Title
Maximal Inequalities and Mixing Times

Author
Hermon, Jonathan

Publication Date
2016

Peer reviewed|Thesis/dissertation

Maximal Inequalities and Mixing Times

By

Jonathan Hermon

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy
in
Statistics
in the Graduate Division
of
University of California, Berkeley

Committee in charge

Professor Allan Sly, Chair
Professor David Aldous
Professor James Pitman
Professor Fraydoun Rezakhanlou

Fall 2016

Abstract

Maximal Inequalities and Mixing Times

by

Jonathan Hermon

Doctor of Philosophy in Statistics

University of California, Berkeley

Professor Allan Sly, Chair

The aim of this thesis is to present several (co-authored) works of the author concerning applications of maximal inequalities to the theory of Markov chains and put them in a unifying context. Using maximal inequalities we show that different notions of convergence of a Markov chain to its stationary distribution are in some quantitative sense equivalent to some seemingly weaker notions of convergence. In particular, it is shown that the convergence to stationarity of an ergodic reversible Markov chain w.r.t. the L_p distances ($p \in [1,\infty]$) and the relative-entropy distance can be understood (up to a constant factor) in terms of hitting time distributions. We present several applications of these characterizations, mostly ones concerning the cutoff phenomenon and robustness of mixing times.

A sequence of Markov chains is said to exhibit (total variation) cutoff if the convergence to stationarity in total variation distance is abrupt. Though many families of chains are believed to exhibit cutoff, proving the occurrence of this phenomenon is often an extremely challenging task. Verifying the occurrence of the cutoff phenomenon is a major area of modern probability, and despite remarkable progress over the last three decades, there are still only relatively few examples which are completely understood. Although drawing much attention, the progress made in the investigation of the cutoff phenomenon was done mostly through understanding examples and the field suffers from a lack of general theory.

The cutoff phenomenon was given its name by Aldous and Diaconis in their seminal paper [2] from 1986 in which they suggested the following open problem (re-iterated in [14]), which they refer to as "the most interesting problem": "*Find abstract conditions which ensure that the cutoff phenomenon occurs*".

In a joint work with Riddhipratim Basu and Yuval Peres [6] we showed that under reversibility, $t_{\mathrm{mix}}(\varepsilon)$ (the ε total variation mixing time) can be approximated up to an additive term, proportional to the inverse of the spectral gap, by the minimal time required for the chain to escape from every (fixed) set of stationary probability at most $1/2$ w.p. at least $1-\varepsilon$. This substantially refines earlier works which only characterized t_{mix} up to a constant factor. As a consequence, (under reversibility) we derive a necessary and sufficient condition for cutoff in terms of concentration of hitting times of large sets which are "worst" in some sense.

As an application, we show that a sequence of (possibly weighted) nearest neighbor random walks on finite trees exhibits cutoff if and only if it satisfies a spectral condition known as the product condition. We obtain an optimal bound on the size of the cutoff window, and establish sub-gaussian convergence within it. Our proof is robust, allowing us to extend the analysis to weighted random walks with bounded jumps on intervals.

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, τ_2 (while there are sophisticated analytic tools to bound τ_2, in general they do not determine τ_2 up to a constant factor and they lack a probabilistic interpretation). We show that τ_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_{LS}, as a weighted version of the spectral gap. This characterization yields a probabilistic interpretation of c_{LS} in terms of a hitting time version of hypercontractivity. As applications of our results, we show that (1) for every reversible Markov chain, τ_2 is robust under addition of self-loops with bounded weights, and (2) for weighted nearest neighbor random walks on trees, τ_2 is robust under bounded perturbations of the edge weights.

Let $(X_t)_{t=0}^{\infty}$ be an irreducible reversible discrete-time Markov chain on a finite state space Ω. Denote its transition matrix by P. To avoid periodicity issues one often considers the continuous-time version of the chain, whose kernel is given by $H_t := e^{-t}\sum_k \frac{1}{k!}(tP)^k$. Since reversible chains are either aperiodic or have period 2, it is plausible that a single lazy step suffices to eliminate near periodicity issues. This motivates looking at the associated averaged chain, whose distribution at time $t \geq 1$ is obtained by replacing P^t with $A_t := \frac{1}{2}(P^{t-1} + P^t)$. We confirm a conjecture by Aldous and Fill [3, Open Problem 4.17] by showing that under reversibility, for all $t, M > e$ and $x \in \Omega$

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

We deduce that given a sequence of irreducible reversible discrete-time Markov chains, the sequence of the continuous-time chains exhibits cutoff around time t_n iff the same holds for the sequence of averaged chains. Moreover, we show that the size of the cutoff window of the sequence of averaged chains is at most that of the sequence of the continuous-time chains and that the latter can be determined in terms of the former.

Contents

Acknowledgements:

I start by expressing my profound gratitude towards my advisor for the past four years Allan Sly and towards my main collaborator Yuval Peres, who has also been my mentor during two summer internships at Microsoft Research. Without them this dissertation would not exist. It has been an absolute privilege working with them. I thank them for believing in me and exposing me to a broad variety of interesting problems and providing key ideas on how to tackle them. I am very grateful for their kindness and encouragement and sharing their wealth of experience which taught me a lot. Working with Allan and Yuval has been a wonderful learning experience. I thank them both and also David Aldous for supporting me as a GSR for several semesters. I thank Yuval also for his generous hospitality during the two summers he served as my mentor in MSR as well as in various visits to MSR as his guest. Most of all, I want to thank Allan for his infinite patience and willingness to help and provide guidance.

In the Fall of 2013, I started working with Yuval Peres; a collaboration that continues till date. Everything in this dissertation is based on joint works with Yuval, one of which is joint also with Riddhipratim Basu. I thank them for permitting the use of the co-authored material in my dissertation.

I feel extremely fortunate to have Riddhipratim Basu as a friend and collaborator. Working with him and discussing math together as been an absolute pleasure. I thank him for his constant willingness to help, and especially for the guidance he provided to me during the process of postdoc applications.

Thanks are due to David Aldous, Jim Pitman and Fraydoun Rezakhanlou for their help as members of my qualifying examination committee as well as my dissertation committee. I have also learned a lot by taking classes with each of them, as well as with Steve Evans, Will Fithian and Alan Hammond.

I am also grateful to my other co-authors and collaborators, at and outside UC Berkeley, including Itai Benjamini, Uri Feige, Luiz Renato Fontes, Gady Kozma, Hubert Lacoin, Fabio Pablo Machado, Ben Morris, Daniel Reichman and Yumeng Zhang. Itai Benjamini, who has served as my advisor during my masters's study, continues to provide me with guidance and serves as a source of inspiration. I found the collaborations with Hubert and Yumeng especially enjoyable and I thank them for all of their hard work in our joint projects. I also want to thank Perla Sousi for making many useful suggestions regarding several of my works.

I spent two wonderful summers at Microsoft Research, Redmond in 2014 and 22015 as a research intern; I am grateful to the Theory Group for their hospitality. I am grateful to everyone in the Statistics department at UC Berkeley for being extremely kind and helpful and making the department feel like home. The department staff have been exemplary in their help; La Shana Porlaris deserves a special heartfelt thanks for always being present to sort out many a procedural hurdle and answer numerous questions.

Chapter 1

Introduction

Led by the pioneer works of Aldous and Diaconis and driven by applications such as Monte Carlo simulations and approximate counting of large combinatorial sets (many counting problems in theoretical computer science can be reduce to the problem of sampling from the stationary distribution of an auxiliary Markov chain, see [50]) and also by its connections with statistical physics[1], the modern theory of mixing times of Markov chains became in the last few decades a lively and central part of modern probability theory. Among the most fundamental quantities associated with an ergodic Markov chain is its mixing time, which is the number of time units required for it to get within some target distance from the stationary distribution of the chain, denoted by π. In the aforementioned applications the running time of the algorithm/simulation can be bounded in terms of the mixing time (in other words, a rigorous justification of simulation results requires a theoretical bound on the mixing time).

1.1 Basic definitions and notation

Generically, we shall denote the state space of a Markov chain by Ω and its stationary distribution by π. We say that the chain is finite, whenever Ω is finite. Let $(X_t)_{t=0}^{\infty}$ be an irreducible Markov chain on a finite state space Ω with transition matrix P and stationary distribution π. We denote such a chain by (Ω, P, π). The ***hitting time*** of a set $A \subset \Omega$ is defined to be $T_A := \inf\{t : X_t \in A\}$.

1.1.1 Three ways of avoiding periodicity or near-periodicity issues

We call a chain ***lazy***, if $P(x,x) \geq 1/2$ for all $x \in \Omega$. To avoid periodicity and near-periodicity issues, one often considers the lazy version of a discrete time Markov chain, $(X_t^{\mathrm{L}})_{t=0}^{\infty}$, obtained by replacing P with $P_{\mathrm{L}} := \frac{1}{2}(I+P)$. Similarly, one can consider the δ-lazy version of the chain, obtained by replacing P with $\delta I + (1-\delta)P$, in which we refer to δ as the *holding probability*. Periodicity issues can be avoided also by considering the continuous-time version of the chain, $(X_t^{\mathrm{c}})_{t \geq 0}$. This is a continuous-time Markov chain whose heat kernel is given by $H_t = e^{-t(I-P)}$ (that is, $H_t(x,y) := \sum_{k=0}^{\infty} \frac{e^{-t}t^k}{k!} P^k(x,y)$). It is a classic result of probability

[1]Many of the Markov chains for which we desire to estimate the mixing time are often of a statistical mechanics nature, such as the Glauber dynamics for the Ising model.

theory that for any initial condition the distribution of both X_t^{L} and X_t^{c} converge to π when t tends to infinity. The object of the theory of Mixing times of Markov chains is to study the characteristic of this convergence (see [36] for a self-contained introduction to the subject).

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 $t \geq 0$, 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 $t \geq 0$ is $A_t(x,\cdot)$, where $A_t(x,y) := (P^t(x,y) + P^{t+1}(x,y))/2$. Equivalently, $(X_t^{\mathrm{ave}})_{t=0}^{\infty} := (X_{t+\xi})_{t=0}^{\infty}$, where ξ is a Bernoulli(1/2) random variable, independent of $(X_t)_{t=0}^{\infty}$.

We denote by P_μ^t (resp. P_μ) the distribution of X_t (resp. $(X_t)_{t=0}^{\infty}$), given that the initial distribution is μ. To avoid ambiguity we introduce a separate notation for the continuous-time and lazy versions of the chain. We denote by H_μ^t (resp. H_μ) the distribution of X_t^{c} (resp. $(X_t^{\mathrm{c}})_{t\geq 0}$) given that $X_0^{\mathrm{c}} \sim \mu$. Finally, we denote by $\mathrm{P}_{\mathrm{L},\mu}^t$ (resp. $\mathrm{P}_{\mathrm{L},\mu}$) the distribution of X_t^{L} (resp. $(X_t^{\mathrm{L}})_{t=0}^{\infty}$), given that $X_0^{\mathrm{L}} \sim \mu$. When $\mu(\cdot) = 1_{\cdot=x}$, for some $x \in \Omega$, we simply write P_x^t (similarly, H_x^t, and $\mathrm{P}_{\mathrm{L},x}^t$) and P_x (similarly, H_x and $\mathrm{P}_{\mathrm{L},x}$). In the setup where we consider only the continuous-time chain and there is no ambiguity, we usually write X_t, P_x and P_x^t instead of X_t^{c}, H_x and H_x^t.

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 P is ***reversible*** if $P = P^*$. The additive symmetrization is defined as $Q := (P + P^*)/2$. Note that $Q = Q^*$.

1.1.2 Different notions of distance and mixing times and their relations

There are different notions of distance w.r.t. which one can measure the distance from stationarity. Each of which gives rise to a corresponding notion of mixing time and cutoff (to be defined shortly). The most popular is the total-variation distance. We denote the set of probability distributions on a (finite) set B by $\mathscr{P}(B)$. For any $\mu,\nu \in \mathscr{P}(B)$, their ***total-variation distance*** is defined to be $\|\mu - \nu\|_{\mathrm{TV}} := \frac{1}{2}\sum_x |\mu(x) - \nu(x)| = \sum_{x\in B:\,\mu(x)>\nu(x)} \mu(x) - \nu(x)$. The worst-case total variation distance at time t is defined as

$$d(t) := \max_{x\in\Omega} d_x(t), \text{ where for every } x \in \Omega,\ d_x(t) := \|\mathrm{P}_x(X_t \in \cdot) - \pi\|_{\mathrm{TV}},$$

The ***separation*** distance of μ w.r.t. π is $1 - \min_{x\in\Omega} \frac{\mu(x)}{\pi(x)}$. Thus the worst case separation distance at time t is $d_{\mathrm{sep}}(t) := 1 - \min_{x,y} P^t(x,y)/\pi(y)$. The **$\varepsilon$-total variation and separation mixing times** are defined (resp.) as

$$t_{\mathrm{mix}}(\varepsilon) := \inf\{t : d(t) \leqslant \varepsilon\} \quad \text{and} \quad t_{\mathrm{sep}}(\varepsilon) := \inf\{t : d_{\mathrm{sep}}(t) \leq \varepsilon\}.$$

When $\varepsilon = 1/4$ we omit it from the above notation. It is a classical result (e.g. [36, Lemmas 6.13 and 19.3]) that under reversibility the separation and total-variation distances and mixing times can be compared as follows (the second line being an easy consequence of the first)

$$\begin{aligned} &\forall t \geq 0, \quad d(t) \leq d_{\mathrm{sep}}(t) \leq 1 - (1 - \min(2d(t/2),1))^2 \leq 4d(t/2), \\ &\forall a \in (0,1), \qquad t_{\mathrm{mix}}(a) \leq t_{\mathrm{sep}}(a) \leq 2t_{\mathrm{mix}}(a/4). \end{aligned} \tag{1.1.1}$$

Another important family of distances is the family of L_p distances ($1 \le p \le \infty$). 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 $\mathbb{E}_\pi[g] := \sum_x \pi(x)g(x)$) and $\|f\|_\infty := \max_x |f(x)|$. The L_p norm of a signed measure σ 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 $d_{p,x}(t) := \|\mathrm{H}_x^t - \pi\|_{p,\pi}$. When considering the L_p distance for $p > 1$ we always consider continuous-time chains, unless otherwise is specified (and thus may write P instead of H, etc...). Denote $h_t(x,y) := H_t(x,y)/\pi(y)$. Under reversibility for all $x \in \Omega$ and $t \ge 0$ (e.g. (2.2) in [23])

$$d_{2,x}^2(t) = h_{2t}(x,x) - 1, \quad d_\infty(t) = \max_y h_t(y,y) - 1. \tag{1.1.2}$$

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

$$\tau_p(\varepsilon) := \max_x \tau_{p,x}(\varepsilon), \qquad \text{where } \tau_{p,x}(\varepsilon) := \min\{t : d_{p,x}(t) \le \varepsilon\}. \tag{1.1.3}$$

When $\varepsilon = 1/2$ we omit it from this notation (this is consistent with the notation $t_{\mathrm{mix}} = t_{\mathrm{mix}}(1/4)$ since $\tau_1(\varepsilon) = t_{\mathrm{mix}}(2\varepsilon)$). Let $m_p := 1 + \lceil (2-p)/(2(p-1)) \rceil$. It follows from (1.1.2), 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. [49, Lemma 2.4.6]):

$$\begin{aligned} \tau_2(a) \le &\tau_p(a) \le 2\tau_2(\sqrt{a}) = \tau_\infty(a) && \text{for all } p \in (2,\infty] \text{ and } a > 0, \\ \frac{1}{m_p}&\tau_2(a^{m_p}) \le \tau_p(a) \le \tau_2(a) && \text{for all } p \in (1,2) \text{ and } a > 0, \end{aligned} \tag{1.1.4}$$

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

The ***relative entropy*** of a distribution μ w.r.t. π is defined as

$$D(\mu||\pi) := \sum_x \mu(x)\log(\mu(x)/\pi(x)), \tag{1.1.5}$$

The mixing time in relative entropy is defined as

$$\tau_{\mathrm{Ent},x} := \inf\{t : D(\mathrm{P}_x^t||\pi) \le 1/2\} \quad \text{and} \quad \tau_{\mathrm{Ent}} = \max_x \tau_{\mathrm{Ent},x}. \tag{1.1.6}$$

The relative entropy distance can be compared with the L_1 and L_2 distances as follows: [32] $2D(\mu||\pi) \ge \|\mu - \pi\|_{1,\pi}^2$ and ([22, Theorem 5])

$$D(\mu||\pi) \le \log(1 + \|\mu - \pi\|_{2,\pi}^2). \tag{1.1.7}$$

1.1.3 The spectral gap and the Log-Sobolev constant

Denote $S_t := e^{-(I-Q)t} = \sum_{k=0}^\infty \frac{e^{-t}t^k}{k!}Q^t$. When considering S_t instead of H_t we write $\mathbb{P}_x^t$, $\mathbb{P}_x$ and Y_t instead of H_x^t, H_x and X_t^{c}, respectively. We start with a few basic definitions.

Definition 1.1.1. *Let* (Ω, P, π) *be a finite chain. For* $f, g \in \mathbb{R}^\Omega$*, let* $\mathrm{Var}_\pi f := \|f - \mathbb{E}_\pi f\|_2^2$ *and* $\langle f, g\rangle_\pi := \mathbb{E}_\pi[fg]$*. We identify* P^t*,* P_{L}^t*,* A_t*,* H_t *and* S_t *with the linear operators on* $L^p(\mathbb{R}^\Omega, \pi)$

$$A_t f(x) := \sum_{y\in\Omega} A_t(x,y)f(y) = \mathbb{E}_x[f(X_t^{\mathrm{ave}})], \quad H_t f(x) := \sum_{y\in\Omega} H_t(x,y)f(y) = \mathbb{E}_x[f(X_t^{\mathrm{c}})],$$
$$P^t f(x) := \sum_{y\in\Omega} P^t(x,y)f(y) = \mathbb{E}_x[f(X_t)], \quad P_{\mathrm{L}}^t f(x) := \sum_{y\in\Omega} P_{\mathrm{L}}^t(x,y)f(y) = \mathbb{E}_x[f(X_t^{\mathrm{L}})],$$

and $S_t f(x) := \sum_{y\in\Omega} S_t(x,y)f(y) = \mathbb{E}_x[f(Y_t)]$*. Similarly,* $Q^t f(x) := \sum_{y\in\Omega} Q^t(x,y)f(y)$*.*

Recall that if (Ω, P, π) is a finite irreducible chain, then Q is self-adjoint w.r.t. $\langle\cdot,\cdot\rangle_\pi$, the standard inner product induced by π, and hence has $|\Omega|$ real eigenvalues. Throughout we shall denote them by $1 = \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_{|\Omega|} \geq -1$ (where $\lambda_2 < 1$ since the chain is irreducible and for a lazy chain we also have that $\lambda_{|\Omega|} \geqslant 0$). The ***spectral-gap*** and ***relaxation-time*** of P are defined as $\lambda := (1-\lambda_2)$ and $t_{\mathrm{rel}} = 1/\lambda$. The ***absolute relaxation-time*** is $t_{\mathrm{rel}}^{\mathrm{absolute}} := \max\{(1-\lambda_2)^{-1}, (1-|\lambda_{|\Omega|}|)^{-1}\}$.[2] Under reversibility, the following general relation holds for continuous-time chains and for discrete-time lazy chains

$$(t_{\mathrm{rel}} - 1)|\log(2\varepsilon)| \leq t_{\mathrm{mix}}(\varepsilon) \leq \tau_2(2\varepsilon) \leq t_{\mathrm{rel}}|\log(\varepsilon\pi_*)|, \tag{1.1.8}$$

where $\pi_* := \min_{x\in\Omega}\pi(x)$. (see [36] Theorems 12.3 and 12.4). The following fact (often referred to as the Poincaré inequality or L_2-contraction) is standard (e.g. [23]).

Fact 1.1.2. *Let* (Ω, P, π) *be a finite irreducible Markov chain. Let* $\mu \in \mathscr{P}(\Omega)$*. Then*

$$\|\mu H_s - \pi\|_{2,\pi} \leq e^{-s/t_{\mathrm{rel}}}\|\mu - \pi\|_{2,\pi} \quad \text{and} \quad \|\mu P^s - \pi\|_{2,\pi} \leq e^{-s/t_{\mathrm{rel}}^{\mathrm{absolute}}}\|\mu - \pi\|_{2,\pi}, \tag{1.1.9}$$

for all $s \geq 0$*. In particular, for all* $x \in \Omega$ *and* $M \geq 1$*,*

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

Similarly, for all $f \in \mathbb{R}^\Omega$ *and* $t \geq 0$*, we have* $\mathrm{Var}_\pi H_t f \leq e^{-2t/t_{\mathrm{rel}}}\mathrm{Var}_\pi f$ *and*

$$\mathrm{Var}_\pi P^k f \leq e^{-2k/t_{\mathrm{rel}}^{\mathrm{absolute}}}\mathrm{Var}_\pi f, \text{ for all } k \in \mathbb{Z}_+. \tag{1.1.10}$$

Another important parameter is the *Log-Sobolev constant* (resp. time) defined as

$$c_{\mathrm{LS}} := \inf\{\mathcal{E}(f)/\mathrm{Ent}_\pi(f^2) : f \text{ is non-constant}\}, \quad (\text{resp. } t_{\mathrm{LS}} := 1/c_{\mathrm{LS}}) \tag{1.1.11}$$

where (with the convention $0\log 0 = 0$)

$$\mathrm{Ent}_\pi(f) := \mathbb{E}_\pi[f\log f] - \mathbb{E}_\pi[f]\log\mathbb{E}_\pi[f] = \mathbb{E}_\pi[f\log(f/\mathbb{E}_\pi[f])],$$
$$\mathcal{E}(f,g) := \langle (I-Q)f, g\rangle_\pi \quad \text{and} \quad \mathcal{E}(f) := \mathcal{E}(f,f).$$

Note that $D(\mu||\pi) = \mathrm{Ent}_\pi(\mu/\pi)$. It is always the case that $t_{\mathrm{LS}} \geq 2t_{\mathrm{rel}}$ ([16, Lemma 3.1]).

[2] Note that for lazy chains $t_{\mathrm{rel}} = t_{\mathrm{rel}}^{\mathrm{absolute}}$.

There are numerous works aiming towards general geometric upper bounds on τ_∞. Among the most advanced techniques are the spectral profile [23] and Logarithmic Sobolev inequalities (see [16] for a survey on the topic). It is classical (e.g. [16, Corollary 3.11]) that for reversible chains

$$t_{\mathrm{LS}}/2 \le \tau_2(1/e) \le t_{\mathrm{LS}}(1+\frac{1}{4}\log\log(1/\pi_*)). \tag{1.1.12}$$

There are examples demonstrating both bounds can be attained up to a constant factor.

The Log-Sobolev constant has a useful characterization in terms of hypercontractivity. Let $1 \le p_1, p_2 \le \infty$. The $p_1 \to p_2$ norm of a linear operator $\mathbf{A}$ is given by

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

If $\|\mathbf{A}\|_{p_1\to p_2} \le 1$ for some $1 \le p_2 < p_1 \le \infty$ we say that $\mathbf{A}$ is a hypercontraction. For all p_1, p_2, we have that $\|H_t\|_{p_1\to p_2}$ is non-increasing in t. It is a classic result (e.g. [16, Theorem 3.5]) that the Log-Sobolev time can be characterized in terms of hypercontrativity.

Fact 1.1.3. *Let (Ω, P, π) be a finite reversible chain. Let $s_q := \inf\{t : \|H_t\|_{2\to q} \le 1\}$. Then $t_{\mathrm{LS}} = 4\sup_{q:2<q<\infty} s_q/\log(q-1)$.*

1.2 Maximal inequalities

In this section we present some maximal inequality which shall be central in what comes.

Theorem 1.2.1 (Starr's Maximal inequality [51])**.** *Let (Ω, P, π) 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 $1 < p < \infty$

$$\|f^*\|_p \le p^*\|f\|_p, \quad \textit{where} \quad p^* := p/(p-1) \textit{ is the conjugate exponent of } p. \tag{1.2.1}$$

Moreover, under reversibility, for $f_{*,\mathrm{even}} := \sup_{k\in\mathbb{Z}_+} |P^{2k}f(x)|$ we have that (1.2.1) holds also with $f_{*,\mathrm{even}}$ in the role of f^* and hence $f_* := \sup_{k\in\mathbb{Z}_+} |P^k f(x)|$ satisfies for all $1 < p < \infty$

$$\|f_*\|_p^p \le \|f_{*,\mathrm{even}}\|_p^p + \|(Pf)_{*,\mathrm{even}}\|_p^p \le (p^*)^p(\|f\|_p + \|Pf\|_p^p) \le 2(p^*)^p\|f\|_p. \tag{1.2.2}$$

For a short proof of Starr's inequality (based on his original argument) see [6, Theorem 2.3]. The following lemma is essentially due to Norris, Peres and Zhai [45].

Lemma 1.2.2. *Let (Ω, P, π) be a finite irreducible Markov chain. Let $f_A(x) := 1_{x\in A}/\pi(A)$.*

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

Proof. By (1.2.1) 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 ε to 0 (noting that the r.h.s. is continuous w.r.t. p_*) concludes the proof. $\square$

We note that by [51, Theorem 2] $(1-e^{-1})\|f_A^*\|-1 \le \|f_A \log[\max(1,|f_A|)]\|_1 = |\log \pi(A)|$.

Definition 1.2.3. *Let P be a linear operator and $k \in \mathbb{Z}_+$. We define $\triangle P^k := P^{k+1} - P^k = P^k(P-I)$. For $r > 1$, we define inductively $\triangle^r P^k := \triangle(\triangle^{r-1}P^k) = \triangle^{r-1}P^{k+1} - \triangle^{r-1}P^k = P^k(P-I)^r$. Similarly, we define $\triangle A_k := A_{k+1} - A_k = \frac{1}{2}P^k(P^2-I)$.*

Let (Ω,μ) be a probability space. Let $P : L^2(\Omega,\mu) \to L^2(\Omega,\mu)$ be a positive (i.e. $f \ge 0 \implies Pf \ge 0$) self-adjoint linear operator whose spectrum is contained in the interval $[0,1]$. It is noted in [52] that for all $r \ge 1$, there exists a constant C_r (independent of (Ω,μ) and P), such that for every $f \in L^2(\Omega,\mu)$

$$\|\sup_{t\ge 0}(t+1)^r \triangle^r P^t f\|_2 \le C_r\|f\|_2. \tag{1.2.3}$$

In [35] Stein's argument is extended to the setup where P is a positive contraction with $M(P) := \sup_t t\|P^{t+1} - P^t\|_2 < \infty$ without the assumptions that P is self-adjoint and that its spectrum is contained in $[0,1]$. In this more general setup C_r depends also on $M(P)$.

Corollary 1.2.4. *There exists an absolute constant C such that for every finite irreducible reversible Markov chain, (Ω,P,π) and every $f \in \mathbb{R}^\Omega$*

$$\left\|\sup_{t\ge 0}(t+1)\triangle P_{\mathrm{L}}^t f\right\|_2^2 \le C\mathrm{Var}_\pi f \quad \text{and} \quad \left\|\sup_{t\ge 0}(t+1)\triangle A_t f\right\|_2^2 \le C\mathrm{Var}_\pi f. \tag{1.2.4}$$

Proof: Note that $\triangle A_{2t} f = \frac{P^{2t+2}-P^{2t}}{2} f = \frac{1}{2}\triangle(P^2)^t f$ and $\triangle A_{2t+1} f = \frac{1}{2}\triangle(P^2)^t(Pf)$. Hence (1.2.4) follows from (1.2.3) applied to P_{L} and P^2 by noting that $\triangle P_{\mathrm{L}}^t f = \triangle P_{\mathrm{L}}^t(f - \mathrm{E}_\pi[f])$, $\triangle A_t f = \triangle A_t(f - \mathrm{E}_\pi[f])$ and $\mathrm{Var}_\pi(Pf) \le \mathrm{Var}_\pi f$. ☐

1.3 The cutoff phenomenon

Next, consider a sequence of such chains, $((\Omega_n, P_n, \pi_n) : n \in \mathbb{N})$, each with its corresponding worst-distance from stationarity $d^{(n)}(t)$, its mixing-time $t_{\mathrm{mix}}^{(n)}$, etc.. Loosely speaking, the (total variation) ***cutoff phenomenon*** occurs when over a negligible period of time, known as the ***cutoff window***, the (worst-case) total variation distance (of a certain finite Markov chain from its stationary distribution) drops abruptly from a value close to 1 to near 0. In other words, one should run the n-th chain until the cutoff point for it to even slightly mix in total variation, whereas running it any further is essentially redundant. Formally, we say that the sequence exhibits a total variation ***cutoff*** (resp. ***pre-cutoff***) if the following sharp transition in its convergence to stationarity occurs:

$$\lim_{n\to\infty}\frac{t_{\mathrm{mix}}^{(n)}(\varepsilon)}{t_{\mathrm{mix}}^{(n)}(1-\varepsilon)} = 1, \text{ for all } 0<\varepsilon<1 \quad (\text{resp.} \sup_{0<\varepsilon\le 1/2}\limsup_{n\to\infty}\frac{t_{\mathrm{mix}}^{(n)}(\varepsilon)}{t_{\mathrm{mix}}^{(n)}(1-\varepsilon)} < \infty). \tag{1.3.1}$$

We say that the sequence has a ***cutoff window*** w_n, if $w_n = o(t_{\mathrm{mix}}^{(n)})$ and for every $\varepsilon \in (0,1)$ there exists $c_\varepsilon > 0$ such that for all n

$$t_{\mathrm{mix}}^{(n)}(\varepsilon) - t_{\mathrm{mix}}^{(n)}(1-\varepsilon) \le c_\varepsilon w_n. \tag{1.3.2}$$

Similarly, we later consider the corresponding notions of the mixing times and cutoffs for the sequences of the continuous-time, lazy and averaged (resp.) versions of the chain[3]

One can define the notions of separation and L_p cutoffs and cutoff windows in an analogous manner, by replacing $t_{\mathrm{mix}}^{(n)}(\cdot)$ in (1.3.1)-(1.3.2) with $t_{\mathrm{sep}}^{(n)}(\cdot)$ and $\tau_p^{(n)}(\cdot)$, resp.. However, whenever we use the term "cutoff" without specifying the relevant notion of distance we always consider the total variation distance.

1.3.1 Historical review of the cutoff phenomenon

The cutoff phenomenon was first identified for random transpositions in the ingenious work of Diaconis and Shahshahani [18] and later by Aldous for random walk on the hypercube [1] (see [15] for a survey on the topic). The cutoff phenomenon was given its name by Aldous and Diaconis in their seminal paper [2] from 1986 in which they verified cutoff for several card shuffling schemes and suggested the following open problem (re-iterated in [14]), which they refer to as "the most interesting problem": "*Find abstract conditions which ensure that the cutoff phenomenon occurs*". Though many families of chains are believed to exhibit cutoff, proving the occurrence of this phenomenon is often an extremely challenging task. Verifying the occurrence of the cutoff phenomenon is a major area of modern probability, and despite remarkable progress over the last three decades, there are still only relatively few examples which are completely understood.

We now give a short review of some large families of chains for which the cutoff phenomenon is well-understood. The cutoff phenomenon for the glauber dynamics of the Ising model was studied by several authors with the state of the art being Lubetzky and Sly work on the universality of cutoff (for the Ising model) including some extremely precise results concerning the lattice case, using their Information Percolation approach [38, 39, 40]. They also proved that simple random walk on a sequence of random regular graphs exhibits cutoff (with high probability over the choice of the graph) [41]. Together with Berestycki and Peres they later generalized this result to the setup of random graphs with a given degree sequence [7] (with minimal degree at least 3). Bordenave et al. [10] treated the random digraph setup. Cutoff for the random adjacent transposition was verified by Lacoin [34] who very recently verified cutoff also for the biased setup together with with Labbé [33]. Cutoff for simple random walk on lamplighter graphs whose base graphs are locally uniformly transient[4] was verified by Miller and Peres [43].

Although drawing much attention, the progress made in the investigation of the cutoff phenomenon has been achieved mostly through understanding examples and the field suffers from a lack of general theory. In 2004 [47], during an AIM workshop on the cutoff phenomenon, Peres introduced the so called ***product condition***:

$$t_{\mathrm{rel}}^{(n)} = o(t_{\mathrm{mix}}^{(n)}) \quad (\text{equivalently } \lim_{n\to\infty} \lambda^{(n)} t_{\mathrm{mix}}^{(n)} = \infty), \tag{1.3.3}$$

a necessary condition for pre-cutoff[5] (by (1.1.8); e.g. [36, Proposition 18.4]), and suggested that it is also a sufficient condition for cutoff for many "nice" families of reversible chains.

[3] We defer the presentation of the relevant notation until the point in which it is used.

[4] That is, the effective resistance between any pair of vertices is uniformly bounded from above.

[5] Using the product condition it is not hard to verify that simple random walk on a sequence of tori of increasing side lengths and fixed dimension does not exhibit pre-cutoff (e.g. [36, Chapter 12]).

In general, the product condition does not always imply cutoff. Aldous and Pak have constructed relevant (reversible) examples (see [36, Chapter 18]). This left open the problem of identifying general classes of chains for which the product condition is indeed sufficient for cutoff. This was verified e.g. for lazy birth and death chains, first for separation cutoff by Diaconis and Saloff-Coste [17] and later for total variation by Ding et al. [19]. We note that Chen and Saloff-Coste [13] proved that for a sequence of reversible chains the condition $t_{\mathrm{rel}}^{(n)} = o(\tau_2^{(n)})$ is equivalent to cutoff in L_2 (and in fact to cutoff in L_p for all $p > 1$).

1.3.2 The author's contribution

Most of the author's graduate research has been centered on the problem of developing general theory of Markov chains, with a special emphasis on the cutoff phenomenon. Together with Basu and Peres we gave the first general characterization of the cutoff phenomenon, in terms of concentration of hitting times of sets which are "worst" in some sense[6] [6, Theorem 3]. This may be considered as an answer to the aforementioned problem of Aldous and Diaconis of finding an abstract sufficient condition for cutoff. As an application, we show that a sequence of (possibly weighted) nearest neighbor random walks on finite trees exhibits cutoff if and only if it satisfies the product condition [6, Theorem 1]. We obtain an optimal bound on the size of the cutoff window, and establish sub-gaussian convergence within it. Our proof is robust and can be extended to weighted random walks with bounded jumps on intervals [6, Theorem 2].

The novelty of our approach is the usage of Starr's L^p maximal inequality [51] in order to deduce that for all $A \subset \Omega$ and $\varepsilon \in (0,1)$ there exists a set G_A of large stationary probability, say $\pi(G_A) \geq 1-\varepsilon/2$, such that $\sup_{t:t \geq Ct_{\mathrm{rel}}|\log \varepsilon|} |P^t(g,A)-\pi(A)| \leq \varepsilon$, for all $g \in G_A$. Hence by the Markov property, hitting G_A by time $t - Ct_{\mathrm{rel}}|\log \varepsilon|$ serves as a certificate that the chain is ε-mixed w.r.t. A at time t. As a consequence, we show that under reversibility, $t_{\mathrm{mix}}(\varepsilon)$ can be approximated up to an additive term proportional to the inverse of the spectral gap, by the minimal time required for the chain to escape from every set of stationary probability at most β (where $\beta \in (0,1)$ is arbitrary) w.p. at least $1-\varepsilon$, denoted by $\mathrm{hit}_\beta(\varepsilon)$. In particular, we show that for all $\varepsilon \in (0,1)$ and $\delta \leq \frac{1}{2}(\varepsilon \wedge (1-\varepsilon))$

$$\mathrm{hit}_{1/2}(\varepsilon+\delta) - 4t_{\mathrm{rel}}|\log \delta| \leq t_{\mathrm{mix}}(\varepsilon) \leq \mathrm{hit}_{1/2}(\varepsilon-\delta) + 4t_{\mathrm{rel}}|\log \delta|, \tag{1.3.4}$$

This substantially refines earlier works [4, 37, 48, 46] which only characterized t_{mix} up to a constant factor (and only for a fixed ε). Indeed, under reversibility $(t_{\mathrm{rel}}-1)|\log(2\varepsilon)| \leq t_{\mathrm{mix}}(\varepsilon)$, however under the product condition, the terms involving t_{rel} in (1.3.4) are negligible.

Through more sophisticated applications of Starr's inequality, in a joint work with Peres [29] we obtained a characterization of L_2 mixing and hypercontractivity in terms of hitting times distributions. We show that (under reversibility) τ_2 (resp. τ_{Ent}) is roughly the minimal time required for the chain to escape from every set A of stationary mass at most $1/2$ w.p. at least $1 - \frac{1}{2}\sqrt{\pi(A)\pi(A^c)}$, denoted by ρ (resp. $1 - \frac{C}{|\log \pi(A)|}$ denoted by ρ_{RE}). More precisely,

$$\rho \leq \tau_2 \leq \rho + C_1/c_{\mathrm{LS}} \leq C_2\rho \quad \text{and} \quad \rho_{\mathrm{RE}} \leq \tau_{\mathrm{Ent}} \leq C_2\rho_{\mathrm{RE}}. \tag{1.3.5}$$

[6] In [25] we gave certain extensions of this result, showing in particular that under transitivity, "worst" can be interpreted as maximizing the expected hitting time among all large sets.

In contrast with the aforementioned characterization of cutoff for reversible chains, some of the difficulties in developing a "general theory of cutoff" are demonstrated in [27] (joint with Hubert Lacoin and Peres) and [30] (joint with Peres). In the former we provided a negative answer to a question asked by Ding et al. [19] by showing that separation cutoff need not imply total variation cutoff and vice-versa. This is surprising considering the aforementioned equivalence of L_p cutoffs for all $p > 1$ and the fact that the total variation and separation distances are intimately related (1.1.1) in a way which resemblance the relation between the L_p distances ((1.1.4)). In the latter we show that (in the setup of simple random walk on unweighted bounded degree graphs) multiplication of the edge weights by a $1 + o(1)$ factor may increase the order of the mixing time and lead to cutoff, despite the fact that before the perturbations the sequence of simple random walks did not exhibit even pre-cutoff! Moreover, we show that the occurrence of separation cutoff may depend on the choice of the holding probability (in the discrete-time setup; Similarly, the corresponding sequence of discrete-time lazy chains may exhibit separation cutoff, while the sequence of the associated continuous-time chains does not). This is in sharp contrast with the total-variation case, due to Chen and Saloff-Coste [12].

All of these examples are either simple random walks on bounded degree (unweighted) graphs, or can be transformed into this setup using machinery from [30]. Since the analysis of these examples is very subtle[7] we chose not to include them in this thesis, despite their compatibility with its overall theme.

In a joint work with Yuval Peres [28], we confirm a conjecture by Aldous and Fill [3, Open Problem 4.17] by showing that under reversibility, for all $t, M > e$ and $x \in \Omega$

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

We deduce that given a sequence of irreducible reversible discrete-time Markov chains, the sequence of the continuous-time chains exhibits cutoff around time t_n iff the same holds for the sequence of averaged chains. Moreover, we show that the size of the cutoff window of the sequence of averaged chains is at most that of the sequence of the continuous-time chains and that the latter can be determined in terms of the former.

The main technical tool we utilize is a maximal inequality due to Stein (namely, Corollary 1.2.4). As demonstrated by the above results, the idea of using ergodic theoretical maximal inequalities to study Markov chains holds much promise. In the next chapter we revisit the above results in more details and make various comments and remarks.

[7] Here "subtle" serves as a euphemism for "lengthy and tedious".

Chapter 2

The main results

2.0.3 A historical review on the connections between hitting times and mixing

The idea that expected hitting times of sets which are "worst in expectation" (in the sense of (2.0.1) below) could be related to the mixing time is quite old and goes back to Aldous' 1982 paper [4]. A similar result was obtained later by Lovász and Winkler ([37] Proposition 4.8) as part of their elegant body of work on stopping rules.

The aforementioned connection was substantially refined recently by Peres and Sousi ([48] Theorem 1.1) and independently by Oliveira ([46] Theorem 2). Peres and Sousi [48] considered the mixing times of the associated lazy and averaged chains: denoted by $t_{\mathrm{L}} := \inf \left\{ t : \max_x \|\mathrm{P}_{x,\mathrm{L}}^t - \pi\|_{\mathrm{TV}} \leqslant 1/4 \right\}$ and $t_{\mathrm{ave}} := t_{\mathrm{ave}}(1/4)$, where $t_{\mathrm{ave}}(\varepsilon) := \inf \{t : d_{\mathrm{ave}}(t) \leqslant \varepsilon\}$, and $d_{\mathrm{ave}}(t) := \max_x \|A_t(x,\cdot) - \pi(\cdot)\|_{\mathrm{TV}}$. Their approach relied on the theory of random times to stationarity combined with a certain involved "de-randomization" argument (which is the discrete-time analog of a similar argument due to Aldous [4]) which shows that for every finite irreducible reversible Markov chain and every stopping time T such that $X_T \sim \pi$, we have that $t_{\mathrm{ave}} \leq 220 \max_{x \in \Omega} \mathbb{E}_x[T]$. As a consequence, they showed that for all $\alpha \in (0, 1/2)$ (this was extended to $\alpha = 1/2$ in [24]), there exist constants $c_\alpha, c'_\alpha > 0$ such that for every finite irreducible reversible chain

$$c'_\alpha t_{\mathrm{H}}(\alpha) \leqslant t_{\mathrm{ave}} \leqslant c_\alpha t_{\mathrm{H}}(\alpha), \quad \text{where}$$

$$t_{\mathrm{H}}(\alpha) := \max_{x \in \Omega, A \subset \Omega : \pi(A) \geq \alpha} \mathbb{E}_x[T_A].$$

Using this, they showed that there exist some absolute constants $c_1, C_1, c_2, C_2 > 0$ such that

$$c_1 t_{\mathrm{L}} \leqslant t_{\mathrm{ave}} \leqslant C_1 t_{\mathrm{L}} \quad \text{and so} \quad c_2 t_{\mathrm{H}}(1/4) \leq t_{\mathrm{L}} \leq C_2 t_{\mathrm{H}}(1/4). \tag{2.0.1}$$

Implicitly, they showed that for every $0 < \varepsilon \leq 1/4$ and $0 < \alpha \leq 1/2$,

$$t_{\mathrm{ave}}(\varepsilon) \leqslant c_\alpha \varepsilon^{-4} t_{\mathrm{H}}(\alpha). \tag{2.0.2}$$

2.1 Results concerning hitting times and the cutoff phenomenon.

It is natural to ask whether (2.0.1) and (2.0.2) could be further refined so that the cutoff phenomenon could be characterized in terms of concentration of the hitting times of a sequence of sets $A_n \subset \Omega_n$ which attain the maximum in the definition of $t_{\mathrm{H}}^{(n)}(1/2)$ (starting from the worst initial states). Corollary 1.5 in [25] asserts that this is indeed the case in the transitive setup. More generally, Theorem 2 in [25] asserts that this is indeed the case for any fixed sequence of initial states $x_n \in \Omega_n$ if one replaces $t_{\mathrm{H}}^{(n)}(1/2)$ and $d^{(n)}(t)$ by $t_{\mathrm{H},x_n}^{(n)}(1/2)$ and $d_{x_n}^{(n)}(t)$ (i.e. when the hitting times and the mixing times are defined only w.r.t. this sequence of starting states). Alas, Proposition 1.6 in [25] asserts that in general cutoff could not be characterized in this manner. It turns out that the following hitting parameter is better suited (than $t_{\mathrm{H},x}(\cdot)$) for the purpose of studying the profile of convergence to stationarity (and cutoff).

Definition 2.1.1. *Let (Ω, P, π) be an irreducible chain. For any $x \in \Omega$, $\alpha, \varepsilon \in (0,1)$ and $t \geq 0$, define $p_x(\alpha,t) := \max_{A \subset \Omega:\, \pi(A) \geq \alpha} \mathrm{P}_x[T_A > t]$. Set $p(\alpha,t) := \max_x p_x(\alpha,t)$. We define*

$$\mathrm{hit}_{\alpha,x}(\varepsilon) := \min\{t : p_x(\alpha,t) \leq \varepsilon\} \text{ and } \mathrm{hit}_{\alpha}(\varepsilon) := \min\{t : p(\alpha,t) \leq \varepsilon\}.$$

Definition 2.1.2. *Let (Ω_n, P_n, π_n) be a sequence of irreducible chains and let $\alpha \in (0,1)$. We say that the sequence exhibits a hit_α-cutoff, if for all $\varepsilon \in (0,1/4)$*

$$\mathrm{hit}_{\alpha}^{(n)}(\varepsilon) - \mathrm{hit}_{\alpha}^{(n)}(1-\varepsilon) = o\left(\mathrm{hit}_{\alpha}^{(n)}(1/4)\right).$$

All of the results in the remainder of this subsection are taken from [6] (joint work with Riddhipratim Basu and Yuval Peres). The following is the main abstract theorem from [6].

Theorem 2.1.1 ([6] Theorem 3). *Let (Ω_n, P_n, π_n) be a sequence of lazy reversible irreducible finite chains. The following are equivalent:*

1) *The sequence exhibits a cutoff.*

2) *The sequence exhibits a hit_α-cutoff for some $\alpha \in (0,1/2]$.*

3) *The sequence exhibits a hit_α-cutoff for some $\alpha \in (1/2,1)$ and $t_{\mathrm{rel}}^{(n)} = o(t_{\mathrm{mix}}^{(n)})$.*

Remark 2.1.3. *In Example 3.1.8 we show that there exists a sequence of lazy reversible irreducible finite Markov chains, (Ω_n, P_n, π_n), such that the product condition fails, yet for all $1/2 < \alpha < 1$ there is hit_α-cutoff. Thus the assertion of Theorem 2.1.1 is sharp.*

Remark 2.1.4. *The proof of Theorem 2.1.1 can be extended to the continuous-time case. The necessary adaptations can be found in § 4 of [6].*

At first glance $\mathrm{hit}_\alpha(\varepsilon)$ may seem like a rather weak notion of mixing compared to $t_{\mathrm{mix}}(\varepsilon)$, especially when α is close to 1 (say, $\alpha = 1-\varepsilon$). The following proposition gives a quantitative version of Theorem 2.1.1 (for simplicity we fix $\alpha = 1/2$ in (2.1.1) and (2.1.2)).

Proposition 2.1.5. *For every reversible irreducible finite lazy chain and every $\varepsilon \in (0, \frac{1}{4}]$,*

$$\mathrm{hit}_{1/2}(3\varepsilon/2) - \lceil 2t_{\mathrm{rel}} |\log \varepsilon| \rceil \leq t_{\mathrm{mix}}(\varepsilon) \leq \mathrm{hit}_{1/2}(\varepsilon/2) + \lceil t_{\mathrm{rel}} |\log(\varepsilon/4)| \rceil \text{ and} \tag{2.1.1}$$

$$\mathrm{hit}_{1/2}(1-\varepsilon/2) - \lceil 2t_{\mathrm{rel}} |\log \varepsilon| \rceil \leq t_{\mathrm{mix}}(1-\varepsilon) \leq \mathrm{hit}_{1/2}(1-2\varepsilon) + 1_{\varepsilon > 1/18} \left\lceil \frac{1}{2} t_{\mathrm{rel}} \log 8 \right\rceil. \tag{2.1.2}$$

Moreover,

$$\max\{\mathrm{hit}_{1-\varepsilon/4}(5\varepsilon/4), (t_{\mathrm{rel}}-1)|\log(2\varepsilon)|\} \leq t_{\mathrm{mix}}(\varepsilon) \leq \mathrm{hit}_{1-\varepsilon/4}(3\varepsilon/4) + \left\lceil \frac{3t_{\mathrm{rel}}}{2} |\log(\varepsilon/4)| \right\rceil. \tag{2.1.3}$$

Remark 2.1.6. *Our only use of the laziness assumption is to argue that* $t_{\mathrm{rel}} = t_{\mathrm{rel}}^{\mathrm{absolute}}$*. In particular, Proposition 2.1.5 holds also without the laziness assumption if one replaces* t_{rel} *by* $t_{\mathrm{rel}}^{\mathrm{absolute}}$*. Similarly, without the laziness assumption the assertion of Theorem 2.1.1 should be transformed as follows. A sequence of finite irreducible aperiodic reversible Markov chains exhibits cutoff iff* $(t_{\mathrm{rel}}^{\mathrm{absolute}})^{(n)} = o(t_{\mathrm{mix}}^{(n)})$ *and there exists some* $0 < \alpha < 1$ *such that the sequence exhibits* hit_α*-cutoff.*

Note that for any finite irreducible reversible chain, (Ω, P, π)*, it suffices to consider a* δ*-lazy version of the chain,* $P_\delta := (1-\delta)P + \delta I$*, for some* $\delta \geq \frac{1-\max\{\lambda_2, 0\}}{2}$*, to ensure that* $t_{\mathrm{rel}} = t_{\mathrm{rel}}^{\mathrm{absolute}}$ *(which by the previous paragraph, guarantees that all near-periodicity issues are completely avoided).*

Corollary 2.1.7. *Let* (Ω_n, P_n, π_n) *be a sequence of irreducible reversible discrete-time chains. Let* $\lambda_{\min}^{(n)}$ *be the smallest eigenvalue of* P_n*. Let* $t_{\mathrm{L}}^{(n)}$ *be the mixing time of the lazy version of the* n*th chain in the sequence. Assume that* $t_{\mathrm{L}}^{(n)} = o(t_{\mathrm{mix}}^{(n)})$*. Then it must be the case that*

$$\lambda_{\min}^{(n)} t_{\mathrm{mix}}^{(n)} = \Theta(1).$$

Loosely speaking, we show that the mixing of a lazy reversible Markov chain can be partitioned into two stages as follows. The first is the time it takes the chain to escape from some small set with sufficiently large probability. In the second stage, the chain mixes at a rate which is governed by its relaxation-time. This estimate is sharp is some cases (i.e. there are examples in which the above description is accurate and the rate of convergence in the "second stage" is also lower bounded by the relaxation time).

It follows from Proposition 3.1.2 that the ratio of the LHS and the RHS of (2.1.3) is bounded by an absolute constant independent of ε. Moreover, (2.1.3) bounds $t_{\mathrm{mix}}(\varepsilon)$ in terms of hitting distribution of sets of π measure tending to 1 as ε tends to 0. In (3.1.2) we give a version of (2.1.3) for sets of arbitrary π measure.

Either of the two terms appearing in the sum in RHS of (2.1.3) may dominate the other. For lazy simple random walk on two n-cliques connected by a single edge, the terms in (2.1.3) involving $\mathrm{hit}_{1-\varepsilon/4}$ are negligible. For a sequence of chains satisfying the product condition, all terms in Proposition 2.1.5 involving t_{rel} are negligible. Hence the assertion of Theorem 2.1.1, for $\alpha = 1/2$, follows easily from (2.1.1) and (2.1.2), together with the fact that $\mathrm{hit}_{1/2}^{(n)}(1/4) = \Theta(t_{\mathrm{mix}}^{(n)})$. In Proposition 3.1.5, under the assumption that the product condition holds, we prove this fact and show that in fact, if the sequence exhibits hit_α-cutoff for some $\alpha \in (0,1)$, then it exhibits hit_β-cutoff for all $\beta \in (0,1)$.

2.2 Results concerning characterization of τ_2 and hypercontractivity in terms of hitting times distributions

All of the results from this section are taken from [29] (joint work with Yuval Peres) unless otherwise specified.

There are numerous essentially equivalent characterizations of mixing in L_1 (e.g. [3, Theorem 4.6] and [48]) of a finite reversible Markov chain. Some involve natural probabilistic concepts such as couplings, stopping times and hitting times. In contrast, (paraphrasing Aldous and Fill [3] last sentence of page 155, which mentions that there is no L_2 counterpart to [3, Theorem 4.6]) while there are several sophisticated analytic and geometric tools for bounding the L_2 mixing time, τ_2, none of them has a probabilistic interpretation, and none of them determines τ_2 up to a constant factor.

In a joint work with Peres [29] 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, τ_{Ent} of a reversible Markov chain (Theorem 2.2.1).

While the spectral gap is a natural and simple parameter, the *Log-Sobolev constant*, c_{LS}, is a more involved quantity. When one first encounters c_{LS}, it may seem like an artificial parameter that "magically" gives good bounds on τ_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_{LS} and τ_2 (answering a question asked by James Lee, see Remark 2.2.2) and can be interpreted probabilistically as a hitting-time version of hypercontractivity (see the discussion in § 4.3).

2.2.1 A characterization of τ_2 and τ_{Ent}

All of the results in this subsection are in the continuous-time setup. The discrete-setup is described in the following subsection. 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 Con_δ the collection of all connected sets A satisfying $\pi(A) \le \delta$, where throughout, π shall denote the stationary distribution of the chain. Denote

$$\rho := \max_{x\in\Omega} \rho_x \quad \text{and} \quad \rho_{\mathrm{Ent}} := \max_{x\in\Omega} \rho_{\mathrm{Ent},x}, \quad \text{where} \tag{2.2.1}$$

$$\begin{aligned} \rho_x &:= \min\{t : \mathrm{P}_x[T_{A^c} > t] \le \pi(A) + \frac{1}{2}\sqrt{\pi(A)\pi(A^c)} \text{ for all } A \in \mathrm{Con}_{1/2}\}, \\ \rho_{\mathrm{Ent},x} &:= \min\{t : \mathrm{P}_x[T_{A^c} > t] \le \frac{C}{|\log \pi(A)|}, \text{ for all } A \in \mathrm{Con}_{1/2}\}, \end{aligned} \tag{2.2.2}$$

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

Theorem 2.2.1 ([29] Theorem 1)**.** *There exist absolute constants C_1, C_2, C_3 such that for every irreducible reversible Markov chain on a finite state space*

$$\rho \le \tau_2 \le \rho + C_1/c_{\mathrm{LS}} \le C_2\rho. \tag{2.2.3}$$

$$\rho_{\mathrm{Ent}} \le \tau_{\mathrm{Ent}} \le C_3\rho_{\mathrm{Ent}}. \tag{2.2.4}$$

Note that in the definitions of ρ and ρ_{Ent}, the smaller A is, the smaller we require the chance of not escaping it by time ρ or ρ_{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 § 4.0.4, the first inequalities in (2.2.3)-(2.2.4) are easy and even somewhat "naive".

A lot of attention has been focused on inequalities that interpolate between the Log-Sobolev inequality and the Poincaré (spectral gap) inequality (e.g. [8, 44]). We prove a new extremal characterization (up to a constant factor) of the *Log-Sobolev constant* (Theorems 2.2.2). 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.

Recall that $Q := (P + P^*)/2$. 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, $\lambda(A)$, is defined as the minimal eigenvalue of $I - Q_A$. Denote $t_{\text{rel}}(A) := 1/\lambda(A)$. Denote

$$\kappa := 1/\alpha, \quad \alpha := \min_{A \in \text{con}_{1/2}} \alpha(A), \quad \text{where} \quad \alpha(A) := \lambda(A)/|\log \pi(A)|. \tag{2.2.5}$$

As mentioned earlier, α is a weighted version of λ since ([3, Lemma 4.39] and [23, (1.4)])

$$\lambda/2 \le \min_{A \in \text{con}_{1/2}} \lambda(A) \le \lambda, \quad \text{and so} \quad t_{\text{rel}} \log 2 \le \kappa. \tag{2.2.6}$$

Theorem 2.2.2 ([29] Theorem 2)**.** *For every irreducible Markov chain on a finite state space*

$$\kappa \le t_{\text{LS}} \le 2(\kappa + t_{\text{rel}}(1 + \log 49)) \le 2(1 + (1 + \log 49)/\log 2)\kappa < 17\kappa. \tag{2.2.7}$$

Remark 2.2.1. *The inequality $\kappa \le t_{\text{LS}}$ is easy. See Lemma 4.2 in [23] for a stronger inequality. The harder and more interesting direction is $t_{\text{LS}} \le C\kappa$, which is an improvement over the well-known inequality $t_{\text{LS}} \le t_{\text{rel}} \frac{\log[1/\pi_* - 1]}{1 - 2\pi_*}$, where $\pi_* := \min_{x \in \Omega} \pi(x)$ [16, Corollary A.4].*

Remark 2.2.2. *Despite the fact that t_{LS} is a geometric quantity, Logarithmic Sobolev inequalities have a strong analytic flavor and little probabilistic interpretation. For instance, the proof of the inequality $t_{\text{LS}} \le 2\tau_2(1/e)$ [16, Corollary 3.11] (where $\tau_2(\varepsilon)$ is the L_2 mixing time defined in* (1.1.3)*) relies on Stein's interpolation Theorem for a family of analytic operators. Our analysis yields a probabilistic proof of the fact that $t_{\text{LS}} \le C\tau_2$ for reversible chains (the problem of finding such a proof was posed by James Lee at the Simons institute in 2015). Indeed by Theorem 2.2.2 and* (4.1.2)*, $t_{\text{LS}}/17 \le \kappa \le 3\rho \le 3\tau_2$ (this second inequality is relatively easy, and is obtained by analyzing hitting times, rather than by analytic tools). As we show in § 4.0.4, the inequality $\rho \le \tau_2$ also has a probabilistic interpretation.*

2.2.2 Discrete-time and averaged chain analogs

Let (Ω, P, π) be a finite irreducible chain. We may consider the L_p-mixing times of the discrete-time and averaged chains $\tau_p^{\text{discete}}(\cdot)$ and $\tau_p^{\text{ave}}(\cdot)$, 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.. Similarly 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.. Define ρ_{discete} and $\rho_{\mathrm{Ent}}^{\mathrm{discete}}$ in an analogous manner to ρ, where now the hitting times are defined w.r.t. the discrete-time chain. Define $t_{\mathrm{rel}}^{\mathrm{absolute}} := \max\{(1-\lambda_2)^{-1}, (1-|\lambda_{|\Omega|}|)^{-1}\}$. The following observation is new to this thesis.

Theorem 2.2.3. *There exist absolute constants C_1, C_2, C_3 such that for every irreducible reversible Markov chain on a finite state space*

$$\rho_{\mathrm{discete}} - 1 \le \tau_2^{\mathrm{ave}} \le \rho_{\mathrm{discete}} + C_1 t_{\mathrm{LS}} \le C_2 \rho_{\mathrm{discete}}. \tag{2.2.8}$$

$$\rho_{\mathrm{Ent}}^{\mathrm{discete}} - 1 \le \tau_{\mathrm{Ent}}^{\mathrm{ave}} \le C_3 \rho_{\mathrm{Ent}}^{\mathrm{discete}}. \tag{2.2.9}$$

$$\rho_{\mathrm{discete}} \le \tau_2^{\mathrm{discete}} \le \rho_{\mathrm{discete}} + C_1(t_{\mathrm{LS}} + t_{\mathrm{rel}}^{\mathrm{absolute}}) \le C_2(\rho_{\mathrm{discete}} + t_{\mathrm{rel}}^{\mathrm{absolute}}). \tag{2.2.10}$$

$$\rho_{\mathrm{Ent}}^{\mathrm{discete}} \le \tau_{\mathrm{Ent}}^{\mathrm{discrete}} \le C_3(\rho_{\mathrm{Ent}}^{\mathrm{discete}} + t_{\mathrm{rel}}^{\mathrm{absolute}}). \tag{2.2.11}$$

In conjunction with Theorem 2.2.1, and Lemma 4.1.4, 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 2.2.4. *There exists an absolute constant $C > 0$ such that for every irreducible reversible Markov chain on a finite state space*

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

To see that the reverse inequalities are false consider simple random walk on the n-clique, for which $\tau_2^{\mathrm{ave}} \le 2$ while $\tau_2 = \Theta(\log n)$ and $\tau_{\mathrm{Ent}} = \Theta(\log\log n)$.

2.3 Applications

2.3.1 Robustness of τ_2 under addition of self-loops of bounded weights.

The following corollary (taken from [29], joint work with Peres), proved in § 4.4, is an analog of [48, Corollary 9.5], which gives the corresponding statement for τ_1. While the statement is extremely intuitive, surprisingly, it was recently shown that it may fail for simple random walk on an Eulerian digraph [9, Theorem 1.5].

Corollary 2.3.1. *[[29] Corollary 1.3] Let (X_t) be a reversible irreducible continuous-time Markov chain on a finite state space Ω with generator G. Let $(\tilde{X}_t)$ be a chain with generator $\tilde{G}$ obtained by multiplying for all $x \in \Omega$ the xth row of G by some $r_x \in (1/M, M)$ (for some $M \geq 1$). Then for some absolute constant C the corresponding L_2 mixing times satisfy*

$$\tilde{\tau}_2/(CM \log M) \leq \tau_2 \leq (CM \log M)\tilde{\tau}_2. \tag{2.3.1}$$

Observe that the generator G of a reversible chain on a finite state space Ω, 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 xth 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 2.3.2. *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 only differ from the L_2 mixing time of the chain with $a(x) = 1/2$ for all x, by a factor of $C\delta^{-1}|\log \delta|$.*

2.3.2 A characterization of cutoff for trees.

We start with a few definitions. Let $T := (V, E)$ be a finite tree. Fix some Markov chain, (V, P, π), on a finite tree $T := (V, E)$. That is, a chain with stationary distribution π and state space V such that $P(x, y) > 0$ iff $\{x, y\} \in E$ or $y = x$. Then P is reversible by Kolmogorov's cycle condition.

The following theorem generalizes previous results concerning birth and death chains [19]. The relevant setup is weighted nearest neighbor random walks on finite trees.

Theorem 2.3.1 ([6] (5.15))**.** *There exists an absolute constant C such that for every lazy reversible Markov chain on a tree $T = (V, E)$ with $|V| \geq 3$,*

$$t_{\mathrm{mix}}(\varepsilon) - t_{\mathrm{mix}}(1 - \varepsilon) \leq C\sqrt{t_{\mathrm{rel}} t_{\mathrm{mix}} |\log \varepsilon|} + C t_{\mathrm{rel}} |\log \varepsilon|, \quad \text{for all } 0 \leq \varepsilon \leq 1/4. \tag{2.3.2}$$

In particular, if the product condition holds for a sequence of lazy reversible Markov chains (V_n, P_n, π_n) on finite trees $T_n = (V_n, E_n)$, then the sequence exhibits a cutoff with a cutoff window $w_n = \sqrt{t_{\mathrm{rel}}^{(n)} t_{\mathrm{mix}}^{(n)}}$.

In [17], Diaconis and Saloff-Coste showed that a sequence of birth and death (BD) chains exhibits separation cutoff if and only if $t_{\mathrm{rel}}^{(n)} = o(t_{\mathrm{mix}}^{(n)})$. In [19], Ding et al. extended this also to the notion of total-variation cutoff and showed that the cutoff window is always at most $\sqrt{t_{\mathrm{rel}}^{(n)} t_{\mathrm{mix}}^{(n)}}$ and that in some cases this is tight (see Theorem 1 and Section 2.3 ibid). As BD chains are a particular case of chains on trees, the bound on w_n in Theorem 2.3.1 is tight.

We note that the bound we get on the rate of convergence ((2.3.2)) is better than the estimate in [19, Theorem 2.2] (even for BD chains), which is $t_{\mathrm{mix}}(\varepsilon) - t_{\mathrm{mix}}(1-\varepsilon) \le C\varepsilon^{-1}\sqrt{t_{\mathrm{rel}} t_{\mathrm{mix}}}$.

Concentration of hitting times was a key ingredient both in [17] and [19] (as it shall be here). Their proofs relied on several properties which are specific to BD chains. Our proof of Theorem 2.3.1 can be adapted to the following setup. Denote $[n] := \{1, 2, \dots, n\}$.

Definition 2.3.3. *For $n \in \mathbb{N}$ and $\delta, r > 0$, we call a finite lazy reversible Markov chain, $([n], P, \pi)$, a (δ, r)-**semi birth and death (SBD) chain** if*

(i) For all $i, j \in [n]$ such that $|i - j| > r$, we have $P(i, j) = 0$.

(ii) For all $i, j \in [n]$ such that $|i - j| = 1$, we have that $P(i, j) \ge \delta$.

This is a natural generalization of the class of birth and death chains. Conditions (i)-(ii) tie the geometry of the chain to that of the path $[n]$. We have the following theorem.

Theorem 2.3.2 ([6] Theorem 2)**.** *Let $([n_k], P_k, \pi_k)$ be a sequence of (δ, r)-semi birth and death chains, for some $\delta, r > 0$, satisfying the product condition. Then it exhibits a cutoff with a cutoff window $w_k := \sqrt{t_{\mathrm{mix}}^{(k)} t_{\mathrm{rel}}^{(k)}}$.*

2.3.3 Robustness of τ_∞ for trees.

Recall that for reversible chains the L_2 mixing time, τ_2, determines the L_p-mixing time up to a factor c_p for all $1 < p \le \infty$ (see (1.1.4)). Denote the L_p mixing time of simple random walk on a finite connected simple graph G by $\tau_p(G)$. Kozma [31] made the following conjecture:

Conjecture 2.3.4 ([31])**.** *Let G and H be two finite K-roughly isometric graphs of maximal degree $\le d$. Then*

$$\tau_\infty(G) \le C(K, d)\tau_\infty(H). \tag{2.3.3}$$

It is well-known that (2.3.3) is true if one replaces τ_∞ with t_{LS} (e.g. [16, Lemma 3.4]). Ding and Peres [20] showed that (2.3.3) is false if one replaces τ_∞ with τ_1 (various improvements and extensions of their result can be found in [30], joint work with Yuval Peres). 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 τ_∞ in terms of hitting times might shed some light on Conjecture 2.3.4. Indeed this was one of the main motivations behind [29]. In [26] the author of this thesis constructed a counterexample to Conjecture 2.3.4, where also there the key was sensitivity of hitting times.

Peres and Sousi [48, Theorem 9.1] showed that for weighted nearest neighbor random walks on trees, τ_1 can change only by a constant factor, as a result of a bounded perturbation of the edge weights. As an application of Theorem 2.2.1 we prove the L_2 analog.

Theorem 2.3.3 ([29] Theorem 3)**.** *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*

$$\max(\tau_1, t_{\mathrm{LS}}/4) \le \tau_2 \le \tau_1 + C\max(t_{\mathrm{LS}}, \sqrt{t_{\mathrm{LS}}\tau_1}), \tag{2.3.4}$$

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

$$\tau'_\infty/C_M \le \tau_\infty \le C_M\tau'_\infty. \tag{2.3.5}$$

Remark 2.3.5. *Since t_{LS} is robust under a bounded perturbation of the edge weights (e.g. [16, Lemma 3.3]), indeed* (2.3.5) *follows from* (2.3.4) *in conjunction with the aforementioned L_1 robustness of trees (and the fact that $\tau_2 \le \tau_\infty \le 2\tau_2$, see* (1.1.4)*).*

2.4 Results concerning the averaged chain

All of the results presented in this section are taken from [28] (joint work with Yuval Peres).

Recall that in order to avoid near-periodicity issues one can consider the continuous-time, the lazy and the averaged versions of the chain. In order to compare the (total variation) mixing times of these chains with no ambiguity, we introduce some additional notation.

$$d_{\mathrm{c}}(t) := \max_{x \in \Omega} d_{\mathrm{c}}(t,x), \quad d_{\mathrm{L}}(t) := \max_{x \in \Omega} d_{\mathrm{L}}(t,x), \quad d_{\mathrm{ave}}(t) := \max_{x \in \Omega} d_{\mathrm{ave}}(t,x),$$

where for every $\mu \in \mathscr{P}(\Omega)$,

$$\begin{aligned} d_{\mathrm{c}}(t,\mu) &:= \|\mathrm{P}_{\mu}(X_t^{\mathrm{c}} \in \cdot) - \pi\|_{\mathrm{TV}} = \|\mathrm{H}_{\mu}^t - \pi\|_{\mathrm{TV}}, \\ d_{\mathrm{L}}(t,\mu) &:= \|\mathrm{P}_{\mu}(X_t^{\mathrm{L}} \in \cdot) - \pi\|_{\mathrm{TV}} = \|\mathrm{P}_{\mathrm{L},\mu}^t - \pi\|_{\mathrm{TV}} \text{ and} \\ d_{\mathrm{ave}}(t,\mu) &:= \left\|(\mathrm{P}_{\mu}^t + \mathrm{P}_{\mu}^{t+1})/2 - \pi\right\|_{\mathrm{TV}} = \|\mu(P^{t+1} + P^t)/2 - \pi\|_{\mathrm{TV}}. \end{aligned}$$

The corresponding ε-**total variation mixing times** are, resp.,

$$t_{\mathrm{c}}(\varepsilon) := \inf\{t : d_{\mathrm{c}}(t) \leqslant \varepsilon\},$$

$$t_{\mathrm{L}}(\varepsilon) := \inf\{t : d_{\mathrm{L}}(t) \leqslant \varepsilon\} \quad t_{\mathrm{ave}}(\varepsilon) := \inf\{t : d_{\mathrm{ave}}(t) \leqslant \varepsilon\}.$$

We also define the corresponding mixing-times w.r.t. initial distribution μ to be

$$t_{\mathrm{c}}(\varepsilon,\mu) := \inf\{t : d_{\mathrm{ct}}(t,\mu) \leqslant \varepsilon\}, \quad t_{\mathrm{L}}(\varepsilon,\mu) := \inf\{t : d_{\mathrm{L}}(t,\mu) \leqslant \varepsilon\},$$

$$t_{\mathrm{ave}}(\varepsilon,\mu) := \inf\{t : d_{\mathrm{ave}}(t,\mu) \leqslant \varepsilon\}.$$

When $\varepsilon = 1/4$, it is we omitted. We denote $\mathbb{Z}_+ := \{n \in \mathbb{Z} : n \geq 0\}$ and $\mathbb{R}_+ := \{t \in \mathbb{R} : t \geq 0\}$. Let $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ and $\psi : (0,1] \to (0,1]$. We write $\phi(t) \sim t$ if $\lim_{t\to\infty} \phi(t)/t = 1$. We write $\psi = o(1)$ if $\lim_{\varepsilon\to 0} \psi(\varepsilon) = 0$. In [3] Aldous and Fill raised the following question:

Question (Open Problem 4.17 [3]). *Show that there exist $\psi : (0,1] \to (0,1]$ and $\phi : \mathbb{R}_+ \to \mathbb{Z}_+$ satisfying $\psi = o(1)$ and $\phi(t) \sim t$ such that for every finite irreducible reversible Markov chain,*

$$\forall t \geq 0, \quad d_{\mathrm{ave}}(\phi(t)) \leq \psi(d_{\mathrm{c}}(t)).$$

The first progress towards resolving Aldous and Fill's Open Problem is due to Peres and Sousi (2.0.1)-(2.0.2). Alas, their result is too coarse for the purpose of resolving it. Theorem 2.4.2, which is in fact a weaker version of Theorem 2.4.1, solves Aldous and Fill's Problem. Denote $a \vee b := \max\{a,b\}$, $a \wedge b := \min\{a,b\}$. For every $t \in \mathbb{R}$ we denote the ceiling of t by $\lceil t \rceil := \min\{z \in \mathbb{Z} : z \geq t\}$.

Definition 2.4.1. *Let $0 < \alpha < 1/2$, $C > 0$, $t \geq 1$ and $x \in (0,1)$. We define*

$$\psi_{\alpha,C}(x) := 1 \wedge (x + C|\log(2x)|^{-\alpha}) \quad \text{and} \quad \phi_{\alpha,C}(t) := t + \lceil Ct^{\frac{1+2\alpha}{2}}\sqrt{\alpha \log t}\rceil.$$

Remark 2.4.2. *Note that $\phi_{\alpha,C}(t) \sim t$ and $\psi_{\alpha,C} = o(1)$, for all $C > 0$ and $0 < \alpha < 1/2$.*

Theorem 2.4.1 ([28] Theorem 1.1)**.** *There exist absolute constants $C_1, C_2, C_3 > 0$ such that for every finite irreducible reversible Markov chain, (Ω, P, π), $\mu \in \mathscr{P}(\Omega)$, $0 < \alpha \le 1/2$ and $t \ge 1$,*

$$d_{\mathrm{L}}(\phi_{\alpha,C_1}(t), \mu) \le d_{\mathrm{c}}(t/2, \mu) + C_2 t^{-\alpha}. \tag{2.4.1}$$

$$d_{\mathrm{ave}}(\phi_{\alpha,C_1}(t), \mu) \le d_{\mathrm{L}}(2t, \mu) + C_2 t^{-\alpha}. \tag{2.4.2}$$

$$d_{\mathrm{ave}}(\phi_{\alpha,C_3}(t), \mu) \le d_{\mathrm{c}}(t, \mu) + 2C_2 t^{-\alpha}. \tag{2.4.3}$$

Moreover, (4.2.1)-(4.2.2) remain valid when μ is omitted from both sides.

Note that (4.2.2) follows from (4.2.1)-(2.4.2) by picking some C_3 so that $\phi_{\alpha,C_3}(t) \ge \phi_{\alpha,C_1}(\lceil \frac{1}{2}\phi_{\alpha,C_1}(2t)\rceil)$.

Remark 2.4.3. *The converse inequality $d_{\mathrm{c}}(t + 2t^{3/4}) \le d_{\mathrm{ave}}(t) + e^{-\sqrt{t}}$ is easy ((2.4.9)). Combined with* (4.2.2) *one can readily see that $d_{\mathrm{c}}(\cdot)$ exhibits an abrupt transition iff $d_{\mathrm{ave}}(\cdot)$ exhibits an abrupt transition (in which case, both occur around the same time).*

Theorem 2.4.2. *[[28] Theorem 1.2] There exist absolute constants $C_1, C_2 > 0$ such that for every finite irreducible reversible Markov chain*

$$d_{\mathrm{ave}}(\phi_{\alpha,C_1}(t)) \le \psi_{\alpha,C_2}(d_{\mathrm{c}}(t)), \text{ for every } 0 < \alpha < 1/2 \text{ and } t \ge 2. \tag{2.4.4}$$

Remark 2.4.4. *Theorem 2.4.2 can be rephrased as follows. There exist absolute constants $C_1, C_2 > 0$ such that for every finite irreducible reversible Markov chain,*

$$t_{\mathrm{ave}}(\psi_{\alpha,C_2}(\varepsilon)) \le \phi_{\alpha,C_1}(t_{\mathrm{c}}(\varepsilon)), \text{ for all } 0 < \alpha < 1/2 \text{ and } 0 < \varepsilon < 1. \tag{2.4.5}$$

Theorem 2.4.2 is an immediate consequence of (4.2.2) together with the "worst-case" estimate $d_{\mathrm{c}}(t) \ge (e^{-2t}/2)1_{|\Omega|>1}$ (e.g. [36, Lemma 20.11]). We omit the details. Theorem 2.4.1 follows in turn as the particular case $s := 2 \vee t^{\alpha}\sqrt{\alpha \log t}$ of the following proposition.

Proposition 2.4.5 ([28] Proposition 1.5)**.** *There exists an absolute constant C such that for every finite irreducible reversible chain, (Ω, P, π), every $\mu \in \mathscr{P}(\Omega)$, $t \ge 2$ and $s \in [2, e^t]$ we have that*

$$d_{\mathrm{L}}(t + \lceil s\sqrt{t}\rceil, \mu) \le d_{\mathrm{c}}(t/2, \mu) + Cs^{-1}\sqrt{\log s}. \tag{2.4.6}$$

$$d_{\mathrm{ave}}(t + \lceil s\sqrt{t}\rceil, \mu) \le d_{\mathrm{L}}(2t, \mu) + Cs^{-1}\sqrt{\log s}. \tag{2.4.7}$$

We now make two remarks regarding the sharpness of (2.4.7). The first concerns the error term $Cs^{-1}\sqrt{\log s}$ (and also the "error term", $\psi_{\alpha,C_2}(d_{\mathrm{c}}(t)) - d_{\mathrm{c}}(t)$, in (2.4.4)). The second concerns the "time-shift" term $\lceil s\sqrt{t}\rceil$.

Remark 2.4.6. *Denote $s = s_{n,\alpha} := \lceil n^{0.5+\alpha}\rceil$ and $t = t_{n,\alpha} := 4n + s$. In § 6.4 we construct for every $0 < \alpha \le 1/2$ a sequence of chains with $t_{\mathrm{c}}^{(n)} = (4 \pm o(1))n$ such that for some absolute constants $c_1, c_2 > 0$ the n-th chain in the sequence satisfies that*

$$d_{\mathrm{ave}}(t + s) - d_{\mathrm{c}}(t) \ge \frac{c_1}{s} \ge \frac{c_2}{[\log(1/d_{\mathrm{c}}(t))]^{\frac{1+2\alpha}{4\alpha}}}. \tag{2.4.8}$$

Thus the inverse polynomial decay (w.r.t. s) in (2.4.7) is the correct order of decay, up to the value of the exponent.

Remark 2.4.7. *When s is fixed, the "time-shift" term $s\sqrt{t}$ in* (2.4.7) *is of order $\sqrt{t}$. This cannot be improved. To see this, consider a birth and death chain on $[n] := \{1, 2, \dots, n\}$ with $P(i+1, i) = e^{-n} = 1 - P(i+1, i+2)$ for $i \in [n-2]$ and $P(1, 2) = 1 = P(n, n-1)$. Then if $r_n = o(\sqrt{n})$ we have that $d_{\mathrm{L}}(2n - r_n) = 1/2 \pm o(1)$, while $d_{\mathrm{ave}}(n-3) = 1 - o(1)$.*

The following proposition offers a converse to Theorem 2.4.1. The argument in the proof of (2.4.9) is due to Peres and Sousi ([48, Lemma 2.3]).

Proposition 2.4.8 ([28] Proposition 1.8)**.** *Let (Ω, P, π) be a finite irreducible Markov chain. Then for every $t \in \mathbb{N}$, $0 < s \leq \sqrt{t}$ and $\mu \in \mathscr{P}(\Omega)$,*

$$\begin{aligned} d_{\mathrm{c}}(t + s\sqrt{t}, \mu) &\leq d_{\mathrm{ave}}(t, \mu) + e^{-s^2/4}. \\ d_{\mathrm{L}}(2t + \lceil 2s\sqrt{t} \rceil, \mu) &\leq d_{\mathrm{ave}}(t, \mu) + e^{-s^2/4}. \end{aligned} \tag{2.4.9}$$

$$d_{\mathrm{c}}(t + s\sqrt{t}, \mu) \leq d_{\mathrm{L}}(2t, \mu) + e^{-s^2/2}. \tag{2.4.10}$$

Remark 2.4.9. *In [53] p. 195, it is written: "a theorem is Abelian if it says something about an average of a sequence from a hypothesis about its ordinary limit; it is Tauberian if conversely the implication goes from average to limit".*

Proposition 2.4.8 is easier and more general than Theorem 2.4.1 (as it does not assume reversibility) because it is an Abelian theorem, while Theorem 2.4.1 is Tauberian, hence requires the reversibility assumption, as we now demonstrate. One (non-reversible) instance in which (2.4.7) fails is a biased random walk on the n-cycle with $P(i, i-1) = n^{-\ell} = 1 - P(i, i+1)$, where $i - 1$ and $i + 1$ are defined modulo n and $\ell > 0$ is arbitrary. In this example $t_{\mathrm{L}}(\varepsilon)/(n^2 |\log \varepsilon|) = \Theta(1)$, however $t_{\mathrm{ave}}(\varepsilon)/(n^{\ell+2} |\log \varepsilon|) = \Theta(1)$ (uniformly in $\varepsilon \in (0, 1/2]$).

Next, consider a sequence of such chains, $((\Omega_n, P_n, \pi_n) : n \in \mathbb{N})$, each with its corresponding worst-distance from stationarity $d_n(t)$, its mixing-time $t_{\mathrm{c}}^{(n)}$, etc.. We say that a sequence of chains exhibits a ***continuous-time cutoff*** if the following sharp transition in its convergence to stationarity occurs:

$$\lim_{n \to \infty} t_{\mathrm{c}}^{(n)}(\varepsilon)/t_{\mathrm{c}}^{(n)}(1 - \varepsilon) = 1, \quad \text{for every } 0 < \varepsilon < 1.$$

We say that a sequence of chains exhibits an ***averaged cutoff*** (resp. ***lazy cutoff***) if

$$\lim_{n \to \infty} t_{\mathrm{ave}}^{(n)}(\varepsilon)/t_{\mathrm{ave}}^{(n)}(1 - \varepsilon) = 1 \ \ \text{(resp.,} \ \lim_{n \to \infty} t_{\mathrm{L}}^{(n)}(\varepsilon)/t_{\mathrm{L}}^{(n)}(1 - \varepsilon) = 1), \ \text{for every } 0 < \varepsilon < 1.$$

The following corollary follows at once from Theorem 2.4.1 together with Proposition 2.4.8.

Corollary 2.4.10 ([28] Corollary 1.10)**.** *Let (Ω_n, P_n, π_n) be a sequence of finite irreducible reversible Markov chains. Then the following are equivalent*

(i) The sequence exhibits a continuous-time cutoff.

(ii) The sequence exhibits a lazy cutoff.

(iii) The sequence exhibits an averaged cutoff.

Moreover, if (i) holds, then $\lim_{n\to\infty} t_{\mathrm{ave}}^{(n)}/t_{\mathrm{c}}^{(n)} = \lim_{n\to\infty} t_{\mathrm{L}}^{(n)}/(2t_{\mathrm{c}}^{(n)}) = 1$.

Remark 2.4.11. *The equivalence between (i) and (iii) was previously unknown. In [12] it was shown that (i) and (ii) are equivalent even without the assumption of reversibility.*

Our last point of comparison is related to the width of the ***cutoff window***. We say that a sequence of chains exhibits a continuous-time (resp. averaged) cutoff with a cutoff window w_n if $w_n = o(t_{\mathrm{c}}^{(n)})$ (resp. $w_n = o(t_{\mathrm{ave}}^{(n)})$) and for every $0 < \varepsilon \le 1/4$ there exists some constant $C_\varepsilon > 0$ (depending only on ε) such that

$$\forall n, \quad t_{\mathrm{c}}^{(n)}(\varepsilon) - t_{\mathrm{c}}^{(n)}(1-\varepsilon) \le C_\varepsilon w_n \quad (\text{resp. } t_{\mathrm{ave}}^{(n)}(\varepsilon) - t_{\mathrm{ave}}^{(n)}(1-\varepsilon) \le C_\varepsilon w_n).$$

One can define the notion of a cutoff window for a sequence of associated lazy chains in an analogous manner. Note that the window defined in this manner is not unique.

Theorem 2.4.3 ([28] Theorem 1.3)**.** *Let* (Ω_n, P_n, π_n) *be a sequence of finite irreducible reversible Markov chains.*

(i) Assume that the sequence exhibits a continuous-time cutoff with a window w_n*. Then it exhibits also an averaged cutoff with a window* w_n*.*

(ii) Assume that the sequence exhibits an averaged cutoff with a window w_n*. Then it exhibits also a continuous-time cutoff with a window* $w_n' := w_n \vee \sqrt{t_{\mathrm{c}}^{(n)}}$*.*

Theorem 2.4.3 follows easily from Propositions 2.4.5 and 2.4.8 in conjunction with the following result. We prove Theorem 2.4.3 in § 6.3 for the sake of completeness.

Proposition 2.4.12 ([12] Chen and Saloff-Coste)**.** *Let* (Ω_n, P_n, π_n) *be a sequence of finite irreducible reversible Markov chains. The sequence exhibits a continuous-time cutoff with a window* w_n *iff it exhibits a lazy cutoff with a window* w_n*, in which case* $w_n = \Omega\left(\sqrt{t_{\mathrm{c}}^{(n)}}\right)$*.*

Remark 2.4.13. *There are cases in which the cutoff window for the sequence of the associated averaged chains can be much smaller than that of the associated continuous-time chains. For instance, let* G_n *be a sequence of random* n*-vertex* d_n*-regular graphs, for some* d_n *such that* $\log n \ll d_n \log d_n = n^{o(1)}$*. Let* $(X_t^{(n)})_{t\in\mathbb{Z}_+}$ *be the sequence of discrete-time simple random walks on* G_n*. Then [41] w.h.p. (i.e. with probability* $1-o(1)$*, over the choice of the graphs)*

$$|t_{\mathrm{ave}}^{(n)}(\varepsilon) - \lceil \log_{d_n-1}(d_n n) \rceil| \le 1, \quad \textit{for every } \varepsilon \in (0,1),$$

while the cutoff window for the sequence of associated continuous-time chains is $\sqrt{\log_{d_n-1} n}$*.*

Chapter 3

Characterization of cutoff for reversible Markov chains

In this Chapter we present the proofs of the results from Section 2.1.

3.0.1 An overview

Definition 3.0.14. *Let (Ω, P, π) be a finite reversible irreducible lazy chain. Let $A \subset \Omega$, $s \geq 0$ and $m > 0$. Denote $\rho(A) := \sqrt{\mathrm{Var}_\pi 1_A} = \sqrt{\pi(A)(1-\pi(A))}$. Set $\sigma_s := e^{-s/t_{\mathrm{rel}}}\rho(A)$. We define*

$$G_s(A,m) := \left\{ y : |\mathrm{P}_y^k(A) - \pi(A)| < m\sigma_s \text{ for all } k \geq s \right\}. \tag{3.0.1}$$

We call the set $G_s(A,m)$ the good set for A from time s within m standard-deviations.

As a simple corollary of Starr's L^2 maximal inequality and the L^2-contraction lemma we show in Corollary 3.0.15 that for any non-empty $A \subset \Omega$ and any $m, s \geq 0$ that $\pi(G_s(A,m)) \geq 1 - 8/m^2$. To demonstrate the main idea of our approach we now prove the following inequalities.

$$t_{\mathrm{mix}}(2\varepsilon) \leq \mathrm{hit}_{1-\varepsilon}(\varepsilon) + \left\lceil \frac{t_{\mathrm{rel}}}{2} \log\left(\frac{2}{\varepsilon^3}\right) \right\rceil. \tag{3.0.2}$$

$$\mathrm{hit}_{1-\varepsilon}(1-2\varepsilon) \geq t_{\mathrm{mix}}(1-\varepsilon) - \left\lceil \frac{t_{\mathrm{rel}}}{2} \log\left(\frac{8}{\varepsilon^2}\right) \right\rceil. \tag{3.0.3}$$

We first prove (3.0.2). Let $A \subset \Omega$ be non-empty. Let $x \in \Omega$. Let $s, t, m \geq 0$ to be defined shortly. Denote $G := G_s(A,m)$. We want this set to be of size at least $1 - \varepsilon$. By Corollary 3.0.15 we know that $\pi(G) \geq 1 - 8/m^2$. Thus we pick $m = \sqrt{8/\varepsilon}$. The precision in (3.0.1) is $m\sigma_s \leq \sqrt{8/\varepsilon}(\sqrt{\mathrm{Var}_\pi 1_A}e^{-s/t_{\mathrm{rel}}}) \leq \sqrt{2/\varepsilon}e^{-s/t_{\mathrm{rel}}}$. As we want to have ε precision, we pick $s := \left\lceil \frac{t_{\mathrm{rel}}}{2} \log\left(\frac{2}{\varepsilon^3}\right) \right\rceil$.

We seek to bound $|\mathrm{P}_x^{t+s}(A) - \pi(A)|$. If $|\mathrm{P}_x^{t+s}(A) - \pi(A)| \leq 2\varepsilon$, then the chain is "$2\varepsilon$-mixed w.r.t. A". This is where we use the set G. We now demonstrate that for any $t \geq 0$, hitting G by time t serves as a "certificate" that the chain is ε-mixed w.r.t. A at time $t + s$. Indeed, from the Markov property and the definition of G,

$$|\mathrm{P}_x[X_{t+s} \in A \mid T_G \leq t] - \pi(A)| \leq \max_{g \in G} \sup_{s' \geq s} |\mathrm{P}_g^{s'}(A) - \pi(A)| \leq \varepsilon.$$

In particular,

$$|\mathrm{P}_x^{t+s}(A) - \pi(A)| \le \mathrm{P}_x[T_G > t] + |\mathrm{P}_x[X_{t+s} \in A \mid T_G \le t] - \pi(A)| \le \mathrm{P}_x[T_G > t] + \varepsilon. \quad (3.0.4)$$

We seek to have the bound $\mathrm{P}_x[T_G > t] \le \varepsilon$. Recall that by our choice of m we have that $\pi(G) \ge 1-\varepsilon$. Thus if we pick $t := \mathrm{hit}_{1-\varepsilon}(\varepsilon)$, we guarantee that, regardless of the identity of A and x, we indeed have that $\mathrm{P}_x[T_G > t] \le \varepsilon$. Since x and A were arbitrary, plugging this into (3.0.4) yields (3.0.2). We now prove (3.0.3).

We now set $r := t_{\mathrm{mix}}(1-\varepsilon) - 1$. Then there exist some $x \in \Omega$ and $A \subset \Omega$ such that $\pi(A) - \mathrm{P}_x^r(A) > 1-\varepsilon$. In particular, $\pi(A) > 1-\varepsilon$. Consider again $G_2 := G_{s_2}(A,m)$. Since again we seek the size of G_2 to be at least $1-\varepsilon$, we again choose $m = \sqrt{8/\varepsilon}$. The precision in (3.0.1) is $m\sigma_{s_2} \le \sqrt{8/\varepsilon}(\sqrt{\mathrm{Var}_\pi 1_A}e^{-s_2/t_{\mathrm{rel}}}) \le \sqrt{8/\varepsilon}(\sqrt{1-\pi(A)}e^{-s_2/t_{\mathrm{rel}}}) \le \sqrt{8}e^{-s_2/t_{\mathrm{rel}}}$. We again seek ε precision. Hence we pick $s_2 := \left\lceil \frac{t_{\mathrm{rel}}}{2} \log\left(\frac{8}{\varepsilon^2}\right)\right\rceil$. As in (3.0.4) (with $r - s_2$ in the role of t and s_2 in the role of s) we have that

$$\mathrm{P}_x[T_{G_2} > r - s_2] \ge \pi(A) - \mathrm{P}_x^r(A) - \varepsilon > 1 - 2\varepsilon.$$

Hence it must be the case that $\mathrm{hit}_{1-\varepsilon}(1-2\varepsilon) > r - s_2 = t_{\mathrm{mix}}(1-\varepsilon) - 1 - \left\lceil \frac{t_{\mathrm{rel}}}{2} \log\left(\frac{8}{\varepsilon^2}\right)\right\rceil$.

We now prove the claim concerning the size of the good set.

Corollary 3.0.15. *Let (Ω, P, π) be a finite reversible irreducible lazy chain. As in Definition 3.0.14, define $\rho(A) := \sqrt{\pi(A)(1-\pi(A))}$, $\sigma_t := \rho(A)e^{-t/t_{\mathrm{rel}}}$ and*

$$G_t(A,m) := \left\{ y : |\mathrm{P}_y^k(A) - \pi(A)| < m\sigma_t \text{ for all } k \ge t \right\}.$$

Then

$$\pi(G_t(A,m)) \ge 1 - 8m^{-2}, \text{ for all } A \subset \Omega,\ t \ge 0 \text{ and } m > 0. \quad (3.0.5)$$

Proof. For any $t \ge 0$, let $f_t(x) := P^t(1_A - \pi(A))(x) = \mathrm{P}_x^t(A) - \pi(A)$. Then in the notation of Theorem 1.2.1,

$$(f_t)_*(x) := \sup_{k\ge 0} |P^k f_t(x)| = \sup_{k \ge 0} |\mathrm{P}_x^{k+t}(A) - \pi(A)|,$$

Hence $G_t \supseteq \{x \in \Omega : (f_t)_*(x) < m\sigma_t\}$. Whence

$$1 - \pi(G_t) \le \pi\left\{x : (f_t)_*(x) \ge m\sigma_t\right\}. \quad (3.0.6)$$

Note that since $\pi P^t = \pi$ we have that $\mathbb{E}_\pi(f_t) = \mathbb{E}_\pi(f_0) = \mathbb{E}_\pi(1_A - \pi(A)) = 0$. Now (1.1.10) implies that

$$\|Pf_t\|_2^2 \le \|f_t\|_2^2 = \mathrm{Var}_\pi P^t f_0 \le e^{-2t/t_{\mathrm{rel}}}\mathrm{Var}_\pi f_0 = e^{-2t/t_{\mathrm{rel}}}\rho^2(A) = \sigma_t^2. \quad (3.0.7)$$

Hence by Markov inequality and (1.2.1) we have

$$\pi\left\{x : (f_t)_*(x) \ge m\sigma_t\right\} = \pi\left\{x : ((f_t)_*(x))^2 \ge m^2\sigma_t^2\right\} \le 8m^{-2}. \quad (3.0.8)$$

The corollary now follows by substituting the last bound in (3.0.6). □

3.1 Inequalities relating $t_{\mathrm{mix}}(\varepsilon)$ and $\mathrm{hit}_\alpha(\delta)$

Our aim in this section is to obtain inequalities relating $t_{\mathrm{mix}}(\varepsilon)$ and $\mathrm{hit}_\alpha(\delta)$ for suitable values of α, ε and δ using Corollary 3.0.15.

The following corollary uses the same reasoning as in the proof of (3.0.2)-(3.0.3) with a slightly more careful analysis.

Corollary 3.1.1. *Let (Ω, P, π) be a lazy reversible irreducible finite chain. Let $x \in \Omega$, $\delta, \alpha \in (0,1)$, $s \geq 0$ and $A \subset \Omega$. Denote $t := \mathrm{hit}_{1-\alpha,x}(\delta)$. Then*

$$\mathrm{P}_x^{t+s}(A) \geq (1-\delta)\left[\pi(A) - e^{-s/t_{\mathrm{rel}}}\left[8\alpha^{-1}\pi(A)(1-\pi(A))\right]^{1/2}\right]. \tag{3.1.1}$$

Consequently, for any $0 < \varepsilon < 1$ we have that

$$\mathrm{hit}_{1-\alpha}((\alpha+\varepsilon)\wedge 1) \leq t_{\mathrm{mix}}(\varepsilon) \text{ and } t_{\mathrm{mix}}((\varepsilon+\delta)\wedge 1) \leq \mathrm{hit}_{1-\alpha}(\varepsilon) + \left\lceil \frac{t_{\mathrm{rel}}}{2}\log^+\left(\frac{2(1-\varepsilon)^2}{\alpha\varepsilon\delta}\right)\right\rceil, \tag{3.1.2}$$

where $a \wedge b := \min\{a,b\}$ and $\log^+ x := \max\{\log x, 0\}$. In particular, for any $0 < \varepsilon \leq 1/2$,

$$\mathrm{hit}_{1-\varepsilon/4}(5\varepsilon/4) \leq t_{\mathrm{mix}}(\varepsilon) \leq \mathrm{hit}_{1-\varepsilon/4}(3\varepsilon/4) + \left\lceil \frac{3t_{\mathrm{rel}}}{2}\log(4/\varepsilon)\right\rceil, \tag{3.1.3}$$

$$t_{\mathrm{mix}}(\varepsilon) \leq \mathrm{hit}_{1/2}(\varepsilon/2) + \lceil t_{\mathrm{rel}}\log(4/\varepsilon)\rceil \text{ and } t_{\mathrm{mix}}(1-\varepsilon/2) \leq \mathrm{hit}_{1/2}(1-\varepsilon) + 1_{\varepsilon>1/9}\left\lceil \frac{1}{2}t_{\mathrm{rel}}\log 8\right\rceil. \tag{3.1.4}$$

Proof. We first prove (3.1.1). Fix some $x \in \Omega$. Consider the set

$$G = G_s(A) := \left\{y : |\mathrm{P}_y^k(A) - \pi(A)| < e^{-s/t_{\mathrm{rel}}}\left(8\alpha^{-1}\pi(A)(1-\pi(A))\right)^{1/2} \text{ for all } k \geq s\right\}.$$

Then by Corollary 3.0.15 we have that

$$\pi(G) \geq 1-\alpha.$$

By the Markov property and conditioning on T_G and on X_{T_G} we get that

$$\mathrm{P}_x[X_{t+s} \in A \mid T_G \leq t] \geq \pi(A) - e^{-s/t_{\mathrm{rel}}}\left[8\alpha^{-1}\pi(A)(1-\pi(A))\right]^{1/2}.$$

Since $\pi(G) \geq 1-\alpha$ we have that $\mathrm{P}_x[T_G \leq t] \geq 1-\delta$ for $t := \mathrm{hit}_{1-\alpha,x}(\delta)$. Thus

$$\mathrm{P}_x^{t+s}(A) \geq \mathrm{P}_x[T_G \leq t]\mathrm{P}_x[X_{t+s} \in A \mid T_G \leq t] \geq (1-\delta)\left[\pi(A) - e^{-s/t_{\mathrm{rel}}}\left[8\alpha^{-1}\pi(A)(1-\pi(A))\right]^{1/2}\right],$$

which concludes the proof of (3.1.1). We now prove (3.1.2). The first inequality in (3.1.2) follows directly from the definition of the total variation distance. To see this, let $A \subset \Omega$ be an arbitrary set with $\pi(A) \geq 1-\alpha$. Let $t_1 := t_{\mathrm{mix}}(\varepsilon)$. Then for any $x \in \Omega$,

$$\mathrm{P}_x[T_A \leq t_1] \geq \mathrm{P}_x[X_{t_1} \in A] \geq \pi(A) - \|\mathrm{P}_x^{t_1} - \pi\|_{\mathrm{TV}} \geq 1-\alpha-\varepsilon.$$

In particular, we get directly from Definition 2.1.1 that $\mathrm{hit}_{1-\alpha}(\alpha+\varepsilon) \le t_1 = t_{\mathrm{mix}}(\varepsilon)$. We now prove the second inequality in (3.1.2).

Set $t := \mathrm{hit}_{1-\alpha}(\varepsilon)$ and $s := \left\lceil \frac{1}{2} t_{\mathrm{rel}} \log^{+}\left(\frac{2(1-\varepsilon)^2}{\alpha\varepsilon\delta}\right)\right\rceil$. Let $x \in \Omega$ be such that $d(t+s) = \|\mathrm{P}_x^{t+s} - \pi\|_{\mathrm{TV}}$ and set $A := \{y \in \Omega : \pi(y) > \mathrm{P}_x^{t+s}(y)\}$. Observe that by the choice of t, s, x and A together with (3.1.1) we have that

$$
\begin{aligned}
d(t+s) &= \pi(A) - \mathrm{P}_x^{t+s}(A) \le \varepsilon\pi(A) + (1-\varepsilon)e^{-s/t_{\mathrm{rel}}}\left[8\alpha^{-1}\pi(A)(1-\pi(A))\right]^{1/2} \\
&\le \varepsilon[\pi(A) + 2\sqrt{\delta/\varepsilon}\sqrt{\pi(A)(1-\pi(A))}] \le \varepsilon[1 + (2\sqrt{\delta/\varepsilon})^2/4] = \varepsilon + \delta,
\end{aligned}
\tag{3.1.5}
$$

where in the last inequality we have used the easy fact that for any $c > 0$ and any $x \in [0,1]$ we have that $x + c\sqrt{x(1-x)} \le 1 + c^2/4$. Indeed, since $x \in [0,1]$ it suffices to show that $x + c\sqrt{(1-x)} \le 1 + c^2/4$. Write $\sqrt{1-x} = y$ and $c/2 = a$. By subtracting x from both sides, the previous inequality is equivalent to $2ay \le y^2 + a^2$. This concludes the proof of (3.1.2).

For the second inequality of (3.1.3), apply (3.1.2) with $(\alpha, \varepsilon, \delta)$ being $(\varepsilon/4, 3\varepsilon/4, \varepsilon/4)$. Similarly, to get (3.1.4) apply (3.1.2) with $(\alpha, \varepsilon, \delta)$ being $(1/2, \varepsilon/2, \varepsilon/2)$ or $(1/2, 1-\varepsilon, \varepsilon/2)$, respectively. □

Let $\alpha \in (0,1)$. Observe that for any $A \subset \Omega$ with $\pi(A) \ge \alpha$, any $x \in \Omega$ and any $t, s \ge 0$, by the Markov property we have that $\mathrm{P}_x[T_A > t+s] \le \mathrm{P}_x[T_A > t]\big(\max_z \mathrm{P}_z[T_A > s]\big) \le p(\alpha,t)p(\alpha,s)$. Maximizing over x and A yields that $p(\alpha, t+s) \le p(\alpha,t)p(\alpha,s)$, from which the following proposition follows.

Proposition 3.1.2. *For any $\alpha, \varepsilon, \delta \in (0,1)$ we have that*

$$
\mathrm{hit}_\alpha(\varepsilon\delta) \le \mathrm{hit}_\alpha(\varepsilon) + \mathrm{hit}_\alpha(\delta). \tag{3.1.6}
$$

In the next corollary, we establish inequalities between $\mathrm{hit}_\alpha(\delta)$ and $\mathrm{hit}_\beta(\delta')$ for appropriate values of α, β, δ and δ'.

Corollary 3.1.3. *For any reversible irreducible finite chain and $0 < \varepsilon < \delta < 1$,*

$$
\mathrm{hit}_\beta(\delta) \le \mathrm{hit}_\alpha(\delta) \le \mathrm{hit}_\beta(\delta-\varepsilon) + \left\lceil \alpha^{-1} t_{\mathrm{rel}} \log\left(\frac{1-\alpha}{(1-\beta)\varepsilon}\right)\right\rceil, \text{ for any } 0 < \alpha \le \beta < 1. \tag{3.1.7}
$$

The general idea behind Corollary 3.1.3 is as follows. Loosely speaking, we show that any set $A \subset \Omega$ has a "blow-up" set $H(A)$ (of large π-measure), such that starting from any $x \in H(A)$, the set A is hit "quickly" (in time proportional to $t_{\mathrm{rel}}/\pi(A)$) with large probability.

In order to establish the existence of such a blow-up, it turns out that it suffices to consider the hitting time of A starting from the initial distribution π, which is well-understood.

Lemma 3.1.4. *Let (Ω, P, π) be a finite irreducible reversible Markov chain. Let $A \subsetneq \Omega$ be non-empty. Let $\alpha > 0$ and $w \ge 0$. Let $B(A, w, \alpha) := \left\{y : \mathrm{P}_y\left[T_A > \left\lceil \frac{t_{\mathrm{rel}} w}{\pi(A)}\right\rceil\right] \ge \alpha\right\}$. Then*

$$
\mathrm{P}_\pi[T_A > t] \le \pi(A^c)\left(1 - \frac{\pi(A)}{t_{\mathrm{rel}}}\right)^t \le \pi(A^c)\exp\left(-\frac{t\pi(A)}{t_{\mathrm{rel}}}\right), \text{ for any } t \ge 0. \tag{3.1.8}
$$

In particular,

$$
\pi\left(B(A,w,\alpha)\right) \le \pi(A^c)e^{-w}\alpha^{-1} \text{ and } \pi(A)\mathbb{E}_\pi[T_A] \le t_{\mathrm{rel}}\pi(A^c). \tag{3.1.9}
$$

The proof of Lemma 3.1.4 is deferred to the end of this section.

Proof of Corollary 3.1.3. Denote $s = s_{\alpha,\beta,\varepsilon} := \left\lceil \alpha^{-1} t_{\mathrm{rel}} \log\left(\frac{1-\alpha}{(1-\beta)\varepsilon}\right)\right\rceil$. Let $A \subset \Omega$ be an arbitrary set such that $\pi(A) \geq \alpha$. Consider the set

$$H_1 = H_1(A, \alpha, \beta, \varepsilon) := \{y \in \Omega : \mathrm{P}_y[T_A \leq s] \geq 1 - \varepsilon\}.$$

Then by (3.1.9)

$$\begin{aligned} \pi(H_1) &\geq 1 - (1 - (1 - \varepsilon))^{-1}(1 - \pi(A)) \exp\left[-\frac{s\pi(A)}{t_{\mathrm{rel}}}\right] \\ &\geq 1 - \varepsilon^{-1}(1 - \alpha) \exp\left[-\log\left(\frac{1-\alpha}{(1-\beta)\varepsilon}\right)\right] = \beta. \end{aligned}$$

By the definition of H_1 together with the Markov property and the fact that $\pi(H_1) \geq \beta$, for any $t \geq 0$ and $x \in \Omega$,

$$\begin{aligned} \mathrm{P}_x[T_A \leq t + s] &\geq \mathrm{P}_x[T_{H_1} \leq t, T_A \leq t + s] \geq (1 - \varepsilon)\mathrm{P}_x[T_{H_1} \leq t] \\ &\geq (1 - \varepsilon)(1 - p_x(\beta, t)) \geq 1 - \varepsilon - \max_{y \in \Omega} p_y(\beta, t). \end{aligned} \tag{3.1.10}$$

Taking $t := \mathrm{hit}_\beta(\delta - \varepsilon)$ and minimizing the LHS of (3.1.10) over A and x gives the second inequality in (3.1.7). The first inequality in (3.1.7) is trivial because $\alpha \leq \beta$. $\square$

3.1.1 Proofs of Proposition 2.1.5 and Theorem 2.1.1

Now we are ready to prove our main abstract results.

Proof of Proposition 2.1.5. First note that (2.1.3) follows from (3.1.3) and the first inequality in (1.1.8). Moreover, in light of (3.1.4) we only need to prove the first inequalities in (2.1.1) and (2.1.2). Fix some $0 < \varepsilon \leq 1/4$ and $t \geqslant 0$. Take any set A with $\pi(A) \geqslant \frac{1}{2}$ and $x \in \Omega$. Denote $s_\varepsilon := \lceil 2t_{\mathrm{rel}} |\log \varepsilon| \rceil$. Consider a coupling $(\mathbb{P}, (Y_k, Z_k)_{k \geq 0})$ of the chain $(Y_k)_{k \geq 0}$ with initial distribution $Y_0 \sim \mathrm{P}_x^t$ with the stationary chain $(Z_k)_{k \geq 0}$ so that $\mathbb{P}[(Y_k)_{k \geq 0} \neq (Z_k)_{k \geq 0}] = d_x(t)$ (cf. the proofs of Proposition 4.7 and of Theorem 5.2 in [36] for the existence of such a coupling). By the Markov property

$$\begin{aligned} \mathrm{P}_x[T_A > t + s_\varepsilon] &\leq \mathrm{P}_x[X_k \notin A \text{ for all } t \leq k \leq t + s_\varepsilon] = \mathbb{P}[Y_k \notin A \text{ for all } k \leq s_\varepsilon] \\ &\leq \mathbb{P}[(Y_k)_{k \geq 0} \neq (Z_k)_{k \geq 0}] + \mathbb{P}[Z_k \notin A \text{ for all } k \leq s_\varepsilon] = d_x(t) + \mathrm{P}_\pi[T_A > s_\varepsilon]. \end{aligned}$$

Hence by (3.1.8)

$$\mathrm{P}_x[T_A > t + s_\varepsilon] \leqslant d_x(t) + \frac{1}{2} e^{-s_\varepsilon / 2t_{\mathrm{rel}}} \leqslant d(t) + \frac{\varepsilon}{2}.$$

Putting $t = t_{\mathrm{mix}}(\varepsilon)$ and $t = t_{\mathrm{mix}}(1 - \varepsilon)$ successively in the above equation and maximizing over $x \in \Omega$ and A such that $\pi(A) \geqslant \frac{1}{2}$ gives

$$\mathrm{hit}_{1/2}(3\varepsilon/2) \leqslant t_{\mathrm{mix}}(\varepsilon) + s_\varepsilon \text{ and } \mathrm{hit}_{1/2}(1 - \varepsilon/2) \leqslant t_{\mathrm{mix}}(1 - \varepsilon) + s_\varepsilon,$$

which completes the proof. $\square$

Before completing the proof of Theorem 2.1.1, we prove that under the product condition if a sequence of reversible chains exhibits hit_α-cutoff for some $\alpha \in (0,1)$, then it exhibits hit_α-cutoff for all $\alpha \in (0,1)$.

Proposition 3.1.5. *Let (Ω_n, P_n, π_n) be a sequence of lazy finite irreducible reversible chains for which the product condition holds. Then (1) and (2) below are equivalent:*

(1) There exists $\alpha \in (0,1)$ for which the sequence exhibits a hit_α-cutoff.

(2) The sequence exhibits a hit_α-cutoff for any $\alpha \in (0,1)$.

Moreover,

$$\mathrm{hit}_\alpha^{(n)}(1/4) = \Theta(t_{\mathrm{mix}}^{(n)}), \text{ for any } \alpha \in (0,1). \tag{3.1.11}$$

Furthermore, if (2) holds then

$$\lim_{n\to\infty} \mathrm{hit}_\alpha^{(n)}(1/4)/\mathrm{hit}_{1/2}^{(n)}(1/4) = 1, \text{ for any } \alpha \in (0,1). \tag{3.1.12}$$

Proof. We start by proving (3.1.11). Assume that the product condition holds. Fix some $\alpha \in (0,1)$. Note that we have

$$\mathrm{hit}_\alpha^{(n)}(1/4) \le 4\alpha^{-1}\mathrm{hit}_\alpha^{(n)}\left(1-\frac{3\alpha}{4}\right) \le 4\alpha^{-1}t_{\mathrm{mix}}^{(n)}\left(\frac{\alpha}{4}\right) \le 4\alpha^{-1}(2+\lceil \log_2(1/\alpha)\rceil)t_{\mathrm{mix}}^{(n)}.$$

The first inequality above follows from (3.1.6) and the fact that $(1-3\alpha/4)^{4\alpha^{-1}-1} \le 4e^{-3} \le 1/4$. The second one follows from (3.1.2)(first inequality). The final inequality above is a consequence of the sub-multiplicativity property: for any $k,t \ge 0$, $d(kt) \le (2d(t))^k$ (e.g. [36], (4.24) and Lemma 4.12).

Conversely, by (3.1.6) (second inequality) and the second inequality in (3.1.2) with $(\alpha, \varepsilon, \delta)$ here being $(1-\alpha, 1/8, 1/8)$ (first inequality)

$$\frac{t_{\mathrm{mix}}^{(n)}}{2} - \left\lceil \frac{t_{\mathrm{rel}}^{(n)}}{4} \log\left(\frac{100}{1-\alpha}\right)\right\rceil \le \frac{\mathrm{hit}_\alpha^{(n)}(1/8)}{2} \le \mathrm{hit}_\alpha^{(n)}(1/4).$$

This concludes the proof of (3.1.11). We now prove the equivalence between (1) and (2) under the product condition. It suffices to show that (1) $\Longrightarrow$ (2), as the reversed implication is trivial. Fix $0<\alpha<\beta<1$. It suffices to show that hit_α-cutoff occurs iff hit_β-cutoff occurs.

Fix $\varepsilon \in (0,1/8)$. Denote $s_n = s_n(\alpha,\beta,\varepsilon) := \left\lceil t_{\mathrm{rel}}^{(n)}\alpha^{-1}\log\left(\frac{1-\alpha}{(1-\beta)\varepsilon}\right)\right\rceil$. By the second inequality in Corollary 3.1.3

$$\mathrm{hit}_\alpha^{(n)}(1-\varepsilon) \le \mathrm{hit}_\beta^{(n)}(1-2\varepsilon) + s_n \text{ and } \mathrm{hit}_\alpha^{(n)}(2\varepsilon) \le \mathrm{hit}_\beta^{(n)}(\varepsilon) + s_n. \tag{3.1.13}$$

By the first inequality in Corollary 3.1.3

$$\mathrm{hit}_\beta^{(n)}(2\varepsilon) \le \mathrm{hit}_\alpha^{(n)}(2\varepsilon) \le \mathrm{hit}_\alpha^{(n)}(\varepsilon) \text{ and } \mathrm{hit}_\beta^{(n)}(1-\varepsilon) \le \mathrm{hit}_\beta^{(n)}(1-2\varepsilon) \le \mathrm{hit}_\alpha^{(n)}(1-2\varepsilon). \tag{3.1.14}$$

Hence

$$\begin{aligned}\mathrm{hit}_\beta^{(n)}(2\varepsilon) - \mathrm{hit}_\beta^{(n)}(1-2\varepsilon) &\le \mathrm{hit}_\alpha^{(n)}(\varepsilon) - \mathrm{hit}_\alpha^{(n)}(1-\varepsilon) + s_n,\\ \mathrm{hit}_\alpha^{(n)}(2\varepsilon) - \mathrm{hit}_\alpha^{(n)}(1-2\varepsilon) &\le \mathrm{hit}_\beta^{(n)}(\varepsilon) - \mathrm{hit}_\beta^{(n)}(1-\varepsilon) + s_n.\end{aligned} \tag{3.1.15}$$

Note that by the assumption that the product condition holds, we have that $s_n = o(t_{\mathrm{mix}}^{(n)})$. Assume that the sequence exhibits hit$_\alpha$-cutoff. Then by (3.1.11) the RHS of the first line of (3.1.15) is $o(t_{\mathrm{mix}}^{(n)})$. Again by (3.1.11), this implies that the RHS of the first line of (3.1.15) is $o(\mathrm{hit}_\beta^{(n)}(1/4))$ and so the sequence exhibits hit$_\beta$-cutoff. Applying the same reasoning, using the second line of (3.1.15), shows that if the sequence exhibits hit$_\beta$-cutoff, then it also exhibits hit$_\alpha$-cutoff.

We now prove (3.1.12). Let $a \in (0,1)$. Denote $\alpha := \min\{a, 1/2\}$ and $\beta := \max\{a, 1/2\}$. Let $s_n = s_n(\alpha, \beta, \varepsilon)$ be as before. By the second inequality in Corollary 3.1.3

$$\mathrm{hit}_\alpha^{(n)}(1/4+\varepsilon) - s_n \leq \mathrm{hit}_\beta^{(n)}(1/4) \leq \mathrm{hit}_\alpha^{(n)}(1/4). \tag{3.1.16}$$

By assumption (2) together with the product condition and (3.1.11), the LHS of (3.1.16) is at least $(1-o(1))\mathrm{hit}_\alpha^{(n)}(1/4)$, which by (3.1.16), implies (3.1.12). □

The following proposition shows that for all $\alpha \leqslant 1/2$ the occurrence of hit$_\alpha$-cutoff implies that the product condition holds. In particular, this implies the equivalence of 2) and 3) in Theorem 2.1.1.

Proposition 3.1.6. *Let (Ω_n, P_n, π_n) be a sequence of lazy finite irreducible reversible chains. Assume that the product condition fails. Then for any $\alpha \leq 1/2$ the sequence does not exhibit* hit$_\alpha$*-cutoff.*

Before providing the proof of Proposition 3.1.6, we complete the proof of Theorem 2.1.1.

Proof of Theorem 2.1.1. By Proposition 3.1.6 it suffices to consider the case in which the product condition holds. By Propositions 3.1.5 it suffices to consider the case $\alpha = 1/2$ (that is, it suffices to show that under the product condition the sequence exhibits cutoff iff it exhibits hit$_{1/2}$-cutoff). This follows at once from (2.1.1), (2.1.2) and (3.1.12). □

Proof of Proposition 3.1.6. Fix some $0 < \alpha \leq 1/2$. We first argue that for all n, $k \geq 1$

$$\mathrm{hit}_\alpha^{(n)}([1-\alpha/2]^k) \leq k\lceil |\log_2(\alpha/2)|\rceil t_{\mathrm{mix}}^{(n)}. \tag{3.1.17}$$

By the submultiplicativity property (3.1.6), it suffices to verify (3.1.17) only for $k = 1$. As in the proof of Proposition 3.1.5, by the submultiplicativity property $d(mt) \leq (2d(t))^m$, together with (3.1.2), we have that $\mathrm{hit}_\alpha^{(n)}(1-\alpha/2) \leq t_{\mathrm{mix}}^{(n)}(\alpha/2) \leq \lceil |\log_2(\alpha/2)|\rceil) t_{\mathrm{mix}}^{(n)}$.

Conversely, by the laziness assumption, we have that for all n,

$$\mathrm{hit}_\alpha^{(n)}(\varepsilon/2) \geq |\log_2 \varepsilon|, \text{ for all } 0 < \varepsilon < 1. \tag{3.1.18}$$

To see this, consider the case that $X_0^{(n)} = y_n$, for some $y_n \in \Omega_n$ such that $\pi_n(y_n) \leq 1/2 \leq 1-\alpha$, and that the first $\lfloor |\log_2 \varepsilon|\rfloor$ steps of the chain are lazy (i.e. $y_n = X_1^{(n)} = \cdots = X_{\lfloor |\log_2 \varepsilon|\rfloor})$.

By (3.1.17) in conjunction with (3.1.18) we may assume that $\lim_{n\to\infty} t_{\mathrm{mix}}^{(n)} = \infty$, as otherwise there cannot be hit$_\alpha$-cutoff. By passing to a subsequence, we may assume further that there exists some $C > 0$ such that $t_{\mathrm{mix}}^{(n)} < Ct_{\mathrm{rel}}^{(n)}$. In particular, $\lim_{n\to\infty} t_{\mathrm{rel}}^{(n)} = \infty$ and we may assume without loss of generality that $(\lambda_2^{(n)})^{t_{\mathrm{mix}}^{(n)}} \geq e^{-C}$ for all n, where $\lambda_2^{(n)}$ is the second largest eigenvalue of P_n.

For notational convenience we now suppress the dependence on n from our notation. Let $f_2 \in \mathbb{R}^{\Omega}$ be a non-zero vector satisfying that $Pf_2 = \lambda_2 f_2$. By considering $-f_2$ if necessary, we may assume that $A := \{x \in \Omega : f_2 \leq 0\}$ satisfies $\pi(A) \geq 1/2$. Let $x \in \Omega$ be such that $f_2(x) = \max_{y \in \Omega} f_2(y) =: L$. Note that $L > 0$ since $\mathbb{E}_{\pi}[f_2] = 0$.

Consider $N_k := \lambda_2^{-k} f_2(X_k)$ and $M_k := N_{k \wedge T_A}$, where $X_0 = x$. Observe that $(N_k)_{k \geq 0}$ is a martingale and hence so is $(M_k)_{k \geq 0}$ (w.r.t. the natural filtration induced by the chain). As $M_k \leq 0$ on $\{T_A \leq k\}$ and $M_k \leq \lambda_2^{-k} L$ on $\{T_A > k\}$, we get that for all $k > 0$, $M_k \leq \lambda_2^{-k} L 1_{T_A > k}$, and so

$$L = \mathbb{E}_x[M_0] = \mathbb{E}_x[M_k] \leq \mathbb{E}_x[\lambda_2^{-k} L 1_{T_A > k}] = \lambda_2^{-k} L \mathrm{P}_x[T_A > k]. \tag{3.1.19}$$

Thus $\mathrm{P}_x[T_A > k] \geq \lambda_2^k$, for all k. Consequently, for all $a > 0$,

$$\mathrm{P}_x[T_A > at_{\mathrm{mix}}] \geq \lambda_2^{at_{\mathrm{mix}}} \geq e^{-aC}. \tag{3.1.20}$$

Thus

$$\mathrm{hit}_{\alpha}(\varepsilon/2) \geq \mathrm{hit}_{1/2}(\varepsilon/2) \geq C^{-1} t_{\mathrm{mix}} |\log \varepsilon|, \text{ for any } 0 < \varepsilon < 1.$$

This, in conjunction with (3.1.17), implies that $\frac{\mathrm{hit}_{\alpha}(\varepsilon)}{\mathrm{hit}_{\alpha}(1-\varepsilon)} \geq \frac{|\log \varepsilon|}{C \lceil \log_2(\alpha/2) \rceil}$, for all $0 < \varepsilon \leq \alpha/2$. Consequently, there is no hit_{α}-cutoff. $\square$

3.1.2 Proof of Lemma 3.1.4

Now we prove Lemma 3.1.4. As mentioned before, the hitting time of a set A starting from stationary initial distribution is well-understood (see [21]; for the continuous-time analog see [3], Chapter 3 Sections 5 and 6.5 or [11]). Assuming that the chain is lazy, it follows from the theory of complete monotonicity together with some linear-algebra that this distribution is dominated by a distribution which gives mass $\pi(A)$ to 0, and conditionally on being positive, is distributed as the Geometric distribution with parameter $\pi(A)/t_{\mathrm{rel}}$. Since the existing literature lacks simple treatment of this fact (especially for the discrete-time case) we now prove it for the sake of completeness. We shall prove this fact without assuming laziness. Although without assuming laziness the distribution of T_A under P_{π} need not be completely monotone, the proof is essentially identical as in the lazy case.

For any non-empty $A \subset \Omega$, we write π_A for the distribution of π conditioned on A. That is, $\pi_A(\cdot) := \frac{\pi(\cdot) 1_{\cdot \in A}}{\pi(A)}$.

Lemma 3.1.7. *Let (Ω, P, π) be a reversible irreducible finite chain. Let $A \subsetneq \Omega$ be non-empty. Denote its complement by B and write $k = |B|$. Consider the sub-stochastic matrix P_B, which is the restriction of P to B. That is $P_B(x, y) := P(x, y)$ for $x, y \in B$. Assume that P_B is irreducible, that is, for any $x, y \in B$, exists some $t \geq 0$ such that $P_B^t(x, y) > 0$. Then*

(i) P_B has k real eigenvalues $1 - \pi(A)/t_{\mathrm{rel}} \geq \gamma_1 > \gamma_2 \geq \cdots \geq \gamma_k \geq -\gamma_1$.

(ii) There exist some non-negative $a_1, \ldots, a_k$ satisfying $\sum_{i=1}^k a_i = 1$ such that for any $t \geq 0$,

$$\mathrm{P}_{\pi_B}[T_A > t] = \sum_{i=1}^k a_i \gamma_i^t. \tag{3.1.21}$$

(iii)

$$\mathrm{P}_{\pi_B}[T_A > t] \le \gamma_1^t \le \left(1 - \frac{\pi(A)}{t_{\mathrm{rel}}}\right)^t \le \exp\left(-\frac{t\pi(A)}{t_{\mathrm{rel}}}\right), \text{ for all } t \ge 0. \tag{3.1.22}$$

Proof. We first note that (3.1.22) follows immediately from (3.1.21) and (i). Indeed, by (i), $|\gamma_i| \le \gamma_1 \le 1 - \frac{\pi(A)}{t_{\mathrm{rel}}}$ for all i, and so (3.1.21) implies that $\mathrm{P}_{\pi_B}[T_A > t] \le \gamma_1^t \le \left(1 - \frac{\pi(A)}{t_{\mathrm{rel}}}\right)^t$ for all $t \ge 0$. We now prove (i).

Consider the following inner-product on $\mathbb{R}^B$, $\langle f, g\rangle_{\pi_B} := \sum_{x \in B} \pi_B(x) f(x) g(x)$. Since P is reversible, P_B is self-adjoint w.r.t. this inner-product. Hence indeed P_B has k real eigenvalues $\gamma_1 > \gamma_2 \ge \cdots \ge \gamma_k$ and there is a basis of $\mathbb{R}^B$, $g_1, \ldots, g_k$ of orthonormal vectors w.r.t. the aforementioned inner-product, such that $P_B g_i = \gamma_i g_i$ ($i \in [k]$). By the Perron-Frobenius Theorem $\gamma_1 > 0$ and $\gamma_1 \ge -\gamma_k$.

By the Courant-Fischer variational characterization of eigenvalues we have

$$\frac{1}{t_{\mathrm{rel}}(B)} = 1 - \gamma_1 = \inf\left\{ \frac{\langle (I-P)g, g\rangle_\pi}{\langle g, g\rangle_\pi} : g \geqslant 0, g = 0 \text{ on } A, g \text{ non-constant} \right\}. \tag{3.1.23}$$

Also observe that for all $g \geqslant 0$ such that $g = 0$ on A we have by the Cauchy-Schwarz inequality that $\mathbb{E}_{\pi_B} g^2 \geqslant (\mathbb{E}_{\pi_B} g)^2$ (where for $f \in \mathbb{R}^\Omega$ we denote $\mathbb{E}_{\pi_B} f := \sum_b \pi_B(b) f(b)$) which rearranges to

$$\mathrm{Var}_\pi g = \langle g - \mathbb{E}_\pi g, g - \mathbb{E}_\pi g\rangle_\pi \geqslant \pi(A)\langle g, g\rangle_\pi.$$

Thus by (3.1.23) $1 - \gamma_1 \ge \pi(A) \inf\{\langle (I-P)g, g\rangle_\pi / \mathrm{Var}_\pi g : g \geqslant 0, g = 0 \text{ on } A, g \text{ non-constant}\}$, which in comparison with the variational characterization of t_{rel} (e.g. [36, Remark 13.13])

$$1/t_{\mathrm{rel}} = \inf\{\langle (I-P)g, g\rangle_\pi / \mathrm{Var}_\pi g : g \text{ non-constant}\},$$

yields that $1 - \gamma_1 \ge \pi(A)/t_{\mathrm{rel}}$. This concludes the proof of part (i). We now prove part (ii).

By summing over all paths of length t which are contained in B we get that

$$\mathrm{P}_{\pi_B}[T_A > t] = \sum_{x,y \in B} \pi_B(x) P_B^t(x,y). \tag{3.1.24}$$

By the spectral representation (cf. [36, Lemma 12.2] and Section 4 of Chapter 3 in [3]) for any $x, y \in B$ and $t \in \mathbb{N}$ we have that $P_B^t(x,y) = \sum_{i=1}^k \pi_B(y) g_i(x) g_i(y) \gamma_i^t$. So by (3.1.24)

$$\mathrm{P}_{\pi_B}[T_A > t] = \sum_{x,y \in B} \pi_B(x) \sum_{i=1}^k \pi_B(y) g_i(x) g_i(y) \gamma_i^t = \sum_{i=1}^k a_i \gamma_i^t,$$

where $a_i := \left(\sum_{x \in B} \pi_B(x) g_i(x)\right)^2$. Plugging $t = 0$ shows that indeed $\sum_{i=1}^k a_i = 1$, as desired. □

Using the same argument for the continuous-time setup, it follows that

$$\mathrm{H}_{\pi_B}[T_A > t] = \sum_{x,y \in B} \pi_B(x) \sum_{i=1}^k \pi_B(y) g_i(x) g_i(y) e^{-(1-\gamma_i)t} = \sum_{i=1}^k a_i e^{-(1-\gamma_i)t} \le e^{-t\pi(A)/t_{\mathrm{rel}}}.$$

Proof of Lemma 3.1.4. We first note that (3.1.9) follows easily from (3.1.8). For the first inequality in (3.1.9) set $t = t(A,w) := \lceil t_{\mathrm{rel}} w/\pi(A) \rceil$ and $B := B(A,w,\alpha) = \{y : \mathrm{P}_y [T_A > t] \geq \alpha\}$. Then by (3.1.8)

$$\alpha\pi(B) \leq \pi(B)\mathrm{P}_{\pi_B}[T_A > t] \leq \mathrm{P}_\pi[T_A > t] \leq \pi(A^c)\exp\left(-t\pi(A)/t_{\mathrm{rel}}\right) \leq \pi(A^c)e^{-w}.$$

For the first inequality in (3.1.9) recall that $\mathbb{E}_\pi[T_A] = \sum_{t \geq 0} \mathrm{P}_\pi[T_A > t]$ and apply (3.1.8).

We now prove (3.1.8). Denote the connected components of $A^c := \Omega \setminus A$ by $\{C_1, \ldots, C_k\}$. Denote the complement of C_i by C_i^c. By (3.1.22) we have that

$$\mathrm{P}_\pi[T_A > t] = \sum_{i=1}^{k} \pi(C_i)\mathrm{P}_{\pi_{C_i}}[T_A > t] = \sum_{i=1}^{k} \pi(C_i)\mathrm{P}_{\pi_{C_i}}[T_{C_i^c} > t] \leq$$
$$\sum_{i=1}^{k} \pi(C_i)\left(1 - \frac{\pi(C_i^c)}{t_{\mathrm{rel}}}\right)^t \leq \sum_{i=1}^{k} \pi(C_i)\left(1 - \frac{\pi(A)}{t_{\mathrm{rel}}}\right)^t = \pi(A^c)\exp\left(-\frac{t\pi(A)}{t_{\mathrm{rel}}}\right). \quad \square$$

3.1.3 Sharpness of Theorem 2.1.1

Now we give an example to show that in Proposition 3.1.6 (and hence in Theorem 2.1.1) the value $\frac{1}{2}$ cannot be replaced by any larger value.

Example 3.1.8. *Let (Ω_n, P_n, π_n) be the nearest-neighbor weighted random walk from Figure 3.1. Then $t_{\mathrm{rel}}^{(n)} = \Theta(t_{\mathrm{mix}}^{(n)})$, yet for every $1/2 < \alpha < 1$, the sequence exhibits* hit_α*-cutoff.*

Proof. Let $\Phi_n := \min_{A \subset \Omega_n : 0 < \pi_n(A) \leq 1/2} \Phi_n(A)$ be the Cheeger constant of the n-th chain, where $\Phi_n(A) := \frac{\sum_{a \in A, b \in A^c} \pi_n(a)P_n(a,b)}{\pi_n(A)}$. Then by taking A to be either A_1 or A_2, by Cheeger inequality (e.g. [36], Theorem 13.14), we have that $t_{\mathrm{rel}}^{(n)} \geq \frac{1}{2\Phi_n} \geq c_1 n^2 \geq c_2 t_{\mathrm{mix}}^{(n)}$ (it is easy to show that by (2.0.1) and the fact that $\pi_n(A_i) = 1/2 - o(1)$ for $i = 1, 2$ we have that $t_{\mathrm{mix}}^{(n)} \leq Cn^2$). By (1.1.8), indeed $t_{\mathrm{rel}}^{(n)} = \Theta(t_{\mathrm{mix}}^{(n)})$ and it follows that there is no cutoff.

Fix some $1/2 < \alpha < 1$. Let $B \subset \Omega_n$ be such that $\pi_n(B) \geq \alpha$. Denote the set of vertices belonging to the path, but not to A_1 by D. Then $\pi_n(D) = O(n^{-2}) = o(1)$. Consequently, $\pi_n(A_i \cap B) \geq \alpha - 1/2 - o(1)$, for $i = 1, 2$. Using this observation, it is easy to verify that for all $x \in A_1 \cup A_2$ we have that

$$\mathrm{hit}_{\alpha,x}(\varepsilon) \leq c_\alpha \log(1/\varepsilon), \text{ for any } 0 < \varepsilon < 1, \tag{3.1.25}$$

for some constant c_α independent of n.

Let y be the endpoint of the path which does not lie in A_1. Let z be the other endpoint of the path. The hitting time of z under P_y is concentrated around time $6 \log n$. Then by (3.1.25), together with the Markov property (using the same reasoning as in the proof of Lemma 5.1.1) for all sufficiently large n we have that for any $0 < \varepsilon \leq 1/4$

$$\begin{aligned} \mathrm{hit}_{\alpha,y}^{(n)}(2\varepsilon) &\leq (6 + o(1))\log n + \mathrm{hit}_{\alpha,z}^{(n)}(\varepsilon) = (6 + o(1))\log n, \\ \mathrm{hit}_{\alpha,y}^{(n)}(1 - \varepsilon) &\geq (6 - o(1))\log n. \end{aligned} \tag{3.1.26}$$

Similarly to the proof of Lemma 5.1.1, for any $B \subset \Omega_n$ and any $x \in D$, we have that $\mathrm{P}_y[T_{B \setminus D} > t] \geq \mathrm{P}_x[T_B > t]$, for all t. Since $\pi_n(D) = o(1)$, this implies that for all sufficiently

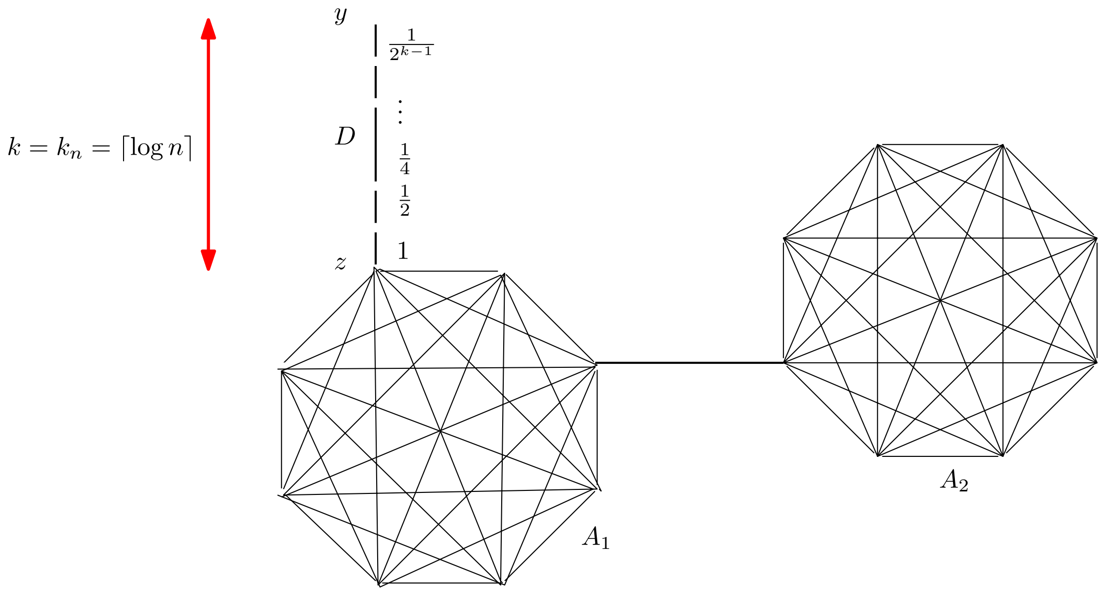

Figure 3.1: We consider a lazy weighted nearest-neighbor random walk on the above graph consisting of two disjoint cliques A_1 and A_2 of size n connected by a single edge and a path of length $k_n = \lceil \log n \rceil$ connected to A_1. The edge weights of all edges incident to vertices in $A_1 \cup A_2$ is 1, while those belonging to the path are indicated in the figure. Inside the path, the walk has a fixed bias towards the clique.

large n, for any $1/2 < \alpha < 1$, there exists some $1/2 < \alpha' < \alpha$ (α' depends on α but not on n), such that for any $x \in D$ we have that $\mathrm{hit}_{\alpha,y}^{(n)}(\varepsilon) \geq \mathrm{hit}_{\alpha',x}^{(n)}(\varepsilon)$, for all $0 < \varepsilon < 1$. This, together with (3.1.25) and the fact that the leftmost terms in both lines of (3.1.26) are up to negligible terms independent of α and ε, implies that the sequence of chains exhibits hit_α-cutoff for all $1/2 < \alpha < 1$. $\square$

Remark 3.1.9. *One can modify the sequence from Example 3.1.8 into a sequence of lazy simple nearest-neighbor random walks on a graph. Construct the n-th graph in the sequence as follows. Start with a binary tree T of depth n. Denote its root by y, the set of its leaves by A_1 and $D := T \setminus A_1$. Turn A_1 into a clique by connecting every two leaves of T by an edge. Take another disjoint complete graph of size $|A_1| = 2^n$ and denote its vertices by A_2. Finally, connect A_1 and A_2 by a single edge. Since the number of edges which are incident to D is at most 2^{n+2}, while the total number of edges of the graph is greater than 2^{2n}, we have that $\pi_n(D) = o(1)$. The analysis above can be extended to this example with minor adaptations (although a rigorous analysis of this example is somewhat more tedious).*

Chapter 4

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

In this Chapter we prove the results from Section 2.2. Hence as in Section 2.2 we concentrate on the continuous-time setup (and write X_t and P_x instead of X_t^c and H_x).

4.0.4 An overview of our approach

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 π 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 $A \subsetneq \Omega$, $x \in \Omega$, $t > 0$ and $\delta \in (0,1)$. Let $\mathscr{P}_{A,\delta}$ be the collection of all distributions μ on Ω, satisfying that $\mu(A) \geq \pi(A) + \delta\pi(A^c)$. 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 π in $\mathscr{P}_{A,\delta}$.

Proposition 4.0.10. *Let (Ω, P, π) be some finite irreducible Markov chain. Let $A \subsetneq \Omega$. Denote $\nu_{A,\delta} := \delta\pi_A + (1-\delta)\pi$. Then for all $\delta \in (0,1)$,*

$$\begin{aligned} \min_{\mu \in \mathscr{P}_{A,\delta}} \|\mu - \pi\|_{2,\pi} &= \|\nu_{A,\delta} - \pi\|_{2,\pi} = \delta\sqrt{\pi(A^c)/\pi(A)}. \\ \min_{\mu \in \mathscr{P}_{A,\delta}} D(\mu\|\pi) &= D(\nu_{A,\delta}\|\pi) = u(\pi(A),\delta), \end{aligned} \tag{4.0.1}$$

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. $\square$

Proposition 4.0.10 motivates the definitions in (2.2.2). We argue that (4.0.1), implies both (2.2.3)-(2.2.4) by making suitable substitutes for δ in (4.0.1). For (2.2.3) substitute $\delta = \frac{1}{2}\sqrt{\frac{\pi(A)}{\pi(A^c)}}$ in the first line of (4.0.1). 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 δ. As mentioned above, this implies that $\mathrm{P}_x^t \in \mathscr{P}_{A,\delta'}$ for some $\delta' > \delta$ and so by (4.0.1) and the choice of δ, we have that $\|\mathrm{P}_x^t - \pi\|_{2,\pi} > 1/2$. For (2.2.4), it is not hard to verify that for some $C', C > 0$, we have that $u(x, \min(\frac{C'}{|\log x|}, 1)) \geq 1/2$ and $x + \frac{C'}{|\log x|}(1-x) \leq \frac{C}{|\log x|}$ for all $x \leq 1/2$. Substituting $\delta = \frac{C'}{|\log \pi(A)|}$ in the second line of (4.0.1) implies (2.2.4) in a similar manner to the above derivation of (2.2.3).

We now explain the idea behind the proof of the upper bound on τ_2 from (2.2.3). Let $x \in \Omega$. Denote $t := \rho_x + 8\kappa + 6t_{\mathrm{rel}}\log 2$. By Theorem 2.2.2 it suffices to bound $d_{2,x}(t)$.

Step 1: Show that (Proposition 4.1.3)

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

Step 2: Show that (Lemma 4.2.1) for $A_s := \{y : h_t(x,y) \geq (s+1)\}$

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

$\implies$ By Poincaré ineq. (1.1.2) it suffices that $s\pi(A_s) \leq 2s^{-3/2}$ for $s \geq M$ (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,

$$\begin{aligned} s\pi(A_s) &\leq H_t(x, A_s) = \mathrm{P}_x[T_{B_s^c} > t, X_t \in A_s] + \mathrm{P}_x[T_{B_s^c} \leq t, X_t \in A_s] \\ &\leq \mathrm{P}_x[T_{B_s^c} > t] + \sup_{y \notin B_s, k \geq 0} H_k(y, A_s) \leq \pi(B_s)^3 + \frac{s}{2}\pi(A_s). \end{aligned} \tag{4.0.2}$$

Step 4: If $\pi(B_s) \leq s^{-1/2}$, 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 1.2.2 indeed $\pi(B_s) \leq s^{-1/2}$. Since $e\sqrt{s}|\log \pi(A_s)|\pi(A_s) \leq s\pi(A_s)/2$, unless $\pi(A_s) \leq Ce^{-\sqrt{s}}$, repeating the reasoning in (4.0.2) with the new choice of B_s concludes the proof. $\square$

The proof of Theorem 2.2.2 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 $T_{B_s^c} \leq t$ (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.

4.1 Bounding escape probabilities using κ

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 μ_A on A, known as the ***quasi-stationary distribution*** of A, satisfying that the escape time from A w.r.t. Q, starting from μ_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 μ_A to denote the quasi-stationary distribution of A. Recall that we denote π conditioned on A by π_A.

Using the spectral decomposition of Q_A (e.g. [6, Lemma 3.8] or [3, (3.87)]) it follows that

$$\forall A \subsetneq \Omega,\ s \geq 0, \quad \mathbb{P}_{\pi_A}[T_{A^c} > s] \leq \mathbb{P}_{\mu_A}[T_{A^c} > s] = \mu_A S_t^A 1_A = e^{-\lambda(A)s}\mu_A(A) = e^{-\lambda(A)s}. \tag{4.1.1}$$

Proposition 4.1.1. *For reversible chains*

$$\kappa \leq 3\rho. \tag{4.1.2}$$

Proof: Let $A \in \mathrm{Con}_{1/2}$ be such that $\kappa = t_{\mathrm{rel}}(A)|\log \pi(A)|$. By (4.1.1) $\mathrm{P}_{\mu_A}[T_{A^c} > \kappa/3] = \pi(A)^{1/3}$. Since $a^{1/3} \geq a + \frac{1}{2}\sqrt{a(1-a)}$, for all $0 \leq a \leq 1/2$, we get that

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

Definition 4.1.2. $\bar{\rho} := \max_x \bar{\rho}_x$ *and* $\bar{\rho}_{\mathrm{Ent}} := \max_x \rho_{\mathrm{Ent},x}$, *where*

$$\begin{aligned} \bar{\rho}_x &:= \min\{t : \mathrm{P}_x[T_{A^c} > t] \leq \pi(A)^3 \textit{ for all } A \in \mathrm{Con}_{1/2}\}. \\ \bar{\rho}_{\mathrm{Ent},x} &:= \min\{t : \mathrm{P}_x[T_{A^c} > t] \leq \frac{1}{16e^2[\log(e^{3/2}/\pi(A))]^3} \textit{ for all } A \in \mathrm{Con}_{1/2}\} \end{aligned} \tag{4.1.3}$$

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

$$\rho \leq \bar{\rho} \leq 8\rho \quad \text{and} \quad \rho_{\mathrm{Ent}} \leq \bar{\rho}_{\mathrm{Ent}} \leq C'\rho_{\mathrm{Ent}}, \tag{4.1.4}$$

for some absolute constant $C' > 0$. The following proposition refines the inequality $\bar{\rho} \leq 8\rho$.

Proposition 4.1.3. *For every reversible chain,*

$$\forall x \in \Omega, \quad \bar{\rho}_x \leq \rho_x + s, \quad \textit{where} \quad s := 8\kappa + 2t_{\mathrm{rel}} \log 8. \tag{4.1.5}$$

Proof: Let $x \in \Omega$ and $A \in \mathrm{Con}_{1/2}$. By (2.2.6) $2t_{\mathrm{rel}} \geq \max_{B \in \mathrm{Con}_{1/2}} t_{\mathrm{rel}}(B)$ and so by (4.1.1)

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

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 ρ_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 4.1.4. *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 P_x and 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

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

$\square$

4.2 An upper bound on τ_2 and τ_{Ent} - Proof of Theorem 2.2.1

4.2.1 A hitting times characterization of mixing in L_2

In this section we prove the following theorem.

Theorem 4.2.1. *For every finite irreducible reversible Markov chain (Ω, P, π) we have that*

$$\forall x, \quad \rho_x \leq \tau_{2,x} \leq \bar{\rho}_x + 4et_{\mathrm{rel}} \leq \rho_x + 8\kappa + (4e + 6\log 2)t_{\mathrm{rel}}. \tag{4.2.1}$$

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

$$\rho \leq \tau_2 \leq (8 + 12e/\log 2)\rho. \tag{4.2.2}$$

Lemma 4.2.1. *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 4.2.1: Let $x \in \Omega$. The inequality $\rho_x \leq \tau_{2,x}$ follows from (4.0.1). Set $t := \bar{\rho}_x$. As above, denote $A_s := \{y : h_t(x,y) \geq s+1\}$. By Fact 1.1.2 and Lemma 4.2.1 it suffices to show that

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

Let $g_s(y) := \sup_k H_k(y, A_s)/\pi(A_s)$. By Lemma 1.2.2 $\|g\|_1 \leq e|\log \pi(A_s)|$. Let

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

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

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

Also, by the definition of B_s we clearly have that

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

Since by the definition of A_s

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

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

$$(s+1)\pi(A_s) \leq 2e\sqrt{s+1}\pi(A_s)|\log \pi(A_s)|,$$

which simplifies as follows

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

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

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

In conclusion, as desired,

$$\int_{e^e}^\infty 2s\pi(A_s)ds \leq \int_{e^e}^\infty \max(2se^{-\sqrt{s+1}+2e}, \frac{4}{(s+1)^{3/2}})ds \leq e^{8e}/4 - e^{2e}. \quad \square$$

4.2.2 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 (2.2.1) and (4.1.3). Recall that by (4.1.4), $\rho_{\mathrm{Ent}} \le \bar{\rho}_{\mathrm{Ent}} \le C\rho_{\mathrm{Ent}}$. The following theorem refines (2.2.4) from Theorem 2.2.1.

Theorem 4.2.2. *Let (Ω, P, π) be a finite irreducible reversible Markov chain. Then*

$$\forall x, \quad \rho_{x,\mathrm{Ent}} \le \tau_{\mathrm{Ent},x} \le \bar{\rho}_{x,\mathrm{Ent}} + 14t_{\mathrm{rel}}. \tag{4.2.3}$$

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

$$\rho_{\mathrm{Ent}} \le \tau_{\mathrm{Ent}} \le C_1\rho_{\mathrm{Ent}}. \tag{4.2.4}$$

4.2.3 Proof of Theorem 4.2.2

Proof of Theorem 4.2.2: Let $x \in \Omega$. The inequality $\rho_{x,\mathrm{Ent}} \le \tau_{\mathrm{Ent},x}$ follows from (4.0.1). The inequality $\tau_{\mathrm{Ent}} \le C_1\rho_{\mathrm{Ent}}$ follows from (4.2.3) and (4.1.4), in conjunction with the fact that (under reversibility) $ct_{\mathrm{rel}} \le \rho_{\mathrm{Ent}}$ for some absolute constant $c > 0$ (c.f. [6, (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 the triangle inequality (which holds for D, by Jensen's inequality applied to each y separately) and (1.1.7)

$$D(\mathrm{P}_x^{r+r'}||\pi) \le \delta D(\mu_{r'}||\pi) + (1-\delta)D(\nu_{r'}||\pi) \le \delta D(\mu_{r'}||\pi) + (1-\delta)\log(1 + \|\nu_{r'} - \pi\|_{2,\pi}^2). \tag{4.2.5}$$

By (1.1.9)

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

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

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

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

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

$$\text{it suffices to show that} \quad \mathbb{E}_\pi[g\log g] \le 1/10 < 1/2 - 1/e - 2e^{-4}. \tag{4.2.6}$$

Similarly to the proof of Theorem 4.2.1, let

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

Then

$$\mathbb{E}_\pi[g \log g] \leq \int_0^\infty \pi(\{y : g(y)\log g(y) > s\})ds = \int_1^\infty (1 + \log s)\pi(A_s)ds. \tag{4.2.7}$$

Note that $(e^{10} + s)\pi(y) \leq H_r(x, y)$ for every $y \in A_s$. Hence as in the proof of Theorem 4.2.1

$$(e^{10} + s)\pi(A_s) \leq H_r(x, A_s) \leq \mathrm{P}_x[T_{B_s^c} > r] + \mathbb{E}_x[X_r \in A_s \mid T_{B_s^c} \leq t]. \tag{4.2.8}$$

By the definition of B_s and the Markov property,

$$\mathbb{E}_x[X_r \in A_s \mid T_{B_s^c} \leq r] \leq \sup_{y \notin B_s, \ell \geq 0} H_\ell(y, A_s) \leq \sqrt{s + e^{10}}\pi(A_s)|\log \pi(A_s)|. \tag{4.2.9}$$

By Lemma 1.2.2 $\pi(B_s) \leq e/\sqrt{s + e^{10}} \leq 1/2$ and hence by the definition of r,

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

As in the proof of Theorem 4.2.1, it follows that for all $s \geq 1$, $(s+e^{10})\pi(A_s) \leq \frac{2}{2e^2(1+\log(s+e^{10}))^3}$, as otherwise by (4.2.8) $(s + e^{10})\pi(A_s) < 2\mathbb{E}_x[X_r \in A_s \mid T_{B_s^c} \leq t]$, which by (4.2.9) implies that

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

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

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

which yields that $\int_1^\infty (1 + \log s)\pi(A_s)ds \leq \int_{1+e^{10}}^\infty \frac{e^{-2}ds}{s(1+\log s)^2} = \frac{e^{-2}}{1+\log(1+e^{10})} < e^{-2}/11$. This concludes the proof using (4.2.6) and (4.2.7).

4.2.4 The discrete-time and averaged chain analogs

The proof of the lower bounds in Theorem 2.2.3 (that is, the first inequality in each of the four equations) is identical to that of Theorem 2.2.1 (namely, these are "naive" bounds that can be proven using the same ideas as in § 4.0.4). The proofs of the upper bounds require the following minor adaptations:

(i) In the definition of the sets A_s we need to replace h_t with k_t and a_t, resp..

(ii) In the applications of Starr's inequality (in the proof of Lemma 1.2.2) one has to work with the discrete-time version, and thus pick up a multiplicative factor of 2 (which is a non-issue).

(iii) One has to replace the Poincaré inequality with the discrete and averaged analogs: For all $\mu \in \mathscr{P}(\Omega)$, $M \ge 1$ and $k \in \mathbb{Z}_+$ we have that

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

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

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

The second inequality in the second line follows from elementary calculus. We now explain why the first inequality in the second row 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,

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

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. $\square$

4.3 A characterization of the Log-Sobolev constant - Proof of Theorem 2.2.2

The following result ([16, Theorem 3.10]) will allow us to bound t_{LS} from above.

Fact 4.3.1. *Let (Ω, P, π) 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*

$$t_{\mathrm{LS}} \le \frac{2q}{q-2} r_q + 2t_{\mathrm{rel}}(1 + \frac{q}{q-2}\log M_q). \tag{4.3.1}$$

Fix some $0 < \varepsilon < 1/2$ and $A \in \mathrm{Con}_{2^{-1/\varepsilon}}$. Assume that $\mathrm{P}_\pi[T_{A^c} > t] \ge 2\pi(A)^{1+\varepsilon}$. Recall that π_A denotes π conditioned on 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 $t_{\mathrm{ht}}/C \le t_{\mathrm{LS}} \le C t_{\mathrm{ht}}$.*

Trivially, $t_{\rm ht} = \min\{t : \mathrm{P}_{\pi_A}[T_{A^c} > t] \le \pi(A)^{1/4}$ for all $A \in \mathrm{Con}_{1/2}\}$. Note that if we replace π_A by the quasi-stationary distribution of A, μ_A, then by (4.1.1) we get precisely $\kappa/4$. This explains why also κ can be interpreted as a hitting time version of hypercontractivity. We note that the above question resembles Open problem 4.38 in [3], which asks whether for reversible chains $t_{\rm rel} \le C \max_{A\in \mathrm{Con}_{1/2}} \mathbb{E}_{\pi_A}[T_{A^c}]$, where indeed [3, Lemma 4.39] $t_{\rm rel} \le \max_{A\in \mathrm{Con}_{1/2}} \mathbb{E}_{\mu_A}[T_{A^c}]$ (the formulation in [3] is slightly different, but it is equivalent to our formulation).

4.3.1 Proof of Theorem 2.2.2

Proof of Theorem 2.2.2: As mentioned in the introduction, it is known that $\kappa \le t_{\rm LS}$. Denote $r := \frac{1}{2}\kappa$. Note that P and $Q = (P + P^*)/2$ have the same $t_{\rm rel}$ and $t_{\rm LS}$. Thus we may work with $S_t = e^{-t(I-Q)}$ instead of H_t. By (4.3.1) 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

$$\text{it suffices to show that} \quad \int_6^\infty 4s^3\pi(A_s)ds \le 16 \le 7^4 - 6^4. \tag{4.3.2}$$

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 (4.3.2) in order to conclude the proof it suffices to show that $\int_6^\infty 4s^3\pi(F_s)ds \le 16$. By Starr's maximal inequality (1.2.1) we know that $\int_0^\infty 4s\pi(B_s)ds = \|f^*\|_2^2 \le 4\|f\|_2^2 = 4$. Thus in order to show that $\int_6^\infty 4s^3\pi(F_s)ds \le 16$, and conclude the proof, it suffices to show that for all $s \ge 6$ we have that $\pi(F_s) \le 4s^{-2}\pi(B_s)$.

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. [6, 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 $\pi(F_s) \le 4s^{-2}\pi(B_s)$. $\square$

4.4 An application to robustness of mixing

Proof of Corollary 2.3.1 It is not hard to verify that Theorem 2.2.1 is still valid in the above setup (this can be formally deduced from Theorem 2.2.1 via the representation of the generator appearing in the paragraph following Corollary 2.3.1). Hence it suffices to verify that (2.3.1) is valid if we replace τ_2 and $\tilde{\tau}_2$ by ρ and $\tilde{\rho}$, resp. (where $\tilde{\rho}$ is the parameter ρ of the chain $(\tilde{X}_t)$). 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) 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 $\tilde{T}_A/M \leqslant_{\mathrm{st}} T_A \leqslant_{\mathrm{st}} M\tilde{T}_A$, where $\leqslant_{\mathrm{st}}$ denotes stochastic domination. Since for all A we have that $\tilde{\pi}(A)/M \leq \pi(A) \leq M\tilde{\pi}(A)$, by the submultiplicity property

$$\forall t \geq 0, m \in \mathbb{Z}_+ \text{ and } A \subset \Omega, \quad \max_x \mathrm{P}_x[T_A > tm] \leq (\max_x \mathrm{P}_x[T_A > t])^m,$$

this implies that $\tilde{\rho}/(C_1 M \log M) \leq \rho \leq (C_1 M \log M)\tilde{\rho}$, as desired. $\square$

Chapter 5

Trees

In this chapter we prove the results from Section 2.3. The results in this chapter are valid both in the discrete-time lazy and in the continuous-time setup.

5.1 Total variation cutoff for trees - Proof of Theorem 2.3.1

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. [36, 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 $\{x, y\} \in E$ or $x = y$) is reversible. Hence we may assume that $\mathcal{T}$ is equipped with edge weights $(c_e)_{e \in E}$. Following [48], 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 $\ell(u) = (u_0 = u, u_1, \ldots, u_k = o)$. 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 $\pi(\mathcal{T}_y) \geq \delta$ and let

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

(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 2.2.2, $\alpha := \sup_{A \in \mathrm{Con}_{1/2}} \alpha(A) \geq c_{\mathrm{LS}}$. Let $D_\beta = D_{\beta,x}$ be the connected component of x in $\mathcal{T} \setminus \{x_\beta\}$. For a leaf x we denote

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

Recall that for any $\emptyset \neq A \subset V$, we write π_A for the distribution of π conditioned on A, $\pi_A(\cdot) := \frac{\pi(\cdot)1_{\cdot \in A}}{\pi(A)}$.

A key observation is that starting from the central vertex o, the chain mixes rapidly (this follows implicitly from the following analysis). Let T_o denote the hitting time of the central vertex. We define the mixing parameter $\tau_o(\varepsilon)$ for $\varepsilon \in (0,1)$ by

$$\tau_o(\varepsilon) := \min\{t : \mathrm{P}_x[T_o > t] \leqslant \varepsilon \ \forall x \in \Omega\}.$$

We show that up to terms of the order of the relaxation-time (which are negligible under the product condition) $\tau_o(\cdot)$ approximates $\mathrm{hit}_{1/2}(\cdot)$ and then using Proposition 2.1.1, the question of cutoff is reduced to showing concentration for the hitting time of the central vertex. Below we make this precise.

Lemma 5.1.1. *Denote* $s_\delta := \lceil 4t_{\mathrm{rel}} |\log(4\delta/9)|\rceil$. *Then*

$$\tau_o(\varepsilon) \leqslant \mathrm{hit}_{1/2}(\varepsilon) \leqslant \tau_o(\varepsilon - \delta) + s_\delta, \ \textit{for every } 0 < \delta < \varepsilon < 1. \tag{5.1.1}$$

Proof. First observe that by the definition of central vertex, for any $x \in V$, $x \neq o$ there exists a set A with $\pi(A) \geqslant \frac{1}{2}$ such that the chain starting at x cannot hit A without first hitting o. Indeed, we can take A to be the union of $\{o\}$ and all components of $T \setminus \{o\}$ not containing x. The first inequality in (5.1.1) follows trivially from this.

To establish the other inequality, fix $A \subseteq V$ with $\pi(A) \geqslant \frac{1}{2}$, $x \in V$ and some $0 < \delta < \varepsilon < 1$. It follows using Markov property and the definition of $\tau_o(\varepsilon - \delta)$ that

$$\mathrm{P}_x[T_A > \tau_o(\varepsilon-\delta) + s_\delta] \leqslant \mathrm{P}_x[T_o > \tau_o(\varepsilon - \delta)] + \mathrm{P}_o[T_A > s_\delta] \leq \varepsilon - \delta + \mathrm{P}_o[T_A > s_\delta].$$

Hence it suffices to show that $\mathrm{P}_o[T_A > s_\delta] \leq \delta$. If $o \in A$ then $\mathrm{P}_o[T_A > s_\delta] = 0$, so without loss of generality assume $o \notin A$. It is easy to see that we can partition $T \setminus \{o\} = T_1 \cup T_2$ such that both T_1 and T_2 are unions of components of $T \setminus \{o\}$ and $\pi(T_1), \pi(T_2) \leqslant 2/3$. For $i = 1,2$, let $A_i := A \cap T_i$ and without loss of generality let us assume $\pi(A_1) \geqslant \frac{1}{4}$. Let $B = T_2 \cup \{o\}$. Clearly the chain started at any $x \in B$ must hit o before hitting A_1. Hence

$$\mathrm{P}_o[T_A > s_\delta] \leqslant \mathrm{P}_o[T_{A_1} > s_\delta] \leqslant \mathrm{P}_{\pi_B}[T_{A_1} > s_\delta] \leqslant \pi(B)^{-1}\mathrm{P}_\pi[T_{A_1} > s_\delta] \tag{5.1.2}$$

Using $\pi(A_1) \geqslant \frac{1}{4}$, $\pi(B) \geqslant \frac{1}{3}$ it follows from (3.1.8) that $\pi(B)^{-1}\mathrm{P}_\pi[T_{A_1} > s_\delta] \leq \delta$. $\square$

In light of Lemma 5.1.1 and Proposition 2.1.5, in order to conclude the proof of Theorem 2.3.1 it suffices to show that $\tau_o(\varepsilon) - \tau_o(1-\varepsilon) \leq C\sqrt{t_{\mathrm{rel}} t_{\mathrm{mix}} \log(1/\varepsilon)} + Ct_{\mathrm{rel}}|\log \varepsilon|$, for all $\varepsilon \in (0, 1/4]$. This follows from Proposition 5.1.5 below.

Let $x, y \in V$ be such that $y \prec x$. 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 P_x, we have that $T_y = \sum_{i=1}^k \xi_i$ and that $\xi_1, \dots, \xi_k$ are independent. It is thus beneficial to investigate the marginal distributions of the ξ's, which we now do.

For any set $A \subset \Omega$, we define $\psi_{A^c} \in \mathscr{P}(A^c)$ as $\psi_{A^c}(y) := \mathrm{P}_{\pi_A}[X_1 = y \mid X_1 \in A^c]$. For $A \subset \Omega$, we denote $T_A^+ := \inf\{t \geq 1 : X_t \in A\}$ and $\Phi(A) := \frac{\sum_{a \in A, b \in A^c} \pi(a)P(a,b)}{\pi(A)} = \mathrm{P}_{\pi_A}[X_1 \notin A]$. Note that

$$\pi(A)\Phi(A) = \sum_{a \in A, b \in A^c} \pi(a)P(a,b) = \sum_{a \in A, b \in A^c} \pi(b)P(b,a) = \pi(A^c)\Phi(A^c). \tag{5.1.3}$$

This is true even without reversibility, since the second term (resp. third term) is the asymptotic frequency of transitions from A to A^c (resp. from A^c to A).

Proposition 5.1.2. *Let (Ω, P, π) be a finite irreducible Markov chain. Let $A \subsetneq \Omega$ be non-empty. Denote the complement of A by B. To avoid ambiguity, let $T_A^{\mathrm{ct}} := \inf\{t : X_t^{\mathrm{c}} \in A\}$.*

$$\begin{aligned} &\mathrm{P}_{\pi_B}[T_A = t]/\Phi(B) = \mathrm{P}_{\psi_B}[T_A \geq t], \text{ for all } t \geq 1. \\ &\frac{d}{dt}\mathrm{H}_{\pi_B}[T_A^{\mathrm{ct}} \leq t]/\Phi(B) = \mathrm{H}_{\psi_B}[T_A^{\mathrm{ct}} > t], \text{ for all } t \geq 0. \end{aligned} \tag{5.1.4}$$

Consequently, $\mathbb{E}_{\psi_B}[T_A] = \frac{1}{\Phi(B)} = \mathbb{E}_{\psi_B}[T_A^{\mathrm{ct}}]$, *and*

$$\begin{aligned} &\mathbb{E}_{\psi_B}[T_A^2] = \mathbb{E}_{\psi_B}[T_A]\left(2\mathbb{E}_{\pi_B}[T_A] - 1\right) \leq 2\mathbb{E}_{\psi_B}[T_A]t_{\mathrm{rel}}(B) \leq \frac{2\mathbb{E}_{\psi_B}[T_A]t_{\mathrm{rel}}}{\pi(A)}. \\ &\mathbb{E}_{\psi_B}[(T_A^{\mathrm{ct}})^2] = 2\mathbb{E}_{\psi_B}[T_A^{\mathrm{ct}}]\mathbb{E}_{\pi_B}[T_A^{\mathrm{ct}}] \leq 2\mathbb{E}_{\psi_B}[T_A^{\mathrm{ct}}]t_{\mathrm{rel}}(B) \leq \frac{2\mathbb{E}_{\psi_B}[T_A^{\mathrm{ct}}]t_{\mathrm{rel}}}{\pi(A)}. \end{aligned} \tag{5.1.5}$$

Moreover, assuming reversibility and in the discrete-time case also laziness, the law of T_A (resp. T_A^{ct}) under P_{ψ_B} (resp. H_{ψ_B}) is a mixture of Geometric (resp. Exponential) distributions.

Proof. We first note that by Wald's equation we have that $\mathbb{E}_{\psi_B}[T_A] = \mathbb{E}_{\psi_B}[T_A^{\mathrm{ct}}]$ and $\mathbb{E}_{\pi_B}[T_A] = \mathbb{E}_{\pi_B}[T_A^{\mathrm{ct}}]$. The inequalities in (5.1.5) follow from the estimate $\mathbb{E}_{\pi_B}[T_A] \leq t_{\mathrm{rel}}(B) \leq t_{\mathrm{rel}}/\pi(A)$.

Summing (5.1.12) over t yields $\mathbb{E}_{\psi_B}[T_A] = \frac{1}{\Phi(B)}$. Multiplying both sides of the first row of (5.1.12) by $2t-1$ and summing over t yields the equality in the first row of (5.1.5). For the equality in the second row, multiply both sides of the second row of (5.1.12) by $2t$ and integrate from 0 to ∞.

We now prove (5.1.12). Let $t \geq 1$. As $\{T_A = t\} = \{X_0 \notin A, \ldots, X_{t-1} \notin A, X_t \in A\}, \{T_A^+ = t+1\} = \{X_1 \notin A, \ldots, X_t \notin A, X_{t+1} \in A\}$ we have by stationarity that $\mathrm{P}_\pi[T_A = t] = \mathrm{P}_\pi[T_A^+ = t+1]$. Thus

$$\begin{aligned} &\pi(B)\mathrm{P}_{\pi_B}[T_A = t] = \mathrm{P}_\pi[T_A = t] = \mathrm{P}_\pi[T_A^+ = t+1] = \mathrm{P}_\pi[X_1 \notin A, \ldots, X_t \notin A, X_{t+1} \in A] \\ &= \mathrm{P}_\pi[X_1 \notin A, \ldots, X_t \notin A] - \mathrm{P}_\pi[X_1 \notin A, \ldots, X_t \notin A, X_{t+1} \notin A] \\ &= \mathrm{P}_\pi[X_1 \notin A, \ldots, X_t \notin A] - \mathrm{P}_\pi[X_0 \notin A, \ldots, X_t \notin A] = \mathrm{P}_\pi[X_0 \in A, X_1 \notin A, \ldots, X_t \notin A] \\ &= \pi(A)\Phi(A)\mathrm{P}_{\psi_B}[X_0 \notin A, \ldots, X_{t-1} \notin A] = \pi(A)\Phi(A)\mathrm{P}_{\psi_B}[T_A \geq t], \end{aligned}$$

which by (5.1.3) implies the first row in (5.1.12). The continuous time follows from the discrete-time case by a standard argument as follows:

$$\frac{d}{dt}\mathrm{H}_{\pi_B}[T_A^{\mathrm{ct}} \leq t] = \sum_{k \geq 0} \mathbb{P}[\mathrm{Pois}(t) = k]\mathrm{P}_{\pi_B}[T_A = k+1]$$

$$= \Phi(B)\sum_{k \geq 0} \mathbb{P}[\mathrm{Pois}(t) = k]\mathrm{P}_{\psi_B}[T_A \geq k+1] = \Phi(B)\mathrm{H}_{\psi_B}[T_A^{\mathrm{ct}} > t],$$

where we used $\mathrm{H}_{\pi_B}[T_A^{\mathrm{ct}} \leq t] = \sum_{k \geq 0} \mathbb{P}[\mathrm{Pois}(t) = k+1]\mathrm{P}_{\pi_B}[T_A = k+1]$, $\frac{d}{dt}\mathbb{P}[\mathrm{Pois}(t) = k+1] = \mathbb{P}[\mathrm{Pois}(t) = k]$. Finally, the claim about the law of T_A under P_{ψ_B} and H_{ψ_B} follows from (5.1.12) together with the fact that the claim is true under P_{π_B} and H_{π_B} (see § 3.1.2). $\square$

Corollary 5.1.3. *If y is the parent of x then the law of T_y under P_x (resp. H_x) is a mixture of Geometric (resp. Exponential) distributions whose maximal mean is at most $t_{\mathrm{rel}}(\mathcal{T}_x) \leq 2t_{\mathrm{rel}}$. Moreover, $\mathbb{E}_x[T_y] = \frac{1}{\Phi(\mathcal{T}_x)}$ and $\mathbb{E}_x[T_y^2] \leq 2t_{\mathrm{rel}}(\mathcal{T}_x)\,\mathbb{E}_x[T_y] \leq 4t_{\mathrm{rel}}\,\mathbb{E}_x[T_y]$ (both holding both in discrete and continuous time).*

Lemma 5.1.4. *If (V,P,π) is a chain on a (weighted) tree (T,o) then (both in discrete-time and in continuous-time)*

$$\mathbb{E}_x[T_o] \le 4t_{\mathrm{mix}}, \text{ for all } x \in V. \tag{5.1.6}$$

Proof. Fix some $x \in V$. Let C_x be the component of $T \backslash \{o\}$ containing x. Denote $B := V \backslash C_x$. Consider $\tau_B := \inf\{k \in \mathbb{N} : X_{kt_{\mathrm{mix}}} \in B\}$. Clearly, $T_o \le \tau_B t_{\mathrm{mix}}$. Since $\pi(B) \geqslant 1/2$, by the Markov property and the definition of the total variation distance, the distribution of τ_B is stochastically dominated by the Geometric distribution with parameter $1/2 - 1/4 = 1/4$. Hence $\mathbb{E}_x[T_o] = \mathbb{E}_x[T_B] \le t_{\mathrm{mix}}\mathbb{E}_x[\tau_B] \le 4t_{\mathrm{mix}}$. □

Proposition 5.1.5. *Let $x, y \in V$ be such that $y \prec x$. Let $(v_0 = x, v_1, \dots, v_k = y)$ be the path from x to y. Define $\xi_i := T_{v_i} - T_{v_{i-1}}$, so that $T_y = \sum_{i=1}^{k} \xi_i$ (starting from x). Denote the connected component of x in $\mathcal{T} \setminus \{y\}$ by $A_{x,y}$. Denote $\sigma_{x,y} := \sqrt{\sum_{i=1}^{k} \mathbb{E}_x[\xi_i^2]}$. Then*

$$\mathrm{Var}_x[T_y] \le \sigma_{x,y}^2 \le 2t_{\mathrm{rel}}(A_{x,y})\mathbb{E}_x[T_y] \le 8t_{\mathrm{rel}}(A_{x,y})t_{\mathrm{mix}} \le 16t_{\mathrm{rel}}t_{\mathrm{mix}}. \tag{5.1.7}$$

For all $t \ge 0$ we have that

$$\mathrm{P}_x[T_y < \mathbb{E}_x[T_y] - t] \le \exp(-\frac{t^2}{2\sigma_{x,y}^2}) \le \exp(-\frac{t^2}{32t_{\mathrm{rel}}t_{\mathrm{mix}}}) \tag{5.1.8}$$

Moreover, if $y = x_\delta$ for some $\delta \le 1/2$ (one can always find such a δ) then

$$\forall t \in [0, 2\mathbb{E}_x[T_{x_\delta}]], \quad \mathrm{P}_x[T_{x_\delta} \ge \mathbb{E}_x[T_{x_\delta}] + t] \le \exp(-\frac{t^2\lambda(D_\delta)}{8\mathbb{E}_x[T_{x_\delta}]}) \le \exp(-\frac{t^2}{64t_{\mathrm{mix}}t_{\mathrm{rel}}}). \tag{5.1.9}$$

$$\forall t \ge 2\mathbb{E}_x[T_{x_\delta}], \quad \mathrm{P}_x[T_{x_\delta} \ge \mathbb{E}_x[T_{x_\delta}] + t] \le \exp[-\lambda(D_\delta)t/4]. \tag{5.1.10}$$

Proof. We first note that (5.1.7) is an immediate consequence of Corollary 5.1.3 and (5.1.6), using independence. The first inequality of (5.1.8) holds by [42] (it only uses the fact that the ξ's are non-negative). The second inequality in (5.1.8) follows from (5.1.7). We now prove (5.1.9)-(5.1.10). We focus on the continuous-time setup.

Claim 5.1.6. *Fix some leaf x and $\delta \in (0, 1/4]$. Let D_δ 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*

$$\mathbb{E}_y[e^{\beta T_z}] \le 1 + \mathbb{E}_y[T_z]\beta(1 + 2\beta/\lambda(D_\delta)) \le e^{\mathbb{E}_y[T_z]\beta(1+2\beta/\lambda(D_\delta))}. \tag{5.1.11}$$

Proof of (5.1.11)*:* Let $\Phi(\mathcal{T}_y) := \frac{\pi(y)P(y,z)}{\pi(\mathcal{T}_y)}$. Let g be the density functions of T_z started from $\pi_{\mathcal{T}_y}$, resp.. By Proposition 5.1.2

$$\forall t \ge 0, \quad g(t) = \Phi(\mathcal{T}_y)\mathrm{P}_y[T_z > t], \quad \text{and hence} \quad \Phi(\mathcal{T}_y)\mathbb{E}_y[T_z] = 1. \tag{5.1.12}$$

Recall that by (4.1.1) 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 (5.1.12)

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

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

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

We now return to conclude the proofs of (5.1.9)-(5.1.10). 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_δ be $(y_1 = x, \dots, y_r = x_\delta)$. 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 (5.1.11), independence and our choice of β

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

The proof of (5.1.10) is analogous, now with the choice $\beta = \lambda(D_\delta)/2$. $\square$

5.2 Robustness of the L_∞ mixing time for trees - Proof of Theorem 2.3.3

Let us first describe the skeleton of the argument of the proof of Theorem 2.3.3.

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

Lemma 5.2.1. *There exists an absolute constant $C > 0$ so that if $y \prec x$ then*

$$\tau_{2,y} \le \tau_{2,x} + C(t_{\mathrm{LS}} + \sqrt{t_{\mathrm{rel}}\tau_1}). \tag{5.2.1}$$

Step 2: Show that for a leaf x we can replace (in (4.1.4)) $\bar{\rho}_x$ (defined in (4.1.3)) with

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

Proposition 5.2.2. *Let x be a leaf. Let $0 < \delta \le 1/4$ 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*

$$\mathrm{P}_x[T_{A^c} > b_x + 3\kappa + 10t_{\mathrm{rel}}] \le \mathrm{P}_x[T_{x_\delta} > b_x] + \mathrm{P}_{x_\delta}[T_{\bar{A}} > 3\kappa + 10t_{\mathrm{rel}}] < \delta^3/2. \tag{5.2.2}$$

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

Proposition 5.2.3. *There exists some $C > 0$ so that for a leaf x and $\delta \in (0, 1/4]$,*

$$b_x(\delta) \le \mathbb{E}_x[T_{x_\delta}] + \max\left(\frac{32}{\alpha_x(\delta)}, 8\sqrt{\mathbb{E}_x[T_{x_\delta}]/\alpha_x(\delta)}\right) \le \tau_1 + C\max(\kappa, \sqrt{\kappa\tau_1}). \tag{5.2.3}$$

The second inequality follows from the first using the fact that $\mathbb{E}_x[T_{x_\delta}] \le \tau_1 + C_5\sqrt{\tau_1 t_{\mathrm{rel}}}$ [6, Corollary 5.5].

Step 4: Similar reasoning as in the proof of (4.1.5) yields that (c.f. [6, Corollary 3.4])

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

By (5.2.1)-(5.2.3) in conjunction with (4.2.1) and (2.2.7) we have that

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

To conclude the proof of Theorem 2.3.3 we now prove Lemma 5.2.1 and Propositions 5.2.2-5.2.3.

Proof of Lemma 5.2.1: Let $y \prec x$. Let $s := \tau_{2,y} - Mt_{\mathrm{LS}}$ for some constant $M > 0$ to be determined later. We may assume $s > 16\sqrt{t_{\mathrm{rel}}\tau_1}$ as otherwise there is nothing to prove. Form the proofs of (4.1.5) and (4.2.1) it follows that we can choose M so that for some $A \in \mathrm{Con}_{1/100}$

$$\mathrm{P}_y[T_{A^c} > s] > 2\pi(A) + \sqrt{\pi(A)\pi(A^c)}. \tag{5.2.4}$$

We leave this as an exercise (the main issue is moving from an estimate for some $B \in \mathrm{Con}_{1/2}$ to one for some $A \in \mathrm{Con}_{1/100}$. This can be done using similar reasoning as in the proof of (4.1.5), c.f. [6, Corollary 3.4]).

Denote the connected component of x in $\mathcal{T} \setminus \{y\}$ by A. By (5.1.7) $\mathrm{Var}_x[T_y] \le 16t_{\mathrm{rel}}\tau_1$. By Chebyshev inequality

$$\mathrm{P}_x[|T_y - \mathbb{E}_x[T_y]| > 8\sqrt{t_{\mathrm{rel}}\tau_1}] \le 1/4. \tag{5.2.5}$$

Let $s' := \max(\mathbb{E}_x[T_y] - 8\sqrt{t_{\mathrm{rel}}\tau_1}, 0)$. By (5.2.4), (5.2.5), $s > 16\sqrt{t_{\mathrm{rel}}\tau_1}$ and the Markov property

$$\mathrm{P}_x[X_{s+s'} \in A] \ge \mathrm{P}_x[|T_y - \mathbb{E}_x[T_y]| \le 8\sqrt{t_{\mathrm{rel}}\tau_1}] \times \mathrm{P}_y[T_{A^c} > s] > \pi(A) + \frac{1}{2}\sqrt{\pi(A)\pi(A^c)}.$$

The proof is concluded using (4.0.1) (in the notation from (4.0.1), $\mathrm{P}_x^{s+s'} \in \mathscr{P}_{A,\delta}$ for some $\delta > \frac{1}{2}\sqrt{\pi(A)/\pi(A^c)}$ and thus $\|\mathrm{P}_x^{s+s'} - \pi\|_{2,\pi} \ge \delta\sqrt{\pi(A^c)/\pi(A)} > 1/2$). $\square$

Proof of Proposition 5.2.2: 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 (4.1.1), the def. of b_x and the fact that $\pi_{V\setminus\bar{A}}(\mathcal{T}_{x_\delta}) > 1/2$ (as $\pi(V \setminus \bar{A}) < 2\delta < 2\pi(\mathcal{T}_{x_\delta})$)

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

Proof of Proposition 5.2.3: By [6, Corollary 5.5] we have that $\mathbb{E}_x[T_{x_\delta}] \le \tau_1 + C_5\sqrt{\tau_1 t_{\mathrm{rel}}}$. If $t_1 := 8\sqrt{\mathbb{E}_x[T_{x_\delta}]/\alpha_x(\delta)} \le 2\mathbb{E}_x[T_{x_\delta}]$ then by (5.1.9) $\mathrm{P}_x[T_{x_\delta} \ge \mathbb{E}_x[T_{x_\delta}] + t_1] \le \delta^3/4$. Otherwise, $t_2 := 32/\alpha_x(\delta) > 2\mathbb{E}_x[T_{x_\delta}]$, and by (5.1.10), $\mathrm{P}_x[T_{x_\delta} \ge \mathbb{E}_x[T_{x_\delta}] + t_2] \le \delta^3/4$. $\square$

5.3 Semi BD chains - Proof of Theorem 2.3.2

In this section we prove Theorem 2.3.2 and establish that product condition is sufficient for cutoff for a sequence of (δ, r)-SBD chains. Although we think of δ as being bounded away from 0, and of r as a constant integer, it will be clear that our analysis remains valid as long as δ does not tend to 0, nor does r to infinity, too rapidly in terms of some functions of $t_{\mathrm{rel}}/t_{\mathrm{mix}}$.

Throughout the section, we use $C_1, C_2, \ldots$ to describe positive constants which depend only on δ and r. Consider a (δ, r)-SBD chain on $([n], P, \pi)$. We call a state $i \in [n]$ a central-vertex if $\pi([i-1]) \vee \pi([n] \setminus [i]) \le 1/2$. As opposed to the setting of Section 5.1, the sets $[i-1]$ and $[n] \setminus [i]$ need not be connected components of $[n] \setminus \{i\}$ w.r.t. the chain, in the sense that it might be possible for the chain to get from $[i-1]$ to $[n] \setminus [i]$ without first hitting i (skipping over i). We pick a central-vertex o and call it the root.

Divide $[n]$ into $m := \lceil n/r \rceil$ consecutive disjoint intervals, $I_1, \ldots, I_m$ each of size r, apart from perhaps I_m. We call each such interval a *block*. Denote by $I_{\tilde{o}}$ the unique block such that the root o belongs to it. Since we are assuming the product condition (and thus $t_{\mathrm{mix}}^{(n)} \to \infty$), in the setup of Theorem 2.3.2 we can assume that $n \gg r$ and hence $I_{\tilde{o}} \neq [n]$ (it is not hard to show that $t_{\mathrm{mix}}^{(n)}$ can be bounded from above in terms of n and δ, and thus we must have $n \to \infty$). Observe the following. Consider some $v \notin I_{\tilde{o}}$ and $u \in I_{\tilde{o}}$ such that $|u - v| = 1$. Then by the definition of a (δ, r)-SBD chain, we have for all $v' \in I_{\tilde{o}}$, $\pi(v) \geqslant \delta^r \pi(v')$. Hence $\pi(I_{\tilde{o}}) \leqslant \frac{r}{r+\delta^r}$. For the rest of this section let us fix $\alpha = \alpha(\delta, r) := 1 - \frac{\delta^r}{4(r+\delta^r)}$.

Recall that in Section 5.1 we exploited the tree structure to reduce the problem of showing cutoff to showing the concentration of the hitting time of the central vertex by showing that starting from the central vertex the chain hits any large set (with large probability) quickly. We argue similarly in this case with the central vertex replaced by the central block. First we need the following lemma.

Lemma 5.3.1. *In the above setup, let $I := \{v, v+1, \ldots, v+r-1\} \subset [n]$. Let $\mu \in \mathscr{P}(I)$. Then*

$$\mathbb{E}_\mu[T_A] \le \max_{y \in I} \mathbb{E}_y[T_A] \le \delta^{-r} \min_{x \in I} \mathbb{E}_x[T_A], \text{ for any } A \subset [n] \setminus I. \tag{5.3.1}$$

Consequently, for any $i \in I$ and $A \subset [v-1]$ (resp. $A \subset [n] \setminus [v+r-1]$) we have that

$$\mathbb{E}_i[T_A] \le \delta^{-r} \mathbb{E}_{\pi_{[n] \setminus [v-1]}}[T_A], \text{ (resp. } \mathbb{E}_i[T_A] \le \delta^{-r} \mathbb{E}_{\pi_{[v+r-1]}}[T_A]). \tag{5.3.2}$$

Proof. We first note that (5.3.2) follows from (5.3.1). Indeed, by condition (i) of the definition of a (δ, r)-SBD chain, if $A \subset [v-1]$ (resp. $A \subset [n] \setminus [v+r-1]$), then under $\mathrm{P}_{\pi_{[n] \setminus [v-1]}}$ (resp. under $\mathrm{P}_{\pi_{[v+r-1]}}$), $T_I \le T_A$. Thus (5.3.2) follows from (5.3.1) by averaging over X_{T_I}. We now prove (5.3.1).

Fix some A such that $A \subset [n] \setminus I$. Fix some distinct $x, y \in I$. Let B_1 be the event that $T_y \le T_A$. One way in which B_1 can occur is that the chain would move from x to y in $|y - x|$ steps such that $|X_k - X_{k-1}| = 1$ for all $1 \le k \le |y - x|$. Denote the last event by B_2. Then

$$\mathbb{E}_x[T_A] \ge \mathbb{E}_x[T_A 1_{B_2}] \ge \mathrm{P}_x[B_2] \mathbb{E}_y[T_A] \ge \delta^r \mathbb{E}_y[T_A].$$

Minimizing over x yields that for any $y \in I$ we have that $\mathbb{E}_y[T_A] \le \delta^{-r} \min_{x \in I} \mathbb{E}_x[T_A]$, from which (5.3.1) follows easily. $\square$

The next proposition reduces the question of proving cutoff for a sequence of (δ,r)-SBD chains under the product condition to that of showing an appropriate concentration for the hitting time of the central block. The argument is analogous to the one in Section 5.1 and hence we only provide a sketch to avoid repetitions. As in Section 5.1, for $\varepsilon \in (0,1)$ let $\tau_C(\varepsilon) := \min\{t : \mathrm{P}_x[T_{I_{\tilde{o}}} > t] \leqslant \varepsilon, \forall x \in [n]\}$. As always, we write $\tau_C^{(k)}(\cdot)$ to indicate that this parameter is taken w.r.t. the k-th chain in a sequence of (δ,r)-SBD chains.

Proposition 5.3.2. *Let $([n_k], P_k, \pi_k)$ be a sequence of (δ,r)-SBD chains. Suppose that there exist constants C_ε for $\varepsilon \in (0,\frac{1}{8})$ and a some sequence $(w_k)_{k=1}^{\infty}$ of numbers such that for all k*

$$\tau_C^{(k)}(\varepsilon) - \tau_C^{(k)}(1-\varepsilon) \leqslant C_\varepsilon w_k \text{ for all } 0 < \varepsilon < 1/8. \tag{5.3.3}$$

Then there exist some constants $C'_\varepsilon, C''_\varepsilon$ such that for all k and all $\varepsilon \in (0,1/8)$

$$\mathrm{hit}_{1/2}^{(k)}(3\varepsilon/2) - \mathrm{hit}_{1/2}^{(k)}(1-3\varepsilon/2) \leqslant C_\varepsilon w_k + C'_\varepsilon t_{\mathrm{rel}}^{(k)} \text{ and} \tag{5.3.4}$$

$$t_{\mathrm{mix}}^{(k)}(2\varepsilon) - t_{\mathrm{mix}}^{(k)}(1-2\varepsilon) \leqslant C_\varepsilon w_k + C''_\varepsilon t_{\mathrm{rel}}^{(k)}. \tag{5.3.5}$$

Proof. Observe that (5.3.5) follows from (5.3.4) using Proposition 2.1.5 and Corollary 3.1.3. To deduce (5.3.4) from (5.3.3), we argue as in Lemma 5.1.1 using Lemma 5.3.3 below, which shows that starting from any vertex in $I_{\tilde{o}}$ the chain hits any set of π-measure at least α in time proportional to t_{rel} with large probability. We omit the details. □

Lemma 5.3.3. *Let $v \in I_{\tilde{o}}$. Let $D \subset [n]$ be such that $\pi(D) \geq \frac{1+\alpha}{2}$. Then $\mathbb{E}_v[T_D] \leq C(\alpha)\delta^{-r} t_{\mathrm{rel}}$ for some constant $C(\alpha)$. In particular, by Markov inequality $\mathrm{hit}_{\alpha,v}(\varepsilon) \leq \varepsilon^{-1} C(\alpha)\delta^{-r} t_{\mathrm{rel}}$.*

Proof. Let $I_{\tilde{o}} = \{v_1, v_1+1, \dots, v_2\}$. Set $A_1 = [v_1 - 1]$ and $A_2 = [n] \setminus [v_2]$. For $i = 1,2$, let $D_i = D \cap A_i$. Using the definition of α, without loss of generality let $\pi(D_1) \geqslant \frac{1-\alpha}{2}$. Set $A = A_2 \cup I_{\tilde{o}}$. By (5.3.2) and the fact that $\pi(A) \geqslant \frac{1}{2}$

$$\mathbb{E}_v[T_D] \leqslant \mathbb{E}_v[T_{D_1}] \leqslant \delta^{-r}\mathbb{E}_{\pi_A}[T_{D_1}] \leq 2\delta^{-r}\mathbb{E}_{\pi}[T_{D_1}].$$

The proof is completed using Lemma 3.1.4. □

Observe that using Cheybeshev inequality it follows that (5.3.3) holds for some constants C_ε if we take $w_n = \max_{x \in [n]} \sqrt{\mathrm{Var}_x[T_{I_{\tilde{o}}}]}$. Theorem 2.3.2 therefore follows at once from Proposition 5.3.2 provided we establish $\mathrm{Var}_x[T_{I_{\tilde{o}}}] \leqslant C_1 \mathbb{E}_x[T_{I_{\tilde{o}}}] t_{\mathrm{rel}}$ for all $x \notin I_{\tilde{o}}$ (e.g. by (2.0.1)). This is what we shall do.

Observe that the root induces a partial order on the blocks. We say that $I_j \prec I_k$ if I_j is a block between I_k and $I_{\tilde{o}}$. For $j \in [m]$, $I_j \neq I_{\tilde{o}}$, we define the parent block of I_j in the obvious manner and denote its index by f_j. We define

$$T(j) := T_{I_j} \text{ and } \bar{\tau}_j := T(f_j) - T(j).$$

Consider some arbitrary $x \in [n]$ and $j \in [m] \setminus \{\tilde{o}\}$ such that $x \in I_j$. Denote $k := |j - \tilde{o}|$, $j_0 = j$ and $j_{i+1} = f_j$ for all $0 < i < k$. Observe that starting from x we have that $T_{I_{\tilde{o}}} = \sum_{\ell=0}^{k-1} \bar{\tau}_{j_\ell}$. As mentioned above, we will bound $\mathrm{Var}_x[\sum_{\ell=0}^{k-1} \bar{\tau}_{j_\ell}]$. As opposed to the situation in Section 5.1, the terms in the sum are no longer independent. We now show that the correlation between them decays exponentially (Lemma 5.3.5) and that for all ℓ we have that $\mathrm{Var}_x[\bar{\tau}_{j_\ell}] \leq C_2 t_{\mathrm{rel}} \mathbb{E}_x[\bar{\tau}_{j_\ell}]$ (Lemma 5.3.6). The desired variance estimate, $\mathrm{Var}_x[\sum_{\ell=0}^{k-1} \bar{\tau}_{j_\ell}] \leqslant C_1 \mathbb{E}_x[T_{I_{\tilde{o}}}] t_{\mathrm{rel}}$, follows by combining these two lemmata. We omit the details.

Lemma 5.3.4. *In the above setup, let $v \in [m] \setminus \{\tilde{o}\}$. Let $(v_0 = v, v_1, \dots, v_s)$ be indices of consecutive blocks. Let $\mu_1, \mu_2 \in \mathscr{P}(I_v)$. Let $k \in [s]$. Denote by $\nu_k^{(j)}$ $(j = 1, 2)$ the hitting distribution of I_{v_k} starting from initial distribution μ_j (i.e. $\nu_k^{(j)}(z) := \mathrm{P}_{\mu_j}[X_{T(v_k)} = z]$). Then $\|\nu_k^{(1)} - \nu_k^{(2)}\|_{\mathrm{TV}} \le (1 - \delta^r)^k$.*

Proof. It suffices to prove the case $k = 1$ as the general case follows by induction using the Markov property. The case $k = 1$ follows from coupling the chain with the two different starting distributions in a way that with probability at least δ^r there exists some $z_v \in I_v$ such that both chains hit z_v before hitting I_{f_v} (not necessarily at the same time) and from that moment on (which may occur at different times for the two chains) they follow the same trajectory. The fact that the hitting time of z_v (and thus also of I_{f_v}) might be different for the two chains makes no difference (as regardless of the hitting time of I_{f_v} w.r.t. the two chains, this coupling is also a coupling of $(\nu_1^{(1)}, \nu_1^{(2)})$, having the desired property). We now describe this coupling more precisely.

Let $\mu_1, \mu_2 \in \mathscr{P}(I_v)$. Let $(X_t^{(1)})_{t\ge 0}$ and $(X_t^{(2)})_{t\ge 0}$ be independent Markov chains where $(X_t^{(i)})_{t\ge 0}$ is distributed as the chain (Ω, P, π) with initial distribution μ_i $(i = 1, 2)$ as follows. Pick $v_1 \sim \mu_1$ and $v_2 \sim \mu_2$ respectively. Run the chain $X_t^{(1)}$ started from v_1. Let $R := \min\{t : X_t^{(1)} = X_0^{(2)}\}$ and $L_i := \min\{t : X_t^{(i)} \in I_{f_v}\}$. Let S denote the event: $R \le L_1$. On S, define $Y_t^{(1)}$ by setting $Y_t^{(1)} = X_t^{(1)}$ for $t < R$ and $Y_{R+t}^{(1)} = X_t^{(2)}$ for any $t \ge 0$, and on S^c, define $Y_t^{(1)} = X_t^{(1)}$ for all t. Denote the joint law of $(Y_t^{(1)}, X_t^{(2)})$ by P_{μ_1,μ_2} and of $(X_t^{(1)}, Y_t^{(1)}, X_t^{(2)})$ by $\mathrm{P}_{\mu_1,\mu_1,\mu_2}$. Clearly P_{μ_1,μ_2} is a coupling with the correct marginals and $\mathrm{P}_{\mu_1,\mu_1,\mu_2}[S] \ge \delta^r$. Let L_2 be as above and $\bar{L}_1 := \min\{t : Y_t^{(1)} \in I_{f_v}\}$. Note that on S, $X_{L_2}^{(2)} = Y_{\bar{L}_1}^{(1)}$. Hence for any $D \subset I_{v_k}$,

$$\begin{aligned}
&\nu_1^{(1)}(D) - \nu_1^{(2)}(D) = \mathrm{P}_{\mu_1,\mu_2}[Y_{\bar{L}_1}^{(1)} \in D] - \mathrm{P}_{\mu_1,\mu_2}[X_{L_2}^{(2)} \in D] \\
&\le \mathrm{P}_{\mu_1,\mu_2}[Y_{\bar{L}_1}^{(1)} \in D, X_{L_2}^{(2)} \notin D] \le 1 - \mathrm{P}_{\mu_1,\mu_1,\mu_2}[S] \le 1 - \delta^r.
\end{aligned}$$

□

Lemma 5.3.5. *In the setup of Lemma 5.3.4, let $0 \le i < j < s$. Let $\mu \in \mathscr{P}(I_v)$. Write $\tau_i := \bar{\tau}_{v_i}$ and $\tau_j := \bar{\tau}_{v_j}$. Then*

$$\mathbb{E}_\mu[\tau_i\tau_j] \le \mathbb{E}_\mu[\tau_i]\mathbb{E}_\mu[\tau_j]\bigg(1 + (1 - \delta^r)^{j-i-1}\delta^{-r}\bigg).$$

Proof. Let μ_{i+1} and μ_j be the hitting distributions of $I_{v_{i+1}}$ and of I_{v_j}, respectively, of the chain with initial distribution μ. By the Markov property, the hitting distribution of I_{v_j} for the chain started with initial distribution either μ or μ_{i+1} is simply μ_j. Again by the Markov property $\mathbb{E}_\mu[\tau_j] = \mathbb{E}_{\mu_{i+1}}[\tau_j] = \mathbb{E}_{\mu_j}[\tau_j]$ and

$$\mathbb{E}_\mu[\tau_i\tau_j] \le \mathbb{E}_\mu[\tau_i] \max_{y \in I_{v_{i+1}}} \mathbb{E}_y[\tau_j]. \tag{5.3.6}$$

Let $y^* \in I_{v_{i+1}}$ be the state achieving the maximum in the RHS above. By Lemma 5.3.4 we can couple successfully the hitting distribution of I_{v_j} (and thus also τ_j) of the chain started from y^* with that of the chain starting from initial distribution μ_{i+1} with probability at least $1-(1-\delta^r)^{j-i-1}$. If the coupling fails, then by (5.3.1) we can upper bound the conditional expectation of τ_j by $\delta^{-r}\,\mathbb{E}_\mu[\tau_j]$. Hence

$$\mathbb{E}_{y^*}[\tau_j] \le \mathbb{E}_{\mu_j}[\tau_j] + (1-\delta)^{j-i-1}\delta^{-r}\mathbb{E}_\mu[\tau_j] = \mathbb{E}_\mu[\tau_j]\Big(1+(1-\delta^r)^{j-i-1}\delta^{-r}\Big).$$

The assertion of the lemma follows by plugging this estimate in (5.3.6). $\square$

Lemma 5.3.6. *Let $j \in [m]\setminus\{o\}$. Let $\nu \in \mathscr{P}([n])$. Then there exists some $C_1, C_2 > 0$ (depending on δ and r) such that $\mathbb{E}_\nu[\bar{\tau}_j^2] \le C_1 t_{\mathrm{rel}}\Phi(I_j)^{-1} \le C_2 t_{\mathrm{rel}}\mathbb{E}_\nu[\bar{\tau}_j]$.*

Proof. Let $\mu := \psi_{I_j}$. By condition (i) in the definition of a (δ, r)-SBD chain, $\mu \in \mathscr{P}(I_j)$. By (5.1.5), $\mathbb{E}_\mu[\bar{\tau}_j^2] \le C_3 t_{\mathrm{rel}}\Phi(I_j)^{-1} \le C_4 t_{\mathrm{rel}}\mathbb{E}_\mu[\bar{\tau}_j]$ for constants C_3, C_4 depending on δ and r. The proof is concluded using the same reasoning as in the proof of (5.3.1) to argue that the first and second moments of $\bar{\tau}_j$ w.r.t. different initial distributions can change by at most some multiplicative constant. $\square$

Chapter 6

The power of averaging at two consecutive time steps: Proof of a mixing conjecture by Aldous and Fill

In this chapter we prove the results from Section 2.4. Accordingly, the proofs in this chapter are taken from [28].

6.1 Proof of Proposition 2.4.5.

In this section we prove Proposition 2.4.5. As noted in the introduction, Theorem 2.4.1 follows as a particular case of Proposition 2.4.5 and Theorem 2.4.2, in turn, follows in a trivial manner from Theorem 2.4.1. We now state large deviation estimates for the Poisson and Binomial distributions. For a proof see e.g. [5, Appendix A].

Fact 6.1.1. *Let $Y \sim \mathrm{Pois}(\mu)$ and let $Y' \sim \mathrm{Bin}(t, 1/2)$. Then for every $\varepsilon > 0$ we have that*

$$\begin{aligned}&\mathbb{P}[Y \le \mu(1-\varepsilon)] \le e^{-\varepsilon^2\mu/2}, \quad \mathbb{P}[Y \ge \mu(1+\varepsilon)] \le \exp\left(-\frac{\varepsilon^2\mu}{2(1+\varepsilon/3)}\right),\\ &\mathbb{P}[Y' \le t(1-\varepsilon)/2] = \mathbb{P}[Y' \ge t(1+\varepsilon)/2] \le e^{-\varepsilon^2 t/4}.\end{aligned} \tag{6.1.1}$$

Let $(N(t))_{t\ge 0}$ and $(M(t))_{t\ge 0}$ be homogeneous Poisson processes with rate 1, such that $(N(t))_{t\ge 0}$, $(M(t))_{t\ge 0}$ and $(X_t)_{t=0}^{\infty}$ are mutually independent. We define

$$N_{\mathrm{L}}(t) := N(t) + M(t) \text{ and } S(\ell) := \sum_{k=1}^{\ell} q_k \sim \mathrm{Bin}(\ell, 1/2),$$

$$\text{where} \quad q_k := 1_{N(T_k) > N(T_{k-1})} \quad \text{and} \quad T_k := \inf\{t : N_{\mathrm{L}}(t) = k\}.$$

Let (Ω, P, π) be a Markov chain. The ***natural coupling*** of $(X_t^{\mathrm{ct}})_{t\ge 0}$, $(X_t)_{t\in\mathbb{Z}_+}$ and $(X_t^{\mathrm{L}})_{t\in\mathbb{Z}_+}$ is defined by setting $X_t^{\mathrm{L}} := X_{S(t)}$ and $X_t^{\mathrm{ct}} := X_{N(t)} = X_{N_{\mathrm{L}}(t)}^{\mathrm{L}}$.

As can be seen from the natural coupling, $H_t = \sum_{k\ge 0} \frac{e^{-2t}(2t)^k}{k!} P_{\mathrm{L}}^k$. This also follows from Poisson thinning. Also, in the natural coupling $(X_t^{\mathrm{L}})_{t\in\mathbb{Z}_+}$ and $(N_{\mathrm{L}}(t))_{t\ge 0}$ are independent. The same holds for $(X_t)_{t\in\mathbb{Z}_+}$ and $(S(t))_{t=0}^{\infty}$. The next lemma follows from the natural coupling by a standard construction (cf. the proofs of Proposition 4.7 and Theorem 5.2 in [36]).

Lemma 6.1.2. *Let (Ω, P, π) be a finite irreducible Markov chain. Let $\mu \in \mathscr{P}(\Omega)$ and $t \in \mathbb{R}_+$.*

(1) There exists a coupling $((Y_i^{\mathrm{L}})_{i\in\mathbb{Z}_+}, (Z_i^{\mathrm{L},\pi})_{i\in\mathbb{Z}_+}, \xi_t)$, such that $(Y_i^{\mathrm{L}})_{i\in\mathbb{Z}_+} \sim \mathrm{P}_{\mathrm{L},\mu}$, $(Z_i^{\mathrm{L},\pi})_{i\in\mathbb{Z}_+} \sim \mathrm{P}_{\mathrm{L},\pi}$ (the law of the stationary lazy chain), $\xi_t \sim \mathrm{Pois}(t)$ in which ξ_t and $(Y_i^{\mathrm{L}})_{i\in\mathbb{Z}_+}$ are independent and

$$\mathrm{P}[Y_{\xi_t}^{\mathrm{L}} = Z_0^{\mathrm{L},\pi}] = \mathrm{P}[Y_{\xi_t+i}^{\mathrm{L}} = Z_i^{\mathrm{L},\pi} \text{ for all } i \geq 0] = 1 - d_{\mathrm{c}}(t/2, \mu).$$

(2) There exists a coupling $((Y_i)_{i\in\mathbb{Z}_+}, (Z_i^{\pi})_{i\in\mathbb{Z}_+}, \xi_t')$, such that $(Y_i)_{i\in\mathbb{Z}_+} \sim \mathrm{P}_{\mu}$, $(Z_i^{\pi})_{i\in\mathbb{Z}_+} \sim \mathrm{P}_{\pi}$ (the law of the stationary chain), $\xi_t' \sim \mathrm{Bin}(2t, 1/2)$ in which ξ_t' and $(Y_i)_{i\in\mathbb{Z}_+}$ are independent and

$$\mathrm{P}[Y_{\xi_t'} = Z_0^{\pi}] = \mathrm{P}[Y_{\xi_t'+i} = Z_i^{\pi} \text{ for all } i \geq 0] = 1 - d_{\mathrm{L}}(2t, \mu).$$

Definition 6.1.3. *Let $t \geq 1$ and $s \in [2, e^t]$. Denote*

$$\begin{aligned} r = r_{s,t} &:= 2\sqrt{2t \log s}, \\ J = J_{s,t} &:= [(t-r) \vee 0, t+r], \\ m = m_{s,t} &:= \lceil r(\sqrt{s}+1) \rceil. \end{aligned} \tag{6.1.2}$$

In the notation of Lemma 6.1.2 (with both couplings taken w.r.t. time t), let G be the event that $Y_{\xi_t+i}^{\mathrm{L}} = Z_i^{\mathrm{L},\pi}$ for all $i \geq 0$ and that $\xi_t \in J$. Similarly, let G' be the event that $Y_{\xi_t'+i} = Z_i^{\pi}$ for all $i \geq 0$ and that $\xi_t' \in J$.

In the following proposition, we only care about (6.1.5) and (6.1.8) (which imply (2.4.6) and (2.4.7), respectively; i.e. the below proposition implies Proposition 2.4.5). We present the rest of the equations in order to make it clear that (6.1.8) is obtained in an analogous manner to (6.1.5). Thus, we shall only prove part (i) of Proposition 6.1.4.

In the notation of Definition 6.1.3, the term $d_{\mathrm{c}}(t/2, \mu) + 2/s^2$ appearing in (6.1.3) and (6.1.5) (resp. $d_{\mathrm{L}}(2t, \mu) + 2/s^2$ appearing in (6.1.6) and (6.1.8)) is an upper bound on the probability that G (resp. G') fails (where the term $2/s^2$ is obtained via Fact 6.1.1).

Proposition 6.1.4. *Let (Ω, P, π) be a finite irreducible reversible chain. Let $\mu \in \mathscr{P}(\Omega)$. Let $B \subset \Omega$. Let $t \geq 1$ and $2 \leq s \leq e^t$. In the notation of Definition 6.1.3,*

(i) Let $\eta^{\mathrm{L}} := 1_{Y_{t+m}^{\mathrm{L}} \in B}$ and $\eta^{\mathrm{L},\pi} := 1_{Z_m \in B}^{\mathrm{L},\pi}$ (where $m = \lceil r(\sqrt{s}+1) \rceil$, $r = 2\sqrt{2t \log s}$). Then

$$\pi(B) - \mathrm{P}_{\mu}[X_{t+m}^{\mathrm{L}} \in B] \leq \frac{2}{s^2} + d_{\mathrm{c}}(t/2, \mu) + \mathbb{E}[(\eta^{\mathrm{L},\pi} - \eta^{\mathrm{L}}) 1_G]. \tag{6.1.3}$$

$$|\mathbb{E}[(\eta^{\mathrm{L}} - \eta^{\mathrm{L},\pi}) 1_G]|^2 \leq s^{-1} \mathbb{E}_{\pi} \left[\sup_{i \geq r\sqrt{s}} i^2 |\triangle P_{\mathrm{L}}^i 1_B|^2 \right] \leq C s^{-1} \mathrm{Var}_{\pi} 1_B \leq \frac{C}{s}. \tag{6.1.4}$$

Consequently,

$$d_{\mathrm{L}}(t+m, \mu) \leq d_{\mathrm{c}}(t/2, \mu) + \frac{2}{s^2} + \sqrt{C/s}. \tag{6.1.5}$$

(ii) Let $w \sim \mathrm{Bernoulli}(1/2)$ *be independent of* $((Y_i)_{i\in\mathbb{Z}_+}, (Z_i^\pi)_{i\in\mathbb{Z}_+}, \xi_t')$*. Let* $\eta = 1_{Y_{t+m+w}\in B}$ *and* $\eta^\pi = 1_{Z^\pi_{m+w}\in B}$*. Then*

$$\pi(B) - \mathrm{P}_\mu[X^{\mathrm{ave}}_{t+m} \in B] \le \frac{2}{s^2} + d_{\mathrm{L}}(2t,\mu) + \mathbb{E}[(\eta^\pi - \eta)1_{G'}]. \tag{6.1.6}$$

$$|\mathbb{E}[(\eta - \eta^\pi)1_{G'}]|^2 \le s^{-1}\mathbb{E}_\pi\left[\sup_{i\ge r\sqrt{s}} i^2 |\triangle A_i 1_B|^2\right] \le Cs^{-1}\mathrm{Var}_\pi 1_B \le \frac{C}{s}. \tag{6.1.7}$$

Consequently,

$$d_{\mathrm{ave}}(t+m,\mu) \le d_{\mathrm{L}}(2t,\mu) + \frac{2}{s^2} + \sqrt{C/s}. \tag{6.1.8}$$

Proof. We first note that (6.1.5) follows from (6.1.3)-(6.1.4) by maximizing over $B \subset \Omega$. We now prove (6.1.3). Let $B \subset \Omega$. Let r, J and m be as in Definition 6.1.3. By Fact 6.1.1 and our assumption that $s \le e^t$ (which implies that $\varepsilon := r/t = 2\sqrt{2t^{-1}\log s} \le 3$),

$$\mathrm{P}[\xi_t \notin J] \le \mathrm{P}[\xi_t < t-r] + \mathrm{P}[\xi_t > t+r] \le e^{-t\varepsilon^2/2} + e^{-\frac{t\varepsilon^2/2}{(1+\varepsilon/3)}} = e^{-4\log s} + e^{-\frac{4\log s}{(1+\varepsilon/3)}} \le 2s^{-2}.$$

Hence $1 - \mathrm{P}[G] \le d_{\mathrm{c}}(t/2,\mu) + 2s^{-2}$, which implies (6.1.3), as

$$\begin{aligned}\pi(B) - \mathrm{P}_\mu[X^{\mathrm{L}}_{t+m} \in B] &\le 1 - \mathrm{P}[G] + \mathrm{P}[G \cap \{Z^{\mathrm{L},\pi}_m \in B\}] - \mathrm{P}[G \cap \{Y^{\mathrm{L}}_{t+m} \in B\}] \\ &= 1 - \mathrm{P}[G] + \mathbb{E}[(\eta^{\mathrm{L},\pi} - \eta^{\mathrm{L}})1_G].\end{aligned}$$

We now argue that for every $x \in \Omega$,

$$|\mathbb{E}[\eta - \eta^{\mathrm{L},\pi} \mid G, Y^{\mathrm{L}}_{\xi_t} = x = Z^{\mathrm{L},\pi}_0]| \le \sqrt{\frac{1}{s}} \sup_{i\ge r\sqrt{s}} i|\triangle P^i_{\mathrm{L}} 1_B(x)|. \tag{6.1.9}$$

Indeed, for every $x \in \Omega$ and $j \in J$

$$\mathbb{E}[\eta^{\mathrm{L}} \mid \xi_t = j, Y^{\mathrm{L}}_j = x = Z^{\mathrm{L},\pi}_0] = P^{t+m-j}_{\mathrm{L}} 1_B(x),$$

$$\mathbb{E}[\eta^{\mathrm{L},\pi} \mid \xi_t = j, Y^{\mathrm{L}}_j = x = Z^{\mathrm{L},\pi}_0] = P^{m}_{\mathrm{L}} 1_B(x).$$

Thus by the triangle inequality

$$\begin{aligned}&|\mathbb{E}[\eta^{\mathrm{L}} - \eta^{\mathrm{L},\pi} \mid \xi_t = j, Y^{\mathrm{L}}_j = x = Z^{\mathrm{L},\pi}_0]| = |P^{t+m-j}_{\mathrm{L}} 1_B(x) - P^m_{\mathrm{L}} 1_B(x)| \\ &\le 1_{j\ne t} \sum_{i=(t+m-j)\wedge m}^{[(t+m-j)\vee m]-1} |\triangle P^i_{\mathrm{L}} 1_B(x)|.\end{aligned} \tag{6.1.10}$$

Note that by the definition of $m = \lceil r(\sqrt{s}+1)\rceil$ and $J = [(t-r)\vee 0, t+r]$, for every $j \in J$ we have that $|j-t| \le r$ and $(t+m-j)\wedge m \ge r\sqrt{s}$. Whence,

$$\begin{aligned}1_{j\ne t} \sum_{i=(t+m-j)\wedge m}^{[(t+m-j)\vee m]-1} |\triangle P^i_{\mathrm{L}} 1_B(x)| &\le r \sup_{i\ge r\sqrt{s}} |\triangle P^i_{\mathrm{L}} 1_B(x)| \\ &\le \frac{r}{r\sqrt{s}} \sup_{i\ge r\sqrt{s}} i|\triangle P^i_{\mathrm{L}} 1_B(x)| = \sqrt{s^{-1}} \sup_{i\ge r\sqrt{s}} i|\triangle P^i_{\mathrm{L}} 1_B(x)|.\end{aligned}$$

Plugging this estimate in (6.1.10) and averaging over j yields (6.1.9).

Since

$$|\mathbb{E}[(\eta^{\mathrm{L}}-\eta^{\mathrm{L},\pi})1_G]| \le \mathbb{E}[|\mathbb{E}[(\eta^{\mathrm{L}}-\eta^{\mathrm{L},\pi})1_G \mid Z_0^{\mathrm{L},\pi},\xi_t]|],$$

averaging (6.1.9) over $Z_0^{\mathrm{L},\pi}$, and using the fact that $\mathrm{P}[G\cap\{Y_{\xi_t}^{\mathrm{L}}=x=Z_0^{\mathrm{L},\pi}\}]\le\pi(x)$, for all x, together with Jensen's inequality and (1.2.4), we get that

$$|\mathbb{E}[(\eta^{\mathrm{L}}-\eta^{\mathrm{L},\pi})1_G]|^2 \le \frac{1}{s}(\mathbb{E}_\pi[\sup_{i\ge r\sqrt{s}} i|\triangle P_{\mathrm{L}}^i 1_B|])^2 \le \frac{1}{s}\mathbb{E}_\pi[\sup_{i\ge r\sqrt{s}} i^2|\triangle P_{\mathrm{L}}^i 1_B|^2] \le Cs^{-1}\mathrm{Var}_\pi 1_B \le C/s.$$

□

6.2 Proof of Proposition 2.4.8

We start the section by stating a standard fact.

Claim 6.2.1. *Let (Ω,P,π) be a finite irreducible chain. Let $\mu\in\mathscr{P}(\Omega)$. Let $(X_t)_{t\in\mathbb{Z}_+}$ be the discrete-time version of the chain. Let T_1,T_2 be independent $\mathbb{Z}_+$ valued random variables independent of $(X_t)_{t\in\mathbb{Z}_+}$. Then $\|\mathrm{P}_\mu[X_{T_1+T_2}\in\cdot]-\pi\|_{\mathrm{TV}}\le\|\mathrm{P}_\mu[X_{T_1}\in\cdot]-\pi\|_{\mathrm{TV}}$, where $\mathrm{P}_\mu[X_{T_1}=y]:=\sum_t \mathrm{P}[T_1=t]\mathrm{P}_\mu^t[X_t=y]$ and $\mathrm{P}_\mu[X_{T_1+T_2}=y]:=\sum_t\mathrm{P}[T_1+T_2=t]\mathrm{P}_\mu^t[X_t=y]$.*

Proof of Proposition 2.4.8: Fix some $t>0$ and $0<s\le\sqrt{t}$. Denote $\tau:=t+s\sqrt{t}$. We first prove (2.4.10). In the notation of the *standard coupling*, $N_{\mathrm{L}}(\tau)\sim\mathrm{Poisson}(2\tau)$ and

$$\mathrm{H}_\mu^\tau-\pi=\sum_{k\ge0}\mathrm{P}[N_{\mathrm{L}}(\tau)=k](\mathrm{P}_{\mathrm{L},\mu}^k-\pi).$$

By the triangle inequality, together with (6.1.1) and the fact that $\|\mathrm{P}_{\mathrm{L},\mu}^k-\pi\|_{\mathrm{TV}}$ is non-decreasing in k and bounded by 1,

$$\begin{aligned}\|\mathrm{H}_\mu^\tau-\pi\|_{\mathrm{TV}} &= \sum_{k\ge0}\mathrm{P}[N_{\mathrm{L}}(\tau)=k]\|\mathrm{P}_{\mathrm{L},\mu}^k-\pi\|_{\mathrm{TV}} \le \mathrm{P}[N_{\mathrm{L}}(\tau)<2t]+\sum_{k\ge2t}\mathrm{P}[N_{\mathrm{L}}(\tau)=k]\|\mathrm{P}_{\mathrm{L},\mu}^k-\pi\|_{\mathrm{TV}}\\ &\le \exp\left[-\frac{4s^2t}{2(2t+2s\sqrt{t})}\right]+\|\mathrm{P}_{\mathrm{L},\mu}^{2t}-\pi\|_{\mathrm{TV}}\le d_{\mathrm{L}}(2t,\mu)+e^{-\delta^2/2},\end{aligned}$$

where in the last inequality we have used the assumption that $s\le\sqrt{t}$. This concludes the proof of (2.4.10). We now prove the first line in (2.4.9). We omit the second line in (2.4.9) as its proof is analogous and as it essentially appears in [48, Lemma 2.3].

As above, denote $\tau:=t+s\sqrt{t}$. Let $Y\sim\mathrm{Poisson}(2\tau)$. Let Z_1 be a random variable whose conditional distribution, given that $Y=n$, is $\mathrm{Bin}((n-1)\vee0,1/2)$. Let η be a Bernoulli random variable with mean $1/2$, independent of Z_1 and Y. Set $Z:=Z_1+\eta 1_{Y>0}$. Let $(X_t)_{t\in\mathbb{Z}_+}$ be the discrete-time version of the chain with $X_0\sim\mu$. Pick Y, Z_1, η and $(X_t)_{t\in\mathbb{Z}_+}$ to be jointly independent. Note that the conditional distribution of Z, given that $Y=n$, is $\mathrm{Bin}(n,1/2)$. Hence by Poisson thinning $Z\sim\mathrm{Poisson}(\tau)$ and so $X_\tau^{\mathrm{ct}}\sim X_Z$.

Let $T:=t+\eta$. Then $Z=(T+Z_1-t)1_{Y>0}$. Thus $Z1_{Z_1\ge t}=(T+(Z_1-t)_+)1_{Z_1\ge t}$, where $a_+:=a\vee0$ (since $Z_1\ge t$ implies that $Y>0$ and $Z_1-t=(Z_1-t)_+$). Consequently,

$$\|\mathrm{P}_\mu(X_Z\in\cdot)-\mathrm{P}_\mu(X_{T+(Z_1-t)_+}\in\cdot)\|_{\mathrm{TV}}\le\|Z-[T+(Z_1-t)_+]\|_{\mathrm{TV}}\le\mathbb{P}[Z_1<t]. \quad (6.2.1)$$

By (6.1.1) and the assumption $s \le \sqrt{t}$,

$$\mathbb{P}[Z_1 < t] \le \mathbb{P}[Z \le t] \le \exp\left[-\frac{s^2 t}{2(t + s\sqrt{t})}\right] \le e^{-s^2/4}. \tag{6.2.2}$$

Finally, by Claim 6.2.1, in conjunction with (6.2.1)-(6.2.2), we get that

$$\begin{aligned}
&d_{\mathrm{c}}(t + s\sqrt{t}, \mu) = \|\mathrm{P}_\mu[X_Z \in \cdot] - \pi\|_{\mathrm{TV}} \le \\
&\|\mathrm{P}_\mu(X_Z \in \cdot) - \mathrm{P}_\mu(X_{T+(Z_1-t)_+} \in \cdot)\|_{\mathrm{TV}} + \|\mathrm{P}_\mu(X_{T+(Z_1-t)_+} \in \cdot) - \pi\|_{\mathrm{TV}} \\
&\le e^{-s^2/4} + \|\mathrm{P}_\mu(X_T \in \cdot) - \pi\|_{\mathrm{TV}} = d_{\mathrm{ave}}(t, \mu) + e^{-s^2/4}. \qquad \square
\end{aligned}$$

6.3 Proof of Theorem 2.4.3

Assume that there is a continuous-time cutoff with a window w_n. Fix some $0 < \varepsilon < 1/4$. By Propositions 2.4.5 (first inequality) and 2.4.12 (second inequality)

$$t_{\mathrm{ave}}^{(n)}(\varepsilon) \le t_{\mathrm{c}}^{(n)}(\varepsilon/2) + C_1(\varepsilon)\sqrt{t_{\mathrm{c}}^{(n)}(\varepsilon/2)} \le t_{\mathrm{c}}^{(n)}(\varepsilon/2) + C_2(\varepsilon) w_n.$$

By Propositions 2.4.8 (first inequality) and 2.4.12 (second inequality) we have that

$$-t_{\mathrm{ave}}^{(n)}(1-\varepsilon) \le -t_{\mathrm{c}}^{(n)}(1-\varepsilon/2) + C_3(\varepsilon)\sqrt{t_{\mathrm{c}}^{(n)}} \le -t_{\mathrm{c}}^{(n)}(1-\varepsilon/2) + C_4(\varepsilon) w_n.$$

Hence

$$t_{\mathrm{ave}}^{(n)}(\varepsilon) - t_{\mathrm{ave}}^{(n)}(1-\varepsilon) \le t_{\mathrm{c}}^{(n)}(\varepsilon/2) - t_{\mathrm{c}}^{(n)}(1-\varepsilon/2) + C_5(\varepsilon) w_n \le C_6(\varepsilon) w_n,$$

as desired. Now assume that the sequence of averaged chains exhibits a cutoff with a window $\tilde{w}_n$. By Proposition 2.4.8

$$t_{\mathrm{c}}^{(n)}(\varepsilon) \le t_{\mathrm{ave}}^{(n)}(\varepsilon/2) + C_7(\varepsilon)\sqrt{t_{\mathrm{c}}^{(n)}}.$$

By Propositions 2.4.5 we have that

$$-t_{\mathrm{c}}^{(n)}(1-\varepsilon) \le -t_{\mathrm{ave}}^{(n)}(1-\varepsilon/2) + C_8(\varepsilon)\sqrt{t_{\mathrm{c}}^{(n)}}.$$

Hence

$$t_{\mathrm{c}}^{(n)}(\varepsilon) - t_{\mathrm{c}}^{(n)}(1-\varepsilon) \le t_{\mathrm{ave}}^{(n)}(\varepsilon/2) - t_{\mathrm{ave}}^{(n)}(1-\varepsilon/2) + C_9(\varepsilon)\sqrt{t_{\mathrm{c}}^{(n)}} \le C_{10}(\varepsilon)(\tilde{w}_n \vee \sqrt{t_{\mathrm{c}}^{(n)}}),$$

as desired. $\square$

6.4 Example

In this section we consider an example which demonstrates that the assertions of Theorems 2.4.1 and 2.4.2 and of Proposition 2.4.5 are in some sense nearly sharp. For notational convenience we suppress the dependence on n in some of the notation below. Throughout this section we write $c_0, c_1, c_2, \ldots$ for positive absolute constants, which are sufficiently small to guarantee that a certain inequality holds.

Equation (6.4.1) below resembles our main results apart from the fact that below the direction of the inequality is reversed, and the exponent of s in the error term of the middle term in (6.4.1) (which decays like an inverse polynomial in s) is larger (compared to the corresponding exponent in Theorem 2.4.1; similarly, the error term on the RHS of (6.4.1) is similar to the one appearing in Theorem 2.4.2, that is to $\psi_{\alpha,C_2}(d_c(t)) - d_c(t)$).

Example 6.4.1. *Fix some $0 < \alpha \leq 1/2$. Let $n \in \mathbb{N}$ be such that $s = s_{n,\alpha} := \lceil n^{0.5+\alpha} \rceil \geq 2$. Consider a nearest-neighbor random walk on the interval $\{0, 1, 2, \ldots, 2n+1\}$, with a bias towards state $2n+1$, whose transition matrix is given by $P(0,1) = 1$, $P(2n+1, 2n) = 1 - \frac{1}{3s}$,*

$$P(i,i) = \begin{cases} \frac{1}{3s} & i \geq 2n - 2s, \\ 0 & \text{otherwise.} \end{cases}$$

Finally, $P(i, i+1) = 3P(i, i-1)$ for all $1 \leq i \leq 2n$ and is given by

$$P(i,i+1) = \begin{cases} \frac{3}{4} - \frac{1}{4s} & i \geq 2n - 2s, \\ 3/4 & \text{otherwise.} \end{cases}$$

By Kolmogorov's cycle condition, this chain is reversible. Both the sequence of the associated continuous-time chains and the sequence of the associated averaged chains exhibit cutoff around time $4n$ with a cutoff window of size $\sqrt{n}$. In particular, prior to time $4n - s$ the worst-case total variation distance from stationarity of both chains tends to 1 as n tends to infinity. Moreover, it is not hard to show that

$$d_c(4n+s) = (1 \pm o(1))\mathrm{H}_0[T_{2n+1} > 4n+s] \leq e^{-c_3 s^2/n} \leq e^{-c_3 n^{2\alpha}}.$$

Conversely, we now show that for $t = 4n + s$, we have that

$$d_{\mathrm{ave}}(t+s) \geq d_c(t) + \frac{c_1}{s} \geq d_c(t) + \frac{c_2}{[\log(1/d_c(t))]^{\frac{1+2\alpha}{4\alpha}}}. \tag{6.4.1}$$

The second inequality in (6.4.1) follows from the choice $s = \lceil n^{\frac{1+2\alpha}{2}} \rceil$ together with $d_c(t) = d_c(4n+s) \leq e^{-c_3 n^{2\alpha}}$. We now prove the first inequality in (6.4.1).

Consider the sets $\mathrm{Even} := \{2i : 0 \leq i \leq n\}$, $\mathrm{Odd} := \{2i+1 : 0 \leq i \leq n\}$ *and* $B := \{i : i \geq 2n - 2s\}$. *It is easy to see that $\pi(B) \geq 1 - 2^{-(2s+1)}$ and that*

$$0 \leq \pi(\mathrm{Even}) - 1/2 \leq \frac{\pi(2n-2s)}{3s} \leq 2^{-2s}. \tag{6.4.2}$$

In order to prove (6.4.1), we shall show that

$$A_{t+s}(0, \mathrm{Even}) \geq \frac{1}{2} + \frac{c_1}{s}. \tag{6.4.3}$$

Let $(X_k)_{k=0}^{\infty}$ be the discrete-time chain with $X_0 = 0$. Note that T_{2n-2s} is even, deterministically. If both X_{4n+2s} and $X_{4n+2s+1}$ lie in B, we define

$$T := \min\{k : T_{2n-2s} \leq k \leq 4n+2s \text{ and } X_\ell \in B \text{ for all } k \leq \ell \leq 4n+2s+1\}.$$

Otherwise, set $T=0$. *It is easy to see that* $\mathrm{P}[T=0] \le Ce^{-c_4 s^2/n}$ *and that*

$$\frac{1}{2}\mathrm{P}_0[X_{4n+2s} \in \mathrm{Even} \mid T=0] + \frac{1}{2}\mathrm{P}_0[X_{4n+2s+1} \in \mathrm{Even} \mid T=0] = 1/2. \tag{6.4.4}$$

Moreover, conditioned on $T>0$, *the number of returns to state* $2n-2s$ *by time* $4n+2s$ *has an exponential tail. Using this fact, it is not hard to verify that*

$$\begin{aligned} &\min_{0 \le r \le 4s} \mathrm{P}[T \textit{ is even} \mid T \neq 0, 4n+2s-T_{2n-2s} = 2r] \ge 1 - \frac{c_5}{s}. \\ &\mathrm{P}[4n+2s-T_{2n-2s} > 8s \mid T \neq 0] \le e^{-c_6 s^2/n}. \end{aligned} \tag{6.4.5}$$

Consider the projected chain $(Y_k)_{k=0}^{4n+2s+1-T}$ *(conditioned on* $T \neq 0$*) on* $\Omega := \{\pm 1\}$ *defined via* $Y_k := 1_{T+k \in \mathrm{Even}} - 1_{T+k \in \mathrm{Odd}}$. *This two state chain whose transition matrix is given by* $P = \begin{pmatrix} \frac{\lambda}{2} & 1-\frac{\lambda}{2} \\ 1-\frac{\lambda}{2} & \frac{\lambda}{2} \end{pmatrix}$, *where* $\lambda := \frac{2}{3s}$, *satisfies* $P\begin{pmatrix} 1 \\ -1 \end{pmatrix} = (\lambda - 1)\begin{pmatrix} 1 \\ -1 \end{pmatrix}$. *Using the spectral decomposition it is easy to verify that* $A_k(1,1) = \frac{1}{2} + \frac{(\lambda-1)^k \lambda}{4}$. *Note that if* $k \le 8s$ *then for even* k*'s we have that* $0 \le A_k(1,1) - \frac{1}{2} = \Theta(s^{-1})$ *and for odd* k*'s* $0 \le \frac{1}{2} - A_k(1,1) = \Theta(s^{-1})$.

Applying this for $k=r$ *when* $T = 4n+2s-r > 0$, *in conjunction with (6.4.4)-(6.4.5) yields (6.4.2) by averaging over* $4n+2s-T$ *and bounding separately the contribution of all even times (i.e.* $4n+2s-T=2k$, $k \le 4s$*) and of all odd times, which are bounded from above by* $8s$ *. We leave the details as an exercise.*

Bibliography

[1] David Aldous. Random walks on finite groups and rapidly mixing markov chains. In *Séminaire de Probabilités XVII 1981/82*, pages 243–297. Springer, 1983.

[2] David Aldous and Persi Diaconis. Shuffling cards and stopping times. *The American Mathematical Monthly*, 93(5):333–348, 1986.

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

[4] David J Aldous. Some inequalities for reversible markov chains. *Journal of the London Mathematical Society*, 2(3):564–576, 1982.

[5] Noga Alon and Joel H Spencer. *The probabilistic method.* John Wiley & Sons, 2004.

[6] Ridhhipratim Basu, Jonathan Hermon, and Yuval Peres. Characterization of cutoff for reversible markov chains. *Ann. Probab., to appear. ArXiv preprint arXiv:arXiv:1409.3250*, 2013.

[7] Nathanael Berestycki, Eyal Lubetzky, Yuval Peres, and Allan Sly. Random walks on the random graph. *arXiv preprint arXiv:1504.01999*, 2015.

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

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

[10] Charles Bordenave, Pietro Caputo, and Justin Salez. Random walk on sparse random digraphs. *arXiv preprint arXiv:1508.06600*, 2015.

[11] Mark Brown. Interlacing eigenvalues in time reversible markov chains. *Mathematics of Operations Research*, 24(4):847–864, 1999.

[12] Guan-Yu Chen and Laurent Saloff-Coste. Comparison of cutoffs between lazy walks and markovian semigroups. *Journal of Applied Probability*, 50(4):943–959, 2013.

[13] Guan-Yu Chen, Laurent Saloff-Coste, et al. The cutoff phenomenon for ergodic markov processes. *Electronic Journal of Probability*, 13(3):26–78, 2008.

[14] Persi Diaconis. The cutoff phenomenon in finite markov chains. *Proceedings of the National Academy of Sciences*, 93(4):1659–1664, 1996.

[15] Persi Diaconis. The cutoff phenomenon in finite markov chains. *Proceedings of the National Academy of Sciences*, 93(4):1659–1664, 1996.

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

[17] Persi Diaconis and Laurent Saloff-Coste. Separation cut-offs for birth and death chains. *The Annals of Applied Probability*, 16(4):2098–2122, 2006.

[18] Persi Diaconis and Mehrdad Shahshahani. Generating a random permutation with random transpositions. *Probability Theory and Related Fields*, 57(2):159–179, 1981.

[19] Jian Ding, Eyal Lubetzky, and Yuval Peres. Total variation cutoff in birth-and-death chains. *Probability theory and related fields*, 146(1-2):61–85, 2010.

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

[21] James Allen Fill and Vince Lyzinski. Hitting times and interlacing eigenvalues: a stochastic approach using intertwinings. *Journal of Theoretical Probability*, pages 1–28, 2012.

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

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

[24] Simon Griffiths, Ross Kang, Roberto Oliveira, and Viresh Patel. Tight inequalities among set hitting times in markov chains. *Proceedings of the American Mathematical Society*, 2014.

[25] Jonathan Hermon. A technical report on hitting times, mixing and cutoff. *arXiv preprint arXiv:1501.01869*, 2015.

[26] 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.

[27] Jonathan Hermon, Hubert Lacoin, and Yuval Peres. Total variation and separation cutoffs are not equivalent and neither one implies the other. *Electronic Journal of Probability*, 21(44):1–36, 2016.

[28] 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.

[29] Jonathan Hermon and Yuval Peres. A characterization of l2 mixing and hypercontractivity via hitting times and maximal inequalities. *arXiv preprint arXiv:1609.07557*, 2016.

[30] Jonathan Hermon and Yuval Peres. On sensitivity of mixing times and cutoff. *arxiv preprint arXiv:1610.04357*, 2016.

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

[32] Solomon Kullback. A lower bound for discrimination information in terms of variation (corresp.). *IEEE Transactions on Information Theory*, 13(1):126–127, 1967.

[33] Cyril Labbé and Hubert Lacoin. Cutoff phenomenon for the asymmetric simple exclusion process and the biased card shuffling. *arXiv preprint arXiv:1610.07383*, 2016.

[34] Hubert Lacoin et al. Mixing time and cutoff for the adjacent transposition shuffle and the simple exclusion. *The Annals of Probability*, 44(2):1426–1487, 2016.

[35] Christian Le Merdy and Quanhua Xu. Maximal theorems and square functions for analytic operators on lp-spaces. *Journal of the London Mathematical Society*, 2012.

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

[37] László Lovász and Peter Winkler. Mixing times. *Microsurveys in discrete probability*, 41:85–134, 1998.

[38] Eyal Lubetzky and Allan Sly. Information percolation for the ising model: cutoff in three dimensions up to criticality. Technical report, Citeseer.

[39] Eyal Lubetzky and Allan Sly. Universality of cutoff for the ising model. *arXiv preprint arXiv:1407.1761*, 2014.

[40] Eyal Lubetzky and Allan Sly. Information percolation and cutoff for the stochastic ising model. *Journal of the American Mathematical Society*, 29(3):729–774, 2016.

[41] Eyal Lubetzky, Allan Sly, et al. Cutoff phenomena for random walks on random regular graphs. *Duke Mathematical Journal*, 153(3):475–510, 2010.

[42] Andreas Maurer et al. A bound on the deviation probability for sums of non-negative random variables. *J. Inequalities in Pure and Applied Mathematics*, 4(1):15, 2003.

[43] Jason Miller, Yuval Peres, et al. Uniformity of the uncovered set of random walk and cutoff for lamplighter chains. *The Annals of Probability*, 40(2):535–577, 2012.

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

[45] 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.

[46] Roberto Imbuzeiro Oliveira. Mixing and hitting times for finite markov chains. *Electron. J. Probab*, 17(70):1–12, 2012.

[47] Y Peres. American institute of mathematics (aim) research workshop "sharp thresholds for mixing times"(palo alto, december 2004). *Summary available at http://www. aimath. org/WWN/mixingtimes.*

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

[49] 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.

[50] Alistair Sinclair. *Algorithms for Random Generation and Counting: A Markov Chain Approach: A Markov Chain Approach*, volume 7. Springer, 1993.

[51] Norton Starr. Operator limit theorems. *Transactions of the American Mathematical Society*, pages 90–115, 1966.

[52] EM Stein. On the maximal ergodic theorem. *Proceedings of the National Academy of Sciences of the United States of America*, 47(12):1894, 1961.

[53] David Vernon Widder. *An introduction to transform theory*, volume 42. Academic Press, 1971.