequality in (5.12) follows from elementary calculus. We now explain why the first inequality in (5.12) holds. Let $f_\mu = \frac{\mu}{\pi}$. By reversibility $\|\mu A_k - \pi\|_{2,\pi} = \|A_k f_\mu - 1\|_2 = \|A_k(f_\mu - \mathbb{E}_\pi[f_\mu])\|_2 = \frac{1}{2}\|P^k(P+I)(f_\mu - \mathbb{E}_\pi[f_\mu])\|_2$. Consider an orthonormal basis of $\mathbb{R}^\Omega$ consisting of eigenvectors $f_1,\ldots,f_{|\Omega|}$ such that $Pf_i = \lambda_i f_i$ for all $i$ (where $f_1 = 1$ and $\lambda_1 = 1$). Denote $b_j := \mathbb{E}_\pi[f_\mu f_j]$. Then,

$$\|\mu A_k - \pi\|_{2,\pi}^2 = \frac{1}{4}\sum_{i=2}^{|\Omega|} b_j^2 \lambda_i^{2k}(1+\lambda_i)^2 \le \max(\lambda_2^{2k}, \frac{1}{4}\lambda_{|\Omega|}^{2k}(1+\lambda_{|\Omega|})^2)\sum_{j=2}^{|\Omega|} b_j^2.$$

Substituting $\|\mu-\pi\|_{2,\pi}^2 = \|f_\mu - \mathbb{E}_\pi[f_\mu]\|_2^2 = \sum_{j=2}^{|\Omega|} b_j^2$ in the r.h.s. concludes the proof of (5.12). For (5.11),

$$\|\mu P^k - \pi\|_{2,\pi}^2 = \sum_{i=2}^{|\Omega|} b_j^2 \lambda_i^{2k} \le \max(\lambda_2, \lambda_{|\Omega|})^{2k} \sum_{i=2}^{|\Omega|} b_j^2 = \|\mu-\pi\|_{2,\pi}^2 e^{-\frac{2k}{t_{\text{rel}}^{\text{absolute}}}}. \quad \square$$

We now prove (2.10). Define $\bar{\rho}_x^{\text{discrete}}$ and $\bar{\rho}^{\text{discrete}}$ in an analogous manner to the definition of $\bar{\rho}_x$ and $\bar{\rho}$. Then, similarly to Proposition 3.8, we have that $\bar{\rho}_x^{\text{discrete}} \le \rho_x^{\text{discrete}} + C_4\kappa_{\text{discrete}}$, for all $x \in \Omega$. Thus it suffices to show that for all $x$

$$\tau_{2,x}^{\text{discete}}(e^9/2) \le \bar{\rho}_x^{\text{discrete}}. \quad (5.13)$$

Denote $t := \bar{\rho}_x^{\text{discrete}}$ and $A_s^{\text{d}} := \{y : k_t(x,y) \ge s+1\}$, where $k_t(x,y) := P^t(x,y)/\pi(y)$. Since, similarly to Lemma 5.1 we have that $\|P^t(x,\cdot) - \pi(\cdot)\|_{2,\pi}^2 \le \int_\ell^\infty 2s\pi(A_s^{\text{d}})ds + \ell^2$, for all $x \in \Omega$ and $\ell \geq 1$, it suffices to show that $\int_{e^8}^{\infty} 2s\pi(A_s^{\mathrm{d}})ds \leq e^{15} \leq e^{18}/4 - e^{16}$. Let $g_s^{\mathrm{d}}(y) := \sup_k P^k(y, A_s^{\mathrm{d}})/\pi(A_s^{\mathrm{d}})$. Similarly to Lemma 3.5 (using the discrete-time version of Starr inequality) $\|g_s^{\mathrm{d}}\|_1 \leq 2e|\log \pi(A_s^{\mathrm{d}})|$. Let

$$B_s^{\mathrm{d}} := \{y : g_s^{\mathrm{d}}(y) > 4e^{-3}\sqrt{s+1}|\log \pi(A_s^{\mathrm{d}})|\} = \{y : \sup_k P^k(y, A_s^{\mathrm{d}}) \geq 4e^{-3}\sqrt{s+1}\pi(A_s^{\mathrm{d}})|\log \pi(A_s^{\mathrm{d}})|\}.$$

Let $s \geq e^8$. By Markov inequality $\pi(B_s^{\mathrm{d}}) \leq e^4/(2\sqrt{s+1}) \leq \frac{1}{2}$ and so by the definition of $\overline{\rho}_x^{\mathrm{discrete}}$

$$\mathrm{P}_x[T_{\Omega \setminus B_s^{\mathrm{d}}} > t, X_t \in A_s^{\mathrm{d}}] \leq \mathrm{P}_x[T_{\Omega \setminus B_s^{\mathrm{d}}} > t] \leq \frac{e^{12}}{8(s+1)^{3/2}}.$$

Since by the definition of $A_s^{\mathrm{d}}$ (first inequality)

$$(s+1)\pi(A_s^{\mathrm{d}}) \leq P^t(x, A_s^{\mathrm{d}}) = \mathrm{P}_x[T_{\Omega \setminus B_s^{\mathrm{d}}} > t, X_t \in A_s^{\mathrm{d}}] + \mathrm{P}_x[T_{\Omega \setminus B_s^{\mathrm{d}}} \leq t, X_t \in A_s^{\mathrm{d}}],$$

we get that if $\mathrm{P}_x[T_{\Omega \setminus B_s^{\mathrm{d}}} > t, X_t \in A_s] \leq \mathrm{P}_x[T_{\Omega \setminus B_s^{\mathrm{d}}} \leq t, X_t \in A_s]$, then by the Markov property and the definition of $B_s^{\mathrm{d}}$

$$(s+1)\pi(A_s^{\mathrm{d}}) \leq 8e^{-3}\sqrt{s+1}\pi(A_s^{\mathrm{d}})|\log \pi(A_s^{\mathrm{d}})|,$$

which simplifies as follows

$$2s\pi(A_s^{\mathrm{d}}) \leq 2se^{-e^3\sqrt{s+1}/8}.$$

while if $\mathrm{P}_x[T_{\Omega \setminus B_s^{\mathrm{d}}} > t, X_t \in A_s^{\mathrm{d}}] > \mathrm{P}_x[T_{\Omega \setminus B_s^{\mathrm{d}}} \leq t, X_t \in A_s^{\mathrm{d}}]$, then we have that

$$2s\pi(A_s^{\mathrm{d}}) < 4\mathrm{P}_x[T_{\Omega \setminus B_s^{\mathrm{d}}} > t, X_t \in A_s^{\mathrm{d}}] \leq \frac{e^{12}}{2(s+1)^{3/2}}.$$

Let $f(s) = e(s+1)^{3/2}e^{-e^3\sqrt{s+1}/8}$. Then $f(e^8) < e^6$ and for $s \geq e^8$ we have that $\frac{d}{ds}(-f(s)) \geq (\frac{e^4}{16} - \frac{3}{2e^3})(s+1)e^{-e^3\sqrt{s+1}/8} \geq 2se^{-e^3\sqrt{s+1}/8}$. Hence indeed

$$\int_{e^8}^{\infty} 2s\pi(A_s^{\mathrm{d}})ds \leq \int_{e^8}^{\infty} \max(2se^{-e^3\sqrt{s+1}/8}, \frac{e^{12}}{2(s+1)^{3/2}})ds \leq f(e^8) + \frac{e^{12}}{(e^8+1)^{\frac{1}{2}}} \leq e^9.$$

This concludes the proof of (5.13) and thus of (2.10). □

*Proof of Proposition 2.9:* Denote $a_m(x,y) := A_m(x,y)/\pi(y)$. Then

$$\|A_m(x,\cdot) - \pi(\cdot)\|_{2,\pi}^2 = \sum_y \pi(y)(a_m(x,y) - 1)^2 = \sum_y \pi(y)(a_m^2(x,y) - 1).$$

By reversibility, $\sum_y \pi(y)(a_m^2(x,y) - 1) = -1 + \sum_y \frac{A_m(x,y)A_m(y,x)}{\pi(x)} = \frac{((\frac{I+P}{2})^2 P^{2m-2})(x,x)}{\pi(x)} - 1$. Denote the eigenvalues of $P$ by $1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_{|\Omega|} \geq -1$ and let $f_1, \ldots, f_{|\Omega|}$ be an orthonormal basis of $\mathbb{R}^\Omega$ such that $Pf_i = \lambda_i f_i$ for all $i$. Denote $r_-(\ell) := \sum_{i:\lambda_i<0} \lambda_i^\ell f_i^2(x)$ and $r_+(\ell) := \sum_{i>1:\lambda_i>0} \lambda_i^\ell f_i^2(x)$. Using the spectral decomposition and $(1+x)x^{k'} \leq \frac{1}{ek'}$ for $-1 \leq x \leq 0$, we have that

$$\|A_{k+k'}(x,\cdot) - \pi(\cdot)\|_{2,\pi}^2 = \sum_{i=2}^{|\Omega|} \lambda_i^{2(k+k'-1)} \left(\frac{1+\lambda_i}{2}\right)^2 f_i^2(x)$$

$$\le \left(\frac{1}{2ek'}\right)^2 r_-(2k-2) + r_+(2k+2k'-2) \le \left(\frac{1}{2ek'}\right)^2 r_-(2k-2) + \lambda_2^{2k'+2}\|\mathrm{P}_x^{k-2} - \pi\|_{2,\pi}^2,$$

where we have used $r_+(2k+2k'-2) \le \lambda_2^{2k'+2} r_+(2k-4)$ and (using (3.1) and $1-x \le e^{-x}$)

$$r_+(2k-4) \le \sum_{i=2}^{|\Omega|} \exp[-(2k-4)(1-\lambda_i)] f_i^2(x) = h_{2k-4}(x,x) - 1 = \|\mathrm{P}_x^{k-2} - \pi\|_{2,\pi}^2.$$

Denote $k_\ell(x,y) = P^\ell(x,y)/\pi(y)$. Then

$$0 \le k_{2\ell+1}(x,x) = \sum_{i=1}^{|\Omega|} \lambda_i^{2\ell+1} f_i^2(x).$$

Thus $r_-(2\ell+2) \le \sum_{i:\lambda_i>0} \lambda_i^{2\ell} f_i^2(x) \le \sum_{i=1}^{|\Omega|} e^{-2\ell(1-\lambda_i)} f_i^2(x) = h_{2\ell}(x,x) = \|\mathrm{P}_x^\ell - \pi\|_{2,\pi}^2 + 1$. Hence

$$\|A_{k+k'}(x,\cdot) - \pi(\cdot)\|_{2,\pi}^2 \le \left(\frac{1}{2ek'}\right)^2 (\|\mathrm{P}_x^{k-2} - \pi\|_{2,\pi}^2 + 1) + \lambda_2^{2k'+2}\|\mathrm{P}_x^{k-2} - \pi\|_{2,\pi}^2. \quad \square$$

# 6 Application to robustness of mixing

## 6.1 Proof of Corollary 2.1

*Proof.* We only prove (2.1) as the proof of (2.2) is analogous. It is not hard to verify that Theorem 1.1 is still valid in the above setup (this can be formally deduced from Theorem 1.1 via the representation of the generator appearing in the paragraph following Corollary 2.1). Hence it suffices to verify that (2.1) is valid if we replace $\tau_2$ and $\tilde{\tau}_2$ by $\rho$ and $\tilde{\rho}$, resp. (where $\tilde{\rho}$ is the parameter $\rho$ of the chain $(\tilde{X}_t)$). Denote by $\pi$, $t_{\mathrm{rel}}$ and $G$ (resp. $\tilde{\pi}$, $\widetilde{t_{\mathrm{rel}}}$ and $\tilde{G}$) the stationary distribution, relaxation time and generator of $(X_t)$ (resp. $(\tilde{X}_t)$), where $t_{\mathrm{rel}}$ is the inverse of the smallest non-zero eigenvalue of $-G$ (equivalently, we can write $G = K(P-I)$ for some transition matrix $P$ and $K>0$, and define $t_{\mathrm{rel}} := t_{\mathrm{rel}}(P)/K$, where $t_{\mathrm{rel}}(P)$ is the relaxation time of $P$). In the notation of Corollary 2.1 we have that $\tilde{G}(x,y) = r_x G(x,y)$ for all $x,y$ and so $\tilde{\pi}(x) = \frac{\pi(x)/r_x}{L}$ for all $x$, where $L := \sum_y \pi(y)/r_y$ and $\max_y \max(r_y, 1/r_y) \le M$. Hence $M^{-2} \le \pi(x)/\tilde{\pi}(x) \le M^2$ for all $x$. It follows from the extremal characterization of the relaxation time that $t_{\mathrm{rel}} \le M\widetilde{t_{\mathrm{rel}}}$. Indeed, using $\mathrm{Var}_\pi f := \mathbb{E}_\pi[(f-\mathbb{E}_\pi f)^2] \le \mathbb{E}_\pi[(f-\mathbb{E}_{\tilde{\pi}} f)^2] \le LM\mathbb{E}_{\tilde{\pi}}[(f-\mathbb{E}_{\tilde{\pi}} f)^2] = LM\,\mathrm{Var}_{\tilde{\pi}} f$, we get that

$$t_{\mathrm{rel}} = \max \frac{\mathrm{Var}_\pi f}{\mathbb{E}_\pi[(-Gf)f]} = \max \frac{\mathrm{Var}_\pi f}{L\,\mathbb{E}_{\tilde{\pi}}[(-\tilde{G}f)f]} \le \max \frac{M\,\mathrm{Var}_{\tilde{\pi}} f}{\mathbb{E}_{\tilde{\pi}}[(-\tilde{G}f)f]} = M\widetilde{t_{\mathrm{rel}}},$$

where the maxima are taken over all non-constant $f \in \mathbb{R}^\Omega$.

Recall that if $Z \sim \mathrm{Exp}(1)$, then $\alpha^{-1}Z \sim \mathrm{Exp}(\alpha)$ for all $\alpha > 0$. Let $Z_1, Z_2, \ldots$ be i.i.d. Exp(1). A straightforward coupling of the chains in which they follow the same trajectory (i.e. they make the same sequence of jumps, possibly at different times) in which if the $(k+1)$th jump is from vertex $x_k$ then the time spent at $x_k$ by the chains between their $k$th and $(k+1)$th jumps is $\frac{Z_k}{-G(x_k,x_k)}$ and $\frac{Z_k}{-r_xG(x_k,x_k)}$, resp., shows that for all $x$ and $A$ the hitting time of $A$ starting from $x$ for the two chains, $T_A$ and $\tilde{T}_A$, resp., satisfy that

$$\tilde{T}_A/M \leqslant_{\mathrm{st}} T_A \leqslant_{\mathrm{st}} M\tilde{T}_A, \tag{6.1}$$

where $\leqslant_{\mathrm{st}}$ denotes stochastic domination.

Let $\rho_{\le\delta,\delta'} := \inf\{t : \max_{x,A:\pi(A)\le\delta} \mathrm{P}_x[T_A > t] \le \delta'[\pi(A) + \frac{1}{2}\sqrt{\pi(A)\pi(A^c)}]\}$ and similarly $\tilde{\rho}_{\le\delta,\delta'} := \inf\{t : \max_{x,A:\tilde{\pi}(A)\le\delta} \mathrm{P}_x[\tilde{T}_A > t] \le \delta'[\tilde{\pi}(A)+\frac{1}{2}\sqrt{\tilde{\pi}(A)\tilde{\pi}(A^c)}]\}$. By the submultiplicity property of hitting times

$$\forall t \ge 0, m \in \mathbb{Z}_+ \text{ and } A \subset \Omega, \quad \max_x \mathrm{P}_x[\tilde{T}_A > tm] \le (\max_x \mathrm{P}_x[\tilde{T}_A > t])^m,$$

we get that for all $\delta' \in (0, 1/2]$

$$\tilde{\rho}_{\le\frac{1}{2},\delta'} \le C_0\tilde{\rho}|\log\delta'|. \tag{6.2}$$

Similar reasoning as in the proof of (3.20) yields that (c.f. [2, Corollary 3.4])

$$\rho \le \rho_{\le\frac{1}{2M^2},\frac{1}{2}} + C_1 t_{\mathrm{rel}} \log M. \tag{6.3}$$

Using (6.1) and the fact that for all $A$ we have that $\tilde{\pi}(A)/M^2 \le \pi(A) \le M^2\tilde{\pi}(A)$ (first inequality) and (6.2) (second inequality) we get that

$$\rho_{\le\frac{1}{2M^2},\frac{1}{2}} \le M\tilde{\rho}_{\le\frac{1}{2},\frac{1}{2M^2}} \le C_2(M\log M)\tilde{\rho}.$$

This, in conjunction with (6.3) and $t_{\mathrm{rel}} \le M\widetilde{t_{\mathrm{rel}}} \le C_3M\tilde{\rho}$ yields that $\rho \le C(M\log M)\tilde{\rho}$, as desired. By symmetry, we also have that $\tilde{\rho} \le C(M\log M)\rho$. □

### 6.2 Robustness of trees

We start with a few definitions. Given a network $(V, E, (c_e)_{e\in E})$, where each edge $\{u, v\} \in E$ is endowed with a conductance (weight) $c_{u,v} = c_{v,u} > 0$, a random walk on $(V, E, (c_e)_{e\in E})$ repeatedly does the following: when the current state is $v \in V$, the random walk will move to vertex $u$ (such that $\{u, v\} \in E$) with probability $c_{u,v}/c_v$, where $c_v := \sum_{w:\{v,w\}\in E} c_{v,w}$. This is a reversible Markov chain whose stationary distribution is given by $\pi(x) := c_x/c_V$, where $c_V := \sum_{v\in V} c_v = 2\sum_{e\in E} c_e$. Conversely, every reversible Markov chain can be presented in this manner by setting $c_{x,y} = \pi(x)P(x, y)$ (e.g. [13, Section 9.1]).

Let $\mathcal{T} := (V, E)$ be a finite tree. By Kolmogorov's cycle condition every Markov chain on $\mathcal{T}$ (i.e. $P(x, y) > 0$ iff $\{x, y\} \in E$) is reversible. Hence we may assume that $\mathcal{T}$ is equipped with edge weights $(c_e)_{e\in E}$. Following [17], we call a vertex $v \in V$ a central-vertex if each connected component of $\mathcal{T} \setminus \{v\}$ has stationary probability at most $1/2$. A central-vertex always exists (and there may be at most two central-vertices). Throughout, we fix a central-vertex $o$ and call it the *root* of the tree. The root induces a partial order $\prec$ on $V$, as follows. For every $u \in V$, we denote the shortest path between $u$ and $o$ by $\ell(u) = (u_0 = u, u_1, \dots, u_k = o)$. We call $u_1$ the *parent* of $u$. We say that $u' \prec u$ if $u' \in \ell(u)$ (i.e. $u$ is a descendant of $u'$ or $u = u'$). The *induced tree* at $u$ is $\mathcal{T}_u := \{v : u \in \ell(v)\} = \{u\} \cup \{v : v \text{ is a descendant of } u\}$. Fix some leaf $x$ and $\delta \in (0, 1/2)$. Let $W_{x,\delta}$ be the collection of all $y \prec x$ such that $\pi(\mathcal{T}_y) \ge \delta$ and let

$$x_\delta := \mathrm{argmin}\{\pi(\mathcal{T}_y) : y \in W_{x,\delta}\}$$

(i.e. $d(x, x_\delta) = \min_{y \in W_{x,\delta}} d(x, y)$, where $d$ denotes the graph distance w.r.t. $\mathcal{T}$). Recall that $\alpha(A) = \lambda(A)/|\log \pi(A)|$ and that by Theorem 1.2, $\alpha := \sup_{A \in \mathrm{Con}_{1/2}} \alpha(A) \geq c_{\mathrm{LS}}$. Let $D_\beta = D_{\beta,x}$ be the connected component of $x$ in $\mathcal{T} \setminus \{x_\beta\}$. For a leaf $x$ we denote

$$\alpha_x(\delta) := \alpha(D_\delta) \quad \text{and} \quad \alpha_x := \max_{\delta \in (0,1/4]} \alpha_x(\delta) \geq \alpha.$$

Let us now describe the skeleton of the argument in the proof of Theorem 2.1.

Step 1: Show that it suffices to consider leafs as initial states. More precisely

**Lemma 6.1.** *There exists an absolute constant $C > 0$ so that if $y \prec x$ then*

$$\tau_{2,y} \leq \tau_{2,x} + C(t_{\mathrm{LS}} + \sqrt{t_{\mathrm{rel}} \tau_1}). \tag{6.4}$$

Step 2: Show that for a leaf $x$ we can replace (in (3.19)) $\bar{\rho}_x$ (defined in (3.18)) with

$$b_x := \sup_{\delta \in (0,1/4]} b_x(\delta) \quad \text{where} \quad b_x(\delta) := \min\{t : \mathrm{P}_x[T_{x_\delta} > t] \leq \delta^3/4\}.$$

**Proposition 6.2.** *Let $x$ be a leaf. Let $0 < \delta \leq 1/4$ and $A \in \mathrm{Con}_\delta$. Denote $\bar{A} = A^c \setminus D_\delta$, where $D_\beta = D_{\beta,x}$ is the connected component of $x$ in $\mathcal{T} \setminus \{x_\beta\}$. Then*

$$\mathrm{P}_x[T_{A^c} > b_x + 3\kappa + 10 t_{\mathrm{rel}}] \leq \mathrm{P}_x[T_{x_\delta} > b_x] + \mathrm{P}_{x_\delta}[T_{\bar{A}} > 3\kappa + 10 t_{\mathrm{rel}}] < \delta^3/2. \tag{6.5}$$

Step 3 For a leaf $x$ and $\delta \in (0, 1/4]$, derive a large deviation estimate for $T_{x_\delta}$:

**Proposition 6.3.** *There exists some $C > 0$ so that for a leaf $x$ and $\delta \in (0, 1/4]$,*

$$b_x(\delta) \leq \mathbb{E}_x[T_{x_\delta}] + \max\left(\frac{32}{\alpha_x(\delta)}, 8\sqrt{\mathbb{E}_x[T_{x_\delta}]/\alpha_x(\delta)}\right) \leq \tau_1 + C \max(\kappa, \sqrt{\kappa \tau_1}). \tag{6.6}$$

The second inequality follows from the first using the fact that $\mathbb{E}_x[T_{x_\delta}] \leq \tau_1 + C_5 \sqrt{\tau_1 t_{\mathrm{rel}}}$ [2, Corollary 5.5].

Step 4 Similar reasoning as in the proof of (3.20) yields that (c.f. [2, Corollary 3.4])

$$\bar{\rho}_x \leq \min\{t : \mathrm{P}_x[T_{A^c} > t] \leq \pi(A)^3/2 \text{ for all } A \in \mathrm{Con}_{1/4}\} + 10 t_{\mathrm{rel}}$$

By (6.4)-(6.6) in conjunction with (5.1) and (1.7) we have that

$$\begin{aligned}
\tau_2 - C_1\sqrt{t_{\mathrm{rel}}\tau_1} &\leq \max_{x: x \text{ a leaf}} \tau_{2,x} + C_1 t_{\mathrm{LS}} \leq \max_{x: x \text{ a leaf}} \bar{\rho}_x + C_2 t_{\mathrm{LS}} \\
&\leq \max_{x: x \text{ a leaf}} b_x + C_3 t_{\mathrm{LS}} \leq \tau_1 + C_4 \max(t_{\mathrm{LS}}, \sqrt{t_{\mathrm{LS}}\tau_1}). \quad \Box
\end{aligned}$$

**Remark 6.4.** *While it is intuitive that "typically" the worst initial state is a leaf (i.e. $\tau_2 = \tau_{2,x}$ for some leaf $x$), it is not clear if this is always the case.*

To conclude the proof of Theorem 2.1 we now prove Lemma 6.1 and Propositions 6.2-6.3.

*Proof of Lemma 6.1:* Let $y \prec x$. Let $s := \tau_{2,y} - Mt_{\mathrm{LS}}$ for some constant $M > 0$ to be determined later. We may assume $s > 64\sqrt{t_{\mathrm{rel}}\tau_1}$ as otherwise there is nothing to prove. By (5.1) it follows that we can choose $M$ so that $\tau_{2,y} - Mt_{\mathrm{LS}} < \rho_y$, and so for some $A \in \mathrm{Con}_{1/2}$

$$\mathrm{P}_y[T_{A^c} > s] > \pi(A) + \frac{1}{2}\sqrt{\pi(A)\pi(A^c)}. \tag{6.7}$$

Denote the connected component of $x$ in $\mathcal{T}\setminus\{y\}$ by $A'$. Since $y \prec x$ we have $\pi(A') \le 1/2$. Hence, for all $z \in A'$ we have $\mathrm{P}_z[T_y > \tau_1] \le H_{\tau_1}(z, A') \le \pi(A') + 1/4 \le 3/4$. Using the Markov property, by induction we get that $\mathrm{P}_z[T_y > k\tau_1] \le (3/4)^k$ for all $k \in \mathbb{N}$ and $z \in A'$ and so $\mathbb{E}_x[T_y] \le 4\tau_1$.

Let $(v_0 = x, v_1, \dots, v_k = y)$ be the path from $x$ to $y$. Define $\xi_i := T_{v_i} - T_{v_{i-1}}$. Then by the tree structure, under $\mathrm{P}_x$, we have that $T_y = \sum_{i=1}^{k} \xi_i$ and that $\xi_1, \dots, \xi_k$ are independent. Denote $\Phi(\mathcal{T}_{v_i}) := \frac{\pi(v_i)P(v_i,v_{i+1})}{\pi(\mathcal{T}_{v_i})}$. By specializing Kac's formula to trees (see [1, (2.23)] for the general Kac's formula we are using and for its specialization for trees see (6.12) below and c.f. [2, Proposition 5.6 and Lemma 5.2]) we have that $\mathbb{E}_{v_{i-1}}[T_{v_i}] = 1/\Phi(\mathcal{T}_{v_i})$ and that $\mathbb{E}_{v_{i-1}}[T_{v_i}^2] \le 2\mathbb{E}_{v_{i-1}}[T_{v_i}]\mathbb{E}_{\pi_{\mathcal{T}_{v_{i-1}}}}[T_{v_i}] \le 4t_{\mathrm{rel}}\mathbb{E}_{v_{i-1}}[T_{v_i}]$. Whence,

$$\mathrm{Var}_x[T_y] = \sum_{i=1}^{k} \mathrm{Var}_{v_{i-1}}[T_{v_i}] \le \sum_{i=1}^{k} \mathbb{E}_{v_{i-1}}[T_{v_i}^2] \le 4t_{\mathrm{rel}} \sum_{i=1}^{k} \mathbb{E}_{v_{i-1}}[T_{v_i}] = 4t_{\mathrm{rel}}\mathbb{E}_x[T_y] \le 16t_{\mathrm{rel}}\tau_1.$$

By Chebyshev inequality

$$\mathrm{P}_x[|T_y - \mathbb{E}_x[T_y]| > 32\sqrt{t_{\mathrm{rel}}\tau_1}] \le 1/64. \tag{6.8}$$

Let $s' := \max(\mathbb{E}_x[T_y] - 32\sqrt{t_{\mathrm{rel}}\tau_1}, 0)$. By (6.7), (6.8), $s > 64\sqrt{t_{\mathrm{rel}}\tau_1}$ and the Markov property

$$\mathrm{P}_x[X_{s+s'} \in A] \ge \mathrm{P}_x[|T_y - \mathbb{E}_x[T_y]| \le 32\sqrt{t_{\mathrm{rel}}\tau_1}] \times \mathrm{P}_y[T_{A^c} > s] > (\pi(A) + \frac{1}{2}\sqrt{\pi(A)\pi(A^c)})\frac{63}{64},$$

and so $\mathrm{P}_x[X_{s+s'} \in A] \ge \pi(A) + \frac{1}{8}\sqrt{\pi(A)\pi(A^c)}$. The proof is concluded using (4.1) as follows. In the notation from (4.1), $\mathrm{P}_x^{s+s'} \in \mathscr{P}_{A,\delta}$ for some $\delta \ge \frac{1}{8}\sqrt{\pi(A)/\pi(A^c)}$ and thus (using the Poincaré inequality) $\|\mathrm{P}_x^{s+s'-2t_{\mathrm{rel}}} - \pi\|_{2,\pi} > 4\|\mathrm{P}_x^{s+s'} - \pi\|_{2,\pi} \ge 4\delta\sqrt{\pi(A^c)/\pi(A)} \ge 1/2$. □

*Proof of Proposition 6.2:* Fix some leaf $x$, $0 < \delta \le 1/4$ and $A \in \mathrm{Con}_\delta$. Recall that $\bar{A} = A^c \setminus D_\delta$. Using the tree structure it is easy to see that for all $s, s' \ge 0$

$$\mathrm{P}_x[T_{A^c} > s+s'] \le \mathrm{P}_x[T_{\bar{A}} > s+s'] \le \mathrm{P}_x[T_{x_\delta} > s] + \mathrm{P}_{x_\delta}[T_{\bar{A}} > s'] \le \mathrm{P}_x[T_{x_\delta} > s] + \mathrm{P}_{\pi_{\mathcal{T}_{x_\delta}}}[T_{\bar{A}} > s']$$

and so by (3.16), the def. of $b_x$ and the fact that $\pi_{V\setminus\bar{A}}(\mathcal{T}_{x_\delta}) > 1/2$ (as $\pi(V\setminus\bar{A}) < 2\delta < 2\pi(\mathcal{T}_{x_\delta})$)

$$\begin{aligned}\mathrm{P}_x[T_{A^c} > b_x + 3\kappa + 10t_{\mathrm{rel}}] &\le \mathrm{P}_x[T_{x_\delta} > b_x] + \mathrm{P}_{\pi_{\mathcal{T}_{x_\delta}}}[T_{\bar{A}} > 3\kappa + 10t_{\mathrm{rel}}] \\ &< \mathrm{P}_x[T_{x_\delta} > b_x] + 2\mathrm{P}_{\pi_{V\setminus\bar{A}}}[T_{\bar{A}} > 3\kappa + 10t_{\mathrm{rel}}] \le \delta^3/4 + \delta^3/4 = \delta^3/2.\end{aligned}$$ □

*Proof of Proposition 6.3:* By [2, Corollary 5.5] we have that $\mathbb{E}_x[T_{x_\delta}] \le \tau_1 + C_5\sqrt{\tau_1 t_{\mathrm{rel}}}$ and hence it suffices to show that

$$\forall t \in [0, 2\mathbb{E}_x[T_{x_\delta}]], \quad \mathrm{P}_x[T_{x_\delta} \ge \mathbb{E}_x[T_{x_\delta}] + t] \le \exp[-t^2\lambda(D_\delta)/(8\mathbb{E}_x[T_{x_\delta}])]. \tag{6.9}$$

$$\forall t \ge 2\mathbb{E}_x[T_{x_\delta}], \quad \mathrm{P}_x[T_{x_\delta} \ge \mathbb{E}_x[T_{x_\delta}] + t] \le \exp[-\lambda(D_\delta)t/4]. \tag{6.10}$$

Indeed, if $t_1 := 8\sqrt{\mathbb{E}_x[T_{x_\delta}]/\alpha_x(\delta)} \le 2\mathbb{E}_x[T_{x_\delta}]$ then by (6.9) $\mathrm{P}_x[T_{x_\delta} \ge \mathbb{E}_x[T_{x_\delta}] + t_1] \le \delta^3/4$. Otherwise, $t_2 := 32/\alpha_x(\delta) > 2\mathbb{E}_x[T_{x_\delta}]$, and by (6.10), $\mathrm{P}_x[T_{x_\delta} \ge \mathbb{E}_x[T_{x_\delta}] + t_2] \le \delta^3/4$.

We note that (6.9) is essentially Lemma 5.8 in [2]. We start with an auxiliary calculation

**Claim 6.5.** *Fix some leaf $x$ and $\delta \in (0, 1/4]$. Let $D_\delta$ be the connected component of $x$ in $\mathcal{T} \setminus \{x_\delta\}$. Let $y \in D_\delta$ and $z$ be its parent. Then for all $\beta \le \lambda(D_\delta)/2$ we have that*

$$\mathbb{E}_y[e^{\beta T_z}] \le 1 + \mathbb{E}_y[T_z]\beta(1 + 2\beta/\lambda(D_\delta)) \le e^{\mathbb{E}_y[T_z]\beta(1+2\beta/\lambda(D_\delta))}. \tag{6.11}$$

*Proof of* (6.11)*:* Let $\Phi(\mathcal{T}_y) := \frac{\pi(y)P(y,z)}{\pi(\mathcal{T}_y)}$. Let $f$ and $g$ be the density functions of $T_z$ started from $y$ and $\pi_{\mathcal{T}_y}$, resp.. By Kac formula (c.f. [2, Proposition 5.6] or [1, (2.23)]),

$$\forall t \ge 0, \quad g(t) = \Phi(\mathcal{T}_y)\mathrm{P}_y[T_z > t], \quad \text{and hence} \quad \Phi(\mathcal{T}_y)\mathbb{E}_y[T_z] = 1. \tag{6.12}$$

Recall that by (3.16) the law of $T_z$ starting from $\pi_{\mathcal{T}_y}$ is stochastically dominated by the Exponential distribution with parameter $\lambda(\mathcal{T}_y) \ge \lambda(D_\delta)$ and so for every non-decreasing function $k$ we have that $\int_0^\infty k(t)g(t)dt \le \int_0^\infty k(t)\lambda(D_\delta)e^{-\lambda(D_\delta)t}dt$. Finally by (6.12)

$$\mathbb{E}_y[e^{\beta T_z}] - 1 = \int (e^{\beta t} - 1)f(t)dt = \int \beta e^{\beta t}\mathrm{P}_y[T_z > t]dt = \mathbb{E}_y[T_z]\int \beta e^{\beta t}g(t)dt$$

$$= \beta\mathbb{E}_y[T_z]\int e^{\beta t}\lambda(D_\delta)e^{-\lambda(D_\delta)t}dt = \frac{\beta\mathbb{E}_y[T_z]\lambda(D_\delta)}{\lambda(D_\delta) - \beta} \le \mathbb{E}_y[T_z]\beta(1 + 2\beta/\lambda(D_\delta)),$$

where we used $\beta \le \lambda(D_\delta)/2$ to deduce that $\frac{\lambda(D_\delta)}{\lambda(D_\delta)-\beta} = 1 + \frac{\beta}{\lambda(D_\delta)-\beta} \le 1 + \frac{2\beta}{\lambda(D_\delta)}$. ☐

We now return to conclude the proofs of (6.9)-(6.10). Let $t \in [0, 2\mathbb{E}_x[T_{x_\delta}]]$. Set $\beta = \frac{t\lambda(D_\delta)}{4\mathbb{E}_x[T_{x_\delta}]}$ (note that $\beta \le \lambda(D_\delta)/2$). Let the path from $x$ to $x_\delta$ be $(y_1 = x, \ldots, y_r = x_\delta)$. Observe that starting from $x$ we have that $T_{x_\delta} = \sum_{i=2}^r T_{y_i} - T_{y_{i-1}}$. By the Markov property the terms in the sum are independent and $T_{y_i} - T_{y_{i-1}}$ is distributed as $T_{y_i}$ started from $y_{i-1}$. Denote $\mu_i := \mathbb{E}_{y_{i-1}}[T_{y_i}]$ and $\mu := \sum_{i=2}^r \mu_i = \mathbb{E}_x[T_{x_\delta}]$. By (6.11), independence and our choice of $\beta$

$$\mathrm{P}_x[T_{x_\delta} \ge \mu + t] \le e^{-\beta(\mu+t)}\prod_{i=2}^r \mathbb{E}_{y_{i-1}}[e^{\beta T_{y_i}}] \le e^{-\beta(\mu+t)}\prod_{i=2}^r e^{\mu_i\beta(1+2\beta/\lambda(D_\delta))} = e^{-t^2\lambda(D_\delta)/(8\mu)}.$$

The proof of (6.10) is analogous, now with the choice $\beta = \lambda(D_\delta)/2$. ☐

## 7 Open Problems

The *modified Log-Sobolev constant* is defined as

$$c_{\mathrm{MLS}} := \inf_{f \in \mathbb{R}^{\Omega}} \mathcal{E}(e^f, f)/\mathrm{Ent}_{\pi}(e^f).$$

The following question suggests a natural extension of Theorem 1.2. Recall that under reversibility $1/c_{\mathrm{LS}} \le 2\tau_{\infty}$ and $\lambda^{-1}\log 2 \le \tau_1$ (e.g. [13, Lemma 20.11]). The following question asks whether a similar relation holds between $c_{\mathrm{MLS}}$ and $\tau_{\mathrm{Ent}}$.

**Question 7.1.** *Is it the case that $1/c_{\mathrm{MLS}} \le C\tau_{\mathrm{Ent}}$ for some absolute constant $C$?*

**Question 7.2.** *Is it the case that under reversibility $1/c_{\mathrm{MLS}} \le C\rho_{\mathrm{Ent}}$ for some absolute constant $C$ (and thus $1/c_{\mathrm{MLS}} \le C\tau_{\mathrm{Ent}}$)?*

**Question 7.3.** *Recall that under reversibility $\tau_2 \le \rho + C/c_{\mathrm{LS}}$. Is it true that under reversibility $\tau_{\mathrm{Ent}} \le \rho_{\mathrm{Ent}} + C/c_{\mathrm{MLS}}$?*

## Acknowledgements

We are grateful to Shirshendu Ganguly, Gady Kozma, James Lee and Prasad Tetali for useful discussions. Most of the work on this paper was done while the first author was an intern at Microsoft Research, Redmond. The first author would like to thank Microsoft Research for two wonderful summers.

## References

[1] David Aldous and Jim Fill. Reversible markov chains and random walks on graphs, 2002.

[2] Ridhhipratim Basu, Jonathan Hermon, and Yuval Peres. Characterization of cutoff for reversible markov chains. *Ann. Probab., to appear.* `http://arxiv.org/abs/1409.3250`, 2013.

[3] Sergey G Bobkov and Prasad Tetali. Modified logarithmic sobolev inequalities in discrete settings. *Journal of Theoretical Probability*, 19(2):289–336, 2006.

[4] Lucas Boczkowski, Yuval Peres, and Perla Sousi. Sensitivity of mixing times in eulerian digraphs. *arXiv preprint arXiv:1603.05639*, 2016.

[5] Patrick Cattiaux and Arnaud Guillin. Hitting times, functional inequalities, lyapunov conditions and uniform ergodicity. *arXiv preprint arXiv:1604.06336*, 2016.

[6] Persi Diaconis and Laurent Saloff-Coste. Logarithmic sobolev inequalities for finite markov chains. *The Annals of Applied Probability*, 6(3):695–750, 1996.

[7] Jian Ding, Yuval Peres, et al. Sensitivity of mixing times. *Electronic Communications in Probability*, 18:1–6, 2013.

[8] Alison L Gibbs and Francis Edward Su. On choosing and bounding probability metrics. *International statistical review*, 70(3):419–435, 2002.

[9] Sharad Goel, Ravi Montenegro, and Prasad Tetali. Mixing time bounds via the spectral profile. *Electronic Journal of Probability*, 11(1):1–26, 2006.

[10] Jonathan Hermon. On sensitivity of uniform mixing times. *Annales de l'Institut Henri Poincaré Probabilités et Statistiques, to appear. ArXiv preprint arXiv:1607.01672*, 2016.

[11] Jonathan Hermon and Yuval Peres. The power of averaging at two consecutive time steps: Proof of a mixing conjecture by aldous and fill. *Annales de l'Institut Henri Poincaré Probabilités et Statistiques, to appear. ArXiv preprint arXiv:1508.04836*, 2015.

[12] Gady Kozma. On the precision of the spectral profile. *Latin American Journal of Probability and Mathematical Statistics*, 3:321–329, 2007.

[13] David Asher Levin, Yuval Peres, and Elizabeth Lee Wilmer. *Markov chains and mixing times.* American Mathematical Society, Providence, 2009.

[14] Arakaparampil M Mathai and PN Rathie. *Basic concepts in information theory and statistics: axiomatic foundations and applications.* Halsted Press, New York, 1975.

[15] Elchanan Mossel, Krzysztof Oleszkiewicz, and Arnab Sen. On reverse hypercontractivity. *Geometric and Functional Analysis*, 23(3):1062–1097, 2013.

[16] James Norris, Yuval Peres, and Alex Zhai. Surprise probabilities in markov chains. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1759–1773. SIAM, 2015.

[17] Yuval Peres and Perla Sousi. Mixing times are hitting times of large sets. *Journal of Theoretical Probability*, 28(2):488–519, 2015.

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

[19] Norton Starr. Operator limit theorems. *Transactions of the American Mathematical Society*, 121(1):90–115, 1966.