# Rapid mixing through decomposition and induction

*Jung-Bae Son*


Doctor of Philosophy

Laboratory for Foundations of Computer Science
Division of Informatics
University of Edinburgh

2004

# Abstract

The concern of this thesis is a performance analysis of certain Markov chain Monte Carlo algorithms. The performance of Markov chain Monte Carlo algorithms in general is determined by the rate of convergence toward stationarity of the Markov chains involved. There now exists a substantial body of work on this subject and to begin with a synopsis of "classical" techniques for bounding convergence rate is given. Such techniques are: coupling, spectral gap, conductance and canonical paths. The focus then shifts to recent improvements on these techniques followed by sample applications: the first one illustrates how the path coupling method simplifies the analysis of a Markov chain for generating random lozenge tilings. The next example highlights the gains made by average conductance over "classical" conductance for the bases-exchange walk on balanced matroids. The result for the latter example is obtained by an inductive argument and is subsequently improved by the use of so-called logarithmic Sobolev constants. Bounding the logarithmic Sobolev constant is here achieved by following a decomposition-cum-induction approach. Such decompositional approaches for analysing Markov chain convergence are another manifestation of the "divide-and-conquer" paradigm. The thesis concludes with the treatment of a novel decomposition method and its application to bounding the convergence rate of a random process on finite, complete $d$-ary trees.

# Acknowledgements

I gratefully acknowledge the assistance of many people in the preparation and writing of this thesis. Their help of various kinds was invaluable to me and ought not to pass unmentioned. In particular, I owe a deep debt of gratitude to my supervisor, Mark Jerrum: for offering his support, for many insightful suggestions along with impeccable guidance and supervision, for providing opportunities to interact with a wider research community and last but not least for being a patient and critical proofreader of the manuscript. Likewise, I thank Mary Cryan and Martin Dyer for numerous suggestions which helped to further clarify the content of the thesis. Credit also goes to Eric Vigoda and Ravi Montenegro. Discussions with Eric and collaborating with Ravi in the early stages contributed to shaping my understanding of research.

I am profoundly indebted to my parents whose initial financial support only made my stay in Edinburgh possible. Furthermore, I thank my family and friends for their never-ending moral support and lastly, I would like to express my gratitude to DAAD, EPSRC(SARA) and the Graduate School for partially funding my studies.

# Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified. However, parts of the thesis, namely Section 4.4 and Chapter 5, have appeared in [96] and [61] respectively in slightly different form.

(*Jung-Bae Son*)

# Table of Contents

# Chapter 1

# Introduction

The focus of this thesis is the treatment of a certain type of randomized algorithm. Namely, we will be studying instances of the Markov chain Monte Carlo method (MCMC-method). It is a supplement to the Monte Carlo method, which many readers might be familiar with, and has over the last 15 years become the subject of extensive research, a good overview of which can be found in the following excellent survey articles by Dyer & Greenhill [40], Jerrum [56], Jerrum & Sinclair [59], Kannan [66], Lovász [81] and Welsh [115]. The versatility of this method stems from its two-prongedness: primarily, it is a method for generating random elements of combinatorial structures but in numerous settings it can be extended to a method for approximately counting the number of elements of combinatorial structures. At the heart of the Markov chain Monte Carlo method lies the simulation of a Markov chain with a stationary limiting distribution[1]. Hence, if the chain is simulated long enough, it is possible to sample an element exactly or almost exactly from the stationary distribution. The proviso of course is to choose a chain whose state space and stationary distribution coincide with the elements and the distribution one wishes to sample from respectively. Obviously this approach is vacuous from a practical point of view if the rate of convergence towards stationarity is slow. Nowadays, there are numerous techniques available for pinning down the rate of convergence. Here, our concern is to demonstrate how

[1]For those unfamiliar with Markov chains, an informal explanation is given later in this chapter. A formal, comprehensive description will be given in chapter 2 or read a book, for example Durrett [35], Norris [98] or Grimmet & Strirzaker [44].

induction, in conjuction with already known techniques (first and foremost decomposition) proves to be successful in tackling the problem of bounding convergence rate. To better understand the problems arising when analysing Markov chain Monte Carlo methods we will embark on a short journey through its history, which will take up most of the remainder of this introduction.

Originally, the Markov chain Monte Carlo method was introduced by Metropolis et al. [90] to overcome shortcomings of the classical Monte Carlo method. Before we address these weaknesses (computationally speaking) of the Monte Carlo method, let us have a closer look at how it works. At the core of the Monte Carlo method is the law of large numbers (in one of its many guises). The most commonly used example application of the Monte Carlo method is approximate integration: Suppose that $g$ is a continuous function from $[0,1]$ to $\mathbb{R}$. The (Riemann-) integral of $g$ is given by $\int_0^1 g(x)\,dx$. If $X$ is a continuous, real-valued random variable uniformly distributed over $[0,1]$, then this integral equals the expectation $\mathrm{E}[g(X)]$ of $X$. Obviously we are only interested in this quantity if the expectation actually exists. Now, let $X_1,\ldots,X_n$ be $n$ independent samples of $X$. According to the strong law of large numbers $\frac{1}{n}\sum_{i=1}^{n} g(X_i) \to \mathrm{E}[g(X)]$ almost surely if $n \to \infty$. Thus to approximately integrate $g$, we only have to take enough samples of the random variable $X$.

The principle of the Monte Carlo method seems simple enough. However, from a computational point of view it can be seriously flawed in two ways. First, there are cases where it might be necessary to take exponentially many samples before the Monte Carlo method will yield any useful information, which renders it useless for practical purposes. The second problem is the procurement of samples: Is it always possible to obtain samples? Or to put it differently: Does there always exist a sampling procedure? Our intention here is not to answer this question in the affirmative or negative in all generality. But what we are going to do is to offer a viable way to generate appropriate samples. Our way of obtaining samples will be via Markov chains and the term Markov chain Monte Carlo algorithm applies to every Monte Carlo algorithm that uses Markov chains to generate samples.

Bringing in Markov chains has proven to be a successful way to tackle both of the above problems and often it is the only method known to deliver. To highlight

both problems and how Markov chains can help, we address the problem of approximating the volume of $d$-dimensional convex objects. Suppose we are dealing with two-dimensional space and wish to know the area of the unit disk, i.e. disk of radius 1. The Monte Carlo approach to estimating its size would be to circumscribe a square of side length 2 around it and carry out what is commonly called "dart throws" at the square, i.e. take points from the square uniformly at random. Since uniformly sampling from the square is straightforward (that is if samples are taken from a discrete space, sampling from continuous spaces case is more involved) and computing the size of the square is trivial, the ratio of the darts in the disk to the darts within the square will be an estimation of the disk's size. Notice that, by a reversal of rôles, we could use the reciprocal of the ratio to estimate the size of the square in terms of the size of the disk. This method works fine in lower dimensions. However, for higher dimensions the volume of a (hyper-) ball, i.e. a ball of dimension $d \geq 3$, is [109][2]: $\frac{r^d \pi^{d/2}}{\Gamma(d/2+1)}$ and hence (using Stirling's formula) the ratio of the volumes of (hyper-) cube, i.e. a cube of dimension $d \geq 3$, to (hyper-) ball deteriorates exponentially in the dimension like

$$\frac{\pi^{d/2}}{\Gamma(d/2+1)\,2^d} \approx \frac{\sqrt{\pi e}^d}{\sqrt{\pi d}\,\sqrt{2d}^d},$$

which means that on average we will have to wait exponentially long before a dart lands within the ball. Conversely, the ratio of the volumes of a ball to an inscribed cube will explode exponentially. Thus, we must rule out this naïve Monte Carlo approach for solving the problem efficiently. A more careful argument, nonetheless, salvages the Monte Carlo method approach to this problem. For a convex body $K$ that allows two non-zero spheres, one inscribed in and the other encompassing the body — which is obviously the case for cubes — Grötschel, Lovász and Schrijver [47] gave a polynomial-time linear transformation of $K$ into a "well-rounded" body, i.e. one that contains the unit ball with the origin in the centre and is contained in a concentric ball of radius $\sqrt{d}(d+1)$, when $d$ is dimension. By abuse of notation let $K$ subsequently

[2]Note that this formula holds for both even and odd dimensions $d$ since the *Gamma-function* $\Gamma(x)$ (see e.g. [12]) is defined as

$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt, \quad x > 0$$

and has the property $\Gamma(x)\,\Gamma\left(x+\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2^{2x-1}}\Gamma(2x)$ for $x > 0$ and $\Gamma(n+1) = n!$ for $n \in \mathbb{N}$.

be the linearly transformed body. We continue by constructing a sequence of concentric balls $B_0, \ldots, B_k$ such that $B_0$ is the unit ball, $B_k$ the ball of radius $\sqrt{d}(d+1)$ and ball $B_i$ is entirely contained in ball $B_j$ for $i < j$. We choose $B_i$ to have radius $\left(\frac{d}{d-1}\right)^i$, which guarantees that the ratio of the volumes of $B_{i+1}$ to $B_i$ is at most $O(1)$. Notice that $B_0 = B_0 \cap K \subseteq B_1 \cap K \subseteq \ldots \subseteq B_k \cap K = K$. As the radius of $B_0$ is known, the ratio of the volumes of $K$ and $B_0$ can be computed by the telescopic product

$$\frac{\mathrm{Vol}(K)}{\mathrm{Vol}(B_0)} = \prod_{i=0}^{k-1} \frac{\mathrm{Vol}(K \cap B_{i+1})}{\mathrm{Vol}(K \cap B_i)}. \tag{1.1}$$

In contrast to the ratio on the left-hand side (lhs) of (1.1), which can be exponentially small for the reason given above, the ratios on the right-hand side (rhs) of (1.1) cannot be greater than a constant. This, together with the fact that $k = \left\lceil \log_{\frac{d}{d-1}} \sqrt{d}(d+1) \right\rceil$ is polynomially bounded in $d$, implies that the rhs of (1.1), and thus the lhs of (1.1), can be estimated by a number of samples polynomial in $d$. In principle it is therefore possible to efficiently estimate the volume of $K$ using the Monte Carlo method. In practice however this is where a pure Monte Carlo approach badly forsakes us and we are forced to bring in Markov chains; for as yet the only way known to sample from convex bodies is by using Markov chains.

One such Markov chain we could adopt is the following co-ordinate walk on a $d$-dimensional hyper-grid of edge-length $\delta$: If the current state is some grid point $x$, choose one of the $2d$ co-ordinate directions uniformly at random and move to the neighbouring grid point if it lies within some convex body, say $A$. If it does not, stay at $x$ (membership of $A$ can for example be decided by an oracle). This walk is potentially ill-behaved: it may not converge very quickly for general convex sets, it may even be unconnected, since $A$ might have sharp corners, so that a grid point contained in $A$ is unreachable from others in $A$ by the prescribed moves. However, if $A$ is well-rounded (as is the case for the $K \cap B_i$'s above) and $\delta$ not too large, then Dyer, Frieze and Kannan [38] have shown that the above co-ordinate walk is rapidly mixing. Their algorithm can be used to estimate the volume of well-rounded, $d$-dimensional convex bodies and their analysis addresses many difficult technical details we chose to omit here.

For those unfamiliar with Markov chains, roughly speaking a Markov chain is a

random process that is defined on a set of states and makes moves from one state to another in accordance with predetermined transition probabilities (the collection of all transition probabilities of a Markov chain is often referred to as the *transition kernel* or short *kernel*). Usually, we define a probability distribution on the set of states. This distribution indicates the probability of the chain being in any one particular state. In most cases these probabilities are not invariant under the transitions of the chain, i.e. the probability of being in a certain state changes with every move of the Markov chain. Nonetheless, there are distinguished probability distributions that remain unchanged under the transitions of the chain. These are called *stationary distributions*. In short, if a Markov chain is in stationarity, the probability of being in a certain state stays the same as the chain evolves.

The Markov chain Monte Carlo method exploits the property that using a "Metropolis-Hastings kernel", for example, a Markov chain can be made to converge towards almost any desired stationary distribution. In other words, for such a chain the probability of being in a certain state will be stationary after sufficiently many steps. Hence, in order to obtain appropriate samples all we have to do is to devise a Markov chain such that its stationary distribution is exactly the distribution we want to sample from, and to ensure that the chain converges towards it. This sampling technique was first used by Metropolis et al. [90] in 1953 and further studied by Hastings [51] in 1970. While the Markov chain Monte Carlo method lets us generate samples, it comes with a cost: Can we make any predictions on how long it will take until the chain is close to the stationary distribution? And how do we measure closeness to stationarity anyway? We shall return to these questions in section 2.1.

So far, we described how the Markov chain Monte Carlo method is primarily a sampling procedure. In the mid 1980's the method was given a new twist when it was applied to counting problems. Counting problems can be viewed as extensions of existence problems. An existence problem asks whether a specific problem has any solutions; probably the most well-known representative of this class of problems is SAT, the problem of deciding whether a Boolean formula in conjunctive normal form is satisfiable. The corresponding counting problem is #SAT, where we are interested in the total number of satisfying assignments. Generally speaking, when dealing with

counting problems we are interested in the number of solutions of a given problem instance. As can be seen from the example of #SAT, in most interesting cases of counting problems the solution spaces will be exponential in the input size and the existence of efficient deterministic counting algorithms highly unlikely. We therefore have to turn to approximate solutions.

The first attempts to analyse Markov chain Monte Carlo algorithms for approximating counting problems in a theoretical computer science context were made by Broder [16] and Jerrum, Valiant and Vazirani [63]. Broder analysed an algorithm for approximating the number of perfect matchings of what he called a dense graph. However, his analysis was slightly flawed and a correct analysis of the algorithm was subsequently provided by Jerrum and Sinclair [58]. In contrast to Broder's work, the result by Jerrum et al. proved the equivalence of almost uniform sampling and approximate counting for self-reducible problems. The terms approximation, almost uniform sampling and self-reducibility might require some explanation. The notion of randomized approximate counting goes back to Karp and Luby [69]. Let $f : \Sigma^* \to \mathbb{N}$, where $\Sigma$ is some countable alphabet, be a counting function: for example, if $x \in \Sigma^*$ were the encoding of a simple, undirected graph, then $f(x)$ could be the number of perfect matchings of $x$.

**Notation 1.0.1** *If $\Sigma$ is an alphabet and $x \in \Sigma^*$, then $|x|$ denotes the* length of $x$.

**Definition 1.0.2** *A* randomized approximation scheme *for $f$ is a randomized algorithm $A$ that for all inputs $\langle x, \varepsilon \rangle \in \Sigma^* \times \mathbb{R}^+$ produces an output $A(x, \varepsilon)$ such that*

$$\mathbb{P}\left[(1+\varepsilon)^{-1} f(x) \leq A(x,\varepsilon) \leq (1+\varepsilon) f(x)\right] \geq 3/4.$$

*Such a scheme is called* polynomial *if the running time of the algorithm is polynomial in $|x|$; it is called* fully polynomial *if the running time is polynomial in both $|x|$ and $1/\varepsilon$.*

Informally, a randomized approximation scheme is a randomized algorithm that will with arbitrarily high probability return a result deviating from the correct value by no more than a given approximation guarantee. The value $3/4$ in definition 1.0.2 is arbitrary since it was shown in [63] that it can be replaced by any value in the range $[3/4, 1)$.

Now assume that $R \subseteq \Sigma^* \times \Sigma^*$ is a binary relation on some countable alphabet $\Sigma$. Continuing with our example from above, we could let a pair $(x,y)$ be in $R$ if $x$ denotes a simple, undirected graph and $y$ a perfect matching in $x$. In general, $R$ encodes a problem, $x$ a problem instance and $(x,y) \in R$ if and only if $y$ is a solution to $x$. The set $R(x) = \{y \in \Sigma^* \mid (x,y) \in R\}$ contains the *solutions* of $x$ given $R$.

**Definition 1.0.3** *A randomized algorithm $A$ is an* almost uniform sampler *for a relation $R \subseteq \Sigma^* \times \Sigma^*$ if and only if there exists a function $\varphi : \Sigma^* \to (0,1]$ such that for all inputs $\langle x, \varepsilon \rangle \in \Sigma^* \times \mathbb{R}^+$ to $A$ and for all words $y \in \Sigma^*$*

$$\begin{aligned} (x,y) \notin R &\Rightarrow \mathbb{P}[A \text{ outputs } y] = 0, \\ (x,y) \in R &\Rightarrow (1+\varepsilon)^{-1}\varphi(x) \leq \mathbb{P}[A \text{ outputs } y] \leq (1+\varepsilon)\varphi(x) \end{aligned}$$

*and for all inputs $\langle x, \varepsilon \rangle$ such that $R(x) \neq \emptyset$*

$$\mathbb{P}[A \text{ outputs some } y] \geq 1/2.$$

*Notice that if $A$ produces an output $y$, then necessarily $y \in R(x)$. An almost uniform sampler is* fully polynomial *if the running time is polynomial in $|x|$ and $\log(1/\varepsilon)$.*

This means an almost uniform sampler is a randomized algorithm that given some input will with arbitrarily high probability return a sample chosen almost uniformly at random from the possible solutions of that input. Jerrum, Valiant and Vazirani's [63] seminal result is that for self-reducible, deterministically polynomial-time checkable problems there exists a fully polynomial randomized approximation scheme if and only if there is a fully polynomial almost uniform sampler. The concept of self-reducibility was first introduced by Schnorr [106] in 1976. His definition is as follows:

**Definition 1.0.4** *Let $\Sigma$ be an alphabet and for $x \in \Sigma^*$ let $|x|$ be the length of $x$. A binary relation $R$ on $\Sigma^* \times \Sigma^*$ is* (polynomial time) self-reducible *if there exist polynomial time computable functions $l : \Sigma^* \to \mathbb{N}$, $\psi : \Sigma^* \times \Sigma^* \to \Sigma^*$ and $\sigma : \Sigma^* \to \mathbb{N}$ such that*

**(SR1)** *$l(x) \in O(|x|)$ and $(x,y) \in R \Rightarrow |y| = l(x)$ and:*
$$R\left(x, y_1 \ldots y_{l(x)}\right) = R\left(\psi\left(x, y_1 \ldots y_{\sigma(x)}\right), y_{\sigma(x)+1} \ldots y_{l(x)}\right)$$

**(SR2)** *$\sigma(x) \in O(\log|x|)$ and $l(x) \geq 0 \Rightarrow \sigma(x) \geq 0$*

**(SR3)** $|\psi(x,w)| \leq |x|$.

N.B.: The result by Jerrum et. al. still holds if $l(x) \in O(|x|)$ in (SR1) is relaxed to $l(x) \in O(|x|^{c_R})$ where $c_R$ is a constant depending on $R$. Call this relaxation (SR1)'.

There are several ways to make self-reducibility more intuitively comprehensible. In [63] and [111] self-reducibility is described as the property that the solution set of an instance of a problem can be expressed as the direct union of the solution sets of smaller instances of the same problem. This conception is founded on the following interpretation of (SR1): if $y$ is a solution of $x$, then it is possible to take (and remove) a prefix of $y$ of length $\sigma(x)$ and use it to derive from $x$ a new (shorter) instance, namely $\psi(x, y_1 \ldots y_{\sigma(x)})$, such that the remainder of $y$ is a solution of the new instance $\psi(x, y_1 \ldots y_{\sigma(x)})$. Condition (SR1) encapsulates the core idea underlying self-reducibility and the meaning of (SR1) becomes much clearer when computations of a Turing machine are considered: Suppose that $R$ is decidable by a deterministic Turing machine $M_R$ such that $(x,y) \in R$ means $M_R$ accepts input $(x,y)$. For the sake of argument suppose further that $M_R$ always reads the entire input $(x,y)$ before it accepts or rejects and that the computation of $M_R$ on $(x,y)$ proceeds in two stages: first $M_R$ "reads" the problem instance $x$, then, in the "checking phase", $M_R$ checks if $y$ is a solution to $x$. Let us call the parameter $y$ the *argument* and call the sequence of configurations of the Turing machine during the checking phase for some argument $y$ the *trace* of $y$. Observe that for every problem instance $x$ there is a bijection between traces and arguments. Appealing to the bijection between arguments and traces but utterly disregarding the semantical difference between them, assume from now on that $y$ is a trace (i.e. the argument in the input $(x,y)$ is being replaced by a trace) and that $M_R$ can interpret this modified input. If $M_R$ is given this (modified) input, the parameter $y$ "hijacks" $M_R$ in the checking phase and instructs $M_R$ explicitly how to carry out the rest of the computation. Now, given an instance $x$ and a trace $y$ let $\psi(x, y_1 \ldots y_{\sigma(x)})$ encode a problem instance which (implicitly) reflects the first $\sigma(x)$ steps of the checking phase of $M_R$ on input $(x,y)$ such that by reading the problem instance $\psi(x, y_1 \ldots y_{\sigma(x)})$ the machine $M_R$ is "fast-forwarded" to the $\sigma(x)+1$-th step of the checking phase of $M_R$ on $(x,y)$; informally, regard the computation of $M_R$ on $(\psi(x, y_1 \ldots y_{\sigma(x)}), y_{\sigma(x)+1} \ldots y_{l(x)})$ as resuming a computation of $M_R$ on $(x,y)$ which

was interrupted after the $\sigma(x)$-th step of the checking phase. It is obvious that $M_R$ accepts $(x,y)$ if and only if $M_R$ accepts $\left(\psi\left(x,y_1\ldots y_{\sigma(x)}\right),y_{\sigma(x)+1}\cdots y_{l(x)}\right)$.

We stress that the above is only intended to shed some light on the principal idea behind self-reducibility and that the complexity issues arising in (SR2) and (SR3) have not been addressed. However, in the appendix a similar line of reasoning will be used to show that for every problem $\mathfrak{p}$ in NP there exists an encoding under which $R_{\mathfrak{p}}$, the relation characterizing $\mathfrak{p}$, is (polynomial time) self-reducible.

As an example of self-reducibility consider the problem of finding perfect matchings in a graph. Let $x$ encode a graph and $y$ a subgraph of $x$, then $R(x,y)$ denotes that $y$ is a perfect matching in $x$. Suppose that $\mu$ is a Turing machine deciding $R$. Essentially, a trace of $y$ in $\mu$ indicates in which order the edges of the matching $y$ are chosen. Hence a prefix of a trace will correspond to a subset of the edges of the matching $y$. With this in mind, we can regard $\psi\left(x,y_1\ldots y_{\sigma(x)}\right)$ as an encoding for finding a perfect matching, which contains the edges $y_1,\ldots,y_{\sigma(x)}$, in $x$. Notice that this is equivalent to finding a perfect matching in the subgraph of $x$ obtained by deleting the subgraph defined by $y_1,\ldots,y_{\sigma(x)}$, which indeed constitutes a smaller instance of the same problem.

Let us now return to the connection between Markov chain Monte Carlo methods/almost uniform sampling and approximate counting as proposed by Broder [16] and Jerrum et al. [63]. Although Jerrum et al.'s result sounds much more general than Broder's the main idea is basically the same. We will illustrate the main idea for approximately counting the number of perfect matchings in a graph. First assume that we know one perfect matching $y = y_1\ldots y_k$ in a graph $x$. Let $|H|$ be the number of all perfect matchings in $x$ and notice that it can be computed as follows:

$$|H| = \frac{|H|}{\left|H_{y_1}\right|} \times \frac{\left|H_{y_1}\right|}{\left|H_{y_1y_2}\right|} \times \cdots \times \frac{\left|H_{y_1\ldots y_{k-1}}\right|}{\left|H_{y_1\ldots y_k}\right|} \times \left|H_{y_1\ldots y_k}\right|, \tag{1.2}$$

where $H_{y_1\ldots y_i}$, $i \leq k$, is the set of all perfect matchings in $x$ that contain the edges $y_1,\ldots,y_i$ and obviously $\left|H_{y_1\ldots y_k}\right| = 1$. Lovász [81] called (1.2) a *product estimator* and we have already seen it in a slightly different form in (1.1). The ratios in (1.2) can be computed using (almost) uniform samplers: Suppose we have access to a uniform sampler for $H$. By taking sufficiently many samples of $H$ and determining the fraction of the samples that lie within $H_{y_1}$, the ratio $|H| / \left|H_{y_1}\right|$ can be accurately estimated. The other ratios in (1.2) can be determined in a similar fashion, i.e. a uniform

sampling procedure for $H_{y_1}$ can be used to estimate $|H_{y_1}| / |H_{y_1 y_2}|$, etc. The above approach using product estimators might seem straightforward in theory, in practice however our choice of $y_1, \ldots, y_k$ has to be made carefully to guarantee that the approximation is within specified error bounds. We will go into no more detail here but refer the reader concerned about approximation guarantees and deviation from the correct result to [16], [56], [59] and [63] where the connection between almost uniform sampling and approximate counting is discussed in much more detail.

In a way the result by Broder [16] opened the floodgates for Markov chain Monte Carlo methods, as since then an abundance of approximation results for counting problems have been found. Perhaps the most noteworthy amongst them are the results by Dyer, Frieze and Kannan [38] for approximating the volume of convex bodies, Jerrum and Sinclair [58] for approximating the permanent of dense 0, 1-matrices and Jerrum, Sinclair and Vigoda [60] for approximating the permanent of any matrix with non-negative entries. Many other applications can be found in [56] and [111]. Let us conclude our short visit to the beginnings of the Markov chain Monte Carlo method with the following remarks: In the setting of almost uniform sampling and randomized approximate counting via Markov chains, the focus of the analysis is the determination of the rate of convergence of the employed Markov chain. This rate, roughly speaking, indicates after how many steps the Markov chain is close to stationarity. Rate of convergence matters in both cases whether our goal is to generate samples or to count approximately. When we use the Markov chain Monte Carlo method for approximate counting it is, in addition, necessary to calculate the number of samples needed for a good approximation. The final running time of the approximation algorithm is then dependent on the convergence rate and the number of necessary samples. In practice however it seems commonplace to only bound the convergence rate, show that the structure in hand is self-reducible and state the existence of a fully polynomial random approximation scheme, appealing to Jerrum et al.'s result in [63].

The rest of this thesis is structured as follows: In chapter 2 we will give basic definitions in the context of the Markov chain Monte Carlo method and describe the "classic" techniques used for bounding convergence rate. Furthermore we will discuss the relationships among several of those techniques.

The focus of chapter 3 is the path coupling-technique developed by Bubley and Dyer [18] in 1997. We will apply this technique to bound the convergence rate of a Markov chain for generating Lozenge tilings on a two-dimensional, triangular lattice.

Chapter 4 is our first example of how induction can be used to analyse the convergence rate of Markov chains. Here, convergence rate will be determined using average conductance. Average conductance was introduced by Kannan and Lovász [68] in 1999 as an improvement on classic conductance which was defined by Sinclair and Jerrum [111] in 1989. First, we establish a bound on average conductance for the random walk on the $d$-dimensional hypercube. Later we employ an inductive argument which shows that the same bound on average conductance holds for the random walk on the bases-exchange graph of balanced matroids.

Chapter 5 is concerned with bounding spectral gap and logarithmic Sobolev constant of a special class of Markov chains. Examples of such chains are the random walks on the $d$-dimensional hypercube and on the bases-exchange graph of balanced matroids discussed in chapter 4. Both spectral gap and log-Sobolev constants can be used to bound convergence rate and will be defined in sections 2.3.1 and 5 respectively. The gist of chapter 5 is to demonstrate how a careful decomposition of (the constituents of) spectral gap and log-Sobolev constant combined with an inductive argument can yield an easy analysis of unwieldy objects like log-Sobolev constants.

In chapter 6 a new decomposition result is proposed. Decomposition as a way of analysing Markov chain Monte Carlo methods is a fairly recent technique by Madras and Randall [85] and simplified by Martin and Randall [86]. Our decomposition technique borrows heavily from the original decomposition techniques by Madras/Randall and Martin/Randall and recycles some of the main ideas of chapter 5. To test the new technique it is applied to the Ising model on finite, complete $d$-ary trees.

# Chapter 2

# The MCMC-approach

As discussed in the introduction, the Markov chain Monte Carlo method is an approach towards approximate sampling and counting. However, throughout the rest of the thesis, our interest in the Markov chain Monte Carlo method will be primarily in its rôle as a sampling procedure. Hence we will focus on determining the rate of convergence, also known as the mixing time, of Markov chains. This chapter will give basic definitions, describe the kind of Markov chains to be studied, and introduce the earliest techniques used for bounding mixing time.

## 2.1 Basic definitions

Generally speaking, a Markov chain is a random process, i.e. a process that lives on a specified state space and moves from one state to another according to so-called transition probabilities. Markov chains come in many different forms and shapes: state spaces can be either finite or continuous, transition probabilities may change over time and transitions can be either discrete-time or continuous-time. Discrete-time means that a transition occurs at every tick of a (usually no further specified) clock. For continuous-time Markov chains time is regarded as progressing continuously and there are several models which can be used to describe the evolution of a continuous time chain. One common way to model continuous-time Markov chains is by using a so-called *jump chain*. Roughly speaking, in the jump-chain model, the continuous-time

chain is modelled by a discrete-time Markov chain (the *jump chain*) which holds for an exponentially distributed amount of time between transitions, i.e. if $i$ is the current state, then a transition from $i$ to state, say, $j \neq i$ occurs after spending an exponentially distributed amount of time at $i$. A more detailed discussion of continuous-time Markov chains can be found in Grimmet & Stirzaker [44] or Norris [98]. Unless stated otherwise, the Markov chains in our setting will be on a finite state space with transitions occurring at discrete time steps. Additionally, we stipulate that the Markov chains are homogeneous, i.e. the transition probabilities are independent of time and hence are constant throughout time.

Let us introduce the following notation to describe the behaviour of a Markov chain as it progresses. We equate a Markov chain with a family of random variables $\{X_t \mid t \in \mathbb{N}_0\}$ indexed by time. The random variable $X_t$ indicates the state the Markov chain is in at time $t$.

**Definition 2.1.1** *A* finite, homogeneous Markov chain $M$ *is a pair* $(\Omega, P)$, *where* $\Omega$ *is a finite* state space *of cardinality* $n$ *and*

$$P = [p_{ij}]$$

*is an* $n \times n$*–matrix called the* transition kernel *(or short* kernel*) of* $M$. *The entries of the kernel are the* transition probabilities *of the chain:*

$$p_{ij} = \mathbb{P}[X_{t+1} = j \mid X_t = i].$$

A transition probability indicates the probability that the Markov chain moves from state $i$ to state $j$ in one step. Notationwise we will more often use $P(i,j)$ than $p_{ij}$.

As $P$ governs the one-step transition probabilities, the matrix $P^k = P \cdots P$, for $k = 1, 2, \ldots$ — the *k-fold product* of $P$ — governs the $k$-step transition probabilities of the chain, i.e.

$$P^k(i,j) = \mathbb{P}[X_{t+k} = j \mid X_t = i]$$

is the probability that the chain starting from state $i$ reaches state $j$ after $k$ steps. Let $\mu$ be an initial probability distribution on $\Omega$, i.e. suppose that $\mathbb{P}[X_0 = i]$ is $\mu(i)$. Then the

probabilities $\mathbb{P}[X_k = i]$, for $k \geq 1$, are given by $\mu^k = \mu P^k$. If the initial distribution is given by

$$\mu_x(z) = \mathbb{P}[X_0 = z] = \begin{cases} 1, & \text{if } z = x \\ 0, & \text{otherwise,} \end{cases}$$

then ${\mu_x}^k(z) = P^k(x,z)$. In general $\mu^l \neq \mu^m$, for $l \neq m$. There are, nonetheless, probability distributions $\pi$ such that $\pi = \pi P^k$. It is obvious that this holds for the following kind of distributions.

**Definition 2.1.2** *A probability distribution $\pi$ such that*

$$\pi = \pi P$$

*is a* stationary distribution *of the Markov chain with kernel $P$.*

In the following it will be shown that every finite Markov chain satisfying certain properties possesses a unique stationary distribution. A necessary prerequisite for understanding a rigorous statement of this result is a classification of Markov chains.

**Definition 2.1.3** *A Markov chain is* irreducible *if for every $x, y \in \Omega$ there exists some $k \in \mathbb{N}$ such that $P^k(x,y) > 0$.*

In other words, if a chain is irreducible, then it is possible to reach any state from any other state of the chain. Another important property of Markov chains is periodicity/aperiodicity.

**Definition 2.1.4** *The* period $d(i)$ *of state $i$ is the greatest common divisor of the points of time of possible re-occurrences of state $i$, i.e. $d(i) = \gcd\{k : P^k(i,i) > 0\}$. A Markov chain is called* aperiodic *if all states have period one.*

With the concepts of irreducibility and aperiodicity we can now formulate one of the most well-known results of Markov chain theory. It is called the *ergodic theorem* and a proof of the version we give here can be obtained by combining several results stated in [44].

**Theorem 2.1.5** *Every finite and irreducible Markov chain $M = (\Omega, P)$ has a unique stationary distribution $\pi$. Furthermore, if $M$ is aperiodic, then the stationary distribution is a* limiting distribution *i.e. :*

$$P^k(i,j) \to \pi(j) \quad as \quad k \to \infty, \qquad for\ all\ i\ and\ j.$$

Finding stationary distributions can be greatly simplified if we require Markov chains to be time-reversible.

**Definition 2.1.6** *An irreducible Markov chain with stationary distribution $\pi$ is called* time-reversible *if its transition probabilities satisfy the* detailed balance *condition:*

$$\pi(i)P(i,j) = \pi(j)P(j,i) \qquad for\ all\ i,j.$$

The following lemma shows how time-reversibility facilitates the task of finding stationary distributions.

**Lemma 2.1.7** *Let $M = (\Omega, P)$ be an irreducible Markov chain and suppose that there exists a distribution $\pi$ that satisfies detailed balance: $\pi(i)P(i,j) = \pi(j)P(j,i)$ for all $i, j \in \Omega$. Then $\pi$ is a stationary distribution of $M$.*

We will not give a proof of Lemma 2.1.7 but refer the interested reader to [44].

After all these definitions, we can now describe in detail how Markov chains are used as sampling procedures. Recall that the probabilities $\mathbb{P}[X_k = i]$ are given by $\mu^k = \mu P^k$ for some initial probability distribution $\mu$ on $\Omega$. If the Markov chain is finite, irreducible and aperiodic, then by Theorem 2.1.5 the chain has a unique stationary distribution $\pi$, which is also a limiting distribution, and hence

$$\mu^k(j) = \sum_{i=1}^{n} \mu(i)P^k(i,j) \to \sum_{i=1}^{n} \mu(i)\pi(j) = \pi(j) \qquad \text{for} \quad k \to \infty.$$

This means that, irrespective of the initial probability distribution, after sufficiently many steps the probabilities of the chain being in a certain state are close to the stationary distribution. Eventually the chain will reach its stationary distribution: it is then said to be in *equilibrium*. It is now fairly straightforward to see how Markov chains give rise to sampling procedures: Devise a finite, irreducible and aperiodic Markov

chain such that its stationary distribution coincides with the distribution you wish to sample from. Our working assumptions, unless stated otherwise, will hence be that the Markov chain is finite, irreducible, aperiodic and time-reversible with stationary limiting distribution $\pi$.

Not having seen any examples this may sound complicated but in most cases constructing such Markov chains is surprisingly simple and we will present examples in later chapters. To verify that the chain has indeed the desired stationary distribution, Lemma 2.1.7 can be employed (i.e. check if the chain is time-reversible). It is then simply a matter of patience to run the Markov chain until it is close to equilibrium. Once the chain is near or in equilibrium, it is stopped and the current state output as a random sample. The crucial question is, how long it will take until the chain is close to stationarity, as surely any instance where this takes exponentially long must be dismissed as impractical. In what follows, our aim will be to rigorously analyse after how many steps a Markov chain is close to equilibrium.

Our measure of closeness will be total variation distance.

**Definition 2.1.8** *Given two probability distributions* $\rho$ *and* $\eta$ *on* $\Omega$ *their* total variation distance *is defined by*

$$\|\rho - \eta\|_{\mathrm{TV}} = \max_{A \subseteq \Omega} |\rho(A) - \eta(A)|, \tag{2.1}$$

*where* $\rho(A) = \sum_{x \in A} \rho(x)$ *and analogously for* $\eta(A)$.

The total variation distance coincides with $\sum_{x \in A} (\rho(x) - \eta(x))$ if $A$ is the entire area in $\Omega$ where the values of $\rho$ are greater then those of $\eta$. Notice that this must be equal to $\sum_{x \in \Omega \setminus A} (\eta(x) - \rho(x))$ since $\rho$ and $\eta$ are probability distributions, which implies $\eta(x) \geq \rho(x)$ for $x \in \Omega \setminus A$. This observation leads to another quantification of total variation distance

$$\|\rho - \eta\|_{\mathrm{TV}} = \frac{1}{2} \sum_{x \in \Omega} |\rho(x) - \eta(x)|. \tag{2.2}$$

Since in our setting the stationary distribution is also a limiting distribution, intuitively

$$\left\| \mu^{k+1} - \pi \right\|_{\mathrm{TV}} \leq \left\| \mu^{k} - \pi \right\|_{\mathrm{TV}}$$

should hold; the proof we give here is copied from [57], p.47:

$$\begin{aligned}
2\left\|\mu^{k+1}-\pi\right\|_{\mathrm{TV}} &= \sum_{x\in\Omega}\left|\mu^{k+1}(x)-\pi(x)\right| \\
&= \sum_{x\in\Omega}\left|\sum_{y\in\Omega}\mu^{k}(y)P(y,x)-\sum_{y\in\Omega}\pi(y)P(y,x)\right| \\
&\leq \sum_{x\in\Omega}\sum_{y\in\Omega}\left|\mu^{k}(y)-\pi(y)\right|P(y,x) \\
&= \sum_{y\in\Omega}\left|\mu^{k}(y)-\pi(y)\right|\sum_{x\in\Omega}P(y,x) \\
&= 2\left\|\mu^{k}-\pi\right\|_{\mathrm{TV}}.
\end{aligned}$$

This shows that the total variation distance between $\mu^k$ and $\pi$ is non-increasing in $k$. Thus we can define our measure of convergence to stationarity as follows:

**Definition 2.1.9** *Assume a finite, irreducible and aperiodic Markov chain with stationary limiting distribution $\pi$ and initial state $x$, i.e. $\mathbb{P}[X_0 = x] = 1$ and $\mathbb{P}[X_0 = y] = 0$ for $y \neq x$. The chain's rate of convergence to stationarity from its initial state $x$, also called* mixing time from $x$, *is defined as*

$$\tau_x(\delta) = \min\left\{t : \left\|\mu_x{}^t - \pi\right\|_{\mathrm{TV}} \leq \delta\right\}, \tag{2.3}$$

*where $\delta$ is an error bound controlling closeness to the stationary distribution. For statements independent of the initial state* mixing time *is defined as*

$$\tau(\delta) = \max_{x\in\Omega}\left\{\tau_x(\delta)\right\}.$$

For practical purposes the mixing time of the chain must be small. Of course this begs the question "small, in what sense?" The problems that we will be dealing with will typically have a huge state space but, in comparison, a very short description of the state space (henceforth we will refer to the length of the description as the *input size*). As an example consider a problem where the state space is the set of all bitvectors of length $n$, then surely $|\Omega| = 2^n$, whereas the description of the state space: "$\Omega = \{(x_1,\ldots,x_n) \mid x_i \in \{0,1\}\}$" is significantly shorter. With this in mind, we define:

**Definition 2.1.10** *A Markov chain is said to be* rapidly mixing *if the mixing time is polynomial in the input size n and the logarithm of the inverse of the error bound* $\delta$*, i.e. if*

$$\tau(\delta) \leq \mathrm{poly}\left(n, \log \delta^{-1}\right).$$

**Notation 2.1.11** *We use* log *to denote logarithms with unspecified base and* $\log_b$ *for the logarithm to the base b. The symbol for the natural logarithm will be* ln.

The choice of the error bound $\delta$ is to some extent purely arbitrary and many authors choose $\delta = 1/e$ since logarithms to the base $e$ feature in the analytic techniques used and then the constants work out nicely. This arbitrariness is due to the fact that once total variation is within $1/e$ of stationarity, it decreases exponentially in linearly many steps. Using [4], Section 3.1, Lemma 5 by Aldous and Fill, we obtain:

**Lemma 2.1.12** *For a Markov chain* $M = (\Omega, P)$ *with stationary limiting distribution* $\pi$ *and for* $t \geq 0$ *the total variation distance is bounded by*

$$\max_{x \in \Omega} \left\{ \left\| \mu_x^t - \pi \right\|_{\mathrm{TV}} \right\} \leq \exp\left( - \left\lfloor t / \tau\left(e^{-1}\right) \right\rfloor \right).$$

One might wonder why total variation distance is favoured over other distances for measuring proximity to equilibrium. This is mostly due to the fact that it bounds the absolute error made when a Markov chain sampler is used for obtaining a random approximation scheme. What we mean is the following: Consider the product estimator (1.2). Let $\mu_i = \left|H_{y_1 \ldots y_{i-1}}\right| / \left|H_{y_1 \ldots y_i}\right|$ be the $i$-th ratio. Let $Z_i$ be a random variable taking values 0 and 1. It is 1 in the event that a matching sampled from a distribution within $\delta$ of uniformity from $H_{y_1 \ldots y_{i-1}}$ lies also in $H_{y_1 \ldots y_i}$ and 0 otherwise. Let $\hat{\pi}_i(0)$ and $\hat{\pi}_i(1)$ be the probabilities that the random sampler generates a matching such that $Z_i = 0$ and $Z_i = 1$ respectively. Notice that these probabilities deviate from $\pi_i(1)$ and $\pi_i(0)$, the probabilities that a perfectly random matching of $H_{y_1 \ldots y_{i-1}}$ is or is not in $H_{y_1 \ldots y_i}$ respectively. Let $\mathrm{E}_{\hat{\pi}_i}[Z_i]$ be the expectation of $Z_i$ with respect to $\hat{\pi}_i$ and assume that $\mathrm{E}_{\pi_i}[Z_i]$ is defined similarly. Observe that $\mathrm{E}_{\pi_i}[Z_i] = 0 \cdot \pi_i(0) + 1 \cdot \pi_i(1) = 1/\mu_i$. Suppose our random sampler generates $r$ independent samples $Z_i^{(1)}, \ldots, Z_i^{(r)}$ to approximate $\mathrm{E}_{\pi_i}[Z_i]$ and let $\bar{Z}_i = r^{(-1)} \sum_{j=1}^{r} Z_i^{(j)}$. Then by linearity of expectation $\mathrm{E}[\bar{Z}_i] = \mathrm{E}_{\hat{\pi}_i}[Z_i]$. Suppose the samples are obtained by

running the Markov chain long enough such that $\|\hat{\pi}_i - \pi_i\|_{\mathrm{TV}} \leq \delta$ for some $\delta \in \mathbb{R}$. Then clearly $\left|\mathrm{E}_{\hat{\pi}_i}[Z_i] - \mathrm{E}_{\pi_i}[Z_i]\right| \leq 2\delta$. If we consider the inverse of (1.2), then total variation distance yields an absolute error bound on the ratios we want to estimate. There are other distances which measure how far two distributions are apart, e.g. Fill [42] and Mihail [91] studied $\chi^2$-distance, which is defined as

$$\chi^2(\mu^t, \pi) = \sum_{x \in \Omega} \frac{(\mu^t(x) - \pi(x))^2}{\pi(x)}.$$

For chains meeting our working assumptions, this distance has the nice property that it contracts by at least a certain factor with each step of the chain

$$\chi^2(\mu^{t+1}, \pi) \leq \beta \chi^2(\mu^t, \pi).$$

The factor $\beta$ is actually the second largest eigenvalue of $P$ and is in $[0,1)$ as will be shown in section 2.3.1. This significantly simplifies the process of deducing bounds on mixing time. In fact, most of the known bounds on mixing time with respect to total variation distance are derived from bounds for $\chi^2$-distance, which are then translated into bounds for total variation distance. Using Jensen's inequality, $f(\mathrm{E}[X]) \leq \mathrm{E}[f(X)]$, for convex $f$, it is easily seen that for $f(x) = x^2$

$$4\|\mu - \pi\|_{\mathrm{TV}}^2 = \left(\sum_{x \in \Omega} \left|\frac{\mu(x)}{\pi(x)} - 1\right| \pi(x)\right)^2 \leq \chi^2(\mu, \pi)$$

and hence, any bound on $\chi^2$-distance will yield an upper bound on total variation distance. It should be remarked that historically (cf. [58], [111]) many results were formulated using *relative pointwise distance*

$$\Delta_U(t) = \max_{i,j \in U} \frac{|P^t(i,j) - \pi(j)|}{\pi(j)},$$

where $U \subseteq \Omega$. For most interesting chains however rapid mixing with respect to total variation distance will imply rapid mixing w.r.t. relative pointwise distance and vice versa. To see this first observe that

$$\Delta_U(t) \leq \max_i \left\|P^t(i,\cdot) - \pi\right\|_{\mathrm{TV}} \left(\min_i \pi(i)\right)^{-1}.$$

Then recall that our goal is almost uniform sampling from exponentially large sample spaces and thus $\min_i \pi(i) = 2^{-O(n)}$. Combining this with Lemma 2.1.12 implies that either choice of distance makes no essential difference to rapid mixing (in the sense that the mixing time will still be polynomial in $n$).

In the rest of this chapter we will outline "traditional" methods used for bounding mixing time and comment on relationships between them.

## 2.2 Coupling

Among the first techniques used to obtain bounds on the mixing time of Markov chains is the "coupling" method. It is a method that comes in many different forms. The term was coined around 1970 by Frank Spitzer although the method itself dates back to 1938 when W. Doeblin [33] used it in the study of Markov chains. Lindvall [80] gives a detailed account of the history and mathematics of coupling. In the current context let coupling be defined as follows:

**Definition 2.2.1** *Let $M$ be a Markov chain on a finite state space $\Omega$ with transition kernel $P$. A* coupled *chain (or* a coupling*) $(X_t, Y_t)$ is a Markov process on $\Omega \times \Omega$ such that each of the processes $(X_t)$ and $(Y_t)$ considered in isolation is a faithful copy of $M$:*

$$\mathbb{P}\left[X_{t+1} = x' \mid X_t = x \wedge Y_t = y\right] = P(x, x') \tag{2.4}$$

*and*

$$\mathbb{P}\left[Y_{t+1} = y' \mid X_t = x \wedge Y_t = y\right] = P(y, y') \tag{2.5}$$

*for all $x, y, x', y' \in \Omega$.*

Note that the method, which uses a coupled chain, is referred to as "coupling" while the coupled chain itself is referred to as "*a* coupling" or "*the* coupling", i.e. is prefixed by an indefinite or definite article. Conditions (2.4) and (2.5) include the possibility that $(X_t)$ and $(Y_t)$ are independent evolutions of $M$ but do not imply it. In fact, we shall use the possibility that $\mathbb{P}[X_{t+1} = x' \wedge Y_{t+1} = y' \mid X_t = x \wedge Y_t = y] \neq P(x, x') \cdot P(y, y')$ to encourage $(X_t, Y_t)$ to coalesce rapidly. *Coalescence at time $T$* means that the chains $(X_t)$, $(Y_t)$ are at the same state at time $T$, i.e. $X_T = Y_T$. Whenever coupling

is used, the aim is to design a (coupled) chain $(X_t, Y_t)$ such that once coalescence has occurred the chains $(X_t)$ and $(Y_t)$ stay together, i.e. $X_{t'} = Y_{t'}$ for all $t' \geq T$ if $(X_t, Y_t)$ has coalesced at time $T$.

If coalescence occurs rapidly independent of the initial states $X_0$ and $Y_0$, we may deduce that $M$ is rapidly mixing. The key result used here is the *coupling lemma*, which was first stated by Aldous [2] (see also Diaconis [29]).

**Lemma 2.2.2** *Suppose that $M$ is a finite, irreducible and aperiodic Markov chain with transition kernel $P$ and let $((X_t, Y_t) \mid t \in \mathbb{N})$ be a coupling as in Definition 2.2.1. Suppose that $\mathfrak{t} : (0,1] \to \mathbb{N}$ is a function such that $\mathbb{P}\left[X_{\mathfrak{t}(\varepsilon)} \neq Y_{\mathfrak{t}(\varepsilon)}\right] \leq \varepsilon$ for all $\varepsilon \in (0,1]$, independent of the choice of initial state $(X_0, Y_0)$. Then the mixing time $\tau(\varepsilon)$ of $M$ is bounded above by $\mathfrak{t}(\varepsilon)$.*

Although Aldous formulated his result for continuous time chains he added that it essentially carries over to discrete time chains. The conceptual simplicity of the coupling method and the advent of new techniques like path coupling (cf. section 3.1), which greatly facilitates the designing of couplings, has made coupling virtually the technique of choice for proving rapid mixing.

## 2.3 Poincaré and Cheeger-type inequalities

This section deals with three closely inter-related techniques for bounding mixing time: canonical paths, conductance and spectral gap. The spectral gap of a Markov chain is the difference between the largest two absolute values of the eigenvalues of the transition kernel of a Markov chain. Classical results from Perron-Frobenius theory (see e.g. Seneta [108]) show that the convergence of a Markov chain is governed by the absolute value of the second largest eigenvalue in absolute value. This was utilized by many authors (cf. Diaconis & Stroock [31], Fill [42], Lovász [81] and Sinclair & Jerrum [111]) to obtain quantitative bounds on the rate of convergence in terms of spectral gap, although in some cases for distance measures different to total variation distance.

It is known from linear algebra (see e.g. Lang [75]) that the eigenvalues of an $n \times n$-matrix $A$ are the roots of its characteristic polynomial, i.e. a number $\lambda$ is an eigenvalue

of $A$ iff $\mathrm{Det}(\lambda I - A) = 0$, where Det denotes the determinant of a matrix and $I$ is the identity matrix. It is also known from linear algebra that a matrix $A$ can be triangulated into a matrix

$$B = \begin{bmatrix} b_{11} & & & & \\ 0 & b_{22} & & * & \\ 0 & 0 & \ddots & & \\ \vdots & \vdots & & \ddots & \\ 0 & 0 & \cdots & 0 & b_{nn} \end{bmatrix},$$

where entries $b_{ij} = 0$, for all $i > j$, such that

$$\mathrm{Det}(A) = \pm \mathrm{Det}(B) = \pm \prod_{i=1}^{n} b_{ii}.$$

Such a triangulation is equivalent to a Gauss-elimination, which can be carried out in time $O(n^3)$ (cf. [112] assuming, of course, that arithmetic operations cost constant time). It can also be shown that the characteristic polynomial can be obtained in time $O(n^3)$. Once we know the characteristic polynomial, we can compute its roots to determine the spectral gap of the matrix. However, finding the roots of a polynomial is not always straightforward: If one is lucky, finding the roots of a polynomial can be easy, otherwise, especially for polynomials of high degree, it can be very expensive. An exposition of available techniques for computing eigenvalues of matrices in general can be found in [23].

In the setting of the Markov chain Monte Carlo method, transition matrices are invariably huge; state spaces of exponential size are the norm. Hence, computing the spectral gap by determining the roots of the characteristic polynomial is inefficient (in our setting, even computing the characteristic polynomial itself is hugely inefficient). Instead, our goal will be to give estimates of the spectral gap. One approach is by using conductance. This quantity can be linked to spectral gap (cf. Aldous [3], Alon [5], Alon & Milman [6], Dodziuk [32] and Lawler & Sokal [76]) via a discrete version of Cheeger's inequality [25] from differential geometry. But since even conductance could be hard to compute, early bounds on it were obtained using so-called canonical paths (see Jerrum & Sinclair [58]). The concept of canonical paths was therefore predominantly regarded as a tool for bounding conductance. However, subsequent

results showed that canonical paths can be directly related to spectral gap via Poincaré-type inequalities and the canonical paths method has since been acknowledged as a technique in its own right (see Diaconis & Stroock [31] and Sinclair [110]). We start with a discussion of spectral gap since it forms the theoretical backbone of the sequel.

### 2.3.1 Spectral gap

Recall our working assumption that the Markov chain $M = (\Omega, P)$, $\Omega = \{x_1, \ldots, x_n\}$, is finite, irreducible, aperiodic and time-reversible. It hence possesses a unique stationary distribution $\pi$.

The Perron-Frobenius Theorem (see Seneta [107]) states that for *primitive*[1] matrices there exists a real eigenvalue $r > 0$, which can be associated to strictly positive left and right eigenvectors, s.t. $r > |\lambda|$ for any eigenvalue $\lambda \neq r$. It follows further from Perron-Frobenius Theory that for stochastic matrices the value of $r = 1$ (see Seneta [107]).

Notice that $\pi$ is a left eigenvector of $P$ with eigenvalue $\beta_1 = 1$. Hence, by the Perron-Frobenius Theorem, the remaining eigenvalues satisfy $1 > |\beta_i|$, for $1 < i \leq n$. It is known (see e.g. Lang [75]) that for an invertible $n \times n$-matrix, say $B$, the characteristic polynomials

$$\mathrm{Det}\,(tI - P) = \mathrm{Det}\left(tI - B^{-1}PB\right), \quad t \in \mathbb{C}$$

and thus the eigenvalues are the same. If we choose $B^{-1}$ to be the diagonal matrix

$$B^{-1} = \mathrm{diag}\left(\sqrt{\pi(x_1)}, \ldots, \sqrt{\pi(x_n)}\right),$$

then, by time-reversibility, $B^{-1}PB$ is symmetric with entries $B^{-1}PB(x,y) = \sqrt{\frac{\pi(x)}{\pi(y)}}P(x,y)$. As symmetric matrices have only real eigenvalues (cf. Lang [75] or Horn & Johnson [54]), all eigenvalues of $P$ must be real.

**Definition 2.3.1** *The* spectrum *of $M$ consists of the eigenvalues $\beta_i$, $i = 1, \ldots, n$, of $M$. The difference between the two largest absolutes of eigenvalues, in this case*

$$\lambda = 1 - \max\{|\beta_2|, |\beta_n|\},$$

[1] All irreducible and acyclic matrices are primitive (cf. Seneta [107]).

*is known as the* spectral gap.

In the context of time-reversibility spectral gap can equivalently be defined in terms of *Poincaré inequalities*:

$$\lambda = \min_{f:\Omega\to\mathbb{R}} \left\{ \frac{\mathcal{E}(f,f)}{\mathrm{Var}(f)} : \mathrm{Var}(f) \neq 0 \right\} \tag{2.6}$$

$$= \min_{f:\Omega\to\mathbb{R}} \{\mathcal{E}(f,f) : \|f\|_2 = 1, \mathrm{E}[f] = 0\}, \tag{2.7}$$

where

$$\mathcal{E}(f,f) = \frac{1}{2} \sum_{x,y\in\Omega} |f(x) - f(y)|^2 \pi(x) P(x,y) \tag{2.8}$$

is a so-called *Dirichlet form*,

$$\|f\|_p = \left( \sum_{x\in\Omega} |f(x)|^p \pi(x) \right)^{1/p}$$

the $L_p(\pi)$-norm[2], $\mathrm{Var}(f)$ the variance of $f$ and $\mathrm{E}[f]$ its expectation. Sometimes it is helpful to use a slightly non-standard definition of variance

$$\mathrm{Var}(f) = \frac{1}{2} \sum_{x,y\in\Omega} |f(x) - f(y)|^2 \pi(x)\pi(y). \tag{2.9}$$

This alternative definition of spectral gap will play an important rôle in chapters 5 and 6.

Many authors have observed that the spectral gap describes the long-term convergence behaviour of Markov chains. Here we use a slightly modified result from the one stated by Sinclair in [110] (similarly, cf. Diaconis & Stroock [31], Fill [42], Lovász [81] and Sinclair & Jerrum [111])

**Lemma 2.3.2** *For $\varepsilon > 0$ and spectral gap $\lambda$ the mixing time is bounded by*

$$\tau_x(\varepsilon) \leq \frac{1}{\lambda} \left( \ln \frac{1}{\pi(x)} + \ln \frac{1}{\varepsilon} \right).$$

[2] $L_p(\mu)$ is the space of real-valued functions $f$ such that $|f|^p$ is integrable w.r.t. measure $\mu$, i.e. $\int |f|^p d\mu < \infty$ (cf. [34]).

It should be remarked that the bound stated in Lemma 2.3.2 is an implication of a result by Diaconis and Stroock ([31], proposition 3) for continuous time chains. However, Diaconis and Stroock give a bound for discrete chains too and moreover the resulting bounds on mixing time are such that the mixing times of continuous time chains are larger than those of discrete chains and the formulation of Lemma 2.3.2 takes this into account.

If we are prepared to use a slightly slower chain, then we are allowed to dispose of $\beta_n$ and concentrate on $\beta_2$ for determining spectral gap. The slower chain is often called the "lazy chain" because it has self-loop probabilities of at least $1/2$ and its other transition probabilities are diminished appropriately. One way to obtain such a lazy chain is through the following definition:

$$M_{Zz} = (\Omega, P_{Zz}) \quad \text{, where}$$
$$P_{Zz} = \tfrac{1}{2}(I_n + P)$$

and $I_n$ is the $n \times n$ identity matrix. Sinclair and Jerrum [111] observed that a lazy chain defined as above mixes at most twice as slow as the original chain:

**Lemma 2.3.3** *Let $P$ be the transition matrix of a finite, irreducible and time-reversible Markov chain with eigenvalues $1 = \beta_1 > \beta_2 \geq \ldots \geq \beta_n > -1$. Then the lazy chain with transition matrix $P_{Zz}$ is finite, irreducible, aperiodic and time-reversible with the same stationary distribution and its eigenvalues, identically ordered, satisfy $(\beta_{Zz})_i = (1+\beta_i)/2 > 0$.*

Unless stated otherwise, we will in future assume that all self-loops have probability at least $1/2$ and the spectral gap is given by the difference of the two largest eigenvalues.

### 2.3.2 Conductance

Let $M = (\Omega, P)$ be a finite, irreducible, aperiodic and time-reversible Markov chain with stationary distribution $\pi$. In the light of the detailed balance condition (Definition 2.1.6), $M$ can be considered an undirected graph $(\Omega, T)$ with vertex set $\Omega$ and edge set

$$T = \left\{ \{x,y\} \in \Omega^2 \mid \widetilde{P}(x,y) > 0 \right\}, \tag{2.10}$$

where

$$\widetilde{P}(x,y) := \pi(x)P(x,y) = \pi(y)P(y,x). \tag{2.11}$$

Sinclair and Jerrum defined in [111]:

**Definition 2.3.4** *The* conductance *of $M$ is*

$$\Phi = \Phi(M) = \min_{\substack{S\subset\Omega \\ 0<\pi(S)\leq\frac{1}{2}}} \frac{\widetilde{P}(S,\overline{S})}{\pi(S)},$$

*where $\widetilde{P}(S,\overline{S})$ denotes the sum of $\widetilde{P}(x,y)$ over edges $\{x,y\} \in T$ with $x \in S$ and $y \in \overline{S} = \Omega \setminus S$.*

Observe that taking the minimum over $\frac{1}{2} \leq \pi(S) < 1$ in the definition of conductance would be symmetrical because of time-reversibility.

Conductance appears like a weighted version of edge expansion of the graph $(\Omega, T)$ associated with $M$. Intuitively, the quotient in the definition of conductance captures the conditional probability that the chain in equilibrium escapes from a small (w.r.t. probability measure) subset $S$ of the state space in one step, given that it is initially in $S$; thus $\Phi$ measures the tendency of $M$ to escape from any small enough region of the state space and hence to make rapid progress towards equilibrium. This intuitive connection can be given a precise quantitative form using the following *Cheeger-type inequality* (see e.g. Jerrum & Sinclair [111] or Diaconis & Stroock [31]):

**Lemma 2.3.5** *Let $\lambda$ be the spectral gap of a finite, irreducible, aperiodic and time-reversible Markov chain, then*

$$\frac{\Phi^2}{2} \leq \lambda \leq 2\Phi.$$

The upper bound can be obtained straightforwardly: Let $\mathbb{I}_S$ be the characteristic function of $S$, i.e.

$$\mathbb{I}_S(x) = \begin{cases} 1, & \text{if } x \in S \\ 0, & \text{otherwise.} \end{cases}$$

Then using (2.6)

$$\lambda \leq \frac{\mathcal{E}(\mathbb{I}_S, \mathbb{I}_S)}{\mathrm{Var}(\mathbb{I}_S)} = \frac{\widetilde{P}(S,\overline{S})}{\pi(S)\pi(\overline{S})} \leq 2\frac{\widetilde{P}(S,\overline{S})}{\pi(S)}. \tag{2.12}$$

Since (2.12) holds for the characteristic function of any subset $S \subseteq \Omega$, the claimed inequality holds. A proof of the lower bound is more involved and we refer the reader to [31] or [111].

Some authors define conductance as

$$\Phi_{\text{alt}} = \min_{\substack{S \subset \Omega \\ 0 < \pi(S) \leq \frac{1}{2}}} \frac{\widetilde{P}(S, \overline{S})}{\pi(S)\pi(\overline{S})}, \tag{2.13}$$

The subscript "alt" is used to indicate that it is an alternative definition. The rationale behind this is that conductance is then like spectral gap — only where the minimum is taken over $\{0,1\}$-functions (i.e. characteristic functions) instead of all functions from $\Omega$ to $\mathbb{R}$. The last inequality in (2.12) implies that

$$\Phi \leq \Phi_{\text{alt}} \leq 2\Phi$$

and the analogue to Lemma 2.3.5 is

**Corollary 2.3.6**

$$\frac{\Phi_{\text{alt}}^2}{8} \leq \lambda \leq \Phi_{\text{alt}}.$$

Combining Lemmata 2.3.2 and 2.3.5 yields bounds on mixing time in terms of conductance.

**Theorem 2.3.7** *Let $M$ be a finite, irreducible, aperiodic and time-reversible Markov chain with self-loop probabilities $P(x,x) \geq 1/2$ for all states $x$. Let $\Phi$ be the conductance of $M$ as defined in Definition 2.3.4. Then the mixing time of $M$ satisfies*

$$\tau(\varepsilon) \leq \frac{2}{\Phi^2}\left(\ln \frac{1}{\min_{x \in \Omega}\{\pi(x)\}} + \ln \frac{1}{\varepsilon}\right).$$

One way to bound the conductance of a Markov chain $M$ is to exploit the above mentioned connection with edge expansion of a graph. Edge expansion of a graph can be established using so-called *isoperimetric inequalities*. These usually translate immediately into lower bounds on conductance. We will give an example of this in chapter 4.

### 2.3.3 Canonical paths

The last technique for bounding mixing time in this chapter is canonical paths. Consider an irreducible and aperiodic Markov chain $M$ with finite state space $\Omega$, transition probabilities $P(\cdot,\cdot)$ and stationary distribution $\pi$. Assume that $M$ is time-reversible. As mentioned before, $M$ can be considered as an undirected graph $(\Omega, T)$ with vertex set $\Omega$ and edge set defined as in (2.10). For each (ordered) pair $(x,y) \in \Omega^2$, we specify a *canonical path* $\gamma_{xy}$ from $x$ to $y$ in the graph $(\Omega, T)$; the canonical path $\gamma_{xy}$ corresponds to a sequence of legal transitions in $M$ that leads from initial state $x$ to final state $y$ without using any edge more than once. Denote by $\Gamma = \{\gamma_{xy} \mid x,y \in \Omega\}$ the set of all canonical paths. For the method to yield good bounds, it is important to choose a set of paths $\Gamma$ that avoids the creation of *hot spots*: edges of the graph that carry a particularly heavy burden of canonical paths.

In early applications of canonical paths the degree to which an even edge-loading has been achieved was measured by the quantity (see [58])

$$\bar{\rho} = \bar{\rho}(\Gamma) = \max_t \frac{1}{\widetilde{P}(t)} \sum_{\gamma_{xy} \ni t} \pi(x)\pi(y), \tag{2.14}$$

where the maximum is over oriented edges (transitions) $t$ of $(\Omega, T)$. As shown by several authors ([31], [58] and [110]) this choice yields a fairly tight lower bound on conductance:

**Lemma 2.3.8** *For any time-reversible Markov chain and any choice of canonical paths*

$$\Phi \geq \frac{1}{2\bar{\rho}}.$$

Bounds on mixing time can then be obtained by combining Lemma 2.3.8, Cheeger's inequality (Lemma 2.3.5) and Lemma 2.3.2.

Alternative choices of edge-load in conjuction with Poincaré-type inequalities can yield direct bounds on spectral gap. We will give two examples. Sinclair [110] suggested the following choice

$$\overline{\rho}_{\text{Sinc}} = \max_t \frac{1}{\widetilde{P}(t)} \sum_{\gamma_{xy} \ni t} |\gamma_{xy}| \, \pi(x)\pi(y), \tag{2.15}$$

which is almost identical to (2.14) only $|\gamma_{xy}|$ denotes the length of path $\gamma_{xy}$, i.e. the number of edges. With this choice spectral gap is bounded below as follows

**Lemma 2.3.9**

$$\frac{1}{\overline{\rho}_{\text{Sinc}}} \leq \lambda.$$

The proof idea is to deploy the Poincaré inequality characterisation of spectral gap. Details can be found in [110].

Another choice of edge-load was proposed by Diaconis and Stroock [31]

$$\overline{\rho}_{\text{D-St}} = \max_{t} \sum_{\gamma_{xy} \ni t} |\gamma_{xy}|_{\widetilde{P}} \pi(x)\pi(y), \tag{2.16}$$

where $|\gamma_{xy}|_{\widetilde{P}} = \sum_{t \in \gamma_{xy}} \widetilde{P}(t)^{-1}$ and $t$ are oriented edges in $T$. Again, the proof of the following bound (given in [31])

**Lemma 2.3.10**

$$\frac{1}{\overline{\rho}_{\text{D-St}}} \leq \lambda$$

uses the Poincaré inequality characterisation of spectral gap.

Let $\overline{\rho}_{\text{alt}}$ be either as in (2.15) or (2.16), then using Lemma 2.3.2 the following bound on mixing time holds:

**Theorem 2.3.11** *Let $M$ be a finite, irreducible, aperiodic and time-reversible Markov chain with self-loop probabilities $P(x,x) > 1/2$ for all states $x$. Let $\Gamma$ be a set of canonical paths with maximum edge loading $\overline{\rho}_{\text{alt}} = \overline{\rho}_{\text{alt}}(\Gamma)$. Then the mixing time of $M$ satisfies*

$$\tau(\varepsilon) \leq \overline{\rho}_{\text{alt}} \left( \ln \frac{1}{\min_{x \in \Omega}\{\pi(x)\}} + \ln \frac{1}{\varepsilon} \right).$$

We would like to remark that $\overline{\rho}_{\text{Sinc}}$ and $\overline{\rho}_{\text{D-St}}$ coincide in the case of random walks on graphs, whereas in general both quantities can differ significantly. The reason for this seems to be that $\overline{\rho}_{\text{Sinc}}$ has a more "local" nature than $\overline{\rho}_{\text{D-St}}$. This is discussed in more detail and corroborated by examples in [110].

This finishes our recapitulation of "classical" techniques for bounding mixing time. In the following chapters we will focus on recent developments in the area, i.e. improvements in the techniques discussed above and entirely new approaches.

# Chapter 3

# Lozenge tilings and path coupling

The first improvement on a "classical" technique that we are going to introduce is on coupling. Although the coupling technique explained in section 2.2 is conceptually rather simple, in most cases finding the right coupling requires detailed insights into the combinatorics of the given problem and this complexity can make this technique extremely difficult to apply. The main obstacle is the need to design a coupling between all pairs of states which will force $(X_t)$ and $(Y_t)$ to coalesce rapidly while on the other hand meeting conditions (2.4) and (2.5) of Definition 2.2.1. Path coupling, invented by Bubley and Dyer [18], facilitates the combinatorial difficulty of designing a coupling with the desired properties.

## 3.1 Path coupling

With path coupling, we can dispense with devising a coupling for all pairs of states. Instead, a *path*, a sequence of states between an arbitrary pair of states, is defined. Thus, only pairs of states that are adjacent on some path, for which the task of satisfying (2.4) and (2.5) is relatively easy, need to be considered. Note that states adjacent on a path are not necessarily adjacent states in the Markov chain. The path coupling theorem shows that a coupling along a path of adjacent states easily extends to a coupling on arbitrary states. Assume for the following that $M = (\Omega, P)$ is a finite, irreducible and aperiodic Markov chain and all couplings are in the sense of Definition 2.2.1:

**Theorem 3.1.1** *Let $\delta$ be an integer valued metric defined on $\Omega \times \Omega$, which takes values in $\{0, \ldots D\}$. Let $A$ be a subset of $\Omega \times \Omega$ such that for all $(X_t, Y_t) \in \Omega \times \Omega$ there exists a path*

$$X_t = Z_0, Z_1, \ldots, Z_r = Y_t$$

*between $X_t$ and $Y_t$ such that $(Z_l, Z_{l+1}) \in A$ for $0 \leq l < r$ and*

$$\sum_{l=0}^{r-1} \delta(Z_l, Z_{l+1}) = \delta(X_t, Y_t).$$

*Define a coupling $(X,Y) \mapsto (X',Y')$ of the Markov chain $M$ on all pairs $(X,Y) \in A$. Apply[1] this coupling along the given sequence from $X_t$ to $Y_t$ to obtain the new sequence $Z'_0, Z'_1, \ldots, Z'_r$, where $X_{t+1} = Z'_0$ and $Y_{t+1} = Z'_r$. Then $(X_t, Y_t) \mapsto (X_{t+1}, Y_{t+1})$ is a coupling for $M$ on $\Omega \times \Omega$. Moreover, if there exists $\zeta \leq 1$ such that*

$$\mathrm{E}\left[\delta\left(X', Y'\right) \mid (X,Y)\right] \leq \zeta \delta(X,Y)$$

*for all $(X,Y) \in A$, then*

$$\mathrm{E}[\delta(X_{t+1}, Y_{t+1}) \mid (X_t, Y_t)] \leq \zeta \delta(X_t, Y_t).$$

Theorem 3.1.1 and the subsequent Theorem 3.1.2 are taken from [39], which also contains proofs of both. The original formulation of the path coupling theorem can be found in [18]. Theorem 3.1.2, which is a reformulation of the coupling lemma (Lemma 2.2.2), shows how the path coupling theorem (Theorem 3.1.1) translates into mixing times.

**Theorem 3.1.2** *Let $(X_t, Y_t)$ be a coupling for the Markov chain $M$ and let $\delta$ be any integer valued metric defined on $\Omega \times \Omega$. Suppose that there exists $\zeta \leq 1$ such that*

$$\mathrm{E}[\delta(X_{t+1}, Y_{t+1}) \mid (X_t, Y_t)] \leq \zeta \delta(X_t, Y_t)$$

*for all $t$. Let $D$ be the maximum value that $\delta$ achieves on $\Omega \times \Omega$.*

[1] What we mean is the following: First, apply the coupling to obtain $\left(Z'_0, Z'_1\right)$, then for every successive $Z'_{i+1}$, $i = 1, \ldots, r-1$, condition on $Z'_i$ (i.e. assume that $Z_{i+1}$ changes to $Z'_{i+1}$ conditional on $Z_i$ changing to $Z'_i$).

*If* $\zeta < 1$, *then the mixing time* $\tau(\varepsilon)$ *of* $M$ *satisfies*

$$\tau(\varepsilon) \leq \frac{1}{1-\zeta} \ln \frac{D}{\varepsilon}.$$

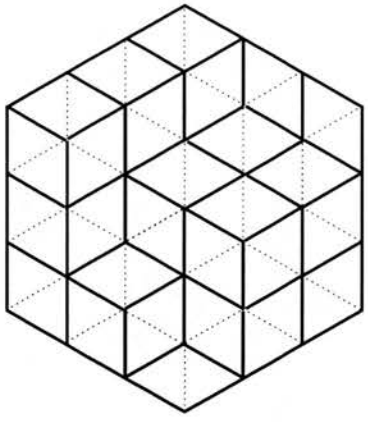
*If* $\zeta = 1$ *and there exists* $\xi > 0$ *such that* $\mathbb{P}\left[\delta(X_{t+1}, Y_{t+1}) \neq \delta(X_t, Y_t)\right] \geq \xi$ *for all* $t$, *then*

$$\tau(\varepsilon) \leq \left\lceil \frac{eD^2}{\xi} \right\rceil \left\lceil \ln \frac{1}{\varepsilon} \right\rceil.$$

## 3.2 Path coupling and lozenge tilings

We apply path coupling to analyse the convergence behaviour of the Markov chain for generating random lozenge tilings on a finite region $S$ of the triangular, 2-dimensional lattice; for simplicity assume throughout that $S$ is hexagonal. A *lozenge* is a rhombus-shaped figure, i.e. it has four equal sides and non-rectangular angles. Here we require the smaller angles to be 60 and the larger 120 degrees such that a lozenge covers two adjacent equilateral (regular) triangles. By *triangular lattice* we denote a lattice structure formed by adjacent, regular triangles. Any covering of a hexagonal region of the triangular lattice with non-overlapping lozenges is called a *lozenge tiling*. Observe that each lozenge in such a tiling covers two adjacent triangles in the lattice and has three possible orientations. An illustration of what we have in mind is illustrated in figure 3.1.

Lozenge tilings are an important tool in statistical physics, where they are equated with configurations of a dimer system on some region $S$ — a dimer being a particle that occupies two adjacent sites in the lattice (for more information see e.g. [52]). Many physical properties of such systems correspond to the expectation w.r.t. the uniform distribution of some function defined on configurations of the system, for example the energy of a configuration depends on the correlation between orientations of tiles. An almost uniform sampler for generating tilings, in conjunction with the classical Monte Carlo method, would allow us to approximate such expectations to arbitrary precision. Moreover, the random sampling of such tilings in itself is worthwhile since it can help us study typical properties of random configurations: for example, a random lozenge tiling on a large hexagonal region will with high probability possess a circular central



Lozenge tiling

Figure 3.1: Lozenge tiling

region in which the lozenges are arranged in a disorderly fashion whereas in the outer region the tiles line up, i.e. have all the same orientation; This property is known as the Artic Circle Theorem [64].

Using two different approaches several authors have worked on the problem of generating (general) random tilings: one approach (cf. Ciucu & Propp [28] and Wilson [116]) uses linear algebra; to give you a flavour: the number of tilings can be calculated exactly using the Gessel-Viennot method[2] [43], this information can then be used to generate a random tiling (cf. Guénoche [48] or Jerrum et al. [63]). The other approach, pursued by Jerrum & Sinclair [58], Kenyon, Randall & Sinclair [71], Luby, Randall & Sinclair [84], Propp & Wilson [100] and Wilson [117], uses Markov chains.

One of the first results for bounding the mixing time of Markov chains for, in particular, randomly sampling lozenge tilings was by Luby, Randall and Sinclair [84]. Their bound on the mixing time was $\tau(\varepsilon) \leq 8en^4\left(\ln \varepsilon^{-1}\right)$, which they subsequently in unpublished work improved to $O\left(n^{3.5}\right)$, where $n$ is asymptotically[3] the number of

[2]The Gessel-Viennot method computes the number of so-called routings by evaluating a suitable determinant and we will see in section 3.2.1 that there exists a bijection between routings and tilings.

[3]To be precise, $n$ is the size of the region on which the so-called routings are defined. In the case of a lozenge tiling on a hexagon of sidelength $l$, the number of lozenges is $3l^2$ and the size of the region on which the routings are defined is $3l^2 - 2l$.

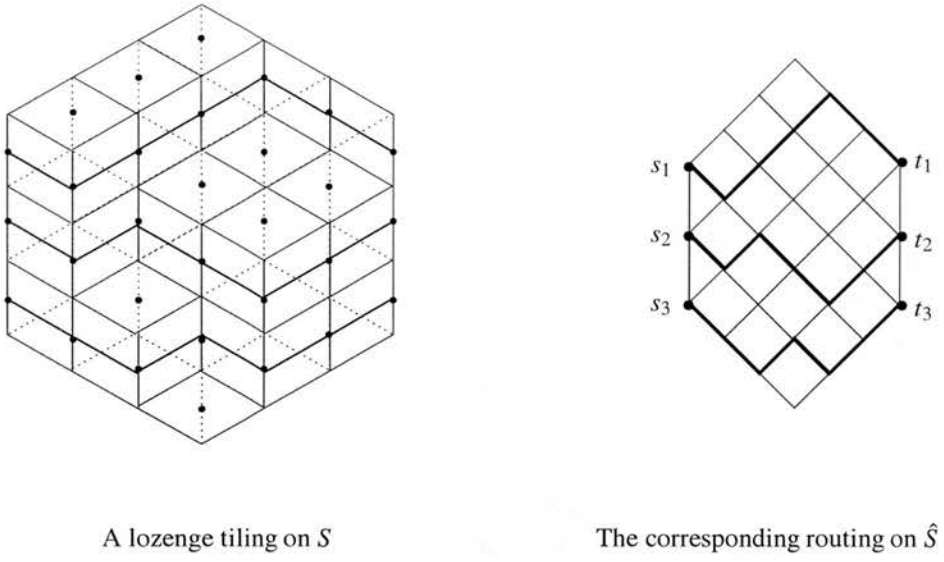
lozenges. In fact, their method of analysing the mixing time of the chain proved to be very general and could be applied to the analysis of Markov chains for generating domino tilings on a region of the Cartesian lattice (e.g. an $n \times n$ chessboard) and also to the analysis of chains for generating Eulerian orientations of a region of the Cartesian lattice (another way for generating Eulerian orientations is given by Mihail & Winkler [94]). Their framework is two-pronged: first, they identify tilings on a lattice with routings on a related lattice. Second, they interpret moves of the Markov chain on tilings as manipulations of routings on the associated lattice. Subsequent work by Wilson [117] improved the upper bound on the mixing time for the lozenge tiling chain to $\Theta\left(n^2 \ln n\right)$.

The framework provided by Luby et al. can be made to work with a path coupling argument, too, with only slight alterations to the original analysis. The bounds on mixing time we deduce are similar to Luby et al.'s, nevertheless the path coupling approach we propose leads to a simplification of the mathematics and case analyses involved. At the end of the day path coupling lives up to its reputation and delivers a good example of how the overall complexity of the analysis is reduced. In the case of domino tilings and Eulerian orientations, unfortunately, a similar approach runs into problems.

### 3.2.1 Lozenge tilings and lozenge routings

We begin our analysis by a recapitulation of the 1-1-correspondence between tilings and routings exhibited by Luby et al. Given a finite, simply connected, hexagonal region $S$ of the 2-dimensional triangular lattice, we define an associated region $\hat{S}$ of the Cartesian lattice as follows. The vertices of $\hat{S}$ correspond to the midpoints of the vertical edges in $S$ and two vertices in $\hat{S}$ are connected if the corresponding points in $S$ lie on adjacent triangles. Figure 3.2 depicts an instance of such an association where $S$ has side length 3.

The vertices of $\hat{S}$ that correspond to vertical edges on the boundary of $S$ are called *sources* and *sinks*: a vertex $v$ is a source if the interior of $\hat{S}$ lies to the right of $v$ and a sink if the interior of $\hat{S}$ lies to its left. As figure 3.2 suggests, sources and sinks can be labelled appropriately by numbering from top to bottom. Let us call the $i$-th source




A lozenge tiling on $S$ The corresponding routing on $\hat{S}$

Figure 3.2: Lozenge tilings and routings

and sink $s_i$ and $t_i$ respectively.

Assume that there are $k$ sinks and sources in $\hat{S}$. A *lozenge routing* of $\hat{S}$ is a set of $k$ non-intersecting (i.e. no repetition of vertices nor edges) shortest paths on the Cartesian lattice within $\hat{S}$ from $s_i$ to $t_i$ for each $i$. Figure 3.2 suggests that there is a correspondence between tilings and routings in lattices of associated regions. In fact this correspondence is a bijection. First observe that the lozenges covering $S$ have only two orientations: they lie either *horizontally* such that there are no vertical edges or *vertically* in which case two edges of the lozenge stand vertically. Given a lozenge tiling, mark every vertical lozenge with a line "connecting" the midpoints of the two vertical edges of the two triangles covered by the lozenge. Vertical edges of lozenges in the tiling can only be adjacent to other vertical lozenges; thus the markings join up to form a continuous line from one vertical boundary of $S$ to the other. Due to the above mentioned rule governing the adjacency of lozenges in a valid tiling it is obvious that a region with $k$ vertical edges on each side can only be covered by tilings with $k$ non-intersecting lines running all the way from boundary to boundary. Each line in a tiling on $S$ gives rise to a shortest path connecting a source to a sink in $\hat{S}$ and hence collectively the $k$ (non-intersecting) lines give rise to a routing in $\hat{S}$. To obtain the converse, cover $S$ with lozenges such that the markings line up in accordance with a

given routing in $\hat{S}$. Observe that the boundary of any uncovered region of $S$ can only be covered by horizontal lozenges, which ultimately forces us to cover all uncovered regions of $S$ with horizontal lozenges only. This bijection between lozenge tilings and routings illustrated in figure 3.2 has been formalised by Sachs et al. (see [1] and [65]):

**Theorem 3.2.1** *The set of lozenge tilings of $S$ corresponds bijectively with the set of lozenge routings of $\hat{S}$.*

### 3.2.2 Generating lozenge tilings

Following Luby et al. we do not analyse a Markov chain proper for generating lozenge tilings but its equivalent chain on lozenge routings. Furthermore we will not analyse the most *natural* Markov chain for generating lozenge tilings, which would be the one progressing from one tiling to another by rotating three neighbouring lozenges forming a hexagon at each step. In the routing this would correspond to inverting a "peak" or "valley". Denying the simplicity of its local moves, unfortunately, this Markov chain turns out to be very hard to analyse directly. Although empirical evidence has been suggesting rapid mixing, this has only recently been proved rigorously by Randall and Tetali [102]. Nonetheless, the difficulty of analysing this chain directly is reflected in the fact that their bounds on mixing time for this chain could only be obtained by comparing it to another chain whose mixing time was known. Our approach will be based on Luby et al.'s Markov chains whose local moves are tower rotations. The moves of the natural chain then simply correspond to the special case of rotating a tower of height 1.

Fundamental to the analysis is the concept of towers and tower rotations. In the routings lattice, define the *cell at* $(x,y)$ to be the edges connecting $(x,y),(x+1,y+1),(x,y+2)$ and $(x-1,y+1)$. A *tower of height* $h$ is a connected set of cells at the points $(x,y),(x,y+2),\ldots,(x,y+2(h-1))$, where either the points $(x,y),(x,y+2),\ldots,(x,y+2(h-1))$ are all valleys and the point $(x,y+2h)$ does not lie on the routing, or the points $(x,y+2h),(x,y+2h-2),\ldots,(x,y+2)$ are all peaks and the point $(x,y)$ does not lie on the routing. We call the points $(x,y)$ and $(x,y+2h)$ the *bottom* or *top* of the tower respectively. A rotation of a tower of height $h$ is then carried out by inverting each of its $h$ valleys into peaks or vice versa (see figure 3.3).

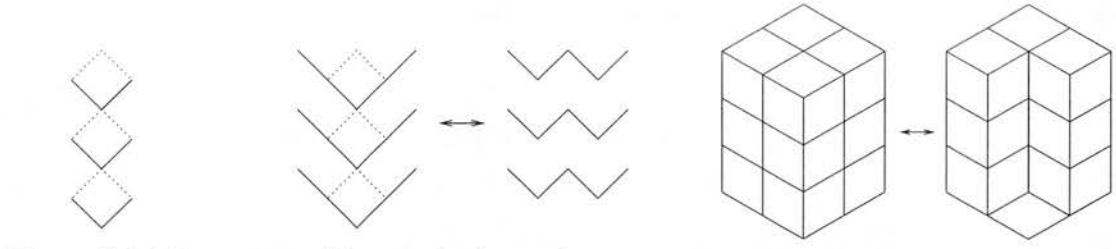



Figure 3.3: tower rotation

Define a Markov chain $M_{\text{loz}}$ for generating random Lozenge tilings as follows. The states $\Omega$ of the chain are lozenge routings on $\hat{S}$ and the transition probabilities $P(\cdot,\cdot)$ are defined by

$$P(R_1,R_2) = \begin{cases} 1/2Nh, & \text{if } R_1,R_2 \text{ differ by a tower rotation of} \\ & \text{height } h \\ 1-\sum_{R\neq R_1} P(R_1,R), & \text{if } R_1 = R_2, \end{cases} \tag{3.1}$$

where $R_1,R_2$ are lozenge routings and $N$ is the total number of non-sink/source vertices on the paths in the routing. Notice that $N$ is constant for all routings $R$ on $S$. The moves of $M_{\text{loz}}$ may be implemented as follows: given a routing $R$

1. choose both a point $p \in \{1,\ldots,N\}$ and number $r \in [0,1]$ u.a.r.

2. If $r \leq 1/2$ and $p$ is a valley, i.e. bottom of a tower of height $h$, then if $r \leq 1/2h$ rotate the tower (observe that owing to the hexagonal shape of $S$ tower rotations are always possible). On the other hand, if $r > 1/2$ check whether $p$ is a peak (and thus the top of a unique tower of height $h$), then if $r > 1-1/2h$ rotate this tower. In all other cases do nothing.

Luby et al. showed that $M_{\text{loz}}$ is finite, irreducible, aperiodic and time-reversible: Since the region $\hat{S}$ or $S$, respectively, is finite, the finiteness of $M_{\text{loz}}$ is obvious. Also, each routing has self-loop probability greater than $1/2$, which ensures aperiodicity. To show irreducibility it suffices to exhibit the existence of a routing that is reachable from every other routing. To this end introduce a partial order "$\preceq$"; a routing $R$ is lower than $R'$,

or $R \preceq R'$ in symbols, if all paths in $R$ are lower than those of $R'$. There exists a lowest routing on $\hat{S}$ (namely the one where all paths go down and then up) and it is not hard to see that it is reachable from every other routing. Then irreducibility follows from the fact that all the moves of $M_{\text{loz}}$ can be reversed. Finally, time-reversibility holds as the transition probabilities and the uniform distribution on the set of routings on $\hat{S}$ meet the detailed balance condition of Definition 2.1.6. Hence $M_{\text{loz}}$ indeed converges towards the uniform distribution on lozenge tilings.

### 3.2.3 $M_{\text{loz}}$ is rapidly mixing

We will bound the mixing time of $M_{\text{loz}}$ via path coupling. To meet the provisos of the path coupling theorem (Theorem 3.1.1), we have to define an adjacency set $A \subseteq \Omega \times \Omega$ and a metric $\delta(\cdot,\cdot)$ on $\Omega \times \Omega$. Define the adjacency set $A$ to be the set of all routings on $\hat{S}$ that differ by exactly one cell, i.e. if $R$ and $R'$ are routings on $\hat{S}$, then $R$ and $R'$ disagree at only one cell, where $R$ has a peak and $R'$ a valley (as depicted in figure 3.4 where $R$ is solid and $R'$ dotted).

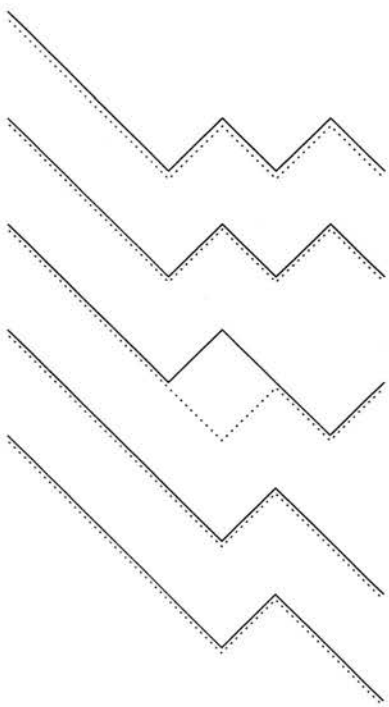


Figure 3.4: Adjacent routings

For two arbitrary routings $R$ and $R'$ define $\delta(R,R')$ as the sum of the number of cells between corresponding paths. Observe that for any two $X_t$, $Y_t \in \Omega$ it is possible to get from $X_t$ to $Y_t$ just by successively inverting one peak into a valley, and vice versa, at a time. Thus, for all pairs of routings $(X_t,Y_t) \in \Omega \times \Omega$ there exists a path $X_t = Z_0, Z_1, \ldots, Z_r = Y_t$ between $X_t$ and $Y_t$ such that $(Z_l, Z_{l+1}) \in A$ for all $0 \leq l < r$ and $\sum_{l=0}^{r-1} \delta(Z_l, Z_{l+1}) = \delta(X_t, Y_t)$.

Designing a coupling for $M_{\text{loz}}$ on $A$ is now very straightforward. Assume that $(R,R') \in A$ and choose the same[4]random point $p$ on both routings, say the $i$-th point of the $j$-th path, and the same random bit $r \in [0,1]$. Then update each routing by rotating at point $p$ with the appropriate probability as determined by the random number $r$. Since all conditions of Theorem 3.1.1 are met, this coupling on $A$ can be extended to a coupling $(X_t,Y_t)$ of $M_{\text{loz}}$ on $\Omega \times \Omega$ as described.

With path coupling, it is imperative to bound $\mathrm{E}[\delta(X_{t+1},Y_{t+1}) \mid (X_t,Y_t)]$ for all $(X_t,Y_t) \in A$. Assume that $R_1$ and $R_2$ are two adjacent routings on $\hat{S}$, i.e. $(R_1,R_2) \in A$, and therefore differ in exactly one cell, say the one at $(x_d,y_d)$. W.l.o.g. there will be two towers above this cell: one of height $d'_u$, in say $R_1$, stretching upward with peaks at points $(x_d,y_d+2)$ to $(x_d,y_d+2d'_u)$, and another of height $d'_u-1$ in $R_2$ starting from $(x_d,y_d+4)$; and two towers below this cell: a tower of height $d'_l$ in $R_2$ facing downward with valleys at $(x_d,y_d)$ down to $\left(x_d,y_d-2\left(d'_l-1\right)\right)$ and another of height $d'_l-1$ in $R_1$ from $(x_d,y_d-2)$ down to $\left(x_d,y_d-2\left(d'_l-1\right)\right)$. Call these the *upper* and *lower primary-disagreement towers.* In figure 3.5 the taller upper primary-disagreement tower is given by $p'_{u_1}\ldots p'_{u_5}$, the shorter by $p'_{u_2}\ldots p'_{u_5}$ and the taller and shorter lower primary-disagreement towers by $p'_{l_1}\ldots p'_{l_4}$ and $p'_{l_2}\ldots p'_{l_4}$ respectively.

The importance of primary-disagreement towers is that if the coupling of $M_{\text{loz}}$ chooses an internal point on one of these towers, then $\delta(R'_1,R'_2)-\delta(R_1,R_2)$, the change in distance between $R_1$ and $R_2$, lies in $\left[-1,\max\left\{d'_u,d'_l\right\}-1\right]$. This is because the shorter or sub-towers of the shorter (upper or lower) primary-disagreement towers may be rotated while the taller tower remains unchanged.

[4]"Same" here means we choose in each routing the $i$-th point of the $j$-th path. We do not mean the same lattice point! Because $R$ and $R'$ differ at exactly one point, there will be a lattice point that is on $R$ but not on $R'$.

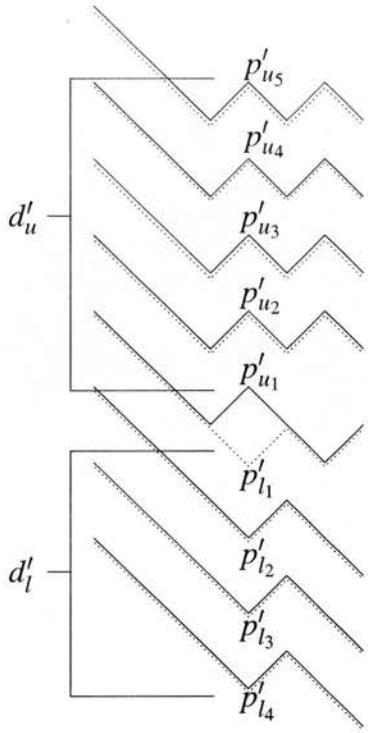



Figure 3.5: Primary-disagreement towers

For a moment suppose that changes in the difference between $R_1$ and $R_2$ can only[5] be caused by tower rotations involving points on the primary-disagreement towers and calculate the expected change in distance $\mathrm{E}[\delta'(R'_1,R'_2)-\delta'(R_1,R_2)]$, here the measure $\delta'$ is only defined on routings that differ by a tower rotation involving points on primary-disagreement towers and $R'_1$, $R'_2$ are obtained from $R_1$ and $R_2$ by applying the coupling on $A$. First, consider the contributions from the upper primary-disagreement towers measured by $\delta'_u$, which is like $\delta'$ only restricted to the upper primary-disagreement towers. Our aim is to compute the following probabilities:

$$\mathbb{P}\left[\Delta\delta'_u=-1\right]=\mathbb{P}\left[\delta'_u\left(R'_1,R'_2\right)-\delta'_u(R_1,R_2)=-1\right]$$

and

$$\mathbb{P}\left[\Delta\delta'_u=i\right], \qquad \text{for } i=1\ldots d'_u-1.$$

We neglect $\mathbb{P}[\Delta\delta'_u=0]$ as this event does not contribute to $\mathrm{E}[\Delta\delta']=$

[5] As we will see later, on p. 42, the total change is caused by primary- and secondary-disagreement tower-rotations.

$\mathrm{E}[\delta'(R_1', R_2') - \delta'(R_1, R_2)]$. The probability $\mathbb{P}[\Delta\delta_u' = -1]$ is nothing else than the probability that the chain "couples", i.e. $R_1' = R_2'$. This occurs when the taller upper primary-disagreement tower or sub-towers of it are rotated as the shorter tower will invariably follow suit or when the tower of height one with peak at $(x_d, y_d + 2)$ is rotated. Hence

$$\mathbb{P}\left[\Delta\delta_u' = -1\right] = \frac{1}{N}\sum_{i=1}^{d_u'} \frac{1}{2i}. \tag{3.2}$$

As for the other probabilities, note that $\mathbb{P}[\Delta\delta_u' = i]$ is the probability of the sub-tower $(x_d, y_d + 4), \ldots, (x_d, y_d + 2(i+1))$ of the shorter upper primary-disagreement tower being rotated while the overlapping sub-tower of the taller upper primary-disagreement tower, $(x_d, y_d + 2), \ldots, (x_d, y_d + 2(i+1))$, is not. With the transition probabilities of $M_{\mathrm{loz}}$ as defined in (3.1), clearly

$$\mathbb{P}\left[\Delta\delta_u' = i\right] = \frac{1}{N}\left(\frac{1}{2i} - \frac{1}{2(i+1)}\right), \quad \text{for } i = 1, \ldots, d_u' - 1. \tag{3.3}$$

Using (3.2) and (3.3), the expected change in distance only taking into account the contribution from rotations of the upper primary-disagreement towers is

$$\begin{aligned}
\mathrm{E}\left[\Delta\delta_u'\right] &= \frac{1}{N}\left(\sum_{i=1}^{d_u'-1}\left(\frac{1}{2i} - \frac{1}{2(i+1)}\right) i - \sum_{i=0}^{d_u'-1} \frac{1}{2(i+1)}\right) \\
&= \frac{1}{N}\left(\sum_{i=1}^{d_u'-1}\left(\frac{1}{2} - \frac{i}{2(i+1)} - \frac{1}{2(i+1)}\right) - \frac{1}{2}\right) \\
&= -\frac{1}{2N}.
\end{aligned}$$

Define $\delta_l'$ as the counterpart of $\delta_u'$ on the lower primary-disagreement towers. The same calculations as above yield

$$\mathrm{E}\left[\Delta\delta_l'\right] = -\frac{1}{2N}.$$

Finally, putting all of the above together the change in distance caused by rotations of primary-disagreement towers is

$$\mathrm{E}\left[\Delta\delta'\right] = \mathrm{E}\left[\Delta\delta_u'\right] + \mathrm{E}\left[\Delta\delta_l'\right] = -\frac{1}{N}.$$

However, changes in distance between $R_1$ and $R_2$ can also be caused by *secondary-disagreement towers*. There are always two of them — one on each side of the cell

where $R_1$ and $R_2$ differ. In figure 3.6 there is a left secondary-disagreement tower, $p''_{l_1} \dots p''_{l_5}$, of height $d''_l$ starting from $(x_d - 1, y_d + 1)$ and a right one, $p''_{r_1} \dots p''_{r_4}$, of height $d''_r$. Furthermore, rotations of secondary-disagreement towers can only change the distance by $d''_l$ or $d''_r$ respectively.

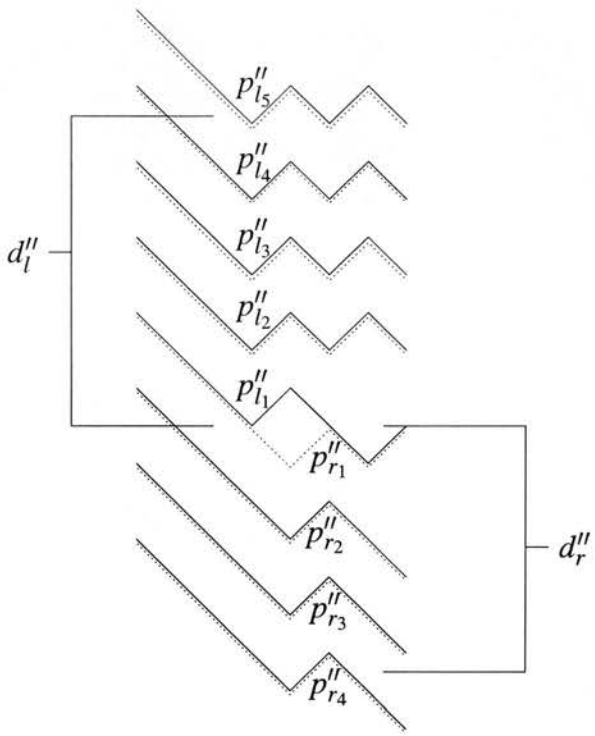



Figure 3.6: Secondary-disagreement towers

To realise that there can be only two secondary-disagreement towers, a case analysis is in order. It will involve a study of the immediate surroundings of the cell where $R_1$ and $R_2$ differ. Drawing attention to the paths right above and below this cell, notice that these can only be laid out in 16 different ways. Identifying symmetrical routings leaves us with seven distinct cases, which are depicted in figure 3.7.

Of these a), c), e) and g) allow two secondary-disagreement towers of arbitrary height, while in b), d) and f) one secondary-disagreement tower will always be restricted to height one. Similarly to $\delta'$, let $\delta''$ be only defined on routings that differ by rotations of secondary-disagreement towers and assume that $\delta''_l$, $\delta''_r$ are the analogue for secondary-disagreement towers to $\delta'_u$ and $\delta'_l$. As mentioned above secondary-disagreement towers can only be rotated in their entirety. Since the probability that

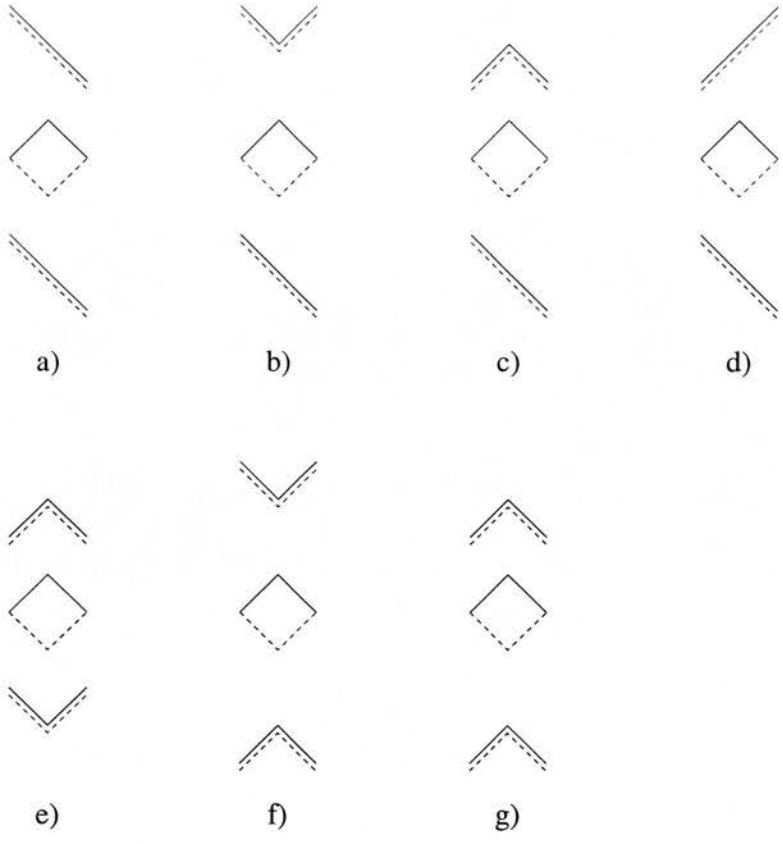



Figure 3.7: Distinct, adjacent routings

the left secondary-disagreement tower is rotated is $1/2Nd_l''$ and the probability that the right one is rotated $1/2Nd_r''$, the expected change caused by secondary-disagreement tower rotations is

$$\mathrm{E}\left[\Delta\delta_l''\right] = \mathrm{E}\left[\Delta\delta_r''\right] = \frac{1}{2N}.$$

As ultimately the overall change in distance between two adjacent routings is caused by rotations of primary- and secondary-disagreement towers, the total expected change in distance for adjacent routings is

$$\mathrm{E}[\Delta\delta] = \mathrm{E}\left[\Delta\delta'\right] + \mathrm{E}\left[\Delta\delta''\right] = 0. \tag{3.4}$$

Now recall the path coupling theorem (Theorem 3.1.1) and note that (3.4) ensures

$$\mathrm{E}[\delta(X_{t+1}, Y_{t+1}) \mid (X_t, Y_t)] \leq \zeta\delta(X_t, Y_t)$$

with $\zeta = 1$. Thus the second case of Theorem 3.1.2 applies. To obtain a bound on mixing time, it remains to estimate $\mathbb{P}[\delta(X_{t+1}, Y_{t+1}) \neq \delta(X_t, Y_t)]$. A very crude lower bound on this probability can be deduced from (3.2) and the following assumptions: first, assume that the number of (full) cells in $\hat{S}$ is $n$ — observe that the half cells at the vertical boundaries of $\hat{S}$ complement each other. Additionally, assume w.l.o.g. that given a pair of routings $(X_t, Y_t)$ on $\hat{S}$ their distance satisfies $\delta(X_t, Y_t) \geq 1$. Then

$$\begin{aligned}\mathbb{P}[\delta(X_{t+1}, Y_{t+1}) \neq \delta(X_t, Y_t)] &\geq \mathbb{P}\left[\Delta\delta'_u = -1\right] + \mathbb{P}\left[\Delta\delta'_l = -1\right] \\ &\geq \frac{1}{n},\end{aligned}$$

where the second inequality follows from the fact that $N \leq n$. The last bit of information required for Theorem 3.1.2 is $D$, the maximal value the metric $\delta$ can achieve. According to Luby et al. [84] this quantity is upper bounded by $n^{3/2}$, a result which they obtain by using so-called *height-functions*. In the rather special case of hexagonal lattice regions a similar bound can achieved with a maximization argument: Note that any hexagonal region $\hat{S}$ (containing $n$ cells) of the Cartesian lattice can be constructed from a square of side length $s = l + k$ by removing a square of side length $l$ as depicted in figure 3.8 (the solid lines represent $\hat{S}$ and the dotted lines the square of side length $l$).

Furthermore, the following hold: $n = (k+l)^2 - l^2$ and, if $h$ is the height of the vertical boundary of $\hat{S}$, then by Pythagoras's Theorem $h^2 = 2l^2$. Observe that in a hexagonal region $\hat{S}$ the maximal distance between two routings is achieved if in one routing all paths go up and then down while in the other all paths go down and then up (the distance between each path is then $k^2$). Since a routing on a hexagonal region of height $h$ has exactly $h$ paths, the maximum of the metric $\delta$ is,

$$\begin{aligned}D &= hk^2 \\ &= \sqrt{2}lk^2 \\ &= \sqrt{2}\frac{n-k^2}{2k}k^2 = \frac{\sqrt{2}}{2}\left(nk - k^3\right).\end{aligned}$$

Since our goal is to find the maximal distance between two routings in a region of size $n$, we can consider $n$ to be a constant and maximise $D$ w.r.t. $k$:

$$D' = \frac{\sqrt{2}}{2}\left(n - 3k^2\right),$$

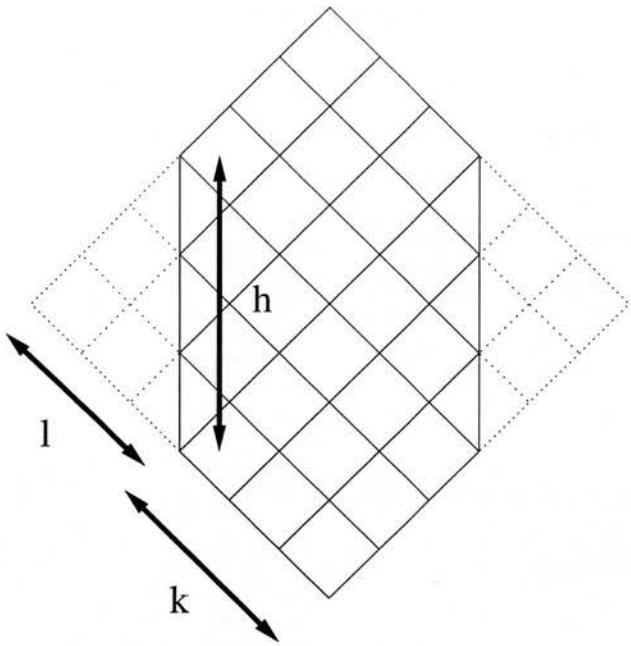



Figure 3.8: constructing a hexagon from a square

$$D'' = -3\sqrt{2}k$$

reveal that the maximum is attained at $k = \sqrt{n/3}$. So that finally,

$$\begin{aligned} D &= \frac{\sqrt{2}}{2}\left(n\left(\sqrt{\frac{n}{3}}\right) - \left(\sqrt{\frac{n}{3}}\right)^3\right) \\ &= \frac{\sqrt{2}}{2}\left(\frac{\sqrt{n^3}}{\sqrt{3}} - \frac{\sqrt{n^3}}{3\sqrt{3}}\right) \\ &= \frac{\sqrt{2n^3}}{3\sqrt{3}}. \end{aligned}$$

Piecing everything together, our bound on the mixing time of $M_{\text{loz}}$ is stated below.

**Theorem 3.2.2** *Let $M_{\text{loz}}$ be a Markov chain on the set of lozenge routings on a region $\hat{S}$ of size $n$ of the Cartesian lattice with transition probabilities defined as in (3.1). Then its mixing time is upper bounded by*

$$\tau(\varepsilon) \leq n^4 \left\lceil \ln \frac{1}{\varepsilon} \right\rceil.$$

Compared to Luby et al.'s original result ([84], Theorem 3.10), which bounded mixing time by $8en^4 \left\lceil \ln \frac{1}{\varepsilon} \right\rceil$, our result is a slight improvement and the analysis is more rigorous. However, as mentioned before Luby et al. managed to improve their result in unpublished work to $O\left(n^{3.5}\right)$. The best known upper bound for the mixing time of this Markov chain was given by Wilson [117]: $\tau(\varepsilon) \in \Theta\left(n^2 \ln n\right)$.

# Chapter 4

# Average conductance and balanced matroids

In this chapter we are going to discuss an improvement on conductance (cf. section 2.3.2) and apply it to obtain better bounds on the mixing time of the random walk on hypercubes and the bases-exchange walk on balanced matroids. As in the past sections our working assumption will be that the Markov chains in question are finite, irreducible, aperiodic and time-reversible.

## 4.1 Average conductance

The notion of conductance, first proposed by Jerrum and Sinclair [111], reflects the tendency of a Markov chain to escape from a set in one step, once it has reached its stationary distribution. In its original formulation, however, the definition is too pessimistic in the sense that it only considers the conductance of the worst subset of states. The key idea of *average conductance* introduced by Kannan and Lovász [68] is that quite often a better inequality holds for very small (or very large) sets. This observation can be expressed in terms of a new concept called the *conductance function*:

**Definition 4.1.1** *Let $\widetilde{P}(\cdot,\cdot)$ be defined as in (2.11) and $\Omega$ be the state space of a Markov*

*chain $M$ with stationary distribution $\pi$. Then*

$$\Phi(x) = \min_{\substack{S\subset\Omega \\ 0<\pi(S)\leq x}} \frac{\widetilde{P}(S,\overline{S})}{\pi(S)}, \tag{4.1}$$

*is called the* conductance function *of $M$.*

The above definition is due to Montenegro (see e.g. [96]). The original definition by Kannan and Lovász [68] was $\Phi_{\text{orig}}(x) = \min_{\substack{S\subset\Omega \\ 0<\pi(S)\leq x}} \frac{\widetilde{P}(S,\overline{S})}{\pi(S)\pi(\overline{S})}$. Note however that both versions of the conductance function are monotonically decreasing in $x$ and $\Phi(x) = \Phi$ (and $\Phi_{\text{orig}}(x) = \Phi_{\text{alt}}$) for $x \geq 1/2$, which reflects the observation we made above.

It does not come as a surprise that average conductance can be used for obtaining bounds on mixing time. Rather unusual though is that Kannan and Lovász express their bounds on mixing time in terms of hitting time, which is a concept that belongs to a somehow different framework for bounding convergence rates of Markov chains from what we have encountered so far. This framework was introduced by Lovász and Winkler [82]. Fundamental is the concept of so called stopping rules. A *stopping rule* is a rule that observes the Markov chain and tells us whether to stop or not, depending only on the past evolution of the chain. Such a decision can for example be made using coin flips. Then the stopping rule has to specify for each finite evolution of the chain the probability of continuing the chain so that with probability 1 the chain is stopped after a finite number of steps.

**Definition 4.1.2** *For two probability distributions $\mu$, $\nu$ on $\Omega$ the* hitting time *from $\mu$ to $\nu$, in symbols $\mathcal{H}(\mu,\nu)$, is the minimum expected number of steps of any stopping rule $\Gamma$ such that if the chain is started with distribution $\mu$ and the stopping rule $\Gamma$ is observed, then the chain stops exactly in distribution $\nu$.*

*The* mixing time[1] *$\mathcal{H}$ of the Markov chain is then defined as*

$$\mathcal{H} = \max_{\mu} \mathcal{H}(\mu,\pi).$$

[1]The mixing time $H$ is not to be confused with the mixing time $\tau$, which was defined in chapter 2, p.18. To avoid possible confusion, we will use the terms mixing time $H$ and mixing time $\tau$ when it is not clear from the context which definition of mixing time is meant.

Next, we will bound mixing time by average conductance. Instead of Kannan and Lovász's original result (see [68]), which contained a slight mistake, we will use a corrected and slightly improved version by Montenegro [96]. Define $\pi_0 = \min_{x\in\Omega} \pi(x)$.

**Theorem 4.1.3** *Define the conductance function $\Phi(x)$ as in (4.1). The mixing time $\mathcal{H}$ of $M$ can be bounded by*

$$\mathcal{H} \leq K\left(14\int_{\pi_0}^{\frac{1}{2}} \frac{dx}{x\Phi^2(x)} + \frac{4}{\Phi}\right),$$

*where $K$ is a constant independent of the Markov chain.*

Since $\Phi(x) \geq \Phi$, it follows immediately that average conductance will yield asymptotically better bounds than conductance:

$$\begin{aligned}
\mathcal{H} &\in O\left(\frac{1}{\Phi^2}\int_{\pi_0}^{1/2}\frac{dx}{x} + \frac{1}{\Phi}\right) && (4.2)\\
&\in O\left(\frac{1}{\Phi^2}\ln\frac{1}{\pi_0} + \frac{1}{\Phi}\right) && (4.3)\\
&\in O\left(\frac{1}{\Phi^2}\ln\frac{1}{\pi_0}\right), && (4.4)
\end{aligned}$$

where the last inequality is due to Lemma 2.3.5. This formalizes the intuition that by averaging conductance, i.e. exploiting the fact that smaller sets may have larger conductance, gains can be made.

At this point the question: "Where has $\tau$ gone?" is truly justified. Lovász and Winkler showed in [82] that $\mathcal{H}$ can be related to $\tau$ using using a "blind" stopping rule. A blind stopping rule is a stopping rule that says "stop" after a specified number of steps, irrespective of the evolution of the chain. They showed that for any initial distribution $\mu$ if $M$ is stopped after $t = c\,\mathcal{H}\ln\frac{1}{\varepsilon}$ steps, where $c$ is a constant independent of the chain, and $\mu^t$ is the distribution reached, then $\|\mu^t - \pi\|_{\mathrm{TV}} \leq \varepsilon$. Thus

$$\tau(\varepsilon) \leq c\,\mathcal{H}\ln\frac{1}{\varepsilon}. \qquad (4.5)$$

To obtain bounds on mixing time $\tau$ using average conductance, we have to combine Theorem 4.1.3 with (4.5). For practical purposes this involves bounding the constants $c$

and $K$, or their product $c \cdot K$ to be more precise. Early bounds suggested that $c \cdot K$ is very large. A taste of how large is given by Kannan [67]:

$$\tau(1/4) \leq 2000 \int_{\pi_0}^{\frac{3}{4}} \frac{dx}{x\Phi^2(x)}.$$

A recent result by Morris and Peres [97] shows that bounds on $c \cdot K$ can be significantly smaller. Over the next few sections we will compare average conductance to "classical" conductance.

## 4.2 Average conductance of a hypercube

We start with a simple example that has been thoroughly studied and is now well understood. Consider the random walk on the $n$-dimensional hypercube[2]. Since the hypercube in $n$ dimensions has got $2^n$ vertices, each vertex corresponds to a bit-vector of length $n$. Thus the random walk on the hypercube can be modelled by a Markov chain $M_{\text{cube}}$ as follows

1. With probability $1/2$ let $X_t = X_{t+1}$, otherwise
2. choose $i \in \{1,\ldots,n\}$ u.a.r. and let $X_{t+1} = (x_1,\ldots,1-x_i,\ldots,x_n)$, where $x_i \in \{0,1\}$.

Obviously, $M_{\text{cube}}$ is finite, aperiodic and irreducible. Furthermore, it is time-reversible as its transition probabilities and the uniform distribution on $\Omega = \{0,1\}^n$ are in detailed balance.

To see how much can be gained by using average conductance, let us first bound the mixing time with conductance. There are several ways to determine the conductance of the cube-walk. Those of us with much time on hand might want to play around with various ways to bipartition the state space until they have convinced themselves that splitting $\Omega$ into two sets, say $S$ and $\overline{S}$, such that the points of each of set can be regarded as vertices of a cube of dimension $n-1$, is the right answer (because then

[2] "Cubes" of dimension $d > 3$ are usually called hypercubes. A $d$-dimensional cube has vertex set $\{0,1\}^d$; two bitvectors $x, y \in \{0,1\}^d$ are adjacent if and only if they differ in exactly one bit position.

each vertex of $S$ contributes only one edge to the cutset of $S$ and $\overline{S}$ and for symmetry reasons anything less than that is not possible. (cf. Harper [50])).

On the other hand, there is a more rigorous approach. As mentioned before, *isoperimetric inequalities* are well-suited to the purpose of bounding conductance. Isoperimetric inequalities are used in geometry to relate the interior of an object to its boundary or *perimeter*. An example of such an inequality, which fits the bill for us, is due to Dyer and Frieze [36]. Before stating the inequality we should introduce the following notations: If $\|\cdot\|$ is a norm, then its *dual*, $\|\cdot\|^*$, is defined by

$$\|\cdot\|^* = \sup\{ax \mid \|a\| = 1\}.$$

Furthermore, a function $f$ is log-concave if $\log f$ is concave. And finally, if $K$ is a body, then $\partial K$ denotes its boundary.

**Lemma 4.2.1** *Suppose $K \subseteq \mathbb{R}^n$ is a convex body and $f$ a* log*-concave function on the interior of $K$. For a set $S \subseteq K$ such that $\sigma = \partial S \setminus \partial K$ is a piecewise smooth surface, define $\kappa(S) = \int_S f(x)\,dx$ and $\kappa'(S) = \int_\sigma f(x)\,\|u(x)\|^*\,dx$, where $u(x)$ is the Euclidean unit normal to $\sigma$ at $x \in \sigma$. If $\kappa(S) \leq \frac{1}{2}\kappa(K)$, then*

$$\frac{\kappa(S)}{\kappa'(S)} \leq \frac{1}{2}\mathrm{diam}(K),$$

*where the diameter is measured w.r.t. the (primal) norm $\|\cdot\|$.*

Jerrum and Sinclair [59] used Lemma 4.2.1 to bound the conductance of $M_{\text{cube}}$. Their approach is as follows: First transform the state space $\Omega$ into a convex body $K$. There are several ways to achieve this. Here consider the following construction: For $b \in \Omega = \{0,1\}^n$ let

$$C(b) = \{x \in \mathbb{R}^n \mid |x_i - b_i| \leq 1/2 \text{ for } i = 1 \ldots n\}.$$

The $C(b)$'s are hypercubes of side length 1 centered on elements of $\Omega$. For $S \subseteq \Omega$ set $C(S) = \bigcup_{b \in S} C(b)$. Then obviously $C(\Omega) = K$ is a side length 2 hypercube. Moreover, since it is made up of the $C(b)$'s, in three dimensions it looks a bit like the Rubik's© cube, only that is a $3 \times 3 \times 3$ cube[3] while $K$ is $2 \times 2 \times 2$ in three dimensions. Observe

[3] We won't even start any discussion about whether there are any variations of the Rubik's© cube. The one and only is the $3 \times 3 \times 3$ one!

that a transition from one state of $M_{\text{cube}}$ to another, say $i$ and $j$, corresponds to the facet, or volume of dimension $n-1$, that $C\left(b^i\right)$ and $C\left(b^j\right)$ share, where $C\left(b^i\right)$ and $C\left(b^j\right)$ are, of course, the cubes associated with $i$ and $j$. As each of these facets has size 1, the number of transitions between $S$ and $\overline{S}$ equals the size of $\partial C(S) \setminus \partial K$. Recall the definition of conductance given in Definition 2.3.4. Since each transition probability is $1/2n$, the conductance is

$$\Phi\left(M_{\text{cube}}\right) = \min_{\substack{S \subset \Omega \\ 0<\pi(S)\leq 1/2}} \frac{1}{2n\,|S|} \text{vol}_{n-1}\left(\partial C(S) \setminus \partial K\right), \tag{4.6}$$

where $\text{vol}_d$ denotes $d$-dimensional volume. We can bound $\text{vol}_{n-1}\left(\partial C(S) \setminus \partial K\right)$ with Lemma 4.2.1. If we choose $f$ to be the constant 1 and the $l_\infty$-norm $\|x\|_\infty = \max\{|x_1|, \ldots, |x_n|\}$ as the primal norm, then the $l_1$-norm $\|x\|_\infty^* = \|x\|_1 = |x_1| + \ldots + |x_n|$ is the dual norm, $\kappa(S) = \text{vol}_n(S) = |S|$, $\kappa'(S) = \text{vol}_{n-1}\left(\partial C(S) \setminus \partial K\right)$ and the diameter of $K$ w.r.t. the $l_\infty$-norm is 2. Thus for $|S| \leq \frac{1}{2}|\Omega|$:

$$\text{vol}_{n-1}\left(\partial C(S) \setminus \partial K\right) \geq |S|.$$

Plugging this into (4.6) yields

$$\Phi\left(M_{\text{cube}}\right) \geq \frac{1}{2n} \tag{4.7}$$

and with (4.4)

$$\mathcal{H} \in O\left(n^3\right). \tag{4.8}$$

Next, we will use average conductance to bound the mixing time of $M_{\text{cube}}$. First, we have to come up with a lower bound for the conductance function. As before isoperimetric inequalities can be used for this purpose.

**Lemma 4.2.2** *Let $\Omega$ be the vertex set of an $n$-dimensional hypercube. Let $S \subseteq \Omega$ be a cut in $\Omega$. Then $|\text{cut}(S)|$, the number of edges spanning the cut, is lower bounded by*

$$|\text{cut}(S)| \geq \log_2\left(\frac{|\Omega|}{|S|}\right)|S|.$$

A proof follows from [15]. With Lemma 4.2.2 we obtain

$$\Phi(x) = \min_{0\leq\pi(S)\leq x} \frac{1}{2n} \frac{|\text{cut}(S)|}{|S|} \geq -\frac{\log_2 x}{2n}.$$

Applying average conductance (Theorem 4.1.3) yields

$$\begin{aligned}\mathcal{H} &\in O\left(n^2 \int_{\frac{1}{2^n}}^{\frac{1}{2}} \frac{dx}{x(\ln x)^2} + n\right)\\ &\in O\left(n^2 \int_{\frac{1}{2^n}}^{\frac{1}{2}} \frac{dx}{x(\ln x)^2}\right)\\ &= O\left(n^2 \left[-\frac{1}{\ln x}\right]_{\frac{1}{2^n}}^{\frac{1}{2}}\right)\\ &\in O\left(n^2\right).\end{aligned}$$

Compared to (4.8) this is an improvement of $O(n)$ in time, which seems to be the typical gain[4] in comparison to "classical" conductance.

## 4.3 Balanced matroids

Next, we will study the bases-exchange walk on balanced matroids, which along with general matroids will be defined later in this section. Matroids in general are algebraic structures that provide an abstract treatment of the concept of linear independence. Owing to the abstractness of this notion, many combinatorial counting problems such as counting spanning trees, forests or connected subgraphs of a graph can be subsumed under the problem of counting matroid bases (for a more detailed discussion see [115]). In fact counting bases of general matroids is an open problem in (approximate) counting. Nonetheless some progress has been made for certain classes of matroids: for *graphic matroids*, i.e. matroids whose bases can be represented as the spanning trees of a graph, many approximation results are known, e.g. [17] and [101] (which also includes a very comprehensive list of related work). Dyer and Frieze [37] showed a similar result for regular matroids, *regular matroids* being matroids that are *vectorial*, i.e. can be represented as a finite matrix, over all fields, and Feder and Mihail [41] used a Markov chain Monte Carlo algorithm to prove that the bases of balanced matroids (which will be defined below) can be counted approximately in poly-

[4]Personal communication with Mark Jerrum and Ravi Montenegro.

nomial time[5]. Another approximation algorithm for counting matroid bases is due to Chávez Lomelí and Welsh [24] (which the authors claim works for the majority[6] of matroids.). On the other hand there is a negative result by Azar, Broder and Frieze [9] saying that no deterministic algorithm relying on probes to some oracle, namely an independence oracle, for counting the bases of a general matroid can yield a good approximation within polynomial time.

The connection between MCMC algorithms (or approximate sampling in general) and approximate counting is via self-reducibility, as already mentioned in chapter 1, and we will exhibit this later on. The achievement of Feder and Mihail's efforts notwithstanding, it should be mentioned that as yet it is unknown how large the class of balanced matroids is. What is known is that it encompasses all regular matroids. However, for regular matroids there already exist efficient, deterministic algorithms for counting bases, which utilize the so-called Binet-Cauchy identity (cf. Dyer & Frieze [37]). The only class of non-regular matroids known to be balanced are *uniform matroids* but counting bases of uniform matroids is trivial. Unless it can be shown that for some balanced matroids counting matroid bases is intractable (and as yet none are known), rapid mixing of Markov chains for sampling balanced-matroid bases might be interesting from a technical point of view but gives us nothing new as far as computability is concerned.

After this motivating (or perhaps de-motivating) introduction, we start the study proper of the bases-exchange walk on balanced matroids by giving a definition of general matroids.

**Definition 4.3.1** *Let $S$ be a finite ground set and $I \subseteq 2^S$ a collection of subsets of $S$ usually referred to as independent sets. The pair $\mathcal{M} = (S, I)$ is called a* matroid *if*

1. $\emptyset \in I$,

2. if $A \in I, B \subseteq A$ then $B \in I$ and

[5]It is known that the class of balanced matroids contains the class of regular matroids which in turn contains the class of graphic matroids.

[6]What is meant here is that the class of matroids for which their algorithm works includes the large class of *paving matroids*. A matroid $\mathcal{M}$ is *paving* if none of its circuits has size less than rank of $\mathcal{M}$. For more detailed information see Oxley [99] or Welsh [114].

3. if $A, B \in I$, $|B| = |A| + 1$ then there exists $e \in B \setminus A$ s.t. $A \cup \{e\} \in I$. (independence augmentation axiom)

A very important subset of $I$ is the set of bases.

**Definition 4.3.2** *Let $\mathcal{M}$ be a matroid. An independent set $X \in I$ is a* basis of $\mathcal{M}$ *if it is maximal in the following sense:*

$$(X \cup \{e\}) \notin I, \text{ for all } e \in S.$$

*The set of bases is denoted by $\mathcal{B}$.*

The bases of a matroid are significant because a matroid can be entirely defined by its set of bases. We can therefore use the following notation.

**Notation 4.3.3** *Given a set of bases $\mathcal{B}$ let $\mathcal{M}(\mathcal{B})$ denote the so defined matroid. Conversely, if we are given a matroid $\mathcal{M}$ we will use $\mathcal{B}(\mathcal{M})$ to refer to its bases. We will use $S(\mathcal{M})$ and $I(\mathcal{M})$ to denote the ground set and the independent sets of $\mathcal{M}$ respectively.*

At this point it is appropriate to mention a few points about matroid representation. One obvious way to represent a matroid is by listing all independent sets. Clearly for such a representation counting the bases is always polynomial in the input size. The only drawback is that in this case the input size is potentially exponential in the size of the ground set $S$ and hence not suitable from a computational point of view. More interesting are matroids allowing a succinct representation. Examples are graphic matroids — matroids that can be represented as a graph such that the ground set corresponds to the edges of the graph and the independent sets are the cycle-free subgraphs — or vectorial matroids, in which case the ground set corresponds to the columns of a finite matrix and an independent set to linearly independent columns. One way to obtain such succinct representations is via *oracle* representations. An oracle representation of a matroid consists of the ground set and an oracle. An oracle can be thought of as a subroutine, which can be invoked to obtain information about a particular subset of the ground set (see [104] for more information on matroid oracles). An *independence oracle* for instance is an oracle which when presented with a set (this process is

called a *probe*) will tell if this set is independent (e.g. by returning a basis that contains the independent set). From a computational point of view oracle presentations are the shortest and most general way to represent matroids: firstly, it can hardly be shorter since every representation must at least include the ground set and a characterization of which subsets constitute the independent sets and secondly, it is general because an oracle can be regarded as a characterization of the independent sets.

Observe that the independence augmentation axiom implies that all bases are of the same size. This leads to

**Definition 4.3.4** *The* rank *of a matroid is the size of any of its bases.*

The independence augmentation axiom also entails the property that for every pair of bases $X, Y \in \mathcal{B}$ and every element $e \in X$ there exists an element $f \in Y$ such that $(X \setminus \{e\}) \cup \{f\} \in \mathcal{B}$. This property suggests a very natural random walk on $\mathcal{B}$. The *bases-exchange graph* $G(\mathcal{M})$ of a matroid $\mathcal{M}$ has vertex set $\mathcal{B}(\mathcal{M})$ and edge set $\{\{X,Y\} \mid X,Y \in \mathcal{B} \text{ and } |X \oplus Y| = 2\}$, where $\oplus$ denotes symmetric difference, i.e. only one element of $X$ is not in $Y$ and vice versa. In future, we will call every random walk on the bases-exchange graph a *bases-exchange walk*. An example of such a walk is given below:

**Algorithm 4.3.5** *Let $X$ be the current basis. We move on to basis $X'$ according to the rule*

*1.) Choose a ground set element $e$ and an element $f \in X$ u.a.r.*

*2.) If $(X \setminus \{f\}) \cup \{e\} \in \mathcal{B}$, then set $X' = (X \setminus \{f\}) \cup \{e\}$, otherwise let $X' = X$ (stay at the current basis).*

Note that this walk is finite, irreducible and aperiodic[7]. Thus using bases-exchange walks it is, in principle, possible to sample u.a.r. from $\mathcal{B}(\mathcal{M})$. Mihail et al. ([92],[93]) conjectured that the random walk on $G(\mathcal{M})$ is rapidly mixing for all matroids $\mathcal{M}$. However, as yet it has only been proved for a limited number of matroid-classes. One

[7]For most matroids the walk will be aperiodic because the exchange step 2.) will in many cases not yield a basis. Nonetheless, there are matroids for which step 2.) will always yield a basis, for uniform matroids for example, in this case we will have to expressly add self-loop probabilities.

of them are so called balanced matroids. This notion, as does the proof of rapid mixing, originates from Feder and Mihail [41]. But before we define balanced matroids, we will introduce a few basic concepts and explain why rapid mixing is enough for efficient, approximate counting of matroid bases.

Two important operations on matroids are contraction and deletion. If $e \in S(\mathcal{M})$ is an element of the ground set of $\mathcal{M}$, then the matroid $\mathcal{M} \setminus e$ obtained by *deleting* $e$ has ground set $S(\mathcal{M} \setminus e) = S(\mathcal{M}) \setminus \{e\}$ and bases $\mathcal{B}(\mathcal{M} \setminus e) = \{X \subseteq S(\mathcal{M} \setminus e) \mid X \in \mathcal{B}(\mathcal{M})\}$. The matroid $\mathcal{M}/e$ obtained by *contracting* $e$ has ground set $S(\mathcal{M}/e) = S(\mathcal{M}) \setminus \{e\}$ and bases $\mathcal{B}(\mathcal{M}/e) = \{X \subseteq S(\mathcal{M}/e) \mid X \cup \{e\} \in \mathcal{B}(\mathcal{M})\}$.

**Definition 4.3.6** *Any matroid obtained from $\mathcal{M}$ by a series of contractions and deletions is a* minor *of $\mathcal{M}$.*

We have now all necessary concepts at our disposal to show that approximately sampling matroid bases is as good as approximately counting them. To this end, we will exhibit that matroids are self-reducible. Suppose $\mathcal{M}$ is a matroid of rank $n$ on ground set $\{s_1, \ldots, s_m\}$. We will represent $\mathcal{M}$ by its ground set and an independence oracle $O_{\mathcal{M}}$ (when presented with a subset of the ground set the oracle answers by either saying that the set is not independent or returning a basis that contains the set[8]). For any minor of $\mathcal{M}$ the ground set will be encoded as a vector $a = (a_1, \ldots, a_m) \in \{0, 1, *\}^m$, where $a_i = *$ means that element $s_i$ is contained in the ground set of the minor, while $a_i = 1$ or $a_i = 0$ denotes that the minor was obtained by contracting or deleting element $s_i$ respectively. Let $o_{\mathcal{M}} \in (\Sigma_{O_{\mathcal{M}}})^*$, where $\Sigma_{O_{\mathcal{M}}}$ is some alphabet, be an encoding of the oracle $O_{\mathcal{M}}$. We will be deliberately vague about how the oracle is being encoded because all we are interested in is its ability to tell us whether a subset of the ground set is independent or not; regard the oracle $O_{\mathcal{M}}$ as some old sage that you happen to know but all you know about that sage is that his name is $o_{\mathcal{M}}$. As said before our representation of a matroid $\mathcal{M}$ consists of the ground set and an independence oracle for $\mathcal{M}$. Using the above notation, $\mathcal{M}$ will be encoded as the pair $(a, o_{\mathcal{M}})$, where $a = \{*\}^m$ and

[8]This choice is arbitrary in the sense that we could have equally well chosen an oracle that answers "yes/no" when presented with a set. However, this definition is neater for starting off the random walk because an initial basis can be found by making a probe with for example the empty set (which by definition is contained in any basis, although in that case the initial basis will always be the same unless the oracle chooses a random answer whenever there are alternatives).

$o_{\mathcal{M}}$ encodes $O_{\mathcal{M}}$. Any basis $B$ of $\mathcal{M}$ will be encoded as a bit vector $b = (b_1, \ldots, b_m)$, where $b_i = 1$ if element $s_i$ is in $B$ and $b_i = 0$ otherwise. Suppose $\mathcal{M}'$ is a minor of $\mathcal{M}$ obtained by contracting or deleting elements $s_1, \ldots, s_k$. Our representation for $\mathcal{M}'$ will be the pair $((a_1, \ldots, a_k, a_{k+1}, \ldots, a_m), o_{\mathcal{M}})$, where $(a_1, \ldots, a_k) \in \{0,1\}^k$ and $a_i = *$ for $i = k+1, \ldots, m$. Any basis $B'$ of $\mathcal{M}'$ will be encoded as a bit vector $b' = (b'_1, \ldots, b'_{m-k})$.

Let $R \subseteq \Sigma^* \times \Sigma^*$, where $\Sigma = \{0,1,*\} \cup \Sigma_{O_{\mathcal{M}}}$, be a binary relation denoting "is basis of": if $(x,y) \in R$ for $x = ((a_1, \ldots, a_k, a_{k+1}, \ldots, a_m), o_{\mathcal{M}})$, where $(a_1, \ldots, a_k) \in \{0,1\}^k$ for some $k \in \{0, \ldots, m\}$, $a_i = *$ for $i = k+1, \ldots, m$, and $y \in \{0,1\}^{m-k}$, then $y$ is a basis of $x$. Observe the following: If $b = (b_1, \ldots, b_m) \in \{0,1\}^m$ is a basis of $\mathcal{M}$, then for $0 \leq k \leq m$ either the minor $((b_1, \ldots, b_k, a_{k+1}, \ldots, a_m), o_{\mathcal{M}})$ of $\mathcal{M}$ obtained by contracting/deleting ground set elements $s_1, \ldots, s_k$ has the set denoted by $(b_{k+1}, \ldots, b_m)$ among its bases or the matroid on the ground set $\{a_{k+1}, \ldots, a_m\}$ is empty. Checking if a set of ground set elements is a basis of such a minor is easy indeed: assume that the description of the minor $\mathcal{M}'$ is $((b_1, \ldots, b_k, a_{k+1}, \ldots, a_m), o_{\mathcal{M}})$ for some $k \in \{0, \ldots, m\}$ and that $b = (b_1, \ldots, b_m)$ denotes some basis of $\mathcal{M}$. To check if $y = (y_1, \ldots, y_{m-k})$ denotes a basis of $\mathcal{M}'$ tag all the 1's in the string $b_1, \ldots, b_k, y_1, \ldots, y_{m-k}$ and make a probe to $O_{\mathcal{M}}$ with the tagged elements; if the answer is positive, then $y$ denotes a basis[9] of $\mathcal{M}'$. This observation is the key to the self-reducibility of our encoding. Define for $x = ((b_1, \ldots, b_k, a_{k+1}, \ldots, a_m), o_{\mathcal{M}})$ and $y \in \{0,1\}^{m-k}$— where $(b_1, \ldots, b_k) \in \{0,1\}^k$ for $k \in \{0, \ldots, m\}$ is the prefix of length $k$ of some basis $b$ of $\mathcal{M}$, $a_i = *$ for $i = k+1, \ldots, m$ and $y$ denotes a basis of the minor defined by $x$ — the function $\psi(x,y) = \left(\left(b_1, \ldots, b_k, y_1, \ldots, y_{\sigma(x)}, a_{k+\sigma(x)+1}, \ldots, a_m\right), o_{\mathcal{M}}\right)$, where $\sigma(x) = \log m$ is a number less than $\log |x|$. Observe that $\psi(x,y)$ encodes a minor of $\mathcal{M}$ on ground set $\left\{s_{k+\sigma(x)+1}, \ldots, s_m\right\}$ and that clearly $(x,y) \in R \Leftrightarrow \left(\psi(x,y), \left(y_{\sigma(x)+1}, \ldots, y_{m-k}\right)\right) \in R$. Hence, all that is left to show is that the length and complexity conditions of the functions stated in (SR1) to (SR3) are met. First notice, that from the above it is obvious that $l(x) = |y| \leq |x|$ and $|\psi(x,y)| = |x|$. Secondly, since $\psi$ and $\sigma$ can clearly be computed in time polynomial in $|x|$ and $|y|$, the conditions (SR1) – (SR2) are satisfied and the relation $R$ is self-reducible.

We continue by defining balanced matroids. If $X$ is a basis chosen uniformly at

[9] In fact, then the answer of the oracle corresponds to $(b_1, \ldots, b_k, y_1, \ldots, y_{m-k})$, which denotes by choice of oracle a basis of $\mathcal{M}$ and thus the answer of the oracle coincides with the query.

random from $\mathcal{B}(\mathcal{M})$ and $e$ is an element of $\mathcal{M}$, then (with a slight abuse of notation) let $e$ denote the event $e \in X$ and $\bar{e}$ the event $e \notin X$. By juxtaposition we express the conjunction of events, i.e. $e\bar{f}$ means $e \in X \wedge f \notin X$. A matroid $\mathcal{M}$ is said to satisfy the *negative correlation* property if the inequality $\mathbb{P}[ef] \leq \mathbb{P}[e]\,\mathbb{P}[f]$ holds for all pairs of distinct elements $e$ and $f$ in $S(\mathcal{M})$.

**Definition 4.3.7** *A matroid $\mathcal{M}$ is* balanced *if $\mathcal{M}$ and all its minors are negatively correlated.*

All regular (and thus graphic) matroids are known to be balanced. Feder and Mihail introduced this notion to show that the bases-exchange graph of a balanced matroid has cutset-expansion 1. They did this via the intermediate step of fractional matchings, which they defined as follows:

**Definition 4.3.8** *Let $G(\mathcal{M})$ be the bases-exchange graph of a matroid $\mathcal{M}$ and $A, \overline{A} \subseteq \mathcal{B}(\mathcal{M})$ be the natural copies of $\mathcal{B}(\mathcal{M} \setminus e)$ and $\mathcal{B}(\mathcal{M}/e)$ respectively in $\mathcal{M}$ for some ground set element $e$, i.e. $A = \{X \in \mathcal{B}(\mathcal{M}) \mid e \in X\}$ and $\overline{A} = \{X \in \mathcal{B}(\mathcal{M}) \mid e \notin X\}$. A* fractional matching *is a function $\mathfrak{f} : A \times \overline{A} \to \mathbb{N}$ assigning non-negative weights to the cutset cut$(A)$ (the edges of $G(\mathcal{M})$ spanning $A$ and $\overline{A}$) such that*

$$\forall x \in A : \sum_{y \in \overline{A}} \mathfrak{f}(x,y) = \left|\overline{A}\right| \qquad \text{and} \qquad \forall y \in \overline{A} : \sum_{x \in A} \mathfrak{f}(x,y) = |A| .$$

They proved that every balanced matroid allows a fractional matching and deployed this to obtain[10]:

**Lemma 4.3.9** *For any balanced matroid $\mathcal{M}(\mathcal{B})$ and $A \subset \mathcal{B}$, $|A| \leq \frac{1}{2}|\mathcal{B}|$:*

$$|\mathrm{cut}(A)| \geq |A| .$$

In fact it is even possible to prove a stronger result (see Lemma 4.4.1) than Lemma 4.3.9. We have already seen that for random walks a lower bound on the size of the cutset bounds conductance, too (cf. section 4.2). With this bound Feder and Mihail showed that the bases-exchange walk on balanced matroids (algorithm 4.3.5) is rapidly mixing (cf. [41], Corollary 3.5)

[10]The same result can be obtained by using concepts studied by Mihail and Sudan in [92].

**Theorem 4.3.10** *For any balanced matroid $\mathcal{M}$ of rank $n$ on a ground set of size $m$, the natural random walk on the bases-exchange graph $G(\mathcal{M})$ is rapidly mixing:*

$$\tau(\varepsilon) \in O\left(n \ln m + \ln \frac{1}{\varepsilon}\right) n^2 m^2.$$

However, they succeed in improving this result by applying a rather complex form of canonical paths ([41], Theorem 5.1).

**Theorem 4.3.11** *For the natural random walk on any balanced matroid of rank $n$ and with ground set of size $m$, the total variation distance can be bounded by $\varepsilon$ in time*

$$\tau(\varepsilon) \in O\left(n \ln m + \ln \frac{1}{\varepsilon}\right) n^2 m.$$

For the random walk on so called *near-bases* their ultimate result is ([41], Theorem 5.2)

**Theorem 4.3.12** *For the modified natural random walk on any balanced matroid of rank $n$ and with ground set of size $m$ the mixing time is bounded by*

$$\tau(\varepsilon) \in O\left(n \ln m + \ln \frac{1}{\varepsilon}\right) n^3.$$

## 4.4 Average conductance of a balanced matroid

In this section we will bound the average conductance of the bases-exchange walk on balanced matroids and shows how it can yield improved bounds for the mixing time. This bound can be further improved by the use of so-called logarithmic Sobolev constants, which will be dealt with in more detail in chapter 5.

We start off with deriving a bound on the mixing time using average conductance. Recall that average conductance gave bounds on mixing time in terms of hitting time. To relate our results to the results of Feder and Mihail we choose $\tau = \tau(1/e)$, then $\tau \in O(\mathcal{H})$. A lower bound on the conductance function $\Phi(x)$ (Definition 4.1.1) is obtained by extending Lemma 4.2.2 to balanced matroids.

**Lemma 4.4.1** *Let $G(\mathcal{M})$ be the bases-exchange graph of any balanced matroid $\mathcal{M}$ with bases $\mathcal{B}$. For all subsets $A \subset \mathcal{B}$ such that $0 < |A| \leq |\mathcal{B}|/2$*

$$\frac{|\mathrm{cut}(A)|}{|A|} \geq \log_2 \left(\frac{|\mathcal{B}|}{|A|}\right).$$

*Proof.* We proceed by induction on the size of the ground set of $\mathcal{M}$. For the base-case, $|S(\mathcal{M})| = 1, 2$, the hypothesis is trivially true.

Induction step, $|S(\mathcal{M})| > 2$: Let $A \subset \mathcal{B}$ be a collection of bases, with $|A| \leq |\mathcal{B}|/2$, defining a cut in the bases-exchange graph of $\mathcal{M}$. Given a ground set element $e \in S(\mathcal{M})$, the set of bases $\mathcal{B}$ may be partitioned as $\mathcal{B} = \mathcal{B}_e \cup \mathcal{B}_{\bar{e}}$, where

$$\mathcal{B}_e = \{X \in \mathcal{B} \mid e \in X\} \qquad \text{and} \qquad \mathcal{B}_{\bar{e}} = \{X \in \mathcal{B} \mid e \notin X\};$$

observe that the induced subgraphs $G(\mathcal{M})_{|\mathcal{B}_e}$ and $G(\mathcal{M})_{|\mathcal{B}_{\bar{e}}}$ are isomorphic to the bases-exchange graphs $G(\mathcal{B}(\mathcal{M}/e))$ and $G(\mathcal{B}(\mathcal{M} \setminus e))$ respectively. Furthermore, let

$$A_e = A \cap \mathcal{B}_e \qquad \text{and} \qquad A_{\bar{e}} = A \cap \mathcal{B}_{\bar{e}}.$$

Let $\alpha|\mathcal{B}|$ and $(1-\alpha)|\mathcal{B}|$ for some $\alpha \in [0,1]$ be the sizes of $\mathcal{B}_e$ and $\mathcal{B}_{\bar{e}}$ respectively and define values $x, y \in [0,1]$ such that $|A_e| = x|\mathcal{B}_e|$ and $|A_{\bar{e}}| = y|\mathcal{B}_{\bar{e}}|$. The edges forming the cut are of three kinds:

**(i)** those whose endpoints are both within $\mathcal{B}_e$,

**(ii)** those whose endpoints are both within $\mathcal{B}_{\bar{e}}$ and

**(iii)** those which span $\mathcal{B}_e$ and $\mathcal{B}_{\bar{e}}$.

Since, as mentioned above, the subgraphs induced by $\mathcal{B}_e$ and $\mathcal{B}_{\bar{e}}$ are isomorphic to the bases-exchange graphs on $\mathcal{B}(\mathcal{M}/e)$ and $\mathcal{B}(\mathcal{M} \setminus e)$, the induction hypothesis becomes applicable. By induction hypothesis, the numbers of edges of kinds (i) and (ii) are at least

$$-\min\{x, 1-x\} \log_2(\min\{x, 1-x\})|\mathcal{B}_e|$$

and

$$-\min\{y, 1-y\} \log_2(\min\{y, 1-y\})|\mathcal{B}_{\bar{e}}|$$

respectively. To lower bound the number of edges of kind (iii), assume first that $x \geq y$. By Lemma 4.4.2[11], which says

[11] Lemma 4.4.2 states that all balanced matroids *enforce ratios*. The concept of ratio-enforcement is due to Mihail and Sudan [92].

**Lemma 4.4.2** [Lemma 3.1 of [41]] *For every balanced matroid $\mathcal{M}(\mathcal{B})$ and any ground set element $e \in S(\mathcal{M})$*

$$\begin{aligned} \forall Q \subseteq \mathcal{B}_e \quad &: \quad \frac{\left|\Gamma_{\mathcal{B}_{\overline{e}}}(Q)\right|}{|\mathcal{B}_{\overline{e}}|} \geq \frac{|Q|}{|\mathcal{B}_e|} \\ \forall \overline{Q} \subseteq \mathcal{B}_{\overline{e}} \quad &: \quad \frac{\left|\Gamma_{\mathcal{B}_e}(\overline{Q})\right|}{|\mathcal{B}_e|} \geq \frac{|\overline{Q}|}{|\mathcal{B}_{\overline{e}}|}, \end{aligned}$$

*where $\Gamma_{\mathcal{B}_{\overline{e}}}(Q)$ is the neighbourhood of $Q$ in $\mathcal{B}_{\overline{e}}$ in the graph $G(\mathcal{M})$ and similarly for $\Gamma_{\mathcal{B}_e}(\overline{Q})$.*

there are at least $x|\mathcal{B}_{\overline{e}}|$ bases in $\mathcal{B}_{\overline{e}}$ adjacent to some bases in $A_e$; of these, at least $(x-y)|\mathcal{B}_{\overline{e}}|$ must lie outside $|A_{\overline{e}}|$. Thus there are at least $(x-y)|\mathcal{B}_{\overline{e}}|$ edges of type (iii). This argument can equally well be applied in the opposite direction, starting at the set $\mathcal{B}_{\overline{e}} \setminus A_{\overline{e}}$, yielding a second lower bound of $(x-y)|\mathcal{B}_e|$. Thus the number of edges of kind (iii) is at least $(x-y)\max\{|\mathcal{B}_e|, |\mathcal{B}_{\overline{e}}|\}$. Since the case $x < y$ is entirely symmetric, we obtain, summing the contributions from edges of kinds (i)-(iii):

$$\begin{aligned} |\mathrm{cut}(A)| \quad \geq \quad & -\min\{x, 1-x\}\log_2(\min\{x, 1-x\})|\mathcal{B}_e| \\ & -\min\{y, 1-y\}\log_2(\min\{y, 1-y\})|\mathcal{B}_{\overline{e}}| \\ & +|x-y|\max\{|\mathcal{B}_e|, |\mathcal{B}_{\overline{e}}|\}. \end{aligned}$$

To complete the proof, we must show that $|\mathrm{cut}(A)|$ is always at least

$$-(x\alpha + y(1-\alpha))\log_2(x\alpha + y(1-\alpha))|\mathcal{B}|,$$

whenever $|A| \leq |\mathcal{B}|/2$. Note that this last condition may be expressed as

$$\left(\frac{1}{2} - x\right)|\mathcal{B}_e| + \left(\frac{1}{2} - y\right)|\mathcal{B}_{\overline{e}}| \geq 0.$$

This implies that only one of $x$ or $y$ can be greater than $1/2$. It remains to establish that

$$\begin{aligned} & -\min\{x, 1-x\}\log_2(\min\{x, 1-x\})|\mathcal{B}_e| \\ & -\min\{y, 1-y\}\log_2(\min\{y, 1-y\})|\mathcal{B}_{\overline{e}}| \\ & +|x-y|\max\{|\mathcal{B}_e|, |\mathcal{B}_{\overline{e}}|\} \\ & \quad \geq \quad -(x\alpha + y(1-\alpha))\log_2(x\alpha + y(1-\alpha))|\mathcal{B}|. \end{aligned}$$

As the cases $0 \leq \alpha \leq 1/2$ and $1/2 \leq \alpha \leq 1$ are entirely symmetrical, we have to scrutinise only four of eight cases. But before, notice that in the degenerate case, $\alpha = 1$ (or equally $\alpha = 0$), the induction hypothesis becomes immediately applicable. Assuming w.l.o.g. that $1/2 \leq \alpha < 1$ the four cases are:

1. $1/2 \geq x \geq y \geq 0$,
2. $1/2 \geq y \geq x \geq 0$,
3. $x > 1/2, y < 1/2$,
4. $y > 1/2, x < 1/2$.

The four lemmata below prove that the induction step is valid for each of the above cases.

**Lemma 4.4.3** *For* $1/2 \leq \alpha < 1$ *and* $1/2 \geq x \geq y \geq 0$

$$\begin{aligned} & f_1(x,y,\alpha) \\ &= -\alpha x \log_2 x - (1-\alpha) y \log_2 y + \alpha(x-y) \\ &\qquad + (\alpha x + (1-\alpha) y) \log_2 (\alpha x + (1-\alpha) y) \\ &\geq 0. \end{aligned}$$

*Proof.* First extend $f_1$ to the boundary by continuity. Treating $\alpha$ like a constant, the Hessian[12] of $f_1(x,y,\alpha)$ is

$$h_1(x,y,\alpha) = \begin{pmatrix} -\frac{\alpha}{x \ln 2} + \frac{\alpha^2}{(x\alpha + y(1-\alpha))\ln 2} & \frac{(1-\alpha)\alpha}{(x\alpha + y(1-\alpha))\ln 2} \\ \frac{(1-\alpha)\alpha}{(x\alpha + y(1-\alpha))\ln 2} & -\frac{1-\alpha}{y \ln 2} + \frac{(1-\alpha)^2}{(x\alpha + y(1-\alpha))\ln 2} \end{pmatrix}.$$

As $\frac{\partial^2}{\partial x^2} f_1(x,y,\alpha)$ and $\frac{\partial^2}{\partial y^2} f_1(x,y,\alpha)$ are less than zero in the interior, the Hessian cannot be positive definite there, i.e. $f_1$ has no local minimum in the interior. Thus, we merely have to check the boundaries: $x = y$, $(y = 0 \wedge 1/2 \geq x \geq 0)$ and $(x = 1/2 \wedge 1/2 \geq y \geq 0)$. But, since $\frac{\partial^2}{\partial x^2} f_1(x,y,\alpha) \leq 0$ and also $\frac{\partial^2}{\partial y^2} f_1(x,y,\alpha) \leq 0$, the claim holds if $f_1(x,y,\alpha) \geq 0$ for $(x = 1/2, y = 0)$ and $x = y$.

For $x = 1/2, y = 0$:

$$-\alpha\frac{1}{2}\log_2\frac{1}{2} + \alpha\frac{1}{2} + \alpha\frac{1}{2}\log_2(\alpha\frac{1}{2}) = \alpha\frac{1}{2}(1 + \log_2 \alpha) \geq 0$$

[12] The second derivative of the vector function $f_1$.

since $-1 \leq \log_2 \alpha \leq 0$. And for $x = y$:

$$-\alpha x \log_2 x - (1-\alpha)x \log_2 x + (\alpha x + x - \alpha x)\log_2(\alpha x + x - \alpha x) = 0.$$

□

**Lemma 4.4.4** *For* $1/2 \leq \alpha < 1$ *and* $1/2 \geq y \geq x \geq 0$

$$\begin{aligned} & f_2(x,y,\alpha) \\ = \; & -\alpha x \log_2 x - (1-\alpha) y \log_2 y + \alpha(y-x) \\ & \quad + (\alpha x + (1-\alpha)y)\log_2(\alpha x + (1-\alpha)y) \\ \geq \; & 0. \end{aligned}$$

*Proof.* Similarly to the first case, extend $f_2$ to the boundary by continuity first and then form the Hessian

$$h_2(x,y,\alpha) = h_1(x,y,\alpha) = \begin{pmatrix} -\frac{\alpha}{x\ln 2} + \frac{\alpha^2}{(x\alpha + y(1-\alpha))\ln 2} & \frac{(1-\alpha)\alpha}{(x\alpha+y(1-\alpha))\ln 2} \\ \frac{(1-\alpha)\alpha}{(x\alpha+y(1-\alpha))\ln 2} & -\frac{1-\alpha}{y\ln 2} + \frac{(1-\alpha)^2}{(x\alpha+y(1-\alpha))\ln 2} \end{pmatrix}.$$

Since the Hessian of $f_2(x,y,\alpha)$ is the same as in the first case, the same arguments apply and it suffices to look at the boundaries. This time these are: $x = y$ and $(x = 0, y = 1/2)$.

For $x = y$:

$$-\alpha x \log_2 x - (1-\alpha)x \log_2 x + (\alpha x + x - \alpha x)\log_2(\alpha x + x - \alpha x) = 0$$

and for $x = 0, y = 1/2$ we set:

$$\frac{\alpha}{2} + \frac{1}{2}(1-\alpha)\log_2(1-\alpha) = g(\alpha).$$

The first derivative of $g$ shows that the only local extremum of $g$ lies outside the allowed range for $\alpha$:

$$\frac{d}{d\alpha} g = \frac{1}{2} - \frac{1}{2\ln 2}(\ln(1-\alpha)+1) = 0 \quad \text{iff} \quad \alpha = 1 - e^{\ln 2 - 1} < \frac{1}{2}.$$

Thus, for the boundaries of $g$, $\alpha = 1/2$ and $\alpha \to 1$, the claim holds. □

For the remaining two cases recollect that only one of $x$ and $y$ can be greater than $1/2$.

**Lemma 4.4.5** *For* $1/2 \leq \alpha < 1$ *and* $x > 1/2, y < 1/2$

$$\begin{aligned} & f_3(x,y,\alpha) \\ = \; & -\alpha(1-x)\log_2(1-x) - (1-\alpha)y\log_2 y + \alpha(x-y) \\ & \quad + (\alpha x + (1-\alpha)y)\log_2(\alpha x + (1-\alpha)y) \\ \geq \; & 0. \end{aligned}$$

*Proof.* Extend $f_3$ by continuity to the boundary. As $|A| \leq |\mathcal{B}|/2$, the valid range for $x$ and $y$ is given by $x\alpha + y(1-\alpha) \leq 1/2$, which in turn yields $x \leq 1/(2\alpha)$. To start with we again form the Hessian:

$$h_3(x,y,\alpha) = \begin{pmatrix} -\frac{\alpha}{(1-x)\ln 2} + \frac{\alpha^2}{(x\alpha+y(1-\alpha))\ln 2} & \frac{(1-\alpha)\alpha}{(x\alpha+y(1-\alpha))\ln 2} \\ \frac{(1-\alpha)\alpha}{(x\alpha+y(1-\alpha))\ln 2} & -\frac{1-\alpha}{y\ln 2} + \frac{(1-\alpha)^2}{(x\alpha+y(1-\alpha))\ln 2} \end{pmatrix}.$$

Noticing that $\frac{\partial^2}{\partial y^2} f_3(x,y,\alpha) \leq 0$ and

$$\frac{\partial^2}{\partial x^2} f_3(x,y,\alpha) = \frac{\alpha(\alpha - 2\alpha x + \alpha y - y)}{(1-x)(x\alpha + y(1-\alpha))\ln 2} \leq 0,$$

since $2x \geq 1$, the Hessian $h_3(x,y,\alpha)$ cannot be positive definite. So, the minima can only be found on the boundaries: $(y = 0, x = 1/2)$ and $\alpha x + (1-\alpha)y = 1/2$.

For $y = 0, x = 1/2$:

$$-\alpha\frac{1}{2}\log_2\frac{1}{2} + \frac{\alpha}{2} + \frac{\alpha}{2}\log_2\frac{\alpha}{2} = \frac{\alpha}{2}(1 + \log_2\alpha) \geq 0,$$

since $1/2 \leq \alpha < 1$. For $y = \frac{1-2x\alpha}{2-2\alpha}$ set:

$$-\alpha(1-x)\log_2(1-x) - (1-\alpha)\left(\frac{1-2x\alpha}{2-2\alpha}\right)\log_2\left(\frac{1-2x\alpha}{2-2\alpha}\right) + \frac{2x\alpha - 1}{2-2\alpha} = g(x).$$

The second derivative of $g(x)$ ,

$$\frac{d^2}{dx^2} g(x) = -\frac{\alpha}{(1-x)\ln 2} - \frac{2\alpha^2}{(1-2x\alpha)\ln 2},$$

is less than zero for $x \leq 1/(2\alpha)$. Checking the boundaries $x = 1/2$ and $x = 1/(2\alpha)$ shows that

$$g\left(\frac{1}{2}\right) = -\frac{\alpha}{2}\log_2\frac{1}{2} - \frac{1-\alpha}{2}\log_2\frac{1}{2} - \frac{1}{2} = 0$$

and

$$g\left(\frac{1}{2\alpha}\right) = -\alpha\left(1-\frac{1}{2\alpha}\right)\log_2\left(1-\frac{1}{2\alpha}\right) \geq 0$$

concluding the proof of Lemma 4.4.5. □

**Lemma 4.4.6** *For* $1/2 \leq \alpha < 1$ *and* $y > 1/2, x < 1/2$

$$\begin{aligned} & f_4(x,y,\alpha) \\ = \; & -\alpha x \log_2 x - (1-\alpha)(1-y)\log_2(1-y) + \alpha(y-x) \\ & \quad + (\alpha x + (1-\alpha)y)\log_2(\alpha x + (1-\alpha)y) \\ \geq \; & 0. \end{aligned}$$

*Proof.* By continuity extend $f_4$ to the boundary. As said before, only one of $x$ or $y$ is greater than 1/2. The legal range for $x$ and $y$ is again $x\alpha + y(1-\alpha) \leq 1/2$, which yields the constraint $y \leq 1/(2(1-\alpha))$. Notice that $1/(2(1-\alpha)) \geq 1$ for $\alpha \geq 1/2$ so that in fact, $y$ is upper bounded by 1. Thus for $\alpha > 1/2$ the boundary is a trapezoid. Looking at the Hessian of $f_4(x,y,\alpha)$,

$$h_4(x,y,\alpha) = \begin{pmatrix} -\frac{\alpha}{x\ln 2} + \frac{\alpha^2}{(x\alpha+y(1-\alpha))\ln 2} & \frac{(1-\alpha)\alpha}{(x\alpha+y(1-\alpha))\ln 2} \\ \frac{(1-\alpha)\alpha}{(x\alpha+y(1-\alpha))\ln 2} & -\frac{1-\alpha}{(1-y)\ln 2} + \frac{(1-\alpha)^2}{(x\alpha+y(1-\alpha))\ln 2} \end{pmatrix},$$

discloses that $\frac{\partial^2}{\partial x^2} f_4(x,y,\alpha) \leq 0$ and

$$\frac{\partial^2}{\partial y^2} f_4(x,y,\alpha) = \frac{(1-\alpha)[(1-2y)(1-\alpha) - x\alpha]}{(1-y)(x\alpha + y(1-\alpha))\ln 2} \leq 0.$$

Consequently, the Hessian is not positive definite, i.e. $f_4$ has no local minima in the interior. The minima must therefore be on the following points of the boundary (trapezoid): $(x = 0, y = 1/2)$, $(x = 0, y = 1)$ and $x\alpha + y(1-\alpha) = 1/2$.

For $x = 0, y = 1/2$:

$$-\frac{1-\alpha}{2}\log_2\frac{1}{2} + \frac{\alpha}{2} + \frac{1-\alpha}{2}\log_2\frac{1-\alpha}{2} = \frac{\alpha}{2} + \frac{1}{2}(1-\alpha)\log_2(1-\alpha) \geq 0,$$

for $x = 0, y = 1$: $\alpha + (1-\alpha)\log_2(1-\alpha) \geq 0$ and for $x = \frac{1-2y(1-\alpha)}{2\alpha}$ set:

$$-\alpha\left(\frac{1-2y(1-\alpha)}{2\alpha}\right)\log_2\left(\frac{1-2y(1-\alpha)}{2\alpha}\right) - (1-\alpha)(1-y)\log_2(1-y) + y - 1 \; = \; g(y).$$

The second derivative of $g(y)$,

$$\frac{d^2}{dy^2}g(y) = -\frac{2(1-\alpha)^2}{(1-2y(1-\alpha))\ln 2} - \frac{1-\alpha}{(1-y)\ln 2},$$

has no local minima for $1/2 \leq y \leq 1$. For the boundaries $y = 1/2$ and $y = 1$ we obtain

$$g\left(\frac{1}{2}\right) = -\frac{\alpha}{2}\log_2\frac{1}{2} - \frac{1-\alpha}{2}\log_2\frac{1}{2} - \frac{1}{2} = 0$$

and

$$g(1) = -\alpha\frac{2\alpha-1}{2\alpha}\log_2\frac{2\alpha-1}{2\alpha} \geq 0.$$

This concludes the proof of Lemma 4.4.6 and thus Lemma 4.4.1. □

Recently Montenegro [95] gave a shorter proof of Lemma 4.4.1. He allowed the $A$ to be any subset of $\mathcal{B}$ which simplifies the case analysis.

Using Lemma 4.4.1 we obtain the following result

**Theorem 4.4.7** *The mixing time of the bases-exchange walk on any balanced matroid of rank $n$ on a ground set of size $m$ is at most $cm^2n^2$ for some constant $c$ independent of the matroid.*

*Proof.* By Lemma 4.4.1 $\frac{|\text{cut}(A)|}{|A|} \geq \log_2\left(\frac{|\mathcal{B}|}{|A|}\right)$. To be on the safe side, let us assume that the walk has self-loop probability of at least $1/2$ at every vertex. Thus

$$\begin{aligned}\Phi(x) &= \min_{0\leq\pi(A)\leq x}\frac{\widetilde{P}(A,\overline{A})}{\pi(A)} \geq \frac{1}{2mn}\min_{0\leq\pi(A)\leq x}\frac{|\text{cut}(A)|}{|A|}\\ &\geq \frac{1}{2mn}\log_2\left(\frac{|\mathcal{B}|}{|A|}\right) \geq -\frac{\log_2 x}{2mn}.\end{aligned}$$

Substituting this into the average conductance theorem (Theorem 4.1.3) gives

$$\begin{aligned}\mathcal{H} &\leq K\left(14\int_{\frac{1}{m^n}}^{\frac{1}{2}}\frac{dx}{x\left(\frac{\log_2 x}{2mn}\right)^2} + 8mn\right)\\ &= K\left(56m^2n^2(\ln 2)^2\int_{m^{-n}}^{2^{-1}}\frac{dx}{x(\ln x)^2} + 8mn\right)\\ &= K\left(56m^2n^2(\ln 2)^2\left[-\frac{1}{\ln x}\right]_{m^{-n}}^{2^{-1}} + 8mn\right)\\ &\in O\left(m^2n^2\right).\end{aligned}$$

□

Theorem 4.4.7 is a substantial improvement on Theorem 4.3.10 — the gain in mixing time is $\Omega(n)$. Note that Theorem 4.4.7 is better than the result of Theorem 4.3.11 if $n \ln m \in \Omega(m)$, e.g. $n \in O(m/\ln m)$. In the case of graphic matroids this would be the case when the average degree of vertices is $O(\ln n)$. However we can get a stronger result with logarithmic Sobolev constants, which will be introduced properly in the next chapter. The following formulation by Montenegro [96] (derived from Houdré [55]) gives the following bound on log-Sobolev constants:

**Theorem 4.4.8** *Let $M$ be a Markov chain and define*

$$l_1^+ = \inf_{\pi_0 \le x \le 1/2} \frac{\Phi(x)}{\ln(1/x)},$$

*then the* log*-Sobolev constant $\alpha$ is bounded by*

$$\alpha \ge \frac{\lambda l_1^+}{2\left(\sqrt{\lambda} + 2l_1^+\right)} \ge \frac{\sqrt{\lambda} l_1^+}{12}.$$

We added the second inequality as a simplification using the fact that $2\lambda \ge \Phi^2 \ge (l_1^+ \ln 2)^2$. Then:

**Theorem 4.4.9** *The log-Sobolev constant and mixing time of the bases-exchange walk on any balanced matroid of rank $n$ on a ground set of size $m$ are bounded by*

$$\alpha \ge \frac{1}{32m^{3/2}n^2} \qquad \textit{and} \qquad \tau \le 32m^{3/2}n^2(\ln n + \ln\ln m).$$

*Proof.* The proof of Theorem 4.4.7 shows that $l_1^+ = 1/[(2\ln 2)mn]$. Montenegro [95] pointed out that Feder and Mihail in [41] implicitly bounded spectral gap by $\lambda \ge 1/(mn^2)$. Thus

$$\alpha \ge \frac{\sqrt{\lambda} l_1^+}{16\sqrt{2}} \ge \frac{1}{32\sqrt{2}\,\ln 2\, m^{3/2}n^2} \ge \frac{1}{32m^{3/2}n^2}.$$

Since log-Sobolev constants can be used to bound mixing time (see Lemma 5.1.2), we obtain

$$\tau \le 16m^{3/2}n^2(2 + \ln\ln(m^n)) \le 32m^{3/2}n^2(\ln n + \ln\ln m).$$

□

This is stronger than Theorem 4.3.11 ($\tau \in O(n^3 m \ln m)$) when $n \ln m \in \Omega(\sqrt{m} \ln n)$, e.g. when $m \in O(n^2)$. According to a result by Heller [53] this is true for *simple*[13] regular matroids: $m \leq n(n+1)$, which is smaller than $2n^2$ for $n \geq 1$, and implies that $m \in O(n^2)$ if the size of all parallel classes is bounded by a constant.

[13] Simple matroids are matroids without loops and parallel elements.

# Chapter 5

# Spectral gap and log-Sobolev constant for $\pi$-recursive Markov chains

In this chapter we will derive direct lower bounds for the spectral gap and log-Sobolev constant for a certain class of Markov chains. An example of such Markov chains is the bases-exchange walk on balanced matroids. The resulting bounds on mixing time are better than those given in the previous chapter.

## 5.1 Logarithmic Sobolev constants

Let $M$ be a time-reversible, irreducible and aperiodic Markov chain with transition kernel $P(x,y)$ on a finite state space $\Omega$ with stationary distribution $\pi$. Recall the definition of spectral gap in terms of Poincaré inequalities:

$$\lambda_M = \min_{f:\Omega\to\mathbb{R}} \left\{ \frac{\mathcal{E}_M(f,f)}{\mathrm{Var}_\pi(f)},\ \mathrm{Var}_\pi(f) \neq 0 \right\},$$

where $\mathcal{E}_M(f,f)$ and $\mathrm{Var}_\pi(f)$ are Dirichlet-form and variance respectively:

$$\begin{aligned} \mathcal{E}_M(f,f) &= \frac{1}{2} \sum_{x,y\in\Omega} (f(x)-f(y))^2 \pi(x) P(x,y) \\ \mathrm{Var}_\pi(f) &= \frac{1}{2} \sum_{x,y\in\Omega} (f(x)-f(y))^2 \pi(x)\pi(y). \end{aligned}$$

The subscript $M$ of $\mathcal{E}$ denotes that it is with respect to the Markov chain $M$ with transition kernel $P$, and thus probability distribution $\pi$, and similarly for variance.

Syntactically similar to Poincaré inequalities are so-called *logarithmic Sobolev inequalities*

$$\frac{\mathcal{E}_M(f,f)}{\mathcal{L}_\pi(f)} \geq c.$$

Here $c$ is a constant and

$$\mathcal{L}_\pi(f) = \sum_{x\in\Omega} |f(x)|^2 \pi(x) \ln \frac{|f(x)|^2}{\|f\|^2_{2,\pi}}$$

an entropy like quantity, with $\|f\|_{2,\pi} = \sqrt{\sum_{x\in\Omega} |f(x)|^2\pi(x)}$. Notice that $\|\cdot\|_{2,\pi}$ is not $L_2$-norm. Furthermore, we have used the absolute value signs in $\mathcal{L}_\pi(f)$ only to be consistent with other authors (cf. [30]).

**Definition 5.1.1** *The* log-Sobolev constant *of M is*

$$\alpha_M = \min_{f:\Omega\to\mathbb{R}} \left\{ \frac{\mathcal{E}_M(f,f)}{\mathcal{L}_\pi(f)} : \mathcal{L}_\pi(f) \neq 0 \right\}.$$

As with spectral gap (cf. (2.7)), there is a normalisation for log-Sobolev constants:

$$\alpha_M = \min_{f:\Omega\to\mathbb{R}} \left\{ \frac{\mathcal{E}_M(f,f)}{\mathcal{L}_\pi(f)} : \mathcal{L}_\pi(f) \neq 0, \|f\|_{2,\pi} = 1 \right\}, \tag{5.1}$$

Logarithmic Sobolev inequalities originate from Sobolev inequalities which were used in the study of elliptic differential operators. However Sobolev inequalities do not perform satisfactorily in higher dimensions. Gross [45] introduced logarithmic Sobolev inequalities to overcome this impasse and obtain good descriptions even when the number of variables is infinite.

Although formally the log-Sobolev constant resembles the spectral gap of the chain, that is where the similarity ends. While the spectral gap can be regarded as the difference between the first two largest eigenvalues of the kernel, there seems to be no analogous interpretation for the log-Sobolev constant. However, the log-Sobolev constant can be related to the hypercontractivity of the continuous-time semigroup [10] associated to the transition kernel of $M$ and more recent research discloses a connection with concentration- and mass transportation functions [11], [77]. Diaconis and Saloff-Coste's [30] fairly self-contained treatise on log-Sobolev constants in the context of finite Markov chains is an ideal starting point for readers unfamiliar with the subject.

Diaconis and Saloff-Coste [30] give the following bounds on mixing time:

**Lemma 5.1.2** *Let $M$ be a Markov chain with spectral gap $\lambda$ and* log*-Sobolev constant $\alpha$. Then the mixing time $\tau = \tau(1/e)$ of $M$ is upper bounded by*

$$\tau \leq \frac{2 + \ln(1/\pi_0)}{2\lambda} \tag{5.2}$$

$$\tau \leq \frac{4 + \ln\ln(1/\pi_0)}{4\alpha}, \tag{5.3}$$

*where* $\pi_0 = \min_{x \in \Omega} \{\pi(x)\}$.

We dropped the subscripts of $\alpha$ and $\lambda$ since here they were inessential. As inequality (5.3) shows, log-Sobolev constants give much better bounds on mixing time than spectral gap whenever $\lambda$ and $\alpha$ are of similar magnitude. Typically, $1/\pi_0 = 2^{O(n)}$ where $n$ measures input size. Thus using (5.3) gains us a factor of $n/\ln n$.

In recent years log-Sobolev constants have attracted a lot of interest especially in the field of statistical physics: log-Sobolev constants seem to be particularly well suited to describe the convergence behaviour of Markov chains for spin systems and research output from that community has been huge (see e.g. Cancrini & Roberto [20], Cancrini, Martinelli & Roberto [19], Cesi [22], Landim, Panizo & Yau [74], Ledoux [78], Martinelli, Sinclair & Weitz [88], Stroock & Zegarliński [113], Yoshida [118] and the references given therein). Other problems for which log-Sobolev constants have been studied are the standard diffusion on the $d$-cycle (Chen & Sheu [26]), the knapsack Markov chain (Mathieu [89]) and various kinds of random walks (Ané & Ledoux [8], Lee & Yau [79]). Also of interest is Roberto's approach to bound log-Sobolev constants via canonical paths (see [103]). Some general results on log-Sobolev constants were given by Ané et al. [7], Bobkov & Ledoux [14] and Guionnet & Zegarliński [49].

Peculiarly, in most cases where both spectral gap and log-Sobolev constant are known explicitly they tend to be very close: for the symmetric two-point space, the Ornstein-Uhlenbeck process on $\mathbb{R}^d$ or the standard diffusion on the $d$-sphere it is known that $\alpha = \lambda/2$ (see [10], [26], [45], [46]), a diffusion where $\alpha = \lambda/4$ is given in [73] and for the knapsack Markov chain [89] and several Markov chains for spin systems it is known that $\lambda$ and $\alpha$ are of the same order (see [19], [74], [88]). Since it has been shown by Rothaus [105] that $\alpha \leq \lambda/2$, one might wonder if $\alpha$ and $\lambda$ are always of the same order. However, there are cases for which very good estimates on

spectral gap and log-Sobolev constant are known where $\alpha \ll \lambda$. Examples are random walks on generic regular $r$-graphs (see [30]) or complete graphs (see [55]).

Recent results by Houdré [55] shows how quantities that are easier to compute, e.g. spectral gap or conductance, can be used to lower bound log-Sobolev constants. We have given an example by Theorem 4.4.8 in the last chapter.

## 5.2 π-recursive Markov chains

Let $M = (\Omega, P)$ be a finite, irreducible, aperiodic and time-reversible Markov chain with stationary distribution $\pi$. Observe that $M$ induces an undirected graph $G = (V, E)$ such that $V = \Omega$ and $E = \{\{x,y\} \mid P(x,y) > 0\}$. Conversely, we could say that $M = (\Omega, P)$ is a Markov chain on an undirected graph $G = (\Omega, E)$ such that for all edges $\{x,y\} \in E$ the transition probabilities $P(x,y)$ are non-zero. Call $G$ the *underlying graph of* $M$. We can regard $M$ as a *generalised* random walk on $G$ because it is not necessarily the case that all transitions from one state have the same probability. Let $A \neq \emptyset, \bar{A} \neq \emptyset$ be a bipartition of $\Omega$.

**Definition 5.2.1** *The set of* crossedges[1] $C(A,\bar{A}) \subseteq E$ *is the set of edges with one endpoint in* $A$ *and the other in* $\bar{A}$.

*An assignment* $w : C(A,\bar{A}) \to \mathbb{R}^+$ *satisfying*

$$\forall x \in A : \sum_{y \in \bar{A}} w(x,y) = \frac{\pi(x)}{\pi(A)} \text{ and } \forall y \in \bar{A} : \sum_{x \in A} w(x,y) = \frac{\pi(y)}{\pi(\bar{A})}. \tag{5.4}$$

*is called a* π-matching.

*The underlying graph* $G$ *is* π-matchable *if there exists a bipartition* $A$, $\bar{A}$ *that allows a* π*-matching and we will call the two subgraphs* $G_A$, $G_{\bar{A}}$ *induced by* $A$ *and* $\bar{A}$ *a* split (of $G$) *(or alternatively we say that* $G$ *is* split *into* $G_A$ *and* $G_{\bar{A}}$*). By a* series of splits *we mean a sequence* $G = G_{S_1}, G_{S_2}, \ldots, G_{S_l} = G_S$, $l \in \mathbb{N}$, *where each* $G_{S_i}$, $i = 2 \ldots l$, *is obtained by splitting* $G_{S_{i-1}}$ *and* $G_S$ *is said to be* obtained *by a series of splits.*

For the next definition we need the concept of restriction chain introduced by Madras and Randall [85]. If $M = (\Omega, P)$ is a time-reversible Markov chain with stationary

[1] Note that the set of crossedges $C(A,\bar{A})$ is the same as cut$(A)$ defined in chapter 4.

distribution $\pi$ and $S \subseteq \Omega$, then let $M_S$ be its restriction on $S$ with stationary distribution $\pi_S = \pi/\pi(S)$ (see section 6.1, p.96 for a justification); think of $M_S$ as a copy of $M$ where transitions leaving the set $S$ have been removed. Readers unfamiliar with the concept of restriction chains will find a definition in section 6.1.

**Definition 5.2.2** *Let $M = (\Omega, P)$ be a finite, irreducible, aperiodic and time-reversible Markov chain with stationary distribution $\pi$ and underlying graph $G = (\Omega, E)$. The chain $M$ is* $\pi$-recursive on $G$ *if:*

1. $\Omega$ is of size 1 or
2. $G$ is $\pi$-matchable with a split, say $G_A$ and $G_{\bar{A}}$, such that $M_A$ and $M_{\bar{A}}$ with stationary distributions $\pi_A$ and $\pi_{\bar{A}}$ are $\pi_A$-recursive on $G_A$ and $\pi_{\bar{A}}$-recursive on $G_{\bar{A}}$ respectively.

Examples of $\pi$-recursive Markov chains (as will be shown in section 5.5) are random walks on hypercubes and bases-exchange walks on balanced matroids. It is unkown whether there are any other $\pi$-recursive Markov chains.

## 5.3 Spectral gap of $\pi$-recursive Markov chains

Suppose that $M = (\Omega, P)$ is a finite, irreducible, aperiodic and time-reversible Markov chain with stationary distribution $\pi$ on an underlying graph $G = (\Omega, E)$. Furthermore assume that $M$ is $\pi$-recursive on $G$. In this section we will combine decomposition and induction to obtain a lower bound on the spectral gap of $M$. The induction hypothesis will be that $\mathcal{E}_M(f,f) \geq \hat{\lambda}\mathrm{Var}_\pi(f)$ for some value $\hat{\lambda}$ and all $f : \Omega \to \mathbb{R}$. If this Poincaré inequality holds, then from the definition of spectral gap $\lambda_M \geq \hat{\lambda}$ is an immediate consequence. The main idea is pretty simple:

Decompose both $\mathcal{E}_M(f,f)$ and $\mathrm{Var}_\pi(f)$ w.r.t. a bipartition $A, \bar{A}$ of $\Omega$ that allows a $\pi$-matching, i.e.

$$\mathcal{E}_M(f,f) = \pi(A)\mathcal{E}_{M_A}(f,f) + \pi(\bar{A})\mathcal{E}_{M_{\bar{A}}}(f,f) + \mathrm{ct}_{\mathcal{E}_M(f,f)}, \tag{5.5}$$

where $M_A$, $M_{\bar{A}}$ are the restriction chains of $M$ on $A$, $\bar{A}$ and $\mathrm{ct}_{\mathcal{E}_M(f,f)}$ is a *crossterm* induced by the crossedges $C(A,\bar{A})$ spanning $A$ and $\bar{A}$ and similarly for variance

$$\mathrm{Var}_\pi(f) = \pi(A)\mathrm{Var}_{\pi_A}(f) + \pi(\bar{A})\mathrm{Var}_{\pi_{\bar{A}}}(f) + \mathrm{ct}_{\mathrm{Var}_\pi(f)}. \tag{5.6}$$

Observe that since $M$ is π-recursive on $G$, the restriction chains $M_A$ and $M_{\bar{A}}$ have stationary distribution $\pi_A$ and $\pi_{\bar{A}}$ respectively. Since $M$ is π-recursive on $G$ and $A$, $\bar{A}$ induces a split, $M_A$ and $M_{\bar{A}}$ will be $\pi_A$- and $\pi_{\bar{A}}$-recursive on $G_A$ and $G_{\bar{A}}$. Hence, the induction hypothesis is applicable and

$$\mathcal{E}_M(f,f) \geq \pi(A)\hat{\lambda}\mathrm{Var}_{\pi_A}(f) + \pi(\bar{A})\hat{\lambda}\mathrm{Var}_{\pi_{\bar{A}}}(f) + \mathrm{ct}_{\mathcal{E}_M(f,f)}. \tag{5.7}$$

With the help of π-matchings we can show that there exists a suitable choice for $\hat{\lambda}$ such that

$$\mathrm{ct}_{\mathcal{E}_M(f,f)} \geq \hat{\lambda}\mathrm{ct}_{\mathrm{Var}_\pi(f)}. \tag{5.8}$$

Thus proving the induction hypothesis

$$\mathcal{E}_M(f,f) \geq \hat{\lambda}\mathrm{Var}_\pi(f). \tag{5.9}$$

In the rest of the section we will make the steps outlined by (5.5) to (5.9) rigorous.

**Lemma 5.3.1** *For a partition $A, \bar{A}$ of $\Omega$, $f : \Omega \to \mathbb{R}$ and a Markov chain with stationary distribution π on $\Omega$*

$$\mathcal{E}_M(f,f) = \pi(A)\mathcal{E}_{M_A}(f,f) + \pi(\bar{A})\mathcal{E}_{M_{\bar{A}}}(f,f) + \mathrm{ct}_{\mathcal{E}_M(f,f)},$$

*where*

$$\mathrm{ct}_{\mathcal{E}_M(f,f)} = \sum_{\substack{x\in A\\ y\in\bar{A}}} (f(x)-f(y))^2\pi(x)P(x,y).$$

*Proof.* Observe that given a partition $A$ and $\bar{A}$ the Dirichlet form can be rewritten as

$$\begin{aligned}\mathcal{E}_M(f,f) &= \frac{1}{2}\sum_{x,y\in A}(f(x)-f(y))^2\pi(x)P(x,y)\\ &\quad + \frac{1}{2}\sum_{x,y\in\bar{A}}(f(x)-f(y))^2\pi(x)P(x,y)\\ &\quad + \sum_{\substack{x\in A\\ y\in\bar{A}}}(f(x)-f(y))^2\pi(x)P(x,y).\end{aligned}$$

Since

$$\pi(A)\mathcal{E}_{M_A}(f,f) = \frac{1}{2}\sum_{x,y\in A}(f(x)-f(y))^2\pi(x)P(x,y)$$

and

$$\pi(\bar{A})\mathcal{E}_{M_{\bar{A}}}(f,f) = \frac{1}{2}\sum_{x,y\in \bar{A}}(f(x)-f(y))^2\pi(x)P(x,y),$$

the lemma holds. □

The next lemma shows that variance can be decomposed.

**Lemma 5.3.2** *Let $A$, $\bar{A}$ be a partition of $\Omega$, $f:\Omega\to\mathbb{R}$ and $M$ a Markov chain on $\Omega$ with stationary distribution $\pi$, then*

$$\mathrm{Var}_\pi(f) = \pi(A)\mathrm{Var}_{\pi_A}(f) + \pi(\bar{A})\mathrm{Var}_{\pi_{\bar{A}}}(f) + \mathrm{ct}_{\mathrm{Var}_\pi(f)},$$

*where*

$$\mathrm{ct}_{\mathrm{Var}_\pi(f)} = \pi(A)\pi(\bar{A})\left(\mathrm{E}_{\pi_A}[f] - \mathrm{E}_{\pi_{\bar{A}}}[f]\right)^2$$

*is a crossterm and*

$$\mathrm{E}_{\pi_A}[f] = \sum_{x\in A}\pi_A(x)f(x),$$

$$\mathrm{Var}_{\pi_A}(f) = \sum_{x\in A}\pi_A(x)\left(f(x)-\mathrm{E}_{\pi_A}[f]\right)^2$$

*are expectation and variance w.r.t. $\pi_A$ and similarly for $\mathrm{E}_{\pi_{\bar{A}}}[f]$ and $\mathrm{Var}_{\pi_{\bar{A}}}(f)$.*

*Proof.* First, we establish that

$$\begin{aligned}\mathrm{Var}_\pi(f) &= \pi(A)\mathrm{Var}_{\pi_A}(f) + \pi(\bar{A})\mathrm{Var}_{\pi_{\bar{A}}}(f)\\ &\quad + \pi(A)(\mathrm{E}_{\pi_A}[f]-\mathrm{E}_\pi[f])^2\\ &\quad + \pi(\bar{A})(\mathrm{E}_{\pi_{\bar{A}}}[f]-\mathrm{E}_\pi[f])^2.\end{aligned}$$

To see this, consider the following:

$$\begin{aligned}&\mathrm{Var}_\pi(f)\\ &= \pi(A)\left(\sum_{x\in A}(f(x)-\mathrm{E}_\pi[f])^2\pi_A(x)\right)\end{aligned}$$

$$
\begin{aligned}
&+\pi\left(\overline{A}\right)\left(\sum_{x\in\overline{A}}\left(f(x)-\mathrm{E}_{\pi}[f]\right)^{2}\pi_{\overline{A}}(x)\right)\\
=\;&\pi(A)\sum_{x\in A}f^{2}(x)\,\pi_{A}(x)-\pi(A)\,2\mathrm{E}_{\pi_{A}}[f]\,\mathrm{E}_{\pi}[f]+\pi(A)\,\mathrm{E}_{\pi}[f]^{2}\\
&+\pi\left(\overline{A}\right)\sum_{x\in\overline{A}}f^{2}(x)\,\pi_{\overline{A}}(x)-\pi\left(\overline{A}\right)2\mathrm{E}_{\pi_{\overline{A}}}[f]\,\mathrm{E}_{\pi}[f]+\pi\left(\overline{A}\right)\mathrm{E}_{\pi}[f]^{2}.
\end{aligned}
$$

By adding two zeros, namely $\pi(A)\,2\mathrm{E}_{\pi_A}[f]^2-\pi(A)\,2\mathrm{E}_{\pi_A}[f]^2$ and $\pi\left(\overline{A}\right)2\mathrm{E}_{\pi_{\overline{A}}}[f]^2-\pi\left(\overline{A}\right)2\mathrm{E}_{\pi_{\overline{A}}}[f]^2$, and re-arranging the summands we obtain

$$
\begin{aligned}
&\pi(A)\sum_{x\in A}f^{2}(x)\,\pi_{A}(x)-\pi(A)\,2\mathrm{E}_{\pi_{A}}[f]^{2}+\pi(A)\,\mathrm{E}_{\pi}[f]^{2}\\
&+\pi(A)\sum_{x\in A}f^{2}(x)\,\pi_{A}(x)-\pi(A)\,2\mathrm{E}_{\pi_{A}}[f]^{2}+\pi(A)\,\mathrm{E}_{\pi}[f]^{2}\\
&+\pi(A)\,\mathrm{E}_{\pi_{A}}[f]^{2}-\pi(A)\,2\mathrm{E}_{\pi_{A}}[f]\,\mathrm{E}_{\pi}[f]+\pi(A)\,\mathrm{E}_{\pi}[f]^{2}\\
&+\pi\left(\overline{A}\right)\mathrm{E}_{\pi_{\overline{A}}}[f]^{2}-\pi\left(\overline{A}\right)2\mathrm{E}_{\pi_{\overline{A}}}[f]\,\mathrm{E}_{\pi}[f]+\pi\left(\overline{A}\right)\mathrm{E}_{\pi}[f]^{2},
\end{aligned}
$$

which is the same as

$$
\begin{aligned}
&\pi(A)\left(\sum_{x\in A}\left(f(x)-\mathrm{E}_{\pi_{A}}[f]\right)^{2}\pi_{A}(x)\right)\\
&\quad+\pi\left(\overline{A}\right)\left(\sum_{x\in\overline{A}}\left(f(x)-\mathrm{E}_{\pi_{\overline{A}}}[f]\right)^{2}\pi_{\overline{A}}(x)\right)\\
&\quad+\pi(A)\left(\mathrm{E}_{\pi_{A}}[f]^{2}-2\mathrm{E}_{\pi_{A}}[f]\,\mathrm{E}_{\pi}[f]+\mathrm{E}_{\pi}[f]^{2}\right)\\
&\quad+\pi\left(\overline{A}\right)\left(\mathrm{E}_{\pi_{\overline{A}}}[f]^{2}-2\mathrm{E}_{\pi_{\overline{A}}}[f]\,\mathrm{E}_{\pi}[f]+\mathrm{E}_{\pi}[f]^{2}\right)\\
=\;&\pi(A)\,\mathrm{Var}_{\pi_{A}}(f)+\pi\left(\overline{A}\right)\mathrm{Var}_{\pi_{\bar{A}}}(f)\\
&\quad+\pi(A)(\mathrm{E}_{\pi_{A}}[f]-\mathrm{E}_{\pi}[f])^{2}\\
&\quad+\pi(\bar{A})(\mathrm{E}_{\pi_{\bar{A}}}[f]-\mathrm{E}_{\pi}[f])^{2}.
\end{aligned}
$$

The lemma then follows from

$$
\begin{aligned}
&\pi(A)(\mathrm{E}_{\pi_{A}}[f]-\mathrm{E}_{\pi}[f])^{2}+\pi(\bar{A})(\mathrm{E}_{\pi_{\bar{A}}}[f]-\mathrm{E}_{\pi}[f])^{2}\\
=\;&\pi(A)(\mathrm{E}_{\pi_{A}}[f]-\pi(A)\mathrm{E}_{\pi_{A}}[f]-\pi(\bar{A})\mathrm{E}_{\pi_{\bar{A}}}[f])^{2}\\
&\quad+\pi(\bar{A})(\mathrm{E}_{\pi_{\bar{A}}}[f]-\pi(A)\mathrm{E}_{\pi_{A}}[f]-\pi(\bar{A})\mathrm{E}_{\pi_{\bar{A}}}[f])^{2}\\
=\;&\pi(A)(\pi(\bar{A})\mathrm{E}_{\pi_{A}}[f]-\pi(\bar{A})\mathrm{E}_{\pi_{\bar{A}}}[f])^{2}
\end{aligned}
$$

$$
\begin{aligned}
&\quad + \pi(\bar{A})(\pi(A)\mathrm{E}_{\pi_{\bar{A}}}[f] - \pi(A)\mathrm{E}_{\pi_A}[f])^2 \\
&= \pi(A)\pi(\bar{A})^2(\mathrm{E}_{\pi_A}[f] - \mathrm{E}_{\pi_{\bar{A}}}[f])^2 \\
&\quad + \pi(\bar{A})\pi(A)^2(\mathrm{E}_{\pi_{\bar{A}}}[f] - \mathrm{E}_{\pi_A}[f])^2 \\
&= \pi(\bar{A})\pi(A)(\mathrm{E}_{\pi_{\bar{A}}}[f] - \mathrm{E}_{\pi_A}[f])^2.
\end{aligned}
$$

□

The following lemma corresponds to (5.8) and is the corner-stone of our argument.

**Lemma 5.3.3** *Assume that $M = (\Omega, P)$ is a finite, irreducible, aperiodic and time-reversible Markov chain with stationary distribution $\pi$ on a $\pi$-matchable graph $G = (\Omega, E)$. Then*

$$\mathrm{ct}_{\mathcal{E}_M(f,f)} \geq 2 \min_{\{x,y\}\in E} \{P(x,y)\}\, \mathrm{ct}_{\mathrm{Var}_\pi(f)}.$$

*Proof.* Since $G$ allows a $\pi$-matching, there exists a bipartition $A$, $\bar{A}$ of $\Omega$, an assignment $w$ such that $\pi(x)/\pi(A) = \sum_{y\in\bar{A}} w(x,y)$ and thus $\mathrm{E}_{\pi_A}[f] = \sum_{\substack{x\in A \\ y\in\bar{A}}} w(x,y)f(x)$; similarly for $\mathrm{E}_{\pi_{\bar{A}}}[f]$. Thus:

$$\mathrm{ct}_{\mathrm{Var}_\pi(f)} = \pi(A)\pi(\bar{A}) \left( \sum_{\substack{x\in A \\ y\in\bar{A}}} w(x,y)(f(x) - f(y)) \right)^2.$$

Observe that $\sum_{\substack{x\in A \\ y\in\bar{A}}} w(x,y) = 1$. Since $g(x) = x^2$ is a convex function, we obtain with Jensen's inequality:

$$\mathrm{ct}_{\mathrm{Var}_\pi(f)} \leq \pi(A)\pi(\bar{A}) \sum_{\substack{x\in A \\ y\in\bar{A}}} w(x,y)(f(x) - f(y))^2.$$

Thus, to complete the proof it suffices to find a factor $\eta$ such that

$$\eta\, \pi(A)\pi(\bar{A}) \sum_{\substack{x\in A \\ y\in\bar{A}}} w(x,y)(f(x) - f(y))^2 \leq \sum_{\substack{x\in A \\ y\in\bar{A}}} \pi(x)P(x,y)(f(x) - f(y))^2.$$

By convention, choose $A$ such that $\pi(\bar{A}) \leq \pi(A)$. Furthermore, $w(x,y) \leq \sum_{y\in\bar{A}} w(x,y) = \frac{\pi(x)}{\pi(A)}$. Thus

$$\pi(A)\pi(\bar{A})w(x,y) \leq \pi(A)\pi(\bar{A})\frac{\pi(x)}{\pi(A)} \leq \frac{\pi(x)P(x,y)}{2\min_{\{x,y\}\in E}\{P(x,y)\}}.$$

This shows that our claim is valid for $\eta = 2\min_{\{x,y\}\in E}\{P(x,y)\}$. □

The significance of Lemma 5.3.3 is that $2\min_{\{x,y\}\in E}\{P(x,y)\}$ is exactly the value $\hat{\lambda}$ we were looking for, i.e. it satisfies $\mathcal{E}_M(f,f) \geq \hat{\lambda}\mathrm{Var}_\pi(f)$ for all $f : \Omega \to \mathbb{R}$. This is shown next.

**Theorem 5.3.4** *Given a finite, irreducible, aperiodic and time-reversible Markov chain $M = (\Omega, P)$ with stationary distribution $\pi$ on a graph $G = (\Omega, E)$ such that $M$ is $\pi$-recursive on $G$. The spectral gap of $M$ is lower bounded by twice the minimum transition probability of $M$, i.e.:*

$$\lambda_M \geq 2 \min_{\{x,y\}\in E}\{P(x,y)\}.$$

*Proof.* It suffices to show that for any $f : \Omega \to \mathbb{R}$ and for all $S \subseteq \Omega$ where $G_S$ is obtained by a series of splits of $G$ the following holds:

$$\mathcal{E}_{M_S}(f,f) \geq 2 \min_{\{x,y\}\in E}\{P(x,y)\}\,\mathrm{Var}_{\pi_S}(f).$$

The proof is by induction on the size of the state space of the restriction chain $M_S = (S, P_S)$.

For $|S| = 1$ the variance

$$\mathrm{Var}_{\pi_S}(f) = \frac{1}{2}\sum_{x,y\in S}(f(x)-f(y))^2\pi_S(x)\pi_S(y) = 0$$

and the induction hypothesis is trivially true.

Assume that $|S| > 1$ and $M_S$ has stationary distribution $\pi_S$. Since $M$ is $\pi$-recursive on $G$, the chain $M_S$ is $\pi_S$-recursive on $G_S$. Thus there is a split of $G_S$ into, say $G_A$, $G_{\bar{A}}$. By Lemma 5.3.1:

$$\mathcal{E}_{M_S}(f,f) = \pi_S(A)\mathcal{E}_{M_A}(f,f) + \pi_S\left(\bar{A}\right)\mathcal{E}_{M_{\bar{A}}}(f,f) + \mathrm{ct}_{\mathcal{E}_{M_S}(f,f)}$$

which is by induction hypothesis greater than

$$2\min_{\{x,y\}}\{P(x,y)\}(\pi_S(A)\mathrm{Var}_{\pi_A}(f) + \pi_S\left(\bar{\mathrm{A}}\right)\mathrm{Var}_{\pi_{\bar{A}}}(f)) + \mathrm{ct}_{\mathcal{E}_{M_S}(f,f)}.$$

By Lemma 5.3.3: $\mathrm{ct}_{\mathcal{E}_{M_S}(f,f)} \geq 2\min_{\{x,y\}\in E}\{P(x,y)\}\mathrm{ct}_{\mathrm{Var}_{\pi_S}(f)}$. Thus, applying Lemma 5.3.2 we obtain:

$$\mathcal{E}_{M_S(f,f)} \geq 2 \min_{\{x,y\}\in E}\{P(x,y)\}\,\mathrm{Var}_{\pi_S(f)}.$$

□

*Remark*: An unnatural feature of Theorem 5.3.4 (and later Theorem 5.4.5) is that the existence of one single low-probability transition is enough to degrade the bound. What we mean is the following: let $M = (\Omega, P)$ be a Markov chain whose transition probabilities are similar in size (i.e. differ by at most a small constant factor). We define a "perturbed chain" $M^\dagger = (\Omega, P^\dagger)$ whose transition probabilities agree with $M$ except between a pair of distinguished states $x$ and $y$. For this pair $x, y$, we arrange that $P(x,y) = P(y,x) = 0$ while $0 < P^\dagger(x,y), P^\dagger(y,x) < \delta$, where $\delta$ is some small value much less than $\min_{\{x,y\}\in E}\{P(x,y)\}$. (The stationary distribution $\pi$ may be retained by choosing $P^\dagger(x,y)$ and $P^\dagger(y,x)$ to satisfy detailed balance.) Then our bound for $\lambda_{M^\dagger}$ is much smaller than that for $\lambda_M$, even though the spectral gap of $M^\dagger$ is actually slightly larger than that of $M$. This is counter–intuitive since the addition of a tiny extra edge should not have such a huge impact. This paradox can be resolved by a closer look at the proof of Lemma 5.3.3 and Theorem 5.3.4. Both continue to hold if $2\min_{\{x,y\}\in E}\{P(x,y)\}$ is replaced by

$$\min_{\{x,y\}\in C(A,\bar{A})} \frac{\pi(x)P(x,y)}{\pi(A)\pi(\bar{A})w(x,y)}. \tag{5.10}$$

Thus, the tiny transition probability can be compensated for as long as the $\pi$-matching does not assign it significant weight. In general, this will apply whenever a vertex is incident to more than one edge of the cut. Thus, an alternative formulation of Lemma 5.3.3 and Theorem 5.3.4 involving (5.10) will be closer to the truth. However, later on we will only be dealing with Markov chains whose transition probabilities are uniform, i.e. all non-zero transition probabilities are identical, and for those chains the current versions of Lemma 5.3.3 and Theorem 5.3.4 are sufficiently well suited.

## 5.4 log-Sobolev constant of π-recursive Markov chains

A similar inductive proof can be used to obtain a lower bound on the log-Sobolev constant of $\pi$-recursive Markov chains. As before we assume that $M = (\Omega, P)$ is a finite, irreducible, aperiodic and time-reversible Markov chain with uniform stationary distribution $\pi$ that is $\pi$-recursive on a graph$G = (\Omega, E)$. The goal will be to identify a

value $\hat{\alpha}$ such that $\mathcal{E}_M(f,f) \geq \hat{\alpha}\mathcal{L}_\pi(f)$ for all $f : \Omega \to \mathbb{R}$. Since the approach is identical to (5.5) to (5.9), only with $\hat{\lambda}$ and terms related to variance replaced by $\hat{\alpha}$ and terms related to the entropy-like quantity $\mathcal{L}_\pi(f)$, we will not restate it but go straight into details. Other authors who exploited decomposition to obtain bounds on log-Sobolev constants were Caputo & Martinelli [21], Cesi [22], Ledoux [77], Lu & Yau [83], Martinelli, Sinclair & Weitz [88] to name a few. Similar to variance and Dirichlet-form, the entropy-like quantity $\mathcal{L}_\pi(f)$ can be split up with respect to a bipartition $A$, $\bar{A}$ of $\Omega$:

**Lemma 5.4.1**

$$\mathcal{L}_\pi(f) = \pi(A)\mathcal{L}_{\pi_A}(f) + \pi(\bar{A})\mathcal{L}_{\pi_{\bar{A}}}(f) + \mathrm{ct}_{\mathcal{L}_\pi(f)},$$

*where*

$$\mathrm{ct}_{\mathcal{L}_\pi(f)} = \pi(A)\,\|f\|_{2,\pi_A}^2 \ln \frac{\|f\|_{2,\pi_A}^2}{\|f\|_{2,\pi}^2} + \pi(\bar{A})\,\|f\|_{2,\pi_{\bar{A}}}^2 \ln \frac{\|f\|_{2,\pi_{\bar{A}}}^2}{\|f\|_{2,\pi}^2},$$

$$\|f\|_{2,\pi_A} = \sqrt{\sum_{x\in A} |f(x)|^2 \frac{\pi(x)}{\pi(A)}} \text{ and } \mathcal{L}_{\pi_A}(f) = \sum_{x\in A} |f(x)|^2 \ln \frac{|f(x)|^2}{\|f\|_{2,\pi_A}^2} \frac{\pi(x)}{\pi(A)}$$

*and accordingly for* $\|f\|_{2,\pi_{\bar{A}}}$ *and* $\mathcal{L}_{\pi_{\bar{A}}}(f)$.

*Proof.* Recall that

$$\mathcal{L}_\pi(f) = \sum_{x\in\Omega} |f(x)|^2 \ln \frac{|f(x)|^2}{\|f\|_{2,\pi}^2} \pi(x). \tag{5.11}$$

Decomposing (5.11) with respect to $A$, $\bar{A}$ yields

$$\mathcal{L}_\pi(f) = \sum_{x\in A} |f(x)|^2 \ln |f(x)|^2 \pi(x) + \sum_{x\in\bar{A}} |f(x)|^2 \ln |f(x)|^2 \pi(x) - \|f\|_{2,\pi}^2 \ln \|f\|_{2,\pi}^2.$$

Notice that $\|f\|_{2,\pi}^2 = \pi(A)\,\|f\|_{2,\pi_A}^2 + \pi(\bar{A})\,\|f\|_{2,\pi_{\bar{A}}}^2$ and

$$\sum_{x\in A} |f(x)|^2 \pi(x) \ln |f(x)|^2 = \pi(A)\mathcal{L}_{\pi_A}(f) + \pi(A)\,\|f\|_{2,\pi_A}^2 \ln \|f\|_{2,\pi_A}^2.$$

Then

$$\begin{aligned}\mathcal{L}_\pi(f) \;=\; & \pi(A)\mathcal{L}_{\pi_A}(f) + \pi(\bar{A})\mathcal{L}_{\pi_{\bar{A}}}(f) + \pi(A)\,\|f\|_{2,\pi_A}^2 \ln \|f\|_{2,\pi_A}^2 \\ & + \pi(\bar{A})\,\|f\|_{2,\pi_{\bar{A}}}^2 \ln \|f\|_{2,\pi_{\bar{A}}}^2 - \ln \|f\|_{2,\pi}^2 \left(\pi(A)\,\|f\|_{2,\pi_A}^2 + \pi(\bar{A})\,\|f\|_{2,\pi_{\bar{A}}}^2\right)\end{aligned}$$

and observe that this is merely an alternative formulation of the claim. □

The next lemma is the log-Sobolev counterpart to Lemma 5.3.3:

**Lemma 5.4.2** *Assume that $M = (\Omega, P)$ is a finite, irreducible, aperiodic and time-reversible Markov chain with stationary distribution $\pi$ on a $\pi$-matchable graph $G = (\Omega, E)$. Then*

$$\mathrm{ct}_{\mathcal{E}_M(f,f)} \geq \frac{1}{2} \min_{\{x,y\}\in E} \{P(x,y)\}\, \mathrm{ct}_{\mathcal{L}_\pi(f)}.$$

*Proof.* Let $A$, $\bar{A}$ be a split of $G$, i.e. a bipartition yielding a $\pi$-matching $w$. Using $\ln x \leq x - 1$ we obtain

$$\mathrm{ct}_{\mathcal{L}_\pi(f)} \leq \pi(A)\,\|f\|^2_{2,\pi_A} \frac{\|f\|^2_{2,\pi_A} - \|f\|^2_{2,\pi}}{\|f\|^2_{2,\pi}} + \pi(\bar{A})\,\|f\|^2_{2,\pi_{\bar{A}}} \frac{\|f\|^2_{2,\pi_{\bar{A}}} - \|f\|^2_{2,\pi}}{\|f\|^2_{2,\pi}}.$$

Since $\|f\|^2_{2,\pi} = \pi(A)\,\|f\|^2_{2,\pi_A} + \pi(\bar{A})\,\|f\|^2_{2,\pi_{\bar{A}}}$, we see that the right-hand side of the above inequality is equal to

$$\pi(A)\pi(\bar{A}) \frac{\left(\|f\|^2_{2,\pi_A} - \|f\|^2_{2,\pi_{\bar{A}}}\right)^2}{\|f\|^2_{2,\pi}}.$$

It is helpful to observe that $\|f\|^2_{2,\pi_A} = \mathrm{E}_{\pi_A}\left[f^2\right]$ and similarly $\|f\|^2_{2,\pi_{\bar{A}}} = \mathrm{E}_{\pi_{\bar{A}}}\left[f^2\right]$. Furthermore, owing to the normalizing condition (5.1) for log-Sobolev constants we can w.l.o.g. assume that$\|f\|_{2,\pi} = 1$ and thus

$$\mathrm{ct}_{\mathcal{L}_\pi(f)} \leq \pi(A)\pi(\bar{A}) \left(\mathrm{E}_{\pi_A}\left[f^2\right] - \mathrm{E}_{\pi_{\bar{A}}}\left[f^2\right]\right)^2.$$

It is then enough to show that

$$\mathrm{ct}_{\mathcal{E}_M(f,f)} \geq \frac{1}{2} \min_{\{x,y\}\in E} \{P(x,y)\}\, \pi(A)\pi(\bar{A}) \left(\mathrm{E}_{\pi_A}\left[f^2\right] - \mathrm{E}_{\pi_{\bar{A}}}\left[f^2\right]\right)^2.$$

This can be done with the help of two intermediate inequalities. The first is

$$\sum_{\substack{x\in A\\ y\in\bar{A}}} (f(x) - f(y))^2 \pi(x)P(x,y) \geq \frac{1}{2} \min_{\{x,y\}\in E} \{P(x,y)\} \sum_{\substack{x\in A\\ y\in\bar{A}}} (f(x) - f(y))^2 w(x,y)$$

and the second one

$$\sum_{\substack{x\in A\\ y\in\bar{A}}} (f(x) - f(y))^2 w(x,y) \geq \pi(A)\pi(\bar{A}) \left(\mathrm{E}_{\pi_A}\left[f^2\right] - \mathrm{E}_{\pi_{\bar{A}}}\left[f^2\right]\right)^2. \tag{5.12}$$

As regards the first inequality, observe the following: As in the proof of Lemma 5.3.3 we can choose $A$ such that $\pi(\bar{A}) \leq \pi(A)$. Since $w$ is a $\pi$-matching between $A$ and $\bar{A}$, we have $w(x,y) \leq \sum_{y\in\bar{A}} w(x,y) = \frac{\pi(x)}{\pi(A)}$ and thus,

$$w(x,y) \leq \frac{\pi(x)}{\pi(A)} \leq 2\pi(x) \leq \frac{2\pi(x)P(x,y)}{\min_{\{x,y\}\in E}\{P(x,y)\}}.$$

This establishes the first inequality. To finish the proof we must show that the second inequality (5.12) is valid. This can be done using another two intermediate inequalities. The first one is an instance of the following multivariate Jensen's inequality. We will state the inequality but not give a proof since it is very similar to that of the one-dimensional case.

**Lemma 5.4.3** *If $g : \mathbb{R}^n \to \mathbb{R}$ is a convex function, $p : \prod_{i=1}^n \Omega_i \to [0,1]$ a multivariate probability distribution with marginals $p_i$, $i = 1,\ldots,n$, and $X_i$ random variables on $\Omega_i$, $i = 1,\ldots,n$, then*

$$g\left(\mathrm{E}_{p_1}[X_1],\ldots,\mathrm{E}_{p_n}[X_n]\right) \leq \mathrm{E}_p\left[g(X_1,\ldots,X_n)\right].$$

Notice that a $\pi$-matching $w$ is a bi-variate distribution on $A\times\bar{A}$ with marginals $\pi_A$ and $\pi_{\bar{A}}$. Following Jerrum et al. [62] (see also Kipnis, Olla & Varadhan [72] and Lee & Yau [79]), we choose $g(x,y) = \left(\sqrt{x}-\sqrt{y}\right)^2$ — the convexity of which can be verified by a simple calculation — and random variables $X_A(x) = f^2(x)$, $X_{\bar{A}}(y) = f^2(y)$ on $A$ and $\bar{A}$ respectively. Then by Lemma 5.4.3

$$\sum_{\substack{x\in A\\ y\in\bar{A}}} (f(x)-f(y))^2 w(x,y) \geq \left(\sqrt{\mathrm{E}_{\pi_A}[f^2]} - \sqrt{\mathrm{E}_{\pi_{\bar{A}}}[f^2]}\right)^2.$$

Inequality (5.12) then follows from

$$\left(\sqrt{\mathrm{E}_{\pi_A}[f^2]} - \sqrt{\mathrm{E}_{\pi_{\bar{A}}}[f^2]}\right)^2 \geq \pi(A)\pi(\bar{A})\left(\mathrm{E}_{\pi_A}[f^2] - \mathrm{E}_{\pi_{\bar{A}}}[f^2]\right)^2. \tag{5.13}$$

Because of the normalisation $\|f\|_{2,\pi}^2 = 1$, inequality (5.13) is implicitly of the form

$$\left(\sqrt{\frac{x}{y}} - \sqrt{\frac{1-x}{1-y}}\right)^2 \geq y(1-y)\left(\frac{x}{y} - \frac{1-x}{1-y}\right)^2.$$

Below, in Lemma 5.4.4 we will show that the above inequality is valid, which concludes the proof of Lemma 5.4.2. □

**Lemma 5.4.4** *Let $x, y \in (0,1)$. Then*

$$\left(\sqrt{\frac{x}{y}} - \sqrt{\frac{1-x}{1-y}}\right)^2 - y(1-y)\left(\frac{x}{y} - \frac{1-x}{1-y}\right)^2 \geq 0.$$

*Proof*[2]. The claimed inequality is equivalent to

$$1 - \frac{y(1-y)(\frac{x}{y} - \frac{1-x}{1-y})^2}{\left(\sqrt{\frac{x}{y}} - \sqrt{\frac{1-x}{1-y}}\right)^2} \geq 0$$

and hence to

$$y(1-y)\left(\sqrt{\frac{x}{y}} + \sqrt{\frac{1-x}{1-y}}\right)^2 \leq 1. \tag{5.14}$$

Since $x, y \in (0,1)$, inequality (5.14) is equivalent to

$$\sqrt{x(1-y)} + \sqrt{y(1-x)} \leq 1.$$

Then use $x = \cos^2 a$ and $y = \sin^2 b$ for $a, b \in \mathbb{R}$ to conclude the proof. □

We can now formulate the main result

**Theorem 5.4.5** *Given a finite, irreducible, aperiodic and time-reversible Markov chain $M = (\Omega, P)$ with stationary distribution $\pi$ on a graph $G = (\Omega, E)$ such that $M$ is $\pi$-recursive on $G$. The logarithmic Sobolev constant of $M$ is lower bounded by half the minimum transition probability of $M$, i.e.:*

$$\alpha_M \geq \frac{1}{2} \min_{\{x,y\} \in E} \{P(x,y)\}.$$

*Proof.* The proof is of a similar flavour as the proof of Theorem 5.3.4. It is carried out by induction on the size of state spaces of restriction chains of $M$. Let $M_S = (S, P_S)$ with stationary distribution $\pi_S$ be a restriction chain on the underlying graph $G_S = (S, E_S)$ obtained by a series of splits of $G$.

[2]The proof given is due to an anonymous reviewer of [61]. The original proof was by a lengthy calculation showing that the lhs of the claimed inequality is the square of a real number.

If $|S| = 1$, then $\pi_S(x) = 1$ for $x \in S$, in which case the hypothesis is trivially true because

$$\begin{aligned} 0 &= \frac{1}{2}\sum_{x,y\in S}(f(x)-f(y))^2\pi_S(x)P_S(x,y) \\ &= \frac{1}{2}\min_{\{x,y\}\in E}\{P(x,y)\}\sum_{x\in S}|f(x)|^2\pi_S(x)\ln\frac{|f(x)|^2}{\|f\|^2_{2,\pi_S}} \\ &= 0. \end{aligned}$$

Inductive step: As $M_S$ is $\pi_S$-recursive, the underlying graph $G_S$ can be split into two non–empty sets $A$ and $\bar{A}$ such that:

$$\mathcal{E}_{M_S}(f,f) = \pi_S(A)\mathcal{E}_{M_A}(f,f) + \pi_S(\bar{A})\mathcal{E}_{M_{\bar{A}}}(f,f) + \mathrm{ct}_{\mathcal{E}_{M_S}(f,f)}.$$

By the inductive hypothesis this is greater than

$$\frac{1}{2}\min_{\{x,y\}\in E}\{P(x,y)\}\left(\pi_S(A)\mathcal{L}_{\pi_A}(f) + \pi_S(\bar{A})\mathcal{L}_{\pi_{\bar{A}}}(f)\right) + \mathrm{ct}_{\mathcal{E}_{M_S}(f,f)}.$$

By Lemma 5.4.2 this in turn is greater than

$$\frac{1}{2}\min_{\{x,y\}\in E}\{P(x,y)\}\left(\pi_S(A)\mathcal{L}_{\pi_A}(f) + \pi_S(\bar{A})\mathcal{L}_{\pi_{\bar{A}}}(f) + \mathrm{ct}_{\mathcal{L}_{\pi_S}(f)}\right)$$

and finally using Lemma 5.4.1:

$$\mathcal{E}_{M_S}(f,f) \geq \frac{1}{2}\min_{\{x,y\}\in E}\{P(x,y)\}\mathcal{L}_{\pi_S}(f).$$

□

*Remark*: The bound on the log-Sobolev constant we state here is nearly tight in that any significant improvement, specifically by a factor greater than 2, would implicitly lead to an improvement in the bound on $\lambda$ stated in Theorem 5.3.4. (This follows from the known inequality $\alpha \leq \lambda/2$ [105], [30].) But the bound in Theorem 5.3.4 is tight, e.g. in the case of the hypercube: suppose we split a $d$-dimensional hypercube into two $d-1$-dimensional subcubes. Choose a function $f$ that assumes $+1$ on one subcube and $-1$ on the other. Since the random walk on this hypercube has constant transition probabilities[3] $P(x,y) = 1/(d+1)$, for adjacent $x$ and $y$, and uniform stationary distribution, we have $\mathcal{E}_M(f,f)/\mathrm{Var}_\pi(f) = 2/(d+1)$, which gives an upper bound of $2/(d+1)$ on $\lambda$. Since by Theorem 5.3.4 the lower bound is $\lambda \geq 2/(d+1)$, our bounds are tight for random walks on hypercubes.

[3]The transition probabilities $1/(d+1)$ are obtained by adding self-loops to address ergodicity issues.

## 5.5 Applications

In this section we will discuss some applications of Theorem 5.3.4 and Theorem 5.4.5. The applications we have in mind are random walks on hypercubes and bases-exchange walks on balanced matroids.

### 5.5.1 Random walk on a hypercube

Consider the $d$-dimensional hypercube. Its vertices can be denoted by $d$-tuples from $\Omega = \{0,1\}^d$. Its edges connect vertices whose Hamming distance is 1, i.e. tuples that differ by exactly one entry. Let $M$ be the random walk with transition probabilities $P(x,y) = 1/(d+1)$ for (undirected) edges and self-loops on such a hypercube. This random process has been studied quite extensively before (see [31] and the references given therein) and has uniform stationary distribution $\pi(x) = 1/2^d$. It is already known that this walk on the hypercube has spectral gap $\lambda = 2/(d+1)$ and log-Sobolev constant $\alpha = 1/(d+1)$ (see [30] and [31] for a discussion).

Here, we will obtain a similar result by using the $\pi$-recursiveness of this walk. The easiest way to show that $M$ is $\pi$-recursive is to split the hypercube into two sub-cubes of dimension $d-1$. Since each vertex of a sub-cube has only one neighbour in the other sub-cube, the $\pi$-matching is given by $w(x,y) = 2/2^d$ if $x$ and $y$ are neighbours. Using Theorem 5.3.4 and Theorem 5.4.5 we obtain the bounds $\lambda \geq 2/(d+1)$ and $\alpha \geq 1/2(d+1)$. While our bound on spectral gap is identical to the aforementioned result, our bound on log-Sobolev constant is off by a factor of 2. However, for the random walk on a hypercube it can be shown that Lemma 5.4.2 holds for $\mathrm{ct}_{\mathcal{E}_M(f,f)} \geq \frac{1}{d+1}\mathrm{ct}_{\mathcal{L}_\pi(f)}$, which recovers the result given in [30] by Diaconis and Saloff-Coste:

**Lemma 5.5.1** *For the $d$-dimensional hypercube*

$$\mathrm{ct}_{\mathcal{E}_M(f,f)} \geq \frac{1}{d+1}\mathrm{ct}_{\mathcal{L}_\pi(f)}.$$

*Proof.* Recall the definitions needed for Lemma 5.4.2. What we will show is

$$\sum_{\substack{x\in A \\ y\in \bar{A}}} (f(x)-f(y))^2 \pi(x)P(x,y)$$

$$\geq \frac{1}{d+1}\left(\pi(A)\,\|f\|^2_{2,\pi_A}\ln\frac{\|f\|^2_{2,\pi_A}}{\|f\|^2_{2,\pi}}+\pi(\bar{A})\,\|f\|^2_{2,\pi_{\bar{A}}}\ln\frac{\|f\|^2_{2,\pi_{\bar{A}}}}{\|f\|^2_{2,\pi}}\right), \qquad (5.15)$$

where $A$ and $\bar{A}$ is a partition of $\Omega$. Since $A$ and $\bar{A}$ are chosen in such a way that they form hypercubes of dimension $d-1$, there is a perfect matching between $A$ and $\bar{A}$, i.e. every vertex $x \in A$ has exactly one neighbour $\tilde{x} \in \bar{A}$. We know that the stationary distribution of this random walk is uniform and so $\pi(x) = 1/|\Omega|$. If we define $\sum_{x\in A} f^2(x) = S_A$ and $\sum_{x\in\bar{A}} f^2(x) = S_{\bar{A}}$, then (5.15) becomes

$$\frac{1}{d+1}\sum_{x\in A}(f(x)-f(\tilde{x}))^2 \geq \frac{1}{d+1}\left(S_A\ln\frac{2S_A}{S_A+S_{\bar{A}}}+S_{\bar{A}}\ln\frac{2S_{\bar{A}}}{S_A+S_{\bar{A}}}\right),$$

Notice that we do not lose any information if we normalise $S_A + S_{\bar{A}}$ to 1. Furthermore, observe that

$$\sum_{x\in A}(f(x)-f(\tilde{x}))^2 = S_A - 2\sum_{x\in A} f(x)f(\tilde{x}) + S_{\bar{A}}.$$

Since by the Cauchy-Schwarz inequality

$$\sum_{x\in A} f(x)f(\tilde{x}) \leq \sqrt{\left(\sum_{x\in A} f^2(x)\right)\left(\sum_{\tilde{x}\in\bar{A}} f^2(\tilde{x})\right)},$$

we see that

$$\sum_{x\in A}(f(x)-f(\tilde{x}))^2 \geq \left(\sqrt{S_A}-\sqrt{S_{\bar{A}}}\right)^2.$$

Thus (5.15) holds if

$$\left(\sqrt{S_A}-\sqrt{S_{\bar{A}}}\right)^2 \geq S_A\ln 2S_A + S_{\bar{A}}\ln 2S_{\bar{A}}.$$

Or equivalently since $S_A + S_{\bar{A}} = 1$,

$$g(x) = \left(\sqrt{x}-\sqrt{1-x}\right)^2 - x\ln 2x - (1-x)\ln 2(1-x) \geq 0.$$

Now observe that

$$\begin{aligned} g'(x) &= \ln\left(\frac{1-x}{x}\right) - \sqrt{\frac{1-x}{x}} + \sqrt{\frac{x}{1-x}} \\ g''(x) &= \frac{1}{x(x-1)} + \frac{1}{2x^2\sqrt{\frac{1-x}{x}}} + \frac{1}{2(1-x)^2\sqrt{\frac{x}{1-x}}} \\ &= \frac{1}{2\sqrt{x(1-x)}}\left(\frac{1}{\sqrt{1-x}}-\frac{1}{\sqrt{x}}\right)^2. \end{aligned}$$

Since $g''(x) \geq 0$, the function $g(x)$ has only one local minimum, which is attained at $x = 1/2$ as $g'(1/2) = 0$ shows. It is easy to verify that $g(1/2) = 0$ and therefore $g(x) \geq 0$. This concludes the proof. □

### 5.5.2 Bases-exchange walk on a balanced matroid

Let $\mathcal{M}$ be a balanced matroid with bases $\mathcal{B}$. Recall the definition of balanced matroids given in section 4.3. We already mentioned that Feder and Mihail [41] discovered that every balanced matroid allows a fractional matching. Recall that for two natural isomorphic copies of $\mathcal{B}(\mathcal{M} \setminus e)$ and $\mathcal{B}(\mathcal{M}/e)$, say $A$ and $\bar{A}$, a fractional matching is a function $\mathfrak{f} : A \times \bar{A} \to \mathbb{N}$ assigning integer weights to the cutset (the edges of $G(\mathcal{M})$ spanning $A$ and $\bar{A}$) such that $\forall x \in A : \sum_{y \in \bar{A}} \mathfrak{f}(x,y) = |\bar{A}|$ and $\forall y \in \bar{A} : \sum_{x \in A} \mathfrak{f}(x,y) = |A|$. Note that the bases-exchange walk on (any) matroid has uniform stationary distribution and hence for balanced matroids the probability distribution $P_{\mathfrak{f}} = \left((x,y) \mapsto \frac{\mathfrak{f}(x,y)}{|A||\bar{A}|}\right) : A \times \bar{A} \to [0,1]$ on the cutset is a $\frac{1}{|\mathcal{B}|}$-matching (because the stationary distribution is $\frac{1}{|\mathcal{B}|}$).

Let $M = (\mathcal{B}, P)$ be a bases-exchange walk with constant transition probabilities (i.e. $P(x,y) = p$ if $x \neq y$ and $x \sim y$, $P(x,y) = 0$ if $x \not\sim y$ and $P(x,x) = 1 - \sum_{\{x,y\} \in E} P(x,y)$) on a matroid $\mathcal{M}(\mathcal{B})$ and $G(\mathcal{M}(\mathcal{B})) = (\mathcal{B}, E)$ the underlying bases-exchange graph. Note that $M$ is aperiodic if the self-loop probabilities are larger than zero. Furthermore, $M$ is finite, irreducible and time-reversible with uniform stationary distribution $\pi(x) = 1/|\mathcal{B}|$ for all $x \in \mathcal{B}$. Observe that if $\mathcal{M}'$ with bases $\mathcal{B}'$ is a minor of $\mathcal{M}(\mathcal{B})$, then the restriction chain $M_{\mathcal{B}'}$ is time-reversible with stationary distribution $1/|\mathcal{B}'|$. Furthermore if $\mathcal{M}$ is a balanced matroid, then every fractional matching gives rise to a $\frac{1}{|\mathcal{B}|}$-matching with weights $P_{\mathfrak{f}}(\cdot,\cdot)$. These observations show that the bases-exchange walk $M = (\mathcal{B}, P)$ with constant transition probabilities is $\frac{1}{|\mathcal{B}|}$-recursive on the bases-exchange graph. It might be that some minors of a matroid are empty. This can happen if the minor is obtained by contracting an element that is in no basis or by deleting an element that is in every basis. In this case the minors do not constitute a split. However, unless a matroid has only one basis, there will always exist non-empty minors and thus a split of the bases-exchange graph.

**Corollary 5.5.2** *Let $M = (\mathcal{B}, P)$ be a bases-exchange walk with constant transition probabilities, i.e. $P(x,y) = p$ if $x \neq y$ and $x \sim y$, $P(x,y) = 0$ if $x \not\sim y$ and $P(x,x) = 1 - \sum_{\{x,y\} \in E} P(x,y)$, and uniform stationary distribution on a balanced matroid $\mathcal{M}(\mathcal{B})$ with bases-exchange graph $G(\mathcal{M}(\mathcal{B})) = (\mathcal{B}, E)$. The spectral gap and* log*-Sobolev constant of $M$ are lower bounded by*

$$\begin{aligned} \lambda_M &\geq 2p \\ \alpha_M &\geq \frac{p}{2}. \end{aligned}$$

*Proof.* Apply Theorems 5.3.4 and 5.4.5. □

With Corollary 5.5.2 we can improve on the previous bounds for Algorithm 4.3.5: Let $M$ be a matroid of rank $n$ on a ground set $S$ of size $m$. Assume that transitions are done as specified in Algorithm 4.3.5. For the sake of simplicity add self-loops with probability at least $1/2$ and reduce all other transition probabilities by $1/2$, i.e. $P(x,y) = 1/2mn$ for $x \sim y$ and $P(x,x) = 1 - \sum_{\{x,y\} \in E} P(x,y) \geq 1/2$, this will ensure that $M$ is finite, irreducible, aperiodic and time-reversible w.r.t. the uniform distribution on $\mathcal{B}$. If $M$ is balanced, then applying Theorem 5.4.5 yields the lower bound $\alpha \geq 1/4mn$ on the log-Sobolev constant. Remember that log-Sobolev constants can be used to upper bound mixing time and thus by Lemma 5.1.2 we obtain

$$\tau \in O(mn(\ln n + \ln\ln m)).$$

This greatly improves on the mixing time for this walk given in Theorem 4.4.9:

$$\tau \in O\left(m^{3/2} n^2 (\ln n + \ln\ln m)\right).$$

Using a result by Heller [53] which says that for *simple* regular matroids (which means regular matroids with no parallel elements) $m < n(n+1)$, we can infer that for regular matroids with only a constant number of parallel elements $m \in O(n^2)$. Thus for such matroids our bound on the log-Sobolev constant yields a mixing time of $O(n^3 \ln n)$ for this walk. This is at present the best bound on mixing time for this random walk on the bases-exchange graph of regular matroids with constant number of parallel elements. However, if the size of the groundset is larger, say $m \in \Omega(n^3)$, then Feder and

Mihail's *modified* random walk on balanced matroids (see [41], Theorem 5.2) with $\tau \in O(n^4 \ln m)$ converges after a fewer number of steps[4].

[4]This bound for the *modified* walk however is misleading because it does not mention implementation costs of the walk which the authors admit to being quite high.

# Chapter 6

# A new decomposition result

Markov chain decomposition is a fairly new technique introduced by Madras and Randall [85]. Their idea was to simplify the computation of the spectral gap of a Markov chain by decomposing it into smaller chains. This is done by first introducing a couple of new Markov chains called *restrictions,* which are like the original chain but only defined on subsets of the original state space. These subsets need not necessarily form a partition of the original state space. In fact it is desirable if they overlap to a certain extent. The purpose of restriction chains is to capture the behaviour of the original chain on smaller regions of the state space. The next step is to define a *projection chain* which roughly describes how the original chain moves from one region of the state space to another. In a later paper, Martin and Randall [86] proposed a slightly different kind of decomposition. Their result uses restriction chains that do not overlap, i.e. restriction chains are defined on partitions of the state space. Both decomposition methods yield bounds on the spectral gap of the original chain by combining bounds on the spectral gap of restriction and projection chains. Both results are based on a result by Caracciolo, Pelisetto and Sokal, which is a decomposition result in the context of simulated tempering. The result by Caracciolo, Pelisetto and Sokal has not been published yet but is treated in detail in [85].

In the rest of this chapter we will describe the decomposition techniques used by Madras, Martin and Randall. After that we will give a different decomposition approach for bounding the spectral gap of a Markov chain and apply it to the Ising process on finite, complete $d$-ary trees.

## 6.1 Madras-Randall versus Martin-Randall decomposition

We start off with a recap of the decomposition methods by Madras-Randall and Martin-Randall. In their original formulation both decomposition methods are applicable to finite and infinite state space Markov chains. However, we will only give a formulation for finite Markov chains. Let $M = (\Omega, P)$ be a finite, time-reversible Markov chain with a stationary distribution $\pi$.

**Definition 6.1.1** *Let $A_1, \ldots, A_m$ be not necessarily disjoint subsets of $\Omega$ such that $\bigcup A_i = \Omega$. For each $i = 1, \ldots, m$ define a* restriction chain *$M_{A_i}$ on $A_i$ by deleting any transitions that would take us out of $A_i$. The transition kernel of the restriction chain is given by*

$$P_{A_i}(x,y) = \begin{cases} P(x,y), & \text{if } x,y \in A_i \text{ and } x \neq y \\ 0, & \text{if } x \notin A_i \text{ or } y \notin A_i \\ 1 - \sum_{\substack{y \in A_i \\ y \neq x}} P(x,y), & \text{if } x = y. \end{cases} \tag{6.1}$$

It is worthwhile mentioning that $M_{A_i}$ has stationary distribution $\pi_{A_i}$ where

$$\pi_{A_i}(x) = \frac{\pi(x)}{\pi(A_i)} \text{ for } x \in A_i \tag{6.2}$$

and $\pi(A_i) = \sum_{x \in A_i} \pi(x)$ because

$$\begin{aligned} & \sum_{y \in A_i} \pi_{A_i}(y) P_{A_i}(y,x) \\ &= \frac{1}{\pi(A_i)} \left( \sum_{\substack{y \in A_i \\ y \neq x}} \pi(y) P(y,x) + \pi(x) \left( 1 - \sum_{\substack{y \in A_i \\ y \neq x}} P(x,y) \right) \right) \\ &= \frac{1}{\pi(A_i)} \left( \sum_{\substack{y \in A_i \\ y \neq x}} \pi(y) P(y,x) + \pi(x) - \sum_{\substack{y \in A_i \\ y \neq x}} \pi(y) P(y,x) \right) \\ &= \pi_{A_i}(x). \end{aligned}$$

Above, the first equality follows from using the definition of the restriction chain and the second follows from the fact that we assumed that $M$ is time-reversible w.r.t. $\pi$. Note that $M_{A_i}$ is time-reversible with respect to $\pi_{A_i}$.

Next define

$$Z := \sum_{i=1}^{m} \pi(A_i), \tag{6.3}$$

and the *maximum overlap* $\Theta$ of the covering $\{A_1, \ldots, A_m\}$ by

$$\Theta := \max_{x \in \Omega} |\{i \mid x \in A_i\}|. \tag{6.4}$$

In other words, $\Theta$ is the maximum number of sets that are involved in a non-empty intersection. Note that

$$Z = \sum_{i=1}^{m} \pi(A_i) \leq \Theta \cdot \pi(\Omega)$$

and thus

$$1 \leq Z \leq \Theta \leq m. \tag{6.5}$$

Next, we define a crude model of how the original chain moves from one region to another. Consider a state space $H = \{a_1, \ldots, a_m\}$ of $m$ points representing the $m$ subsets. The *projection chain* on this discrete state space has transition probabilities

$$P_H(a_i, a_j) := \frac{\pi(A_i \cap A_j)}{\Theta \pi(A_i)} \text{ for } i \neq j \tag{6.6}$$

and $P_H(a_i, a_i) = 1 - \sum_{i \neq j} P_H(a_i, a_j)$.

Madras and Randall, in [85], showed that the spectral gap of the original Markov chain can then be bounded by

**Theorem 6.1.2** *In the framework given by (6.1)–(6.6) the spectral gap of the Markov chain $M$ is lower bounded by*

$$\lambda_M \geq \frac{\lambda_{M_H}}{\Theta^2} \min_i \left\{ \lambda_{M_{A_i}} \right\}.$$

Recently, in [86], Martin and Randall have proved a similar result but for restrictions on partitions of the state space, i.e. the subsets do no longer need to overlap. The definition of restriction chains remains unchanged but since restrictions are defined on a partitioning of the state space, the definition of the projection chain must obviously be revised:

Let $\Omega_1, \ldots, \Omega_n$ be a partitioning of the state space $\Omega$. The projection chain $M_{\hat{H}}$ on $\hat{H} = \{\hat{\omega}_1, \ldots, \hat{\omega}_n\}$ has transition probabilities

$$P_{\hat{H}}(\hat{\omega}_i, \hat{\omega}_j) := \frac{1}{\pi(\Omega_i)} \sum_{\substack{x \in \Omega_i \\ y \in \Omega_j}} \pi(x) P(x, y).$$

**Theorem 6.1.3** *With the definitions of above, the spectral gap of $M$ is bounded by*

$$\lambda_M \geq \frac{\lambda_{M_{\hat{H}}}}{2} \min_i \left\{ \lambda_{M_{\Omega_i}} \right\}.$$

The advantage of this approach as claimed by the authors is that it is more easily applicable than the Madras-Randall method since it dispenses with the need for overlap, which can in some cases pose a technical difficulty.

## 6.2 Another decomposition

In this section we will give a new decomposition method to obtain bounds on the spectral gap of a Markov chain. However, we will only consider decomposition w.r.t. bipartitions. The motivation was to study if the Martin-Randall approach could somehow be related to the result of chapter 5, which is essentially also a form of decomposition, and to possibly bridge the gap between both approaches.

In the following assume that $M = (\Omega, P)$ is a finite, irreducible, aperiodic and time-reversible Markov chain with stationary distribution $\pi$. Furthermore we will keep the notations used in chapter 5. As regards the definition of spectral gap we will again make the simplifying normalisation $\mathrm{Var}(f) = 1$ and thus

$$\lambda_M = \min_f \{ \mathcal{E}_M(f,f) \}. \tag{6.7}$$

As already seen in chapter 5 the Dirichlet-form of the Markov chain can be split up

$$\mathcal{E}_M(f,f) = \pi(A)\mathcal{E}_{M_A}(f,f) + \pi(\bar{A})\mathcal{E}_{M_{\bar{A}}}(f,f) + \mathrm{ct}_{\mathcal{E}_M(f,f)}, \tag{6.8}$$

where $A$, $\bar{A}$ is a bipartition of $\Omega$ and $\mathrm{ct}_{\mathcal{E}_M(f,f)} = \sum_{\substack{x\in A \\ y\in\bar{A}}} (f(x)-f(y))^2 \pi(x)P(x,y)$. Note that (6.8) immediately gives rise to two lower bounds on $\mathcal{E}_M(f,f)$:

$$\mathcal{E}_M(f,f) \geq \mathrm{ct}_{\mathcal{E}_M(f,f)} \tag{6.9}$$

and

$$\begin{aligned}
\mathcal{E}_M(f,f) &\geq \pi(A)\mathcal{E}_{M_A}(f,f) + \pi(\bar{A})\mathcal{E}_{M_{\bar{A}}}(f,f) && (6.10)\\
&\geq \pi(A)\lambda_{M_A}\mathrm{Var}_{\pi_A}(f) + \pi(\bar{A})\lambda_{M_{\bar{A}}}\mathrm{Var}_{\pi_{\bar{A}}}(f) && (6.11)\\
&\geq \min\left\{\lambda_{M_A}, \lambda_{M_{\bar{A}}}\right\}\left(\pi(A)\mathrm{Var}_{\pi_A}(f) + \pi(\bar{A})\mathrm{Var}_{\pi_{\bar{A}}}(f)\right) && (6.12)\\
&= \min\left\{\lambda_{M_A}, \lambda_{M_{\bar{A}}}\right\}\left(1 - \mathrm{ct}_{\mathrm{Var}_\pi(f)}\right), && (6.13)
\end{aligned}$$

where (6.11) is obtained by the definition of spectral gap and (6.13) is due to the normalisation of variance.

Since (6.8) holds for any bipartition, we have:

**Lemma 6.2.1** *For any bipartition $A$, $\bar{A}$ of the state space $\Omega$*

$$\mathcal{E}_M(f,f) \geq \max\left\{\mathrm{ct}_{\mathcal{E}_M(f,f)}, \min\left\{\lambda_{M_A}, \lambda_{M_{\bar{A}}}\right\}\left(1-\mathrm{ct}_{\mathrm{Var}_\pi(f)}\right)\right\}.$$

Unfortunately, in general Lemma 6.2.1 will not be of practical value for computing the spectral gap. For computing the spectral gap, it is furthermore necessary to minimize over all functions $f$. Given a (fixed) function $f$ and bipartition $A, \bar{A}$ computing $\mathrm{ct}_{\mathcal{E}_M(f,f)}$ and $\mathrm{ct}_{\mathrm{Var}_\pi(f)}$ may be straightforward, however minimizing over all $f$ will in general not be. For the above framework to be applicable, we have to avoid minimization over functions. Before we can give a more practical version of Lemma 6.2.1, we have to introduce a couple of new concepts.

**Definition 6.2.2** *Let $M = (\Omega, P)$ be a Markov chain with stationary distribution $\pi$. Let $A$, $\bar{A}$ be a partition of $\Omega$ and $\widetilde{P}(A,\bar{A})$ be defined as on page 27. Then*

$$\hat{\pi}(x) = \frac{\pi(A)}{\widetilde{P}(A,\bar{A})} \sum_{y\in\bar{A}} \pi(x)P(x,y) \quad \text{for } x \in A$$

*and*

$$\hat{\pi}(y) = \frac{\pi(\bar{A})}{\widetilde{P}(A,\bar{A})} \sum_{x\in A} \pi(y)P(y,x) \quad \text{for } y \in \bar{A}$$

*is a* $\pi$-approximation. *If $\pi(x) \geq \alpha\hat{\pi}(x)$ for $x \in \Omega$ and $\alpha \in (0,1]$, then $\hat{\pi}$ is an* $\alpha$-$\pi$-approximation. *Notice that $\alpha$ is not the* log-*Sobolev constant.*

**Note 6.2.3** *Observe that for any $\pi$-approximation $\hat{\pi}(A) = \pi(A)$. Furthermore define $\hat{\pi}_A(x) = \frac{\hat{\pi}(x)}{\hat{\pi}(A)}$, which happens to equal $\frac{\hat{\pi}(x)}{\pi(A)}$.*

We can use $\pi$-approximations to lower bound the Dirichlet-form crossterms of time-reversible Markov chains $M = (\Omega, P)$:

$$\widetilde{P}(A,\bar{A})\,\mathrm{ct}_{\mathcal{E}_M(f,f)} = \sum_{\substack{x\in A\\ y\in\bar{A}}} (f(x)-f(y))^2\pi(x)P(x,y) \sum_{\substack{x\in A\\ y\in\bar{A}}} \pi(x)P(x,y).$$

Using the Cauchy-Schwarz inequality this is larger than

$$\left( \sum_{\substack{x\in A \\ y\in \bar{A}}} (f(x)-f(y))\pi(x)P(x,y) \right)^2 .$$

With time-reversibility (cf. section 2.1) this can be re-expressed as

$$\begin{aligned} &\left( \sum_{x\in A} f(x) \sum_{y\in\bar{A}} \pi(x)P(x,y) - \sum_{y\in\bar{A}} f(y) \sum_{x\in A} \pi(y)P(y,x) \right)^2 \\ = &\left( \frac{\widetilde{P}(A,\bar{A})}{\hat{\pi}(A)} \sum_{x\in A} f(x)\hat{\pi}(x) - \frac{\widetilde{P}(A,\bar{A})}{\hat{\pi}(\bar{A})} \sum_{y\in\bar{A}} f(y)\hat{\pi}(y) \right)^2 \end{aligned}$$

and thus finally

$$\mathrm{ct}_{\mathcal{E}_M(f,f)} \geq \widetilde{P}(A,\bar{A}) \left( \mathrm{E}_{\hat{\pi}_A}[f] - \mathrm{E}_{\hat{\pi}_{\bar{A}}}[f] \right)^2 . \tag{6.14}$$

We will lower bound $\left( \mathrm{E}_{\hat{\pi}_A}[f] - \mathrm{E}_{\hat{\pi}_{\bar{A}}}[f] \right)^2$ via two lemmata. The first one is

**Lemma 6.2.4** *Let $\pi$ be a probability distribution and $\hat{\pi}$ an $\alpha$-$\pi$-approximation. Then*

$$\{\mathrm{E}_\pi[f] - \mathrm{E}_{\hat{\pi}}[f]\}^2 \leq \frac{8}{\alpha}\mathrm{Var}_\pi(f).$$

*Proof.* Since $\hat{\pi}$ is an $\alpha$-$\pi$-approximation and expectation can alternatively be defined as

$$\mathrm{Var}_\pi(f) = \min_\mu \left\{ \sum_{x\in\Omega} \pi(x)(f(x)-\mu)^2 \right\}, \tag{6.15}$$

the following holds:

$$\begin{aligned} \mathrm{Var}_{\hat{\pi}}(f) &= \sum_{x\in\Omega} \hat{\pi}(x) \left( f(x) - \mathrm{E}_{\hat{\pi}}[f] \right)^2 \\ &\leq \sum_{x\in\Omega} \hat{\pi}(x) \left( f(x) - \mathrm{E}_{\pi}[f] \right)^2 \\ &\leq \frac{1}{\alpha}\mathrm{Var}_\pi(f). \end{aligned}$$

Thus

$$2\,\mathrm{Var}_\pi(f) \geq \sum_{x\in\Omega} \pi(x)(f(x)-\mathrm{E}_\pi[f])^2 + \alpha \sum_{x\in\Omega} \hat{\pi}(x)(f(x)-\mathrm{E}_{\hat{\pi}}[f])^2.$$

As $\alpha\hat{\pi} \leq \pi$ and $\alpha \in (0,1]$ this is larger than

$$\begin{aligned} &\sum_{x\in\Omega} \alpha\hat{\pi}(x)(f(x)-\mathrm{E}_{\pi}[f])^2 + \sum_{x\in\Omega} \alpha\hat{\pi}(x)(f(x)-\mathrm{E}_{\hat{\pi}}[f])^2 \\ &= \alpha \sum_{x\in\Omega} \hat{\pi}(x)\left((f(x)-\mathrm{E}_{\pi}[f])^2 + (f(x)-\mathrm{E}_{\hat{\pi}}[f])^2\right). \end{aligned} \tag{6.16}$$

At this point it is helpful to realize that real functions of the form $(x-a)^2+(x-b)^2$ are minimal at $x=(a+b)/2$. Therefore (6.16) is larger than

$$\alpha \sum_{x\in\Omega} \frac{1}{4}(\mathrm{E}_{\pi}[f]-\mathrm{E}_{\hat{\pi}}[f])^2\hat{\pi}(x).$$

Notice that $(\mathrm{E}_{\pi}[f]-\mathrm{E}_{\hat{\pi}}[f])^2$ is a constant and recall that $\hat{\pi}$ is a probability distribution. Thus

$$\frac{8}{\alpha}\mathrm{Var}_{\pi}(f) \geq (\mathrm{E}_{\pi}[f]-\mathrm{E}_{\hat{\pi}}[f])^2.$$

□

Before we proceed any further it is helpful to point out a useful technical fact: remember that we have shown in the proof of Lemma 5.3.2 that the crossterms of variance can be expressed in two ways

$$\begin{aligned} \mathrm{ct}_{\mathrm{Var}_{\pi}(f)} &= \pi(A)(\mathrm{E}_{\pi_A}[f]-\mathrm{E}_{\pi}[f])^2 + \pi(\bar{A})(\mathrm{E}_{\pi_{\bar{A}}}[f]-\mathrm{E}_{\pi}[f])^2 && (6.17) \\ &= \pi(A)\pi(\bar{A})(\mathrm{E}_{\pi_A}[f]-\mathrm{E}_{\pi_{\bar{A}}}[f])^2. && (6.18) \end{aligned}$$

The next lemma enables us to lower bound $\left(\mathrm{E}_{\hat{\pi}_A}[f]-\mathrm{E}_{\hat{\pi}_{\bar{A}}}[f]\right)^2$.

**Lemma 6.2.5** *For any fixed $f:\Omega\to\mathbb{R}$, any fixed $\alpha\in(0,1]$ and $\alpha$-$\pi$-approximation $\hat{\pi}$, if*

$$\mathrm{ct}_{\mathrm{Var}_{\pi}(f)} \geq \frac{8}{\alpha+8},$$

*then*

$$\left(\mathrm{E}_{\hat{\pi}_A}[f]-\mathrm{E}_{\hat{\pi}_{\bar{A}}}[f]\right)^2 \geq \frac{1}{\pi(A)\pi(\bar{A})}\left(\sqrt{\mathrm{ct}_{\mathrm{Var}_{\pi}(f)}} - 2\sqrt{\frac{2}{\alpha}}\sqrt{1-\mathrm{ct}_{\mathrm{Var}_{\pi}(f)}}\right)^2.$$

*Proof.* Regard $|x-y| = d(x,y)$ as a metric. Then by two applications of the triangle-inequality

$$|a-d| \leq |a-b| + |b-c| + |c-d|.$$

Thus

$$\begin{aligned}&\left|\mathrm{E}_{\hat{\pi}_A}[f]-\mathrm{E}_{\hat{\pi}_{\bar{A}}}[f]\right| \\ &\quad\geq \left|\mathrm{E}_{\pi_A}[f]-\mathrm{E}_{\pi_{\bar{A}}}[f]\right|-\left|\mathrm{E}_{\pi_A}[f]-\mathrm{E}_{\hat{\pi}_A}[f]\right|-\left|\mathrm{E}_{\pi_{\bar{A}}}[f]-\mathrm{E}_{\hat{\pi}_{\bar{A}}}[f]\right|.\end{aligned}$$

Using Lemma 6.2.4 and the definition of $\mathrm{ct}_{\mathrm{Var}_\pi(f)}$, this is larger than

$$\sqrt{\frac{\mathrm{ct}_{\mathrm{Var}_\pi(f)}}{\pi(A)\pi(\bar{A})}}-2\sqrt{\frac{2}{\alpha}}\left(\sqrt{\mathrm{Var}_{\pi_A}(f)}+\sqrt{\mathrm{Var}_{\pi_{\bar{A}}}(f)}\right). \tag{6.19}$$

Let $m=\max\{\sqrt{\mathrm{Var}_{\pi_A}(f)}+\sqrt{\mathrm{Var}_{\pi_{\bar{A}}}(f)}\}$ subject to $\mathrm{ct}_{\mathrm{Var}_\pi(f)}$ being an independent and fixed constant (i.e. its value depends only on the initial choice of $A$, $\bar{A}$ and $f$) and $\pi(A)\mathrm{Var}_{\pi_A}(f)+\pi(\bar{A})\mathrm{Var}_{\pi_{\bar{A}}}(f)+\mathrm{ct}_{\mathrm{Var}_\pi(f)}=1$. Then surely (6.19) is larger than

$$\sqrt{\frac{\mathrm{ct}_{\mathrm{Var}_\pi(f)}}{\pi(A)\pi(\bar{A})}}-2\sqrt{\frac{2}{\alpha}}m.$$

In other words, we seek to maximise $\sqrt{x}+\sqrt{y}$ subject to $px+qy+c=1$ and $p+q=1$. This is equivalent to maximising the function

$$h(x)=\sqrt{x}+\sqrt{\frac{1-px-c}{q}}.$$

Since

$$h'(x)=\frac{1}{2}x^{-\frac{1}{2}}-\frac{p}{2q}\left(\frac{1-px-c}{q}\right)^{-\frac{1}{2}}$$

and

$$h''(x)=-\frac{1}{4}\left(x^{-\frac{3}{2}}+\frac{p^2}{\left(\frac{1-px-c}{1-p}\right)^{\frac{3}{2}}(1-p)^2}\right)<0,$$

the maximum is at $x=\frac{q}{p}(1-c)$ (where the value for $y$ is then $\frac{p}{q}(1-c)$). Thus (6.19) must be larger than

$$\sqrt{\frac{\mathrm{ct}_{\mathrm{Var}_\pi(f)}}{\pi(A)\pi(\bar{A})}}-2\sqrt{\frac{2}{\alpha}}\left(\sqrt{\frac{\pi(\bar{A})}{\pi(A)}\left(1-\mathrm{ct}_{\mathrm{Var}_\pi(f)}\right)}+\sqrt{\frac{\pi(A)}{\pi(\bar{A})}\left(1-\mathrm{ct}_{\mathrm{Var}_\pi(f)}\right)}\right)$$

$$= \sqrt{\frac{\mathrm{ct}_{\mathrm{Var}_\pi(f)}}{\pi(A)\pi(\bar{A})}} - 2\sqrt{\frac{2}{\alpha}\left(1-\mathrm{ct}_{\mathrm{Var}_\pi(f)}\right)}\left(\sqrt{\frac{\pi(A)}{\pi(\bar{A})}}+\sqrt{\frac{\pi(\bar{A})}{\pi(A)}}\right).$$

$$= \frac{\sqrt{\mathrm{ct}_{\mathrm{Var}_\pi(f)}} - 2\sqrt{\frac{2}{\alpha}\left(1-\mathrm{ct}_{\mathrm{Var}_\pi(f)}\right)}}{\sqrt{\pi(A)\,\pi(\bar{A})}}. \tag{6.20}$$

Observe that (6.20) is greater than 0 if $\mathrm{ct}_{\mathrm{Var}_\pi(f)} \geq \frac{8}{\alpha+8}$. To conclude the proof, notice that

$$\left(\mathrm{E}_{\hat{\pi}_A}[f] - \mathrm{E}_{\hat{\pi}_{\bar{A}}}[f]\right)^2 \geq \frac{1}{\pi(A)\pi(\bar{A})}\left(\sqrt{\mathrm{ct}_{\mathrm{Var}_\pi(f)}} - 2\sqrt{\frac{2}{\alpha}}\sqrt{1-\mathrm{ct}_{\mathrm{Var}_\pi(f)}}\right)^2.$$

□

With Lemma 6.2.1, (6.14) and Lemma 6.2.5 the spectral gap can be lower bounded as follows: Let $A, \bar{A}$ be any bipartition of $\Omega$. If $\mathrm{ct}_{\mathrm{Var}_\pi(f)} \geq \frac{8}{\alpha+8}$, then

$$\lambda_M \geq \min_f \left\{ \max\left\{ \frac{\widetilde{P}(A,\bar{A})}{\pi(A)\pi(\bar{A})}\left(\sqrt{\mathrm{ct}_{\mathrm{Var}_\pi(f)}} - 2\sqrt{\frac{2}{\alpha}}\sqrt{1-\mathrm{ct}_{\mathrm{Var}_\pi(f)}}\right)^2, \right.\right. \tag{6.21}$$
$$\left.\left. \min\left\{\lambda_{M_A}, \lambda_{M_{\bar{A}}}\right\}\left(1-\mathrm{ct}_{\mathrm{Var}_\pi(f)}\right)\right\}\right\},$$

otherwise

$$\lambda_M \geq \min_f \left\{ \min\left\{\lambda_{M_A}, \lambda_{M_{\bar{A}}}\right\}\left(1-\mathrm{ct}_{\mathrm{Var}_\pi(f)}\right)\right\}. \tag{6.22}$$

The problem with (6.21) and (6.22) is the dependence on $\mathrm{ct}_{\mathrm{Var}_\pi(f)}$, which as said before implies a minimization over $f$. But this problem can be overcome. First notice that for $\frac{8}{\alpha+8} \leq x \leq 1$ the expression

$$\left(\sqrt{x} - 2\sqrt{\frac{2}{\alpha}}\sqrt{1-x}\right)^2 \tag{6.23}$$

is monotone in $x$. Now assume that $c_\alpha$ is a value in the open interval $\left(\frac{8}{\alpha+8}, 1\right)$, then for any $f$ with $\mathrm{ct}_{\mathrm{Var}_\pi(f)} \geq c_\alpha$ by (6.21) and monotonicity of (6.23)

$$\mathcal{E}_M(f,f) \geq \frac{\widetilde{P}(A,\bar{A})}{\pi(A)\pi(\bar{A})}\left(\sqrt{c_\alpha} - 2\sqrt{\frac{2}{\alpha}}\sqrt{1-c_\alpha}\right)^2. \tag{6.24}$$

On the other hand, if $\mathrm{ct}_{\mathrm{Var}_\pi(f)} < c_\alpha$, then using the other bound of (6.21) (or (6.22) if $\mathrm{ct}_{\mathrm{Var}_\pi(f)} < \frac{8}{\alpha+8}$), we have

$$\mathcal{E}_M(f,f) \geq \min\left\{\lambda_{M_A}, \lambda_{M_{\bar{A}}}\right\}(1-c_\alpha). \tag{6.25}$$

With (6.24) and (6.25) we can deploy the following strategy for bounding spectral gap:

$$\begin{aligned}\lambda_M \;\geq\; \min\Bigg\{ &\min_f\left\{\mathcal{E}_M(f,f) : \mathrm{ct}_{\mathrm{Var}_\pi(f)} \geq c_\alpha\right\}, \\ &\min_f\left\{\mathcal{E}_M(f,f) : \mathrm{ct}_{\mathrm{Var}_\pi(f)} < c_\alpha\right\}\Bigg\}.\end{aligned} \tag{6.26}$$

Observe that both bounds (6.24) and (6.25) only depend on $c_\alpha$, which is independent of $f$, so that a minimization over $f$ is no longer necessary. Thus (6.24) to (6.26) imply:

**Theorem 6.2.6** *If $A, \bar{A}$ is a partition of the state space $\Omega$ of a finite, irreducible, aperiodic and time-reversible Markov chain $M$ with stationary distribution $\pi$ such that there exists an $\alpha$-$\pi$-approximation, then for any constant $c_\alpha \in \left(\frac{8}{\alpha+8}, 1\right)$ the spectral gap of $M$ can be bounded from below as*

$$\lambda_M \geq \min\left\{\frac{\widetilde{P}(A,\bar{A})}{\pi(A)\pi(\bar{A})}\left(\sqrt{c_\alpha} - 2\sqrt{\frac{2}{\alpha}}\sqrt{1-c_\alpha}\right)^2, \min\left\{\lambda_{M_A}, \lambda_{M_{\bar{A}}}\right\}(1-c_\alpha)\right\}. \tag{6.27}$$

*Remark*: In general we cannot expect Theorem 6.2.6 to yield tight bounds on the spectral gap. To see this recall equations (6.7) and (6.8): for any bipartition into sets $A$, $\bar{A}$ the spectral gap is

$$\lambda_M = \min_{f:\Omega\to\mathbb{R}}\left\{\pi(A)\mathcal{E}_{M_A}(f,f) + \pi(\bar{A})\mathcal{E}_{M_{\bar{A}}}(f,f) + \mathrm{ct}_{\mathcal{E}_M(f,f)} \mid \mathrm{Var}_\pi(f) = 1\right\}.$$

In most cases we can therefore expect that the optimal bound on the spectral gap $\lambda_M$ of the original chain is a combination of the spectral gaps of the restriction chains and a quantity reflecting the *probabilistic flow* between the two sets of the bipartition. Compare this to Theorem 6.2.6: roughly speaking, Theorem 6.2.6 bounds the spectral gap by the minimum of either $\frac{\widetilde{P}(A,\bar{A})}{\pi(A)\pi(\bar{A})}\left(\sqrt{c_\alpha} - 2\sqrt{\frac{2}{\alpha}}\sqrt{1-c_\alpha}\right)^2$, which in essence depends on the probabilistic flow between the sets of the bipartition, or the minimum

of the spectral gaps of the restriction chains. Consequently, Theorem 6.2.6 will always yield a sub-optimal result. Note that the best case for Theorem 6.2.6 is when all three quantities $\lambda_{M_A}$, $\lambda_{M_{\bar{A}}}$ and the probabilistic flow are of similar order. The worst case is when one of of the three quantities contributes dominantly to $\lambda_M$ because then we bound $\lambda_M$ by the smallest of the three.

However, note that the quantities used in Theorem 6.2.6 are all easy to compute. Thus the emphasis of this new decomposition method is less on obtaining good bounds but on easiness to apply.

## 6.3 Spectral gap of Ising model on $d$-ary trees

In this section we will use Theorem 6.2.6 to bound the spectral gap of a random process for the Ising model on a finite, complete $d$-ary tree, i.e. a rooted tree where every node other than a leaf has exactly $d$ children and where every leaf is at the same distance from the root. This Markov chain has been studied before by Kenyon, Mossel and Peres [70], who showed that it is rapidly mixing; superficially, our analysis looks similar to theirs.

Assume that the nodes and leaves of the tree are denoted by $S$ and constitute the sites of an Ising model. The Ising model originated in physics with the purpose to model ferromagnetism. In its most general form the Ising model is defined as follows: Suppose we are given a finite graph $G = (V, E)$ with vertices $\{v_1, \ldots, v_n\}$ and edges $E$. The vertices are usually referred to as *sites* and we assign each site a *spin*, $-1$ or $1$. The upshot of such an assignment is a *configuration*, a vector $x = (x_{v_1}, \ldots, x_{v_n}) \in \Omega = \{-1, 1\}^{|V|}$. Two configurations $x$ and $y$ are *adjacent* if they differ at exactly one site, i.e. $x_{v_j} = 1$ and $y_{v_j} = -1$ (or vice versa) for one site $j \in \{1, \ldots, n\}$ while $x_{v_i} = y_{v_i}$ for the remaining sites $i \neq j$, $i = 1, \ldots, n$. Each configuration is given a probability

$$\pi(x) = \frac{1}{Z} \exp\left\{ \beta \sum_{\substack{v \neq w \\ v \sim w}} x_v x_w \right\},$$

where the sum is over all pairs $v$, $w$ of distinct neighbours in $G$, the relation $\sim$ denotes adjacency, $\beta$ is a real constant (usually called *inverse temperature*) reflecting the surrounding temperature and $Z$ is the appropriate normalizing function, generally known

as the *partition function*

$$Z = \sum_{x \in \Omega} \exp\left\{ \beta \sum_{\substack{v \neq w \\ v \sim w}} x_v x_w \right\}.$$

Note that in general the so-defined probability distribution on the set of configurations is not uniform. In most applications of the Ising model the goal is to sample configurations from a given probability distribution.

As said before, we will consider the Ising model on a finite, complete $d$-ary tree with sites $S$. Let $r_S$ be the root of the tree of height $h$. In the spirit of MCMC-algorithms we will use a Markov chain $M_{\Omega_S} = \left(\Omega_S, P_{M_{\Omega_S}}\right)$ with stationary distribution $\pi_{M_{\Omega_S}}$ for sampling. The state space $\Omega_S$ is defined as

$$\Omega_S = \{+1, -1\}^{|S|}.$$

Under the simplifying condition that $\beta = 1$, for $x \in \Omega_S$ its probability in equilibrium is defined as

$$\pi_{M_{\Omega_S}}(x) = \frac{1}{Z_{M_{\Omega_S}}} \prod_{\substack{v,w \in S \\ v \sim w \\ v \neq w}} e^{x_v \cdot x_w},$$

where $Z_{M_{\Omega_S}}$ is the partition function on $\Omega_S$. We devise the chain $M_{\Omega_S}$ to be a single-site-update chain, i.e. to make transitions according to the following rule: choose one site uniformly at random, then change the spin at that site[1]. More formally, the transition probabilities for adjacent $x, y \in \Omega_S$ are given by

$$P_{M_{\Omega_S}}(x, y) = \frac{1}{|S|} H_{\Omega_S}(x, y),$$

where $H_{\Omega_S}$ is a heat-bath kernel such that for $v \in S$

$$H^v_{\Omega_S}(x, y) = \frac{\pi_{M_{\Omega_S}}(y)}{\pi_{M_{\Omega_S}}(x_{v=+1}) + \pi_{M_{\Omega_S}}(x_{v=-1})} = \begin{cases} \dfrac{e^{y_v \sum_{i=1}^{d} x_{\gamma_i(v)}}}{e^{\sum_{i=1}^{d} x_{\gamma_i(v)}} + e^{-\sum_{i=1}^{d} x_{\gamma_i(v)}}}, & \text{if } v \text{ is root} \\ \dfrac{e^{y_v \sum_{i=1}^{d+1} x_{\gamma_i(v)}}}{e^{\sum_{i=1}^{d+1} x_{\gamma_i(v)}} + e^{-\sum_{i=1}^{d+1} x_{\gamma_i(v)}}}, & \text{if } v \text{ is node} \\ \dfrac{e^{y_v \cdot x_{\gamma(v)}}}{e^{x_{\gamma(v)}} + e^{-x_{\gamma(v)}}}, & \text{if } v \text{ is leaf.} \end{cases}$$

Here, $x_{v=\sigma}$ for $\sigma \in \{+1, -1\}$ expresses a state that is like state $x$ only has spin $\sigma$ at site $v$ and $x_{\gamma_i(v)}$ denotes the spin of the $i$-th neighbour of site $v$ in state $x$ (if $v$ has only

[1] Such local Markov chains on spin systems are also called *Glauber dynamics*.

one neighbour, then we drop the index $i$). It is clear that with these definitions $M_{\Omega_S}$ is finite, irreducible, aperiodic and time-reversible.

Our strategy to bound the spectral gap of $M_{\Omega_S}$ is best described by figure 6.1.

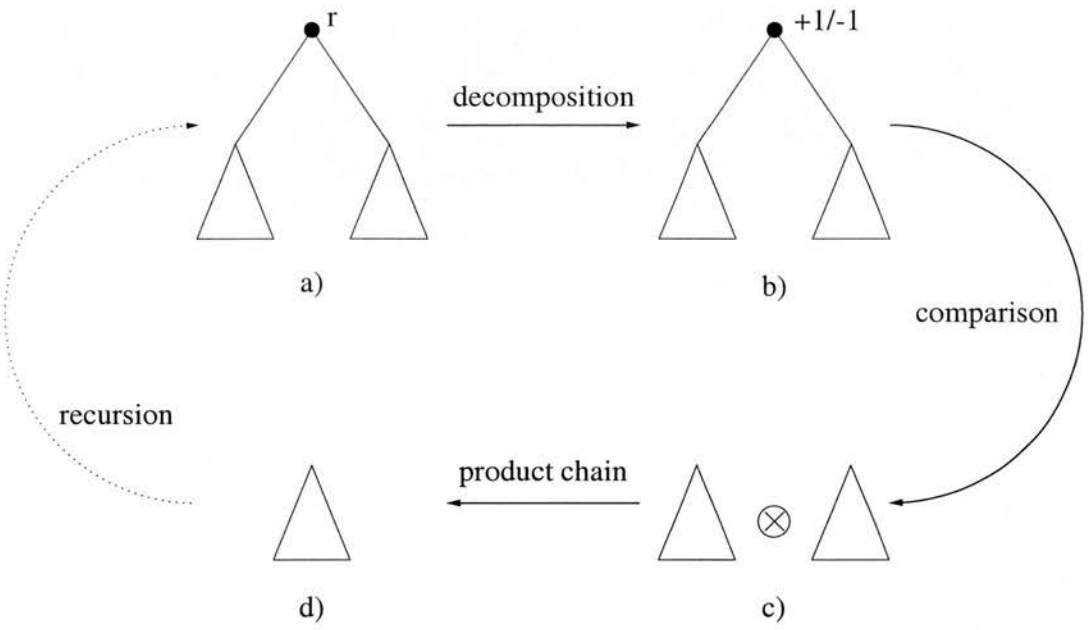



Figure 6.1: strategy for bounding spectral gap of $M_{\Omega_S}$ (depicted for $d = 2$)

Suppose that a) depicts the Ising model $M_{\Omega_S}$ on a tree of height $h$ with root $r$. To invoke the decomposition theorem (Theorem 6.2.6), we partition the state space of $M_{\Omega_S}$ into two sets of equal size: the set $\Omega_{S+1}$ containing all states where the spin at the root $r$ is $+1$ and the other, $\Omega_{S-1}$, states where the spin of $r$ is $-1$. This gives rise to two restriction chains $M_{\Omega_{S+1}}$ and $M_{\Omega_{S-1}}$ (depicted by b)). We then compare $M_{\Omega_{S+1}}$ and $M_{\Omega_{S-1}}$ to an Ising model on two trees of height $h-1$ as depicted by c). It turns out that this Ising model is the product of two Ising models on trees of height $h-1$, which takes us to d). Since d) is an Ising model on a tree of height $h-1$, we can re-use the same chain of steps to recursively compute a lower bound for $M_{\Omega_S}$.

We will now implement the above strategy in detail. To use Theorem 6.2.6 we first have to specify a bipartition. Split $\Omega_S$ into

$$\Omega_{S+1} = \{x \in \Omega_s \mid x_{r_S} = +1\} \quad \text{and} \quad \Omega_{S-1} = \{x \in \Omega_s \mid x_{r_S} = -1\}.$$

Next, we have to find a suitable constant $c_\alpha$.

### 6.3.1 Finding $c_\alpha$

We set off by checking if there exists an $\alpha$-$\pi$-approximation. To determine $\pi_{M_{\Omega_S}}(\Omega_{S+1})$ and $\pi_{M_{\Omega_S}}(\Omega_{S-1})$ observe the following. If $x \in \Omega_S$, then let $-x$ be the state obtained by inverting all spins of $x$. Clearly, if $x \in \Omega_{S+1}$, then $-x \in \Omega_{S-1}$ and vice versa. Furthermore, it is easily verified that $\pi_{M_{\Omega_S}}(x) = \pi_{M_{\Omega_S}}(-x)$. Thus, $\pi_{M_{\Omega_S}}(\Omega_{S+1}) = \pi_{M_{\Omega_S}}(\Omega_{S-1}) = \frac{1}{2}$. As regards $\widetilde{P}_{M_{\Omega_S}}(\Omega_{S+1},\Omega_{S-1})$, notice that structurally $M_{\Omega_S}$ resembles the random walk on an $|S|$-dimensional hypercube. This means every element of $\Omega_{S+1}$ has exactly one neighbour in $\Omega_{S-1}$:

$$\widetilde{P}_{M_{\Omega_S}}(\Omega_{S+1},\Omega_{S-1}) = \sum_{\substack{x\in\Omega_{S+1}\\ y\in\Omega_{S-1}}} \pi_{M_{\Omega_S}}(x)P_{M_{\Omega_S}}(x,y) = \sum_{x\in\Omega_{S+1}} \pi_{M_{\Omega_S}}(x)P_{M_{\Omega_S}}(x,x'),$$

where $x'$ is the sole neighbour of $x$ in $\Omega_{S-1}$. Now observe that for $x \in \Omega_{S+1}$ the $\pi$-approximation can be bounded as follows:

$$\begin{aligned}
\hat{\pi}_{M_{\Omega_S}}(x) &= \frac{1}{2}\frac{P_{M_{\Omega_S}}(x,x')}{\sum_{u\in\Omega_{S+1}}\pi_{M_{\Omega_S}}(u)P_{M_{\Omega_S}}(u,u')}\pi_{M_{\Omega_S}}(x)\\
&= \frac{1}{2}\frac{H_{\Omega_S}(x,x')}{\sum_{u\in\Omega_{S+1}}\pi_{M_{\Omega_S}}(u)H_{\Omega_S}(u,u')}\pi_{M_{\Omega_S}}(x)\\
&\leq \frac{1}{2}\frac{\frac{e^{d+1}}{e^{-(d+1)}+e^{d+1}}}{\sum_{u\in\Omega_{S+1}}\pi_{M_{\Omega_S}}(u)\frac{e^{-(d+1)}}{e^{d+1}+e^{-(d+1)}}}\pi_{M_{\Omega_S}}(x)\\
&= e^{2(d+1)}\pi_{M_{\Omega_S}}(x).
\end{aligned}$$

The very same calculation holds for $y \in \Omega_{S-1}$ and hence $\alpha = e^{-2(d+1)}$. The crucial constant $c_\alpha$ lies in $\left(\frac{8}{e^{-2(d+1)}+8}, 1\right)$.

### 6.3.2 Bounding $\frac{\widetilde{P}_{M_{\Omega_S}}(\Omega_{S+1},\Omega_{S-1})}{\pi_{M_{\Omega_S}}(\Omega_{S+1})\pi_{M_{\Omega_S}}(\Omega_{S-1})}\left(\sqrt{c_\alpha}-2\sqrt{\frac{2}{\alpha}}\sqrt{1-c_\alpha}\right)^2$

Once a value for $c_\alpha$ has been chosen, we have to bound $\widetilde{P}_{M_{\Omega_S}}(\Omega_{S+1},\Omega_{S-1})$. As observed above,

$$\widetilde{P}_{M_{\Omega_S}}(\Omega_{S+1},\Omega_{S-1}) = \sum_{x\in\Omega_{S+1}} \pi_{M_{\Omega_S}}(x)P_{M_{\Omega_S}}(x,x'),$$

where $x'$ is the sole neighbour of $x$ in $\Omega_{S-1}$. By definition of $P_{M_{\Omega_S}}(x,x')$ this equals

$$\sum_{x\in\Omega_{S+1}} \pi_{M_{\Omega_S}}(x)\,\frac{1}{|S|}H_{\Omega_S}(x,x')$$

which is greater than

$$\frac{1}{|S|}\,\frac{e^{-(d+1)}}{e^{d+1}+e^{-(d+1)}}\sum_{x\in\Omega_{S+1}} \pi_{M_{\Omega_S}}(x).$$

Since $\sum_{x\in\Omega_{S+1}} \pi_{M_{\Omega_S}}(x) = 1/2$, we have

$$\frac{\widetilde{P}_{M_{\Omega_S}}(\Omega_{S+1},\Omega_{S-1})}{\pi_{M_{\Omega_S}}(\Omega_{S+1})\,\pi_{M_{\Omega_S}}(\Omega_{S-1})} \geq \frac{2}{|S|}\,\frac{e^{-(d+1)}}{e^{d+1}+e^{-(d+1)}}.$$

### 6.3.3 Bounding $\min\left\{\lambda_{M_{\Omega_{S+1}}},\lambda_{M_{\Omega_{S-1}}}\right\}(1-c_\alpha)$

Now let us turn to bounding $\min\left\{\lambda_{M_{\Omega_{S+1}}},\lambda_{M_{\Omega_{S-1}}}\right\}(1-c_\alpha)$. As $c_\alpha$ is already known, it remains to compute the spectral gaps of $M_{\Omega_{S+1}}$ and $M_{\Omega_{S-1}}$. We will do this using the comparison and product chain techniques of Diaconis and Saloff-Coste [30]. Their comparison lemma is as follows:

**Lemma 6.3.1** [[30], Lemma 3.3] *Let $M=(\Omega,P)$ and $M'=(\Omega,P')$ be two Markov chains on the same finite state space with stationary distributions $\pi$ and $\pi'$ respectively. Assume that there exist $A$, $a>0$ such that*

$$\mathcal{E}_{M'}(f,f) \leq A\,\mathcal{E}_M(f,f), \qquad a\pi \leq \pi'.$$

*Then*

$$\lambda_{M'} \leq \frac{A}{a}\lambda_M.$$

Before we can formulate the other result, which is their so-called product chain lemma, we have to define product chains.

**Definition 6.3.2** *For $i=1,\ldots,n$, let $M_i=(\Omega_i,P_i)$ be a finite Markov chain with stationary distribution $\pi_i$. The* product chain *$M=(\Omega,P)$ of the $M_i$ has state space $\Omega=$*

$\prod_{i=1}^{n} \Omega_i$ *and transition kernel*

$$P(x,y) = \frac{1}{n}\sum_{i=1}^{n} \delta(x_1,y_1)\cdots\delta(x_{i-1},y_{i-1}) \times P(x_i,y_i) \times \delta(x_{i+1},y_{i+1})\cdots\delta(x_d,y_d), \tag{6.28}$$

*where* $x = (x_i)_1^n$, $y = (y_i)_1^n$ *and* $\delta$ *is Kronecker delta, i.e.* $\delta(x,y) = 1$ *if* $x = y$ *and* 0 *otherwise. Observe that M has stationary distribution*

$$\pi(x) = \prod_{i=1}^{n} \pi_i(x_i). \tag{6.29}$$

Informally, a product chain is a Markov chain that makes transitions by first choosing a coordinate and then making a step in that coordinate.

**Lemma 6.3.3** [[30], Lemma 3.2] *Let* $M_i = (\Omega_i, P_i)$, $i = 1,\ldots,n$, *be finite Markov chains with spectral gaps* $\lambda_{M_i}$. *Then the product chain* $M = (\Omega, P)$ *as defined in Definition 6.3.2 satisfies*

$$\lambda_M = \frac{1}{n}\min_i \{\lambda_{M_i}\}.$$

First, we use Lemma 6.3.1 to lower bound the spectral gaps of the restriction chains $M_{\Omega_{S+1}}$ and $M_{\Omega_{S-1}}$ by comparing them to the spectral gap of the yet to be defined chain $M_{\Omega_{S\setminus\{r_S\}}}$ (this corresponds to the step from b) to c) in figure 6.1). We will then show that $M_{\Omega_{S\setminus\{r_S\}}}$ is a product chain in the sense of Lemma 6.3.3, in fact we will show that it consists of $d$ copies of the same chain (this is step c) to d) in figure 6.1). The spectral gap of each one of the $d$ copies can be bounded using a recursive argument (step d) to a) in figure 6.1). Perhaps, we should pause here to comment on the notation used: $M_{\Omega_{S+1}}$ is the chain on $\Omega_{S+1}$, the configurations where the spin on the root of $S$ is $+1$ (and similarly for $M_{\Omega_{S-1}}$); the chain $M_{\Omega_{S\setminus\{r_S\}}}$ will be, as the notation suggests, defined on configurations where the site $r_S$, the root of $S$, is no longer available.

First, take a closer look at the restriction chains $M_{\Omega_{S+1}} = \left(\Omega_{S+1}, P_{M_{\Omega_{S+1}}}\right)$ and $M_{\Omega_{S-1}} = \left(\Omega_{S-1}, P_{M_{\Omega_{S-1}}}\right)$. The stationary distributions are given by

$$\pi_{M_{\Omega_{S+1}}}(x) = \frac{\pi_{M_{\Omega_S}}(x)}{\pi_{M_{\Omega_S}}(\Omega_{S+1})} = \frac{e^{\sum_{i=1}^{d} x_{\gamma_i(r_S)}} \prod_{\substack{v\sim w,\, v\neq w \\ v,w\in S\setminus\{r_S\}}} e^{x_v x_w}}{\sum_{x\in\Omega_{S+1}} e^{\sum_{i=1}^{d} x_{\gamma_i(r_S)}} \prod_{\substack{v\sim w,\, v\neq w \\ v,w\in S\setminus\{r_S\}}} e^{x_v x_w}}, \text{ for } x \in \Omega_{S+1}$$

and

$$\pi_{M_{\Omega_{S-1}}}(x) = \frac{\pi_{M_{\Omega_S}}(x)}{\pi_{M_{\Omega_S}}(\Omega_{S-1})} = \frac{\prod_{\substack{v\sim w,\, v\neq w \\ v,w\in S\setminus\{r_S\}}} e^{x_v x_w} e^{-\sum_{i=1}^d x_{\gamma_i(r_S)}}}{\sum_{x\in\Omega_{S-1}} \prod_{\substack{v\sim w,\, v\neq w \\ v,w\in S\setminus\{r_S\}}} e^{x_v x_w} e^{-\sum_{i=1}^d x_{\gamma_i(r_S)}}}, \text{ for } x\in\Omega_{S-1},$$

where $x_{\gamma_i(r_S)}$ denotes the spin of the $i$-th neighbour of $r_S$. Before we define the transition probabilities $P_{M_{\Omega_{S+1}}}$ and $P_{M_{\Omega_{S-1}}}$, let us examine the possible scenarios. Assume that the current configuration is $x\in\Omega_{S+1}$. Since the Markov chain is a single-site-update chain, a transition to a successor configuration $y$ (which is either an adjacent configuration or, in case of a self-loop, the same configuration) is made as follows:

1. choose a site $v\in S$ u.a.r.
2. update the spin at $v$ (if this is possible at all) according to heat-bath.

There are three possible outcomes to the first step: site $v$ is the root $r_S$, site $v$ is one of the $d$ neighbours of the root or $v$ is none of the aforementioned. If $v$ is the root, then we do not change the configuration because changing the spin at the root would take us out of the state space $\Omega_{S+1}$. If $v$ is a neighbour of the root $r_S$, then the heat-bath transition from $x$ to $y$ is

$$\frac{e^{y_v\left(1+\sum_{i=1}^d x_{\gamma_i(v)}\right)}}{e^{\left(1+\sum_{i=1}^d x_{\gamma_i(v)}\right)}+e^{-\left(1+\sum_{i=1}^d x_{\gamma_i(v)}\right)}}$$

because the spin at the root is fixed to $+1$. If the first two cases do not apply, then the heat-bath kernel on $\Omega_{S+1}$ is the same as the heat-bath kernel on $\Omega_S$. Summing up we have: If $x$ is the current configuration with successor configuration $y$, then the transition probabilities of $M_{\Omega_{S+1}}$ are

$$P_{M_{\Omega_{S+1}}}(x,y) = \begin{cases} \frac{1}{|S|}, & \text{if } v = r_S \\ \frac{1}{|S|}\frac{e^{y_v\left(1+\sum_{i=1}^d x_{\gamma_i(v)}\right)}}{e^{\left(1+\sum_{i=1}^d x_{\gamma_i(v)}\right)}+e^{-\left(1+\sum_{i=1}^d x_{\gamma_i(v)}\right)}}, & \text{if } v\in\{\gamma_1(r_S),\ldots,\gamma_d(r_S)\} \\ \frac{1}{|S|}H^v_{\Omega_S}(x,y), & \text{else} \end{cases}$$

and

$$P_{M_{\Omega_{S-1}}}(x,y) = \begin{cases} \frac{1}{|S|}, & \text{if } v = r_S \\ \frac{1}{|S|}\frac{e^{y_v\left(-1+\sum_{i=1}^d x_{\gamma_i(v)}\right)}}{e^{\left(-1+\sum_{i=1}^d x_{\gamma_i(v)}\right)}+e^{-\left(-1+\sum_{i=1}^d x_{\gamma_i(v)}\right)}}, & \text{if } v\in\{\gamma_1(r_S),\ldots,\gamma_d(r_S)\} \\ \frac{1}{|S|}H^v_{\Omega_S}(x,y), & \text{else} \end{cases}$$

for $M_{\Omega_{S-1}}$.

The Markov chain $M_{\Omega_{S\setminus\{r_S\}}} = \left(\Omega_{S\setminus\{r_S\}}, P_{M_{\Omega_{S\setminus\{r_S\}}}}\right)$ is defined as

$$\Omega_{S\setminus\{r_S\}} = \{+1,-1\}^{|S\setminus\{r_S\}|} = \{+1,-1\}^{|S|-1},$$

$$P_{M_{\Omega_{S\setminus\{r_S\}}}}(x,y) = \frac{1}{|S\setminus\{r_S\}|} H_{\Omega_S}(x,y).$$

Note that $\Omega_{S\setminus\{r_S\}}$ is isomorphic to each of $\Omega_{S+1}$ and $\Omega_{S-1}$. The stationary distribution of the chain for $x \in \Omega_{S\setminus\{r_S\}}$ is

$$\pi_{M_{\Omega_{S\setminus\{r_S\}}}}(x) = \frac{\prod_{\substack{v\sim w, v\neq w \\ v,w\in S\setminus\{r_S\}}} e^{x_v x_w}}{\sum_{x\in\Omega_{S\setminus\{r_S\}}} \prod_{\substack{v\sim w, v\neq w \\ v,w\in S\setminus\{r_S\}}} e^{x_v x_w}}.$$

Now note that for $x_{\Omega_{S+1}} \in \Omega_{S+1}$, $x_{\Omega_{S-1}} \in \Omega_{S-1}$ and $x_{\Omega_{S\setminus\{r_S\}}} \in \Omega_{S\setminus\{r_S\}}$

$$\pi_{M_{\Omega_{S+1}}}\left(x_{\Omega_{S+1}}\right) \leq e^{2d}\,\pi_{M_{\Omega_{S\setminus\{r_S\}}}}\left(x_{\Omega_{S\setminus\{r_S\}}}\right) \tag{6.30}$$

$$\pi_{M_{\Omega_{S-1}}}\left(x_{\Omega_{S-1}}\right) \leq e^{2d}\,\pi_{M_{\Omega_{S\setminus\{r_S\}}}}\left(x_{\Omega_{S\setminus\{r_S\}}}\right) \tag{6.31}$$

Furthermore, we can show

**Proposition 6.3.4**

$$\mathcal{E}_{M_{\Omega_{S\setminus\{r_S\}}}}(f,f) \leq \frac{|S|}{|S\setminus\{r_S\}|} e^{2d+2}\,\mathcal{E}_{M_{\Omega_{S+1}}}(f,f)$$

*and*

$$\mathcal{E}_{M_{\Omega_{S\setminus\{r_S\}}}}(f,f) \leq \frac{|S|}{|S\setminus\{r_S\}|} e^{2d+2}\,\mathcal{E}_{M_{\Omega_{S-1}}}(f,f).$$

*Proof.* We only give a proof for the first inequality as the proof for the second is almost the same. Since $\Omega_{S\setminus\{r_S\}}$ is isomorphic to $\Omega_{S+1}$, it suffices to show that for all $x \neq y \in \Omega_{S\setminus\{r_S\}}$ and $x' \neq y' \in \Omega_{S+1}$, where $x', y'$ are the images of $x, y$ under the "natural" isomorphism from $\Omega_{S\setminus\{r_S\}}$ to $\Omega_{S+1}$,

$$\pi_{M_{\Omega_{S\setminus\{r_S\}}}}(x)\,P_{M_{\Omega_{S\setminus\{r_S\}}}}(x,y) \leq A\,\pi_{M_{\Omega_{S+1}}}\left(x'\right)P_{M_{\Omega_{S+1}}}\left(x',y'\right).$$

Notice that if $v \notin \{r_S, \gamma_1(r_S), \ldots, \gamma_d(r_S)\}$, then

$$P_{M_{\Omega_{S+1}}}\left(x',y'\right) = \frac{|S\setminus\{r_S\}|}{|S|} P_{M_{\Omega_{S\setminus\{r_S\}}}}(x,y).$$

For $v \in \{\gamma_1(r_S), \ldots, \gamma_d(r_S)\}$

$$\begin{aligned} P_{M_{\Omega_{S+1}}}(x', y') &= \frac{1}{|S|} \frac{e^{y'_v} \cdot e^{y'_v\left(\sum_{i=1}^d x'_{\gamma_i(v)}\right)}}{e \cdot e^{\left(\sum_{i=1}^d x'_{\gamma_i(v)}\right)} + e^{-1} \cdot e^{-\left(\sum_{i=1}^d x'_{\gamma_i(v)}\right)}} \\ &\geq \frac{1}{|S|} \frac{e^{-1}}{e} \frac{e^{y_v\left(\sum_{i=1}^d x_{\gamma_i(v)}\right)}}{e^{\left(\sum_{i=1}^d x_{\gamma_i(v)}\right)} + e^{-\left(\sum_{i=1}^d x_{\gamma_i(v)}\right)}} \\ &= \frac{|S \setminus \{r_S\}|}{|S|} e^{-2} P_{M_{\Omega_{S\setminus\{r_S\}}}}(x, y). \end{aligned}$$

Observe that

$$\begin{aligned} \pi_{M_{\Omega_{S+1}}}(x') &= \frac{e^{\sum_{i=1}^d x'_{\gamma_i(r_S)}} \prod_{\substack{v\sim w, v\neq w \\ v,w\in S\setminus\{r_S\}}} e^{x'_v x'_w}}{\sum_{x'\in\Omega_{S+1}} e^{\sum_{i=1}^d x'_{\gamma_i(r_S)}} \prod_{\substack{v\sim w, v\neq w \\ v,w\in S\setminus\{r_S\}}} e^{x'_v x'_w}} \\ &\geq e^{-2d} \pi_{M_{\Omega_{S\setminus\{r_S\}}}}(x). \end{aligned}$$

Thus,

$$\mathcal{E}_{M_{\Omega_{S\setminus\{r_S\}}}}(f, f) \leq \frac{|S|}{|S \setminus \{r_S\}|} e^{2d+2} \mathcal{E}_{M_{\Omega_{S+1}}}(f, f).$$

□

Thus, combining (6.30), (6.31), Proposition 6.3.4 and the comparison lemma (Lemma 6.3.1) we have

$$\lambda_{M_{\Omega_{S\setminus\{r_S\}}}} \leq \frac{|S|}{|S \setminus \{r_S\}|} e^{4d+2} \min\left\{\lambda_{M_{\Omega_{S+1}}}, \lambda_{M_{\Omega_{S-1}}}\right\}.$$

Next, we will show that $M_{\Omega_{S\setminus\{r_S\}}}$ is a product chain in the sense of Definition 6.3.2. The tree $S$ with root $r_S$ has $d$ immediate subtrees with roots $r_{S_{c_1}}, \ldots, r_{S_{c_d}}$, i.e. the $d$ trees "hanging" from $r_S$. Defining Ising models on each of those subtrees gives rise to $d$ identical Markov chains $M_{\Omega_{S_{c_1}}}, \ldots, M_{\Omega_{S_{c_d}}}$. As the chains are identical, we can drop the subscripts and focus on one chain, namely $M_{\Omega_{S_c}} = \left(\Omega_{S_c}, P_{M_{\Omega_{S_c}}}\right)$. Note that $M_{\Omega_{S_c}}$ is a smaller copy of $M_{\Omega_S}$, i.e. an Ising model on a complete $d$-ary tree of height $h-1$. Thus, the chain $M_{\Omega_{S_c}} = \left(\Omega_{S_c}, P_{M_{\Omega_{S_c}}}\right)$ has state space

$$\Omega_{S_c} = \{+1, -1\}^{|S_c|},$$

where $S_c$ denotes the nodes and leaves of the subtree and transition probabilities

$$P_{M_{\Omega_{S_c}}}(x, y) = \frac{1}{|S_c|} H_{\Omega_S}(x, y).$$

It is easily verifiable that the chain is time-reversible with stationary distribution

$$\pi_{M_{\Omega_{S_c}}}(x) = \frac{1}{Z_{M_{\Omega_{S_c}}}} \prod_{\substack{v\sim w \\ v\neq w \\ v,w\in S_c}} e^{x_v x_w},$$

where

$$Z_{M_{\Omega_{S_c}}} = \sum_{x\in S_c} \prod_{\substack{v\sim w \\ v\neq w \\ v,w\in S_c}} e^{x_v x_w}.$$

Now, observe that $\Omega_{S\setminus\{r_S\}} = \prod_{i=1}^{d} \Omega_{S_{c_i}}$, the Cartesian product of the $\Omega_{S_{c_i}}$, and

$$\begin{aligned} P_{M_{\Omega_{S\setminus\{r_S\}}}}(x,y) &= \frac{1}{d}\sum_{i=1}^{d} \delta(x_1,y_1)\cdots\delta(x_{i-1},y_{i-1}) \\ &\quad \times P_{\Omega_{S_{c_i}}}(x_i,y_i) \times \delta(x_{i+1},y_{i+1})\cdots\delta(x_d,y_d), \end{aligned}$$

where $x = (x_i)_1^d$, $y = (y_i)_1^d$ and $\delta$ is Kronecker delta. Furthermore,

$$\begin{aligned} Z_{M_{\Omega_{S\setminus\{r_S\}}}} &= \sum_{x\in\Omega_{S\setminus\{r_S\}}} \prod_{\substack{v\sim w \\ v\neq w \\ v,w\in S\setminus\{r_S\}}} e^{x_v x_w} \\ &= \sum_{(x_1,\ldots,x_d)\in\prod_{i=1}^{d}\Omega_{S_{c_i}}} \prod_{\substack{v\sim w \\ v\neq w \\ v,w\in S_{c_1}}} e^{x_{1v}x_{1w}} \times\cdots\times \prod_{\substack{v\sim w \\ v\neq w \\ v,w\in S_{c_d}}} e^{x_{dv}x_{dw}} \\ &= \sum_{x_1\in\Omega_{S_{c_1}}} \prod_{\substack{v\sim w \\ v\neq w \\ v,w\in S_{c_1}}} e^{x_{1v}x_{1w}} \times\cdots\times \sum_{x_1\in\Omega_{S_{c_d}}} \prod_{\substack{v\sim w \\ v\neq w \\ v,w\in S_{c_d}}} e^{x_{dv}x_{dw}} \\ &= \prod_{i=1}^{d} Z_{M_{\Omega_{S_{c_i}}}}. \end{aligned}$$

Therefore

$$\pi_{M_{\Omega_{S\setminus\{r_S\}}}}(x) = \prod_{i=1}^{d} \pi_{M_{\Omega_{S_{c_i}}}}(x_i)$$

and $M_{\Omega_{S\setminus\{r_S\}}} = \prod_{i=1}^{d} M_{\Omega_{S_{c_i}}}$ as required for Lemma 6.3.3 and hence

$$\lambda_{M_{\Omega_{S\setminus\{r_S\}}}} = \frac{1}{d}\lambda_{M_{\Omega_{S_c}}}.$$

**Fact 6.3.5** *Summing up we have*

$$\begin{aligned} &\min\left\{\lambda_{M_{\Omega_{S+1}}}, \lambda_{M_{\Omega_{S-1}}}\right\}(1-c_\alpha) \\ &\geq \frac{|S\setminus\{r_S\}|}{|S|}\frac{1}{e^{4d+2}}\lambda_{M_{\Omega_{S\setminus\{r_S\}}}}(1-c_\alpha) \quad \textit{(comparison)} \\ &= \frac{|S\setminus\{r_S\}|}{|S|}\frac{1}{e^{4d+2}}\frac{1}{d}\lambda_{M_{\Omega_{S_c}}}(1-c_\alpha). \quad \textit{(product chain)} \end{aligned}$$

Note that at this point we have reached a recursion because we can re-use Theorem 6.2.6 to lower bound $\lambda_{M_{\Omega_{S_c}}}$. The upshot of all the calculations above is:

**Proposition 6.3.6** *If $T$ is a finite, complete $k$-ary tree with root $r_T$ and $M_{\Omega_T} = \left(\Omega_T, P_{M_{\Omega_T}}\right)$ is a time-reversible Markov chain on state space*

$$\Omega_T = \{-1,+1\}^{|T|} \tag{6.32}$$

*with transition probabilities*

$$P_{M_{\Omega_T}}(x,y) = \begin{cases} \frac{1}{|T|}\frac{e^{y_v\left(\Sigma_{i=1}^{k} x_{\gamma_i(v)}\right)}}{e^{\left(\Sigma_{i=1}^{k} x_{\gamma_i(v)}\right)}+e^{-\left(\Sigma_{i=1}^{k} x_{\gamma_i(v)}\right)}}, & \text{if } v = r_T \\ \frac{1}{|T|}\frac{e^{y_v\left(\Sigma_{i=1}^{k+1} x_{\gamma_i(v)}\right)}}{e^{\left(\Sigma_{i=1}^{k+1} x_{\gamma_i(v)}\right)}+e^{-\left(\Sigma_{i=1}^{k+1} x_{\gamma_i(v)}\right)}}, & \text{if } v \text{ is node} \\ \frac{1}{|T|}\frac{e^{y_v x_{\gamma(v)}}}{e^{x_{\gamma(v)}}+e^{-x_{\gamma(v)}}}, & \text{if } v \text{ is leaf} \end{cases} \tag{6.33}$$

*for adjacent $x$, $y$ and stationary distribution*

$$\pi_{M_{\Omega_T}}(x) = \frac{1}{Z_{M_{\Omega_T}}} \prod_{\substack{v\sim w \\ v\neq w \\ v,w\in T}} e^{x_v x_w} \quad \text{where} \quad Z_{M_{\Omega_T}} = \sum_{x\in T} \prod_{\substack{v\sim w \\ v\neq w \\ v,w\in T}} e^{x_v x_w} \tag{6.34}$$

*and $M_{\Omega_{T_c}} = \left(\Omega_{T_c}, P_{M_{\Omega_{T_c}}}\right)$ is a Markov chain identical to $M_{\Omega_T}$ only on an "immediate" subtree of $T$ (i.e. where every occurrence of $T$ in (6.32) to (6.34) is replaced with $T_c$), then*

$$\begin{aligned} \lambda_{M_{\Omega_T}} \geq \min\Bigg\{ & \frac{2}{|T|}\frac{e^{-(k+1)}}{e^{(k+1)}+e^{-(k+1)}}\left(\sqrt{c_\alpha} - 2\sqrt{\frac{2}{\alpha}}\sqrt{1-c_\alpha}\right)^2, \\ & \frac{|T\setminus\{r_T\}|}{|T|}\frac{1}{ke^{4k+2}}\lambda_{M_{\Omega_{T_c}}}(1-c_\alpha)\Bigg\}, \end{aligned}$$

*where $\alpha = e^{-2(k+1)}$ and $c_\alpha \in \left(\frac{8}{\alpha+8}, 1\right)$.*

This pretty much wraps up our calculation for bounding $\lambda_{M_{\Omega_S}}$ because by Theorem 6.2.6, the calculations in sections 6.3.1 and 6.3.2

$$\begin{aligned} \lambda_{M_{\Omega_S}} \geq \min\Bigg\{ & \frac{2}{|S|}\frac{e^{-(d+1)}}{e^{(d+1)}+e^{-(d+1)}}\left(\sqrt{c_\alpha} - 2\sqrt{\frac{2}{\alpha}}\sqrt{1-c_\alpha}\right)^2, \\ & \min\left\{\lambda_{M_{\Omega_{S+1}}}, \lambda_{M_{\Omega_{S-1}}}\right\}(1-c_\alpha)\Bigg\}, \end{aligned} \tag{6.35}$$

where $\alpha = e^{-2(d+1)}$ and $c_\alpha \in \left(\frac{8}{\alpha+8}, 1\right)$. By fact 6.3.5 we have

$$\lambda_{M_{\Omega_S}} \geq \min\left\{ \frac{2}{|S|}\frac{e^{-(d+1)}}{e^{(d+1)}+e^{-(d+1)}}\left(\sqrt{c_\alpha}-2\sqrt{\frac{2}{\alpha}}\sqrt{1-c_\alpha}\right)^2, \right.$$
$$\left. \frac{|S\setminus\{r_S\}|}{|S|}\frac{1}{d\,e^{4d+2}}\lambda_{M_{\Omega_{S_c}}}(1-c_\alpha)\right\}. \tag{6.36}$$

Observe that (6.35) – (6.36) corresponds to one application of Proposition 6.3.6. Another application of Proposition 6.3.6 gives bounds on $\lambda_{M_{\Omega_{S_c}}}$ and using $|S\setminus\{r_S\}| = d\,|S_c|$ we obtain

$$\lambda_{M_{\Omega_S}} \geq \min\left\{ \frac{2}{|S|}\frac{e^{-(d+1)}}{e^{(d+1)}+e^{-(d+1)}}\left(\sqrt{c_\alpha}-2\sqrt{\frac{2}{\alpha}}\sqrt{1-c_\alpha}\right)^2, \right.$$
$$\frac{2}{|S|}\frac{e^{-(d+1)}}{e^{(d+1)}+e^{-(d+1)}}\left(\sqrt{c_\alpha}-2\sqrt{\frac{2}{\alpha}}\sqrt{1-c_\alpha}\right)^2\left(\frac{1-c_\alpha}{e^{4d+2}}\right),$$
$$\left.\frac{|S_c\setminus\{r_{S_c}\}|}{|S|}\frac{1}{d}\lambda_{M_{\Omega_{S_{cc}}}}\left(\frac{1-c_\alpha}{e^{4d+2}}\right)^2\right\}.$$

Notice that at this point that we are effectively bounding the spectral gaps of the chain by recursion. Each level of the recursion is a sequence of Theorem 6.2.6, Markov chain comparison (Lemma 6.3.1) and product chain lemma (Lemma 6.3.3), i.e. one application of Proposition 6.3.6. The recursion comes to a halt when the restriction chains are defined on a tree of height 0, i.e. a tree consisting of a single vertex (namely the root) only. Note that a complete $d$-ary tree of height $h$ has $\frac{d^{h+1}-1}{d-1}$ vertices. Thus the tree on $S$ has height $h_S = \log_d\left(\frac{|S|(d-1)+1}{d}\right)$. Therefore, when the recursion stops $\lambda_{M_{\Omega_S}}$ is bounded by the minimum of

$$\frac{2}{|S|}\frac{e^{-(d+1)}}{e^{(d+1)}+e^{-(d+1)}}\left(\sqrt{c_\alpha}-2\sqrt{\frac{2}{\alpha}}\sqrt{1-c_\alpha}\right)^2\left(\frac{1-c_\alpha}{e^{4d+2}}\right)^k \tag{6.37}$$

for $k = 0\ldots h_S - 1$ and

$$\frac{1}{|S|}\lambda_{M_{\Omega_1}}\left(\frac{1-c_\alpha}{e^{4d+2}}\right)^{h_S}. \tag{6.38}$$

Let $M_{\Omega_1} = \left(\Omega_1, P_{M_{\Omega_1}}\right)$ with

$$\Omega_1 = \{+1, -1\},$$

$$P_{M_{\Omega_1}}(x,y) = \frac{e^y}{e+e^{-1}}$$

and stationary distribution

$$\pi_{M_{\Omega_1}}(x) = \frac{e^x}{e+e^{-1}}.$$

Using $\mathrm{Var}_{\pi_{M_{\Omega_1}}}(f) = 1$, it is easily verified that $\lambda_{M_{\Omega_1}} = 1$. Thus (6.38) equals $\frac{1}{|S|}\left(\frac{|S|(d-1)+1}{d}\right)^{\log_d\left(\frac{1-c_\alpha}{e^{4d+2}}\right)}$. Note that (6.37) is greater than

$$\begin{aligned}
&\frac{2}{|S|}\frac{e^{-(d+1)}}{e^{(d+1)}+e^{-(d+1)}}\left(\sqrt{c_\alpha}-2\sqrt{\frac{2}{\alpha}}\sqrt{1-c_\alpha}\right)^2\left(\frac{1-c_\alpha}{e^{4d+2}}\right)^{\log_d\left(\frac{|S|(d-1)+1}{d}\right)-1}\\
&\geq \frac{1}{1-c_\alpha}\left(\sqrt{c_\alpha}-2\sqrt{\frac{2}{\alpha}}\sqrt{1-c_\alpha}\right)^2\frac{1}{|S|}\left(\frac{|S|(d-1)+1}{d}\right)^{\log_d\left(\frac{1-c_\alpha}{e^{4d+2}}\right)}\\
&= \left(\sqrt{\frac{c_\alpha}{1-c_\alpha}}-2\sqrt{\frac{2}{\alpha}}\right)^2\frac{1}{|S|}\left(\frac{|S|(d-1)+1}{d}\right)^{\log_d\left(\frac{1-c_\alpha}{e^{4d+2}}\right)}. \qquad (6.39)
\end{aligned}$$

Choose $c_\alpha = \frac{1}{2}\left(1+\frac{8}{\alpha+8}\right) = \frac{1}{2}\frac{e^{-2(d+1)}+16}{e^{-2(d+1)}+8}$, then (6.39) equals

$$\begin{aligned}
&\frac{1}{\alpha}\left(\sqrt{\alpha+16}-\sqrt{8}\right)^2\frac{1}{|S|}\left(\frac{|S|(d-1)+1}{d}\right)^{\log_d\left(\frac{1-c_\alpha}{e^{4d+2}}\right)}\\
&= e^{2(d+1)}\left(\sqrt{e^{-(2d+1)}+16}-\sqrt{8}\right)^2\frac{1}{|S|}\left(\frac{|S|(d-1)+1}{d}\right)^{\log_d\left(\frac{e^{-2(3d+2)}}{2e^{-2(d+1)}+16}\right)} \qquad (6.40)
\end{aligned}$$

Since (6.40) is larger than (6.38) (for our choice of $c_\alpha$), we obtain

$$\lambda_{M_{\Omega_S}} \geq \frac{1}{|S|}\left(\frac{|S|(d-1)+1}{d}\right)^{\log_d\left(\frac{e^{-2(3d+2)}}{2e^{-2(d+1)}+16}\right)}. \qquad (6.41)$$

For bounding the mixing time of $M_{\Omega_S}$ it is furthermore necessary to find a lower bound for $\min_{x\in\Omega_S}\{\pi_{\Omega_S}(x)\}$. Since $S$ is a tree, there are exactly $|S|-1$ edges. Hence

$$\prod_{\substack{v,w\in S\\ v\sim w\\ v\neq w}} e^{x_v\cdot x_w} \geq e^{-|S|} \quad \text{and} \quad Z_M \leq 2^{|S|}e^{|S|}. \qquad (6.42)$$

**Theorem 6.3.7** *The mixing time of the single-site-update Markov chain $M_{\Omega_S}$ on the finite, complete d-ary tree S is*

$$\tau(1/e) \in O\left(|S|^{2+\frac{1}{\ln d}(6d+4+\ln 18)}\right).$$

*Proof.* By (6.41)

$$\begin{aligned}
\frac{1}{\lambda_{M_\Omega}} &\leq |S|\left(\frac{|S|(d-1)+1}{d}\right)^{-\log_d\left(\frac{e^{-2(3d+2)}}{2e^{-2(d+1)}+16}\right)} \\
&= |S|\left(\frac{|S|(d-1)+1}{d}\right)^{\log_d\left(2e^{4d+2}+16e^{6d+4}\right)} \\
&\leq |S|\left(\frac{|S|(d-1)+1}{d}\right)^{\log_d\left(18e^{6d+4}\right)} \\
&= |S|\left(\frac{|S|(d-1)+1}{d}\right)^{(6d+4)\log_d e+\log_d 18} \\
&\leq |S|\left(\frac{|S|(d-1)+|S|}{d}\right)^{(6d+4)\log_d e+\log_d 18} \\
&= |S|^{1+(6d+4)\log_d e+\log_d 18} \\
&= |S|^{1+(6d+4)\frac{\ln e}{\ln d}+\frac{\ln 18}{\ln d}} \\
&= |S|^{1+\frac{1}{\ln d}(6d+4+\ln 18)}
\end{aligned}$$

and by (6.42)

$$\ln\frac{1}{\pi(x)} \leq |S|\ln\left(2e^2\right) \qquad \text{for all } x \in \Omega.$$

Plugging the above into Lemma 2.3.2 yields

$$\tau(1/e) \leq c\cdot|S|^{2+\frac{1}{\ln d}(6d+4+\ln 18)}$$

for large $|S|$, where $c$ is a constant independent of the Markov chain. □

As discussed earlier, we do not expect this bound on the mixing time to be close to the truth. In fact, recent results show that our bound is far off: Berger, Kenyon, Mossel and Peres [13] give a bound of $\tau(1/e) \in O(|S|)$ for Glauber dynamics for the Ising model on $d$-ary graphs in the high temperature region[2] with no boundary conditions[3]. Martinelli, Sinclair and Weitz [88] have shown that the log-Sobolev constant of

[2]The high temperature region contains $\beta = 1$.

[3]A boundary condition is a condition imposed on the spins of the sites on the boundary, e.g. +-boundary condition means that the spin of all sites on the boundary is +.

Glauber dynamics for the Ising model on $d$-ary trees with $(+)$-boundary condition[4] is $\Omega\left(n^{-1}\right)$ at all temperatures, which bounds the mixing time by $O(n\ln n)$ at all temperatures[5].

[4] $(+)$-boundary condition means that the spin on the leaves of the tree is $+$.

[5] In a subsequent paper [87] the authors succeed in extending their result to other models of spin systems.

# Chapter 7

# Conclusion

The aim of the research described in this thesis is twofold. First, to apply recent techniques for bounding the mixing time to already solved problems and to examine if or how much gain can be achieved by using these newer techniques. The second goal is to find new techniques for bounding mixing time. With regard to the former: the techniques we look at are path coupling (by Bubley and Dyer [18]), average conductance (due to Kannan and Lovàsz [68]) and log-Sobolev constants (which were originally introduced in an entirely different context by Gross [45]). As regards the latter goal: a new decomposition technique (decomposition in the context of bounding mixing time was trail-blazed by Madras and Randall [85]) is discussed in chapter 6.

In chapter 3 we apply path coupling to analyse the mixing time of a Markov chain for generating random lozenge tilings. The Markov chain under consideration — already proved to be rapidly mixing by Luby et al. [84] — moves from one tiling to another by so-called *tower-rotations* (for details refer to the original paper or section 3.2.2). There is another Markov chain for generating lozenge tilings: the so-called "natural" Markov chain for generating lozenge tilings, which moves from tiling to tiling by rotating a hexagonal cluster of three lozenges at a time[1]. Our original goal was to show rapid mixing for the "natural" Markov chain using path-coupling. Recall the path-coupling theorem (Theorem 3.1.1): to apply path-coupling, it is imperative to

[1]It is "natural" in the sense that the operation is simple and achieves the minimum, in terms of numbers of lozenges that have to be changed to obtain a new lozenge tiling. In terms of "tower rotations" this corresponds to flipping a tower of height 1.

find a metric $\delta$ on $\Omega \times \Omega$ and a neighbourhood relation $A \subseteq \Omega \times \Omega$ such that (using the notation of Theorem 3.1.1) for all $(X,Y) \in A$

$$\mathrm{E}\left[\delta\left(X',Y'\right) \mid (X,Y)\right] \leq \delta(X,Y).$$

However, all attempts to find such a metric and neighbourhood proved to be fruitless. It may well be the case that we merely have not been able to find the right choice of metric and neighbourhood. On the other hand, path-coupling is another example of using *local* information to make a *global* statement. We believe that our failure to exhibit the existence of an appropriate metric and neighbourhood indicates that lozenge tilings are fundamentally non-local, i.e. small local changes like the rotation of a three-cluster of lozenges have far reaching effects, and can therefore not be described by local information alone[2]. Nonetheless, as a "by-product" we obtain a proof of rapid mixing of the tower-rotation chain with a marginally better mixing time than Luby et al. (which is due to the fact that our analysis considers tilings on hexagonal regions whereas theirs is not restricted to hexagonal regions only) and our proof is simpler and more rigorous by comparison. As stated before, Luby et al. in later (unpublished) work improved their mixing time bound to $O\left(n^{3.5}\right)$. The state of the art for the tower-rotation chain is Wilson's $O\left(n^2 \ln n\right)$ in [117].

In chapter 4 we apply average conductance to bound the mixing time of the bases-exchange walk on balanced matroids. Feder and Mihail showed in [41] that the bases-exchange walk on balanced matroids is rapidly mixing. Assuming that the matroid has rank $n$ and is defined on a ground set of size $m$, they give in fact two bounds: a first bound $\tau(1/e) \in O\left(n^3 m^2 \ln m\right)$ obtained by using (classical) conductance and an improved bound $\tau(1/e) \in O\left(n^3 m \ln m\right)$ via canonical paths. The bound we obtain by using average conductance: $\tau(1/e) \in O\left(n^2 m^2\right)$ beats Feder and Mihail's by $\Omega(n)$ (which seems to be the typical gain of average conductance over classical conductance[3]). Using a recent result by Houdrè [55], which allows us to lower bound the log-Sobolev constant (see Definition 5.1.1) by the conductance function (cf. Definition 4.1.1), we succeed in decreasing the upper bound on mixing time to $\tau(1/e) \in O\left(m^{1.5} n^2 (\ln n + \ln\ln m)\right)$. Note that our improved bound is better than

[2]Corroborated by personal communication with David Wilson.

[3]Personal communication with Mark Jerrum and Ravi Montenegro.

Feder and Mihail's canonical-paths bound as long as $n \ln m \in \Omega(\sqrt{m} \ln n)$. This holds for example when $m \in O(n^2)$, which, as discussed at the end of section 4.4, is true for simple regular matroids. Feder and Mihail furthermore introduced a modified bases-exchange walk (a random walk on matroid bases and so-called *near-bases*). At the end of a very terse analysis they bound the number of steps needed until the modified walk is in $1/e$ distance to stationarity by $O(n^4 \ln m)$. Note that this bound is bounding the number of steps only. The cost of actually implementing the modified walk is not stated exactly by the authors, but they believe it to be high; some[4] believe the cost of a single step of the modified walk alone costs $O(n)$. We have not studied the average conductance of the modified walk or the modified walk as such for that matter.

As our final bound on the mixing time of the bases-exchange walk (Theorem 4.4.9) shows log-Sobolev constants can yield better bounds on mixing time than conductance. This, of course, is a consequence of Lemma 5.1.2, which was established by Diaconis and Saloff-Coste [30][5], and Theorem 2.3.7; a brief discussion of Lemma 5.1.2 is given on page 75. Chapter 5 is a natural continuation of chapter 4, because log-Sobolev constants are the next step up[6] from conductance as far as mixing-time techniques are concerned. Furthermore, since we "only" used an estimate of the log-Sobolev constant in Theorem 4.4.9, further motivation was to find out how accurate the estimation had been. We abstract away from bases-exchange walks on balanced matroids and bound instead the log-Sobolev constant of $\pi$-*recursive* Markov chains[7]. Before calculating bounds for the log-Sobolev constant, we calculate lower bounds for the spectral gap. This is motivated by the fact that spectral gap is conceptually and computationally simpler than log-Sobolev constants. But moreover, owing to Rothaus [105], it is known that $\alpha \leq \lambda/2$; hence, a bound on the spectral gap can be used as a benchmark for a bound on the log-Sobolev constant. Similarly to chapter 4, we exploit the inductive nature of the structure we are dealing with[8] to obtain lower bounds on the spectral

[4]Communicated by Martin Dyer.

[5]For random walks on finite graphs the bound has been slightly improved to $\tau(1/e) \leq \frac{\log\log n}{4\alpha}$, where $\alpha$ is log-Sobolev const., by Chung [27].

[6]Strictly speaking, bounding the spectral gap is the immediate next step up from conductance, but we do just that before giving bounds for the log-Sobolev constant.

[7]Although so far, the only known example for a $\pi$-recursive Markov chain is the bases-exchange walk on balanced matroids.

[8]Recall that $\pi$-recursive Markov chains are an abstraction of the bases-exchange walk on balanced

gap $\lambda$. As it turns out, the very same approach can then be used for bounding the log-Sobolev constant $\alpha$. For general $\pi$-recursive Markov chains our bound on the spectral gap is $\lambda \geq 2\min_{\{x,y\}\in E}\{P(x,y) \mid P(x,y) > 0\}$, if $P(\cdot,\cdot)$ is the transition kernel and $E$ the *underlying graph*[9] of the chain, and $\alpha \geq \frac{1}{2}\min_{\{x,y\}\in E}\{P(x,y) \mid P(x,y) > 0\}$ for the log-Sobolev constant[10]. According to Rothaus's result [105] our bound on $\alpha$ is not tight w.r.t. $\lambda$, nonetheless it is of the same order (which is still good enough to yield better bounds on the mixing time than spectral gap). For the bases-exchange walk on a balanced matroid of rank $n$ on a ground set of size $m$ the bounds on the spectral gap and the log-Sobolev constant are $\lambda \geq \frac{2}{mn}$ and $\alpha \geq \frac{1}{2mn}$ respectively. The resulting bound on the mixing time, $\tau(1/e) \in O(mn(\log n + \log\log m))$, is as far as we are aware the best bound for this kind of Markov chain to date; it is a substantial improvement on $\tau(1/e) \in O\left(m^{3/2}n^2(\log n + \log\log m)\right)$ (Theorem 4.4.9). Unfortunately, the methods we used in chapter 5 for bounding the spectral gap and log-Sobolev constant do not seem to carry over to general matroids (in the sense that the resulting bounds on the mixing time are polynomial). Although it is believed (see e.g. [92] and [93]) that the bases-exchange walk on general matroids is rapidly mixing, the jury is still out.

The bounds on spectral gap and log-Sobolev constant we obtained in chapter 5 were computed using a *decompositional* approach: in our case the computation of the spectral gap/log-Sobolev constant was facilitated by decomposing the original Markov chain into two smaller chains[11]. This idea is not new; in fact, as regards computation of log-Sobolev constants decomposition of some sort is commonly used (see e.g. Caputo & Martinelli [21], Cesi [22], Ledoux [77], Lu & Yau [83], Martinelli, Sinclair & Weitz [88]). The use of decomposition for computing the spectral gap of Markov chains first appeared in a paper by Madras and Randall [85]. Since our decomposition technique (from chapter 5) differs noticeably from Madras and Randall's, the research described in chapter 6 was begun with the goal to "bring the two approaches under one umbrella", i.e. to find a "generic" decompositional approach which, depending on some parameters of the Markov chain, would lead to a Madras-

matroids.

[9] See index-entry for a pointer to the definition.

[10] The peculiar form of our bounds on spectral gap and log-Sobolev constant is discussed after the proof of Theorem 5.3.4.

[11] Decomposition is thus another instance of the *divide and conquer* principle.

Randall-type decomposition, a decomposition as seen in chapter 5 or a kind of decomposition lying "in between". The result we state in chapter 6, Theorem 6.2.6, does not achieve the target of the above brief[12]: Firstly, Theorem 6.2.6 only considers decomposition into two restriction chains. Secondly, it states that the spectral gap of a Markov chain is lower bounded by either the minimum of the spectral gaps of the restriction chains or the "probabilistic flow" between the two restriction chains. Hence Theorem 6.2.6 lies somewhere in between the other two decomposition methods; although all in all it seems closer in character to the decomposition seen in chapter 5 than to Madras-Randall decomposition. Since the new decomposition method merely bounds the spectral gap of the original chain by either the minimum of the spectral gaps of the restriction chains or the probabilistic flow between the two chains, in general we do not expect it to yield tight bounds. Its advantage is its easiness to use. We apply the new decomposition method to lower bound the spectral gap of Glauber dynamics for the Ising model with inverse temperature $\beta = 1$ and no boundary conditions on finite, complete $d$-ary trees. The bound on the mixing time that we obtain is $\tau(1/e) \in O\left(|S|^{2+\frac{1}{\ln d}(6d+4+\ln 18)}\right)$, where $S$ is the vertex set of the tree. This is far off from Berger, Kenyon, Mossel and Peres's (optimal) $\tau(1/e) \in O(|S|)$ in [13] for the temperature regime which contains $\beta = 1$. Recently, Martinelli, Sinclair and Weitz [88] have shown that the log-Sobolev constant of Glauber dynamics for the Ising model on $d$-ary trees with $(+)$-boundary condition[13] is $\Omega\left(n^{-1}\right)$ at all temperatures, which translates into a bound of $O(n \ln n)$ on mixing time at all temperatures[14].

[12]However, in [62] the approach used in Theorem 6.2.6 was developed further to obtain a decomposition method which comes much closer to our original goal: The method accommodates partitioning into more than two restriction chains and bounds the spectral gap in terms of the spectral gaps of the projection-, restriction chains and an extra parameter. It yields bounds that are at least as good as the bounds obtained by the other two decomposition methods and depending on the extra parameter its bounds can be much better. Moreover, the method can be applied to obtain bounds on the log-Sobolev constant.

[13]$(+)$-boundary condition means that the spin on the leaves of the tree is $+$.

[14]In a subsequent paper [87] the authors succeed in extending their result to other models of spin systems.

# Appendix A

# All problems in NP are self-reducible

Suppose that $\mathfrak{p}$ is a problem in NP and $R_{\mathfrak{p}} \subseteq \Sigma^* \times \Sigma^*$ some relation (on some alphabet $\Sigma$) perceived to be encoding $\mathfrak{p}$ in a "natural" way, such that for $(x,y) \in R_{\mathfrak{p}}$ the string $x$ encodes a problem instance and $y$ its solution. Our concern is to show that if $R_{\mathfrak{p}}$ is not self-reducible, then we can re-formulate $\mathfrak{p}$ to obtain a self-reducible relation $R'_{\mathfrak{p}} \subseteq (\Sigma')^* \times (\Sigma')^*$ (on some suitably chosen alphabet $\Sigma'$) that is closely related to $R_{\mathfrak{p}}$ in the sense that there exist an injection $f_{\mathfrak{p}} = ((x,y) \mapsto (x',y')) : R_{\mathfrak{p}} \to R'_{\mathfrak{p}}$ such that $R_{\mathfrak{p}}(x,y)$ iff $R'_{\mathfrak{p}}(x',y')$ and a counterpart to $f_{\mathfrak{p}}$, namely a relation (by abuse of notation) $f_{\mathfrak{p}}^{-1} = R'_{\mathfrak{p}} \times R_{\mathfrak{p}}$, which assigns to every pair $(u,v) \in R'_{\mathfrak{p}}$ at least one pair $(x,y) \in R_{\mathfrak{p}}$ such that $R'_{\mathfrak{p}}(u,v)$ iff $R_{\mathfrak{p}}(x,y)$. In other words, although the re-formulation $R'_{\mathfrak{p}}$ of $\mathfrak{p}$ might be "richer" than $R_{\mathfrak{p}}$ (in general $\left|R'_{\mathfrak{p}}\right| \geq |R_{\mathfrak{p}}|$), it includes nothing entirely new; think of $u$ as either a reformulation of the problem instance $x$ or a subproblem of $x$ with associated solution $v$. In the case that $u$ is a reformulation (i.e. $(u,v) = f_{\mathfrak{p}}(x,y)$ for some $x,y$), $v$ will be a reformulation of $y$. In the case that $u$ encodes a subproblem, $v$ will encode a partial solution to $x$. Significantly, the reformulation of problem instances is at most polynomial in the length of the original formulation, i.e. $|f_{\mathfrak{p}}(x,y)| = O\left(|x|^{\xi}\right)$ for a suitably chosen constant $\xi \geq 1$.

Before we can give a more formal formulation of what is stated above, we have to slightly alter the definition of self-reducibility. Recall that the result of Jerrum et al. [63] is still valid if (SR1) is replaced by the relaxation (SR1)'. Thus, we from now on call a relation (polynomial time) self-reducible if (SR1)', (SR2) and (SR3) are satisfied. We then have:

**Theorem A.0.8** *Let* $\mathfrak{p}$ *be a problem in NP and the polynomial-time decidable relation* $R_\mathfrak{p} \subseteq \Sigma^* \times \Sigma^*$ *its encoding on some alphabet* $\Sigma$. *There exists, on a suitably chosen alphabet* $\Sigma'$, *a polynomial-time decidable and (polynomial-time) self-reducible encoding* $R'_\mathfrak{p} \subseteq (\Sigma')^* \times (\Sigma')^*$ *that allows a (polynomial time computable) injective function* $f_\mathfrak{p} = ((x,y) \mapsto (x',y')) : R_\mathfrak{p} \to R'_\mathfrak{p}$ *such that* $(x,y) \in R_\mathfrak{p}$ *iff* $f_\mathfrak{p}(x,y) = (x',y') \in R'_\mathfrak{p}$ *and a (computable) inverse relation* $f_\mathfrak{p}^{-1} = R'_\mathfrak{p} \times R_\mathfrak{p}$ *such that* $(u,v) \in R'_\mathfrak{p}$ *iff* $f_\mathfrak{p}^{-1}(u,v) \subseteq R_\mathfrak{p}$, *where* $f_\mathfrak{p}^{-1}(u,v) = \left\{(x,y) \in \Sigma^* \times \Sigma^* \mid ((u,v),(x,y)) \in f_\mathfrak{p}^{-1}\right\}$. *Furthermore, there exists a constant* $\xi \geq 1$ *such that* $|f_\mathfrak{p}(x,y)| = O\left(|x|^\xi\right)$ *for all* $(x,y) \in R_\mathfrak{p}$.

*Proof.* The idea is to use the computations of a Turing machine as an encoding for $R'_\mathfrak{p}$. This can be done as outlined below: Since $\mathfrak{p}$ is in NP, $R_\mathfrak{p}$ is decidable by a deterministic polynomial-time Turing machine, say $T_{R_\mathfrak{p}}$. This machine accepts a pair $(x,y) \in \Sigma^* \times \Sigma^*$ iff $(x,y) \in R_\mathfrak{p}$. Assume that $T_{R_\mathfrak{p}}$ is a single-tape Turing machine defined by the sextuple $(S, \Sigma, \Gamma, s_0, H, \delta)$, where $S$ is a finite set denoting the states of $T_{R_\mathfrak{p}}$, $\Sigma$ the finite input alphabet, $\Gamma$ the finite tape alphabet (containing the special symbol $\flat$ "blank". Note that $(\Sigma \cup \flat) \subseteq \Gamma$ and $\Gamma \cap S = \emptyset$), $s_0$ the initial state, $H \subseteq S$ the set of halting states and $\delta : (S \setminus H) \times \Gamma \to S \times \Gamma \times \{\leftarrow, \bullet, \rightarrow\}$ the transition function (assume that $\delta$ is given explicitly as the relation $(S \setminus H) \times \Gamma \times S \times \Gamma \times \{\leftarrow, \bullet, \rightarrow\}$). A configuration of $T_{R_\mathfrak{p}}$ is the triple $(s, t_l, t_r)$, where $s$ is the current state, $t_l \in \Gamma^*$ is the string of tape symbols to the left of the tape-head (omitting leading blanks), $t_r \in \Gamma^*$ the string of tape symbols to the right of the tape-head (omitting trailing blanks)[1] and assume that the current position of the tape-head is on the first symbol of $t_r$. As $T_{R_\mathfrak{p}}$ is a deterministic Turing machine, there is a one-to-one correspondence between a pair $(x,y) \in R_\mathfrak{p}$ and accepting computations of $T_{R_\mathfrak{p}}$ (a computation is a sequence of configurations of $T_{R_\mathfrak{p}}$). We can therefore use the computations of $T_{R_\mathfrak{p}}$ to obtain the more expressive encoding $R'_\mathfrak{p}$ for $\mathfrak{p}$. We could encode $y'$ as the sequence of configurations of $T_{R_\mathfrak{p}}$ on input $(x,y)$ such that $y' = c_0 \ldots c_{z-1}$, where $c_i$ with $|c_i| = k$ are configurations (assuming that the accepting computation is of length $z$ and configurations are of length $k$. Should the actual length of a configuration be smaller than $k$, then pad $t_r$ out with blanks). However, with the self-reducibility of $R'_\mathfrak{p}$ in mind (and bearing in mind that $k$ is polynomial in the size

[1] Omitting leading and trailing blanks ensures that configurations will be of finite length if the input and the computation itself are finite.

of $x$) this might not be a suitable choice as we would have to make $x'$ truly long to guarantee that $k \leq \log(|x'|)$ (which is the upper bound for the function $\sigma$ as stated by (SR2)) because if $k > \log(|x'|)$, then the prefix of $y'$ of length $\sigma(x')$ being cut off is not long enough to contain a configuration of $T_{R_\mathfrak{p}}$ and this would leave us with a suffix of $y'$ that begins somewhere in the middle of a configuration, which would make interpreting this suffix rather difficult (as will hopefully become apparent upon further reading). The way around this problem is motivated by the observation that as long as we know the initial configuration $c_0$, the transition function $\delta$ and the current state together with the current symbol under the tape-head for each successor configuration of the computation, it is easy to fully recover each configuration of the computation. Hence, being given the initial configuration $c_0$, $\delta$ and a pair $(s,a) \in S \times \Gamma$ for each of the $z$ configurations of an accepting computation (where $s$ and $a$ are to be interpreted as just discussed) is as good as being given the $z$ configurations explicitly. We will therefore encode $y'$ as a string in $(S \times \Gamma)^*$ such that $y' = (s_0,a_0)\ldots(s_{z-1},a_{z-1})$ where $s_i$ and $a_i$, for $i = 0\ldots z-1$, are the state and the symbol under the tape-head of configuration $c_i$ respectively. Observe that this encoding is substantially shorter than the first one we considered, which used configurations: each string of the form $(s,a)$ has constant[2] length, say the constant is $\rho$, while a configuration has length polynomial in the size of the input $x$. Since it is possible to guarantee that $\rho \leq \log_b(|x'|)$ for a suitably chosen basis $b > 1$ without making the length of $x'$ very long, we will avoid the problem mentioned above.

It remains to define $x'$. To this end let us define the function $f_\mathfrak{p} : R_\mathfrak{p} \to R'_\mathfrak{p}$. Given $(x,y) \in R_\mathfrak{p}$ let $f_\mathfrak{p}(x,y) = (x',y')$, where $x' = (T_{R_\mathfrak{p}}, c_0)$ and $y'$ is as defined above. The configuration $c_0$ is the initial configuration of $T_{R_\mathfrak{p}}$ on input $(x,y)$, i.e. $c_0 = (s_0, \varepsilon, (x,y))$, where $\varepsilon$ denotes the empty string. Recall that $T_{R_\mathfrak{p}}$ is a polynomial-time Turing machine (let us assume its running-time is bounded by some polynomial of degree $\xi \geq 1$). Hence, $f_\mathfrak{p}$ is polynomial-time computable in $|x|$ and the length of $(x',y')$ polynomial in $|x|$ (the length of the description of $T_{R_\mathfrak{p}}$ does not depend on the length of the input $x$

[2]The length of $(s,a)$ is bounded by a constant independent of the length of the input $(x,y) \in R_\mathfrak{p}$ because the finite state set $S$ and the finite tape alphabet $\Gamma$ are independent of the input. Hence, $|S|$ and $|\Gamma|$ are bounded by a constant independent of the size of the input, which implies that $|s|$ and $|a|$, for all $s \in S$ and $a \in \Gamma$, are bounded by a constant.

and is hence a constant).

We will now define $R'_{\mathfrak{p}}$. Let $\Sigma' = \Gamma \cup S \cup \{\leftarrow, \bullet, \rightarrow\} \cup \{,\} \cup \{(\} \cup \{)\}$. The pair of strings $(u,v) \in (\Sigma')^* \times (\Sigma')^*$ is in $R'_{\mathfrak{p}}$ iff $u = \left(T_{R_{\mathfrak{p}}}, (s, t_l, t_r)\right)$ and $v = (s_1, a_1) \ldots (s_\zeta, a_\zeta)$ such that

1. $s = s_1$ and $a_1$ is the leftmost element of $t_r$, i.e. the symbol under the tape-head in configuration $(s, t_l, t_r)$,

2. the configuration with current state $s_{i+1}$ and current symbol $a_{i+1}$ under the tape-head is a valid successor configuration (w.r.t. $T_{R_{\mathfrak{p}}}$) of the configuration with current state $s_i$ and current symbol $a_i$ under the tape-head for $i = 1 \ldots \zeta - 1$ and $s_\zeta$ is an accepting state of $T_{R_{\mathfrak{p}}}$ and

3. if starting from $(s, t_l, t_r)$ the transition function $\delta$ is applied reversely[3], we will after at most polynomially (in the length of the original problem instance $x$) many reverse applications of $\delta$ arrive at a string $\omega = \left(T_{R_{\mathfrak{p}}}, (s_0, \varepsilon, (x,y))\right)$, where $(x,y) \in R_{\mathfrak{p}}$.

Informally, the pairs $(u,v) \in R'_{\mathfrak{p}}$ correspond to the end part of accepting configurations of $T_{R_{\mathfrak{p}}}$; think of an accepting computation of $T_{R_{\mathfrak{p}}}$ that was suspended and then is later on resumed — $u = \left(T_{R_{\mathfrak{p}}}, c\right)$ indicates where the computation was interrupted and $v = v_1 \ldots v_\zeta$ which steps still had to be executed when the computation was interrupted. Condition 3. ensures that only strings $u \in (\Sigma')^*$ that indeed correspond to configurations of an interrupted accepting computation are in $R'_{\mathfrak{p}}$: if $u = \left(T_{R_{\mathfrak{p}}}, c\right)$ corresponds to an intermediate configuration of an accepting computation of $T_{R_{\mathfrak{p}}}$ on some input $(x,y) \in R_{\mathfrak{p}}$, then applying $\delta$ backwards from configuration $c$ will lead us back to the pair $(x,y) \in R_{\mathfrak{p}}$. To see that $R'_{\mathfrak{p}}$ is polynomial-time decidable in the size of the original problem instance $x$ we use a deterministic universal Turing machine, say $T_{R'_{\mathfrak{p}}}$. On input $\left(\left(T_{R_{\mathfrak{p}}}, c\right), v_1 \ldots v_\zeta\right) \in R'_{\mathfrak{p}}$, the machine $T_{R'_{\mathfrak{p}}}$ simulates $T_{R_{\mathfrak{p}}}$ starting from configuration $c$ and accepts if the configuration with current state $s_{i+1}$ and current symbol $a_{i+1}$ under the tape-head is a valid successor configuration (w.r.t. $T_{R_{\mathfrak{p}}}$) of the configuration with

[3]What is meant is that $\delta$ be applied in the reverse direction. Since $\delta$ is no bijection, it might well be the case that there is more than one choice of predecessor, in which case the predecesor can be chosen arbitrarily.

current state $s_i$ and current symbol $a_i$ under the tape-head for $i = 1 \ldots \zeta - 1$ and $s_\zeta$ is an accepting state of $T_{R_\mathfrak{p}}$. Since the simulation of a deterministic Turing machine on a universal deterministic Turing machine can be done in polynomial time and $R_\mathfrak{p}$ itself is polynomial-time decidable, it is obvious that $R'_\mathfrak{p}$ too is polynomial-time decidable. It furthermore is obvious that $f_\mathfrak{p}$ has the above stated properties. It is also clear that the relation $f_\mathfrak{p}^{-1}$, which assigns to every pair $(u,v) \in R'_\mathfrak{p}$ the pairs $(x,y) \in R_\mathfrak{p}$ for which the final part of the computation in $T_{R_\mathfrak{p}}$ coincides with $v$, is computable though not necessarily in polynomial time[4].

It remains to show that $R'_\mathfrak{p}$ is (polynomial-time) self-reducible. First, let us consider the function $\sigma : (\Sigma')^* \to \mathbb{N}$. As mentioned before, if the basis $b$ of the logarithm appearing in (SR2) is carefully chosen, then $\rho \leq \log_b |x'|$ holds. Recall that all strings $(s,a)$ where $s \in S$ and $a \in \Gamma$ have length $\rho$. Let use therefore define $\sigma(x') = \kappa \cdot \rho$ for all $x' \in (\Sigma')^*$ and some non-negative integer constant $\kappa$. This choice of $\sigma$ guarantees that (SR2) is satisfied. Notice that $\sigma$ is obviously polynomial-time computable in the size of $x'$. Next let us construct the function $\psi : (\Sigma')^* \times (\Sigma')^* \to (\Sigma')^*$. Given a pair $\left(\left(T_{R_\mathfrak{p}}, c\right), v_1 \ldots v_\zeta\right) \in R'_\mathfrak{p}$, simulate $T_{R_\mathfrak{p}}$ for $\sigma\left(\left(T_{R_\mathfrak{p}}, c\right)\right) + 1$ steps to obtain the successor configuration $c_{\sigma\left(\left(T_{R_\mathfrak{p}}, c\right)\right)+1}$ of $c$ and define $\psi\left(\left(T_{R_\mathfrak{p}}, c\right), v_1 \ldots v_{\sigma\left(\left(T_{R_\mathfrak{p}}, c\right)\right)}\right) = \left(T_{R_\mathfrak{p}}, c_{\sigma\left(\left(T_{R_\mathfrak{p}}, c\right)\right)+1}\right)$. Declare $\psi$ on $(u,v) \notin R'_\mathfrak{p}$ as undefined. Then $\psi$ satisfies (SR1)' because, as we have already mentioned on page 128, we can assume that all configurations have the same (polynomially bounded) length $k$ and therefore $|c| = \left|c_{\sigma\left(\left(T_{R_\mathfrak{p}}, c\right)\right)+1}\right|$. Since the description of the Turing machine $T_{R_\mathfrak{p}}$ is fixed (i.e. can be regarded as constant for $\mathfrak{p}$), $\left|\psi\left(\left(T_{R_\mathfrak{p}}, c\right), v_1 \ldots v_{\sigma\left(\left(T_{R_\mathfrak{p}}, c\right)\right)}\right)\right| = \left|\left(T_{R_\mathfrak{p}}, c\right)\right|$ and condition (SR3) is met, too. Lastly, observe that $\psi$ is polynomial-time computable because $T_{R_\mathfrak{p}}$ is polynomial-time, which concludes the proof. □

The reader might already have observed that for pairs $(u,v) \in R'_\mathfrak{p}$ with $u = \left(T_{R_\mathfrak{p}}, c\right)$ encoding the problem instance and $v$ encoding a solution the explicit statement of the solution $v$ is actually redundant. This is due to the property that a solution is already incorporated in $u$ albeit in an extremely implicit manner: the configuration $c$ is an

[4]This is because $\mathfrak{p}$ is in NP. Since $\mathfrak{p}$ is in NP, every pair $(u,v) \in R'_\mathfrak{p}$ corresponds to an intermediate configuration of (potentially) exponentially many initial configurations $(x,y) \in R_\mathfrak{p}$. As enumerating exponentially many initial configurations is not possible in polynomial time, computing $f_\mathfrak{p}^{-1}$ cannot be done in polynomial time either.

intermediate configuration of an accepting computation of the deterministic Turing machine $T_{R_p}$ and as such fully predetermines the successor configurations, i.e. the very configurations that are encoded by $v$. The definition of $R'_p$ is as it is merely to make it comply with the definition of self-reducibility.

The meaning of Theorem A.0.8 is that every problem in NP can be encoded in a self-reducible way if the encoding is rich enough. This underlines the suspicion[5] that self-reducibility is less an intrinsic property of a problem but more one of an encoding. To illustrate this point we will show that contingency tables can be encoded in a self-reducible fashion. Let $\mathbb{N}_0$ denote the set of non-negative integers. Given two vectors $r = (r_1, \ldots, r_m) \in (\mathbb{N}_0)^m$ and $c = (c_1, \ldots, c_n) \in (\mathbb{N}_0)^n$, an $m \times n$ matrix $[a_{ij}]$ with non-negative integer entries is a *contingency table* with row sums $r$ and column sums $c$ if $\sum_{j=1}^{n} a_{ij} = r_i$ for every row $i$ and $\sum_{i=1}^{m} a_{ij} = c_j$ for every column $j$: In this formulation, a problem instance is given by the two vectors $r \in (\mathbb{N}_0)^m$, $c \in (\mathbb{N}_0)^n$ and a solution to the problem of finding a contingency table is an $m \times n$ matrix with non-negative integer entries and row and column sums $r$ and $c$ respectively. This encoding of the problem is not self-reducible. Nonetheless, in the spirit of Theorem A.0.8, the problem can be re-phrased to find a self-reducible encoding.

**Proposition A.0.9** *There exists a* non-standard encoding *of Contingency tables which is (polynomial-time) self-reducible.*

*Proof.* Before we can re-phrase the original problem, we need a couple of definitions. Define $(\mathbb{N}_0)^{[k]} = \bigcup_{i=1}^{k} (\mathbb{N}_0)^i$. We say an interval $[a, b]$ is *degenerate* if $a = b$; observe that a degenerate interval can be regarded as a number. Let $L = \{(l_1, \ldots, l_{mn})\}$, where the $l_i$'s are non-negative integer intervals such that all degenerate intervals are on the left and all non-degenerate intervals on the right, i.e. if $l_i$ is a degenerate interval, then all intervals to the left of $l_i$ are degenerate and if it is non-degenerate, then all intervals to its right are non-degenerate. Observe that an element of $L$ corresponds to an $m \times n$ matrix whose entries are non-negative integer intervals. Given two vectors $r \in (\mathbb{N}_0)^m$, $c \in (\mathbb{N}_0)^n$ and a list $l \in L$ (where for some $0 \leq k \leq mn$ only the $mn - k$ leftmost entries are degenerate intervals), a *partially-filled contingency table* is a vector

[5] Personal communication with Mark Jerrum.

$v = (v_1, \ldots, v_k) \in (\mathbb{N}_0)^{[m \cdot n]}$ such that $v_i \in l_{mn-k+i}$, for all $1 \leq i \leq k$, and by abuse of notation[6] the non-negative integer vector $t = (l_1, \ldots, l_{mn-k}, v_1, \ldots, v_k)$ ($t = v$ in case $k = mn$) encodes a contingency table (in the conventional sense) with row sums $r$ and column sums $c$. Notice that contingency tables and partially-filled contingency tables are equivalent: if $[a_{ij}]$ is a contingency table with row and column sums $(r, c)$, then $v = (a_{11}, \ldots, a_{1n}, a_{21}, \ldots, a_{2n}, \ldots, a_{m1}, \ldots, a_{mn})$ is a partially-filled contingency table for $(r, c, l)$ where all entries in $l$ are the same interval, namely $[0, \sum_{i=1}^{n} r_i]$ (observe that for contingency tables $\sum_{i=1}^{n} r_i = \sum_{i,j} a_{ij} = \sum_{j=1}^{m} c_j$). Conversely, if $v = (v_1, \ldots, v_k)$ is a partially-filled contingency table for $(r, c, l)$, then $t = (l_1, \ldots, l_{mn-k}, v_1, \ldots, v_k)$ is as stated above a contingency table with row sums $r$ and column sums $c$. Hence, we can regard partially-filled contingency tables as a non-standard way of encoding contingency tables.

It remains to show that partially-filled contingency tables are (polynomial-time) self-reducible. Let $v = (v_1, \ldots, v_k)$ be a partially-filled contingency table for $(r, c, l)$. Given a prefix $(v_1, \ldots, v_{\sigma((r,c,l))})$, where $\sigma((r, c, l)) \leq \log(|(r, c, l)|)$, define $\psi((r, c, l), (v_1, \ldots, v_{\sigma((r,c,l))})) = (r, c, l')$ where $l' = (l_1, \ldots, l_{mn-k}, [v_1, v_1], \ldots, [v_{\sigma((r,c,l))}, v_{\sigma((r,c,l))}], l_{mn-k+\sigma((r,c,l))+1}, \ldots, l_{mn})$. It is apparent that $v$ is a partially-filled contingency table for $(r, c, l)$ iff $(v_{\sigma((r,c,l))+1}, \ldots, v_k)$ is a partially-filled contingency table for $\psi((r, c, l), (v_1, \ldots, v_{\sigma((r,c,l))}))$. Observe that if we choose a suitable integer encoding, then $|\psi((r, c, l), (v_1, \ldots, v_{\sigma((r,c,l))}))| \leq |(r, c, l)|$. We conclude the proof by noticing that $\psi$ and $\sigma$ are clearly polynomial-time computable. □

[6] It is an abuse of notation because $l_1, \ldots, l_{mn-k}$ are degenerate non-negative integer intervals and not non-negative integer values.

# Bibliography


[1] K. Al-Khnaifes and H. Sachs. Graphs, linear equations, determinants and the number of perfect matchings. In R. Bodendiek, editor, *Contemporary Methods in Graph Theory*, pages 47–71. Wissenschaftsverlag, Mannheim, 1990.

[2] D. Aldous. Random walks on finite groups and rapidly mixing Markov chains. In A. Dols and B Eckmann, editors, *Séminaire de Probabilités XVII 1981/82*, volume 986 of *Springer Lecture Notes in Mathematics*, pages 243–297. Springer, 1986.

[3] D. Aldous. On the Markov chain simulation method for uniform combinatorial distributions and simulated annealing. *Probability in the Engineering and Informational Sciences*, 1:33–46, 1987.

[4] D. Aldous and J. Fill. Chapter 4: Hitting and Convergence Time, and Flow Rate, Parameters for Reversible Markov Chains. Preprint of a book on Markov chains, October 11, 1994 version.

[5] N. Alon. Eigenvalues and expanders. *Combinatorica*, 6:83–96, 1986.

[6] N. Alon and V.D. Milman. $\lambda_1$, isoperimetric inequalities for graphs and superconcentrators. *J. Combin. Theory Ser. B*, 38:73–88, 1985.

[7] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer. Sur les inégalités de Sobolev logarithmiques. *Panoramas et Synthèses [Panoramas and Syntheses]*, 10:xvi+217pp., 2000. in French [Logarithmic Sobolev inequalities]. With a preface by Dominique Bakry and Michel Ledoux.

[8] C. Ané and M. Ledoux. On logarithmic Sobolev inequalities for continuous time random walks on graphs. *Probab. Theory Related Fields*, 116(4):573–602, 2000.

[9] Y. Azar, A.Z. Broder, and A.M. Frieze. On the Problem of approximating the Number of Bases of a Matroid. *Information Processing Letters*, 50:9–11, 1993.

[10] D. Bakry. L'hypercontractivité et son utilisation en théorie des semigroups. In *Ecole d'Été de Saint Flour 1992*, volume 1581 of *Lecture Notes in Mathematics*. Springer, Berlin, 1994.

[11] Franck Barthe. On a reverse form of the Brascamp-Lieb inequality. *Invent. Math.*, 134(2):335–361, 1998.

[12] Hans-Jochen Bartsch. *Taschenbuch mathematischer Formeln*. Verlag Harri Deutsch, Thun and Frankfurt/Main, 7th–13th edition, 1990.

[13] N. Berger, C. Kenyon, E. Mossel, and Y. Peres. Glauber dynamics on trees and hyperbolic graphs. Preprint 2003. Preliminary version: Kenyon, C., Mossel, E. and Peres, Y., Glauber Dynamics on Trees and Hyperbolic Graphs, *Proceedings of the 42nd Annual IEEE Symposium on Foundation of Computer Science* (FOCS), pages 568–578, 2001.

[14] S. Bobkov and M. Ledoux. From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities. *Geom. Funct. Anal.*, 10(5):1028–1052, 2000.

[15] B. Bollobás. *Combinatorics: set systems, hypergraphs, families of vectors, and combinatorial probability*, chapter 16. Cambridge University Press, 1986.

[16] A. Broder. How hard is it to marry at random? (On the approximation of the permanent). In *Proceedings of the 18th Annual ACM Symposium on Foundations of Computer Science (STOC)*, pages 50–58. ACM Press, 1986. Erratum in *Proceedings of the 20th Annual ACM Symposium on Foundations of Computer Science (STOC)*, page 551, 1988.

[17] A. Broder. Generating random spanning trees. In *Proceedings of the 30th Annual ACM Symposium on Foundations of Computer Science (STOC)*, pages 442–447. IEEE, 1989.

[18] R. Bubley and M. Dyer. Path coupling: a technique for proving rapid mixing in Markov chains. In *Proceedings of the 38th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 223–231. IEEE Computer Society Press, 1997.

[19] N. Cancrini, F. Martinelli, and C. Roberto. The logarithmic Sobolev constant of Kawasaki dynamics under a mixing condition revisited. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(4):385–436, 2002.

[20] N. Cancrini and C. Roberto. Logarithmic Sobolev constant for the dilute Ising lattice gas dynamics below the percolation threshold. *Stochastic Process. Appl.*, 102(2):159–205, 2002.

[21] P. Caputo and F. Martinelli. Asymmetric diffusion and the energy gap above the 111 ground state of the quantum $XXZ$ model. *Communications in Mathematical Physics*, 226(2):323–375, 2002.

[22] F. Cesi. Quasi–factorization of the entropy and logarithmic Sobolev inequalities for Gibbs random fields. *Probab. Theory Related Fields*, 120(4):569–584, 2001.

[23] F. Chatelin. *Eigenvalues of Matrices*. Wiley, 1993.

[24] L. Chávez Lomelí and D.J.A. Welsh. Randomized Approximation of the Number of Bases. In J.E. Bonin, J.G. Oxley, and B. Servatius, editors, *Matroid Theory*, volume 197 of *Contemporary Mathematics*, pages 371–376. American Mathematical Society, Providence, Rhode Island, 1996.

[25] J. Cheeger. A lower bound for the smallest eigenvalue of the Laplacian. In R.C. Gunning, editor, *Problems in Analysis*, pages 195–199. Princeton University Press, New Jersey, 1970.

[26] G.-Y. Chen and Y.-C. Sheu. On the log-Sobolev constant for the simple random walk on the $n$-cycle: the even cases. *J. Funct. Anal.*, 202(2):473–485, 2003.

[27] Fan Chung. Logarithmic Sobolev techniques for random walks on graphs. In D.A. Hejhal et al., editors, *Emerging Applications of Number Theory*, volume 109 of *IMA Volumes in Math. and its Applications*, pages 175–186. Springer, 1999.

[28] M. Ciucu and J. Propp. Preprint, 1996.

[29] P. Diaconis. *Group Representations in Probability and Statistics*, volume 11 of *IMS Lecture Notes - Monograph Series*. Institute of Mathematical Statistics, Hayward CA, 1988.

[30] P. Diaconis and L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.*, 6(3):695–750, 1996.

[31] P. Diaconis and D. Stroock. Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.*, 1(1):36–61, 1991.

[32] J. Dodziuk. Difference equations, isoperimetric inequality and transience of certain random walks. *Trans. Amer. Math. Soc.*, 284:787–794, 1984.

[33] W. Doeblin. Exposé de la théorie des chaînes simples constants de Markoff à un nombre fini d'états. *Revue Math. de l'Union Interbalkanique*, 2:77–105, 1938.

[34] J.L. Doob. *Measure theory*, volume 143 of *Graduate texts in mathematics*. Springer-Verlag, New York; London, 1994.

[35] R. Durrett. *Probability: Theory and Examples*. Wadsworth, Pacific Grove, CA, 1991.

[36] M. Dyer and A. Frieze. Computing the volume of convex bodies: a case where randomness provably helps. In *Probabilistic Combinatorics and its Applications*, volume 44 of *Proceedings of AMS Symposia in Applied Mathematics*, pages 1–16, 1991.

[37] M. Dyer and A. Frieze. Random walks, totally unimodular matrices and a randomised dual simplex algorithm. *Mathematical Programming*, pages 1–16, 1994.

[38] M. Dyer, A. Frieze, and R. Kannan. A random polynomial time algorithm for approximating the volume of convex bodies. *Journal of the ACM*, 38:1–17, 1991.

[39] M. Dyer and C. Greenhill. A more rapidly mixing Markov chain for graph colourings. *Random Structures and Algorithms*, 13:285–317, 1998.

[40] M. Dyer and C. Greenhill. Random walks on combinatorial objects. In J.D. Lamb and D.A. Preece, editors, *Surveys in Combinatorics*, volume 267 of *London Mathematical Society Lecture Notes Series*, pages 101–136. Cambridge University Press, Cambridge, 1999.

[41] T. Feder and M. Mihail. Balanced matroids. In *Proceedings of the 24th Annual ACM Symposium on Theory of Computing (STOC)*, pages 26–38. ACM Press, 1992.

[42] J.A. Fill. Eigenvalue Bounds on the Convergence to Stationarity for nonreversible Markov Chains, with an Application to the Exclusion Process. *Annals of Applied Probability*, 1(1):62–87, 1991.

[43] I. Gessel and G. Viennot. Binomial determinants, paths and hook length formulae. *Advances in Mathematics*, 58:300–321, 1985.

[44] G. Grimmet and D. Stirzaker. *Probability and Random Processes*. Oxford University Press, Oxford, third edition, 2001.

[45] L. Gross. Logarithmic Sobolev inequalities. *American Journal of Mathematics*, 97:1061–1083, 1976.

[46] L. Gross. *Logarithmic Sobolev inequalities and contractivity properties of semigroups*, volume 1563 of *Lecture Notes in Mathematics*. Springer, Berlin, 1993.

[47] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization*. Springer Verlag, 1988.

[48] A. Guénoche. Random spanning tree. *Journal of Algorithms*, 4:214–220, 1983. In French.

[49] A. Guionnet and B. Zegarliński. Lectures on logarithmic Sobolev inequalities. In *Séminaire de Probabilités XXXVI*, volume 1801 of *Springer Lecture Notes in Mathematics*, pages 1–134. Springer, Berlin, 2003.

[50] L.H. Harper. Optimal numberings and isoperimetric problems on graphs. *J. Combin. Theory*, 1:385–393, 1966.

[51] W. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109, 1970.

[52] O.J. Heilmann and E.H Lieb. Theory of Monomer-Dimer Systems. *Comm. in Mathematical Physics*, 25:190–232, 1972.

[53] I. Heller. On linear systems with integral valued solutions. *Pacific Journal of Mathematics*, 7:1351–1364, 1957.

[54] R. Horn and C. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[55] C. Houdré. Mixed and isoperimetric estimates on the log-Sobolev constants of graphs and Markov chains. *Combinatorica*, 21(4):489–513, 2001.

[56] M. Jerrum. Mathematical foundations of the Markov chain Monte Carlo method. In M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, and B. Reed, editors, *Probabilistic Methods for Algorithmic Discrete Mathematics*, pages 116–159. Springer Verlag, Berlin, Heidelberg, 1998.

[57] M. Jerrum. *Counting, Sampling and Integrating: Algorithms and Complexity*. Lectures in Mathematics — ETH Zürich. Birkhäuser, Basel, 2002.

[58] M. Jerrum and A. Sinclair. Approximating the permanent. *SIAM Journal on Computing*, 18:1149–1178, 1989.

[59] M. Jerrum and A. Sinclair. The Markov chain Monte Carlo method: an approach to approximate counting and integration. In Dorit Hochbaum, editor, *Approximation Algorithms for NP-hard Problems*, pages 482–520. PWS, 1996.

[60] M. Jerrum, A. Sinclair, and E. Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries. *ECCC Report TR00-079*, 2000. Also in Proceedings of the 33rd Annual ACM Symposium on Theory of Computing (STOC), 2001.

[61] M. Jerrum and J.-B. Son. Spectral gap and log-Sobolev constant for balanced matroids. In *Proceedings of the 43rd Annual IEEE Symposium on Foundation of Computer Science (FOCS)*, pages 721–729. IEEE, 2002.

[62] M. Jerrum, J.-B. Son, P. Tetali, and E. Vigoda. Elementary bounds on Poincaré and log-Sobolev constants for decomposable Markov chains. *Isaac Newton Institute Preprint NI03006-CMP*, 2003.

[63] M. Jerrum, L. Valiant, and V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Science*, 43:169–188, 1986.

[64] W. Jockusch, J. Propp, and P. Shor. Random Domino Tilings and the Arctic Circle Theorem. Preprint, 1995.

[65] P. John and H. Sachs. Calculating the number of perfect matchings and of spanning trees, Pauling's orders, the characteristic polynomial and the eigenvectors of a benzenoid system. In M. Dewar et al., editors, *Advances in the Theory of Benzenoid Hydrocarbons*, volume 153 of *Topics in Current Chemistry*, pages 145–179. Springer Verlag, Berlin, 1990.

[66] R. Kannan. Markov chains and polynomial time algorithms. In *Proceedings of the 35th Annual IEEE Symposium on Foundation of Computer Science (FOCS)*, pages 656–671. IEEE Computer Society Press, 1994.

[67] R. Kannan. Rapid Mixing in Markov Chains. In *Proceedings of the International Congress of Math.*, volume 3, pages 673–683, 2002.

[68] R. Kannan and L. Lovász. Faster mixing via average conductance. In *Proceedings of the 31st Annual ACM Symposium on Theory of Computing (STOC)*, pages 282–287. ACM Press, 1999.

[69] R. Karp and M. Luby. Monte-Carlo algorithms for enumeration and reliability problems. In *Proceedings of the 24th Annual IEEE Symposium on Foundation of Computer Science (FOCS)*, pages 56–64. IEEE Computer Society Press, 1983.

[70] C. Kenyon, E. Mossel, and Y. Peres. Glauber dynamics on trees and hyperbolic graphs. In *Proceedings of the 42nd Annual IEEE Symposium on Foundation of Computer Science (FOCS)*, pages 568–578. IEEE Computer Society Press, 2001.

[71] C. Kenyon, D. Randall, and A. Sinclair. Approximating the number of dimer coverings of a lattice. *Journal of Statistical Physics*, 83:637–659, 1996.

[72] C. Kipnis, S. Olla, and S.R.S. Varadhan. Hydrodynamics and Large Deviation for Simple Exclusion Processes. *Communications on Pure and Applied Mathematics*, 42(2):115–137, 1989.

[73] A. Korzeniowski and D. Stroock. An example in the theory of hypercontractive semigroups. *Proceedings of the American Mathematical Society*, 94:87–90, 1985.

[74] C. Landim, G. Panizo, and H.T. Yau. Spectral gap and logarithmic Sobolev inequality for unbounded conservative spin systems. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(5):739–777, 2002.

[75] S. Lang. *Linear Algebra*. Springer-Verlag, New York, third edition, 1993.

[76] G.F. Lawler and A.D. Sokal. Bounds on the $L_2$ spectrum for Markov chains and Markov processes: A generalization of Cheeger's inequality. *Trans. Amer. Math. Soc.*, 309:557–580, 1988.

[77] M. Ledoux. Concentration of measure and logarithmic Sobolev inequalities. In *Séminaire de Probabilités XXXIII*, volume 1709 of *Springer Lecture Notes in Mathematics*, pages 120–216. Springer, 1999.

[78] M. Ledoux. Logarithmic Sobolev inequalities for unbounded spin systems revisited. In *Séminaire de Probabilités XXXV*, volume 1755 of *Springer Lecture Notes in Mathematics*, pages 167–194. Springer, Berlin, 2001.

[79] Tzong-Yow Lee and Horng-Tzer Yau. Logarithmic Sobolev Inequality for some Models of Random Walks. *The Annals of Probability*, 26(4):1855–1873, 1998.

[80] T. Lindvall. *Lectures on the Coupling Method.* Wiley, New York, 1992.

[81] L. Lovász. Random Walks on Graphs: A Survey. In D. Miklós, Sós V.T., and T. Szönyi, editors, *Combinatorics, Paul Erdös is Eighty*, volume 2, pages 353–398. Bolyai Society Mathematical Studies, Budapest, 1996.

[82] L. Lovász and P. Winkler. Efficient Stopping Rules for Markov Chains. In *Proceedings of the 27th Annual ACM Symposium on Theory of Computing (STOC)*, pages 76–82. ACM Press, 1995.

[83] S.-L. Lu and H.-T. Yau. Spectral gap and logarithmic Sobolev inequality for Kawasaki and Glauber dynamics. *Communications in Mathematical Physics*, 156(2):399–433, 1993.

[84] M. Luby, D. Randall, and A. Sinclair. Markov chain algorithms for planar lattice structures. *SIAM Journal on Computing*, 31(1):167–192, 2001.

[85] N. Madras and D. Randall. Markov chain decomposition for convergence rate analysis. *Annals of Applied Probability*, 12:581–606, 2002.

[86] R. Martin and D. Randall. Sampling adsorbing staircase walks using a new Markov chain decomposition method. In *Proceedings of the 41st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2000.

[87] F. Martinelli, A. Sinclair, and D. Weitz. Fast mixing for independent sets, colorings and other model on trees. Preprint 2003, extended abstract to appear in ACM-SIAM SODA 2004.

[88] F. Martinelli, A. Sinclair, and D. Weitz. The Ising model on trees: Boundary conditions and mixing time. submitted to *Communications in Mathematical*

*Physics*, extended abstract appeared in *Proceedings of the 42nd Annual IEEE Symposium on Foundation of Computer Science* (FOCS), pages 628–639, 2003.

[89] P. Mathieu. Log-Sobolev and spectral gap inequalities for the knapsack Markov chain. *Markov Process. Related Fields*, 8(4):595–610, 2002.

[90] N. Metropolis, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equations of state calculation by fast computing machines. *Journal of Chemical Physics*, 21:1087–1091, 1953.

[91] M. Mihail. *Combinatorial Aspects of Expanders*. PhD thesis, Harvard University, Dept. of Computer Science, 1989.

[92] M. Mihail and M. Sudan. Connectivity properties of matroids. Technical report, U.C. Berkeley, 1989.

[93] M. Mihail and U. Vazirani. On the expansion of $0-1$ polytopes. Technical Report 3, Aitken Laboratory, Harvard University, 1989.

[94] M. Mihail and P. Winkler. On the number of Eulerian orientations of a graph. In *Proceedings of the Third Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 138–145. ACM-SIAM, 1992.

[95] R. Montenegro. personal communication.

[96] R. Montenegro and J.-B. Son. Edge isoperimetry and rapid mixing on matroids and geometric Markov chains. In *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing (STOC)*, pages 704–711. ACM Press, 2001.

[97] B. Morris and Y. Peres. Evolving sets and mixing. In *Proceedings of the 35th Annual ACM Symposium on Theory of Computing (STOC)*. ACM Press, 2003.

[98] J.R. Norris. *Markov Chains*. Cambridge University Press, 1997.

[99] J.G. Oxley. *Matroid theory*. Oxford University Press, New York, 1992.

[100] J. Propp and D. Wilson. Exact Sampling with coupled Markov chains and Applications to Statistical Mechanics. *Random Structures and Algorithms*, 9:223–252, 1996.

[101] J. Propp and D. Wilson. How to Get a Perfectly Random Sample From a Generic Markov Chain and Generate a Random Spanning Tree of a Directed Graph. *Journal of Algorithms*, 27:170–217, 1998.

[102] D. Randall and P. Tetali. Analyzing Glauber dynamics by comparison of Markov chains. *Journal of Mathematical Physics*, 41:1598–1615, 2000.

[103] C. Roberto. A path method for the logarithmic Sobolev constant. *Combin. Probab. Comput.*, 12(4):431–455, 2003.

[104] G.C. Robinson and D.J.A. Welsh. The computational Complexity of Matroid Properties. *Mathematical Proceedings of the Cambridge Philosophical Society*, 87:29–45, 1980.

[105] O. Rothaus. Diffusion on compact Riemannian manifolds and logarithmic Sobolev inequalities. *Journal of Functional Analysis*, 42:102–109, 1981.

[106] C. P. Schnorr. Optimal algorithms for self-reducible problems. In S. Michaelson and R. Milner, editors, *Proceedings of the 3rd International Colloquium on Automata, Languages and Programming*, pages 322–337. Edinburgh University Press, 1976.

[107] E. Seneta. *Non-Negative Matrices: An Introduction to Theory and Applications.* John Wiley and Sons, New York, 1973.

[108] E. Seneta. *Non-negative Matrices and Markov Chains.* Springer Series in Statistics. Springer Verlag, New York, second edition, 1981.

[109] J.C. Simon. *Patterns and Operators: The Foundations of Data Representation.* North Oxford Academic, 1986.

[110] A. Sinclair. Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combinatorics, Probability and Computing*, 1:351–370, 1992.

[111] A. Sinclair and M. Jerrum. Approximate counting, uniform generation and rapidly mixing Markov chains. *Information and Computation*, 82:93–133, 1989.

[112] G.W. Stewart. *Matrix Algorithms*, volume 1: Basic Decomposition. SIAM, 1998.

[113] D. Stroock and B. Zegarliński. The logarithmic Sobolev inequality for discrete spin systems on a lattice. *Comm. Math. Phys.*, 149(1):175–193, 1992.

[114] D.J.A. Welsh. *Matroid theory*. Academic Press, London, 1976.

[115] D.J.A. Welsh. Approximate Counting. In R.A. Bailey, editor, *Surveys in Combinatorics*, volume 241 of *London Mathematical Society Lecture Notes Series*, pages 287–323. Cambridge University Press, 1997.

[116] D. Wilson. Determinant algorithms for random planar structures. In *Proceedings of the Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 258–267. ACM-SIAM, 1997.

[117] D. Wilson. Mixing times of lozenge tiling and card shuffling Markov chains. *Annals of Applied Probability*, to appear.

[118] N. Yoshida. The equivalence of the log-Sobolev inequality and a mixing condition for unbounded spin systems on the lattice. *Ann. Inst. H. Poincaré Probab. Statist.*, 37(2):223–243, 2001.

# Index