Spectral Gap and log–Sobolev Constant for Balanced Matroids

Mark Jerrum*
Division of Informatics
University of Edinburgh
The King's Buildings
Edinburgh, EH9 3JZ, UK
mrj@dcs.ed.ac.uk

Jung–Bae Son†
Division of Informatics
University of Edinburgh
The King's Buildings
Edinburgh, EH9 3JZ, UK
j.b.son@sms.ed.ac.uk

We compute tight lower bounds on the log–Sobolev constant of a class of inductively defined Markov chains, which contains the bases–exchange walks for balanced matroids studied by Feder and Mihail. As a corollary, we obtain improved upper bounds for the mixing time of a variety of Markov chains. An example: the "natural" random walk on spanning trees of a graph G as proposed by Broder — which has been studied by a number of authors — mixes in time $O(mn \log n)$, where n is the number of vertices of G and m the number of edges. This beats the best previous upper bound on this walk by a factor n^2.

1 Introduction

The most successful approach in recent years to sampling combinatorial structures (and hence estimating their number) has been Markov chain simulation. This technique is the basis of the so–called Markov chain Monte Carlo method. In order to obtain useful performance guarantees, it is important to have tight bounds on the convergence rates of the Markov chains in question to equilibrium. Our interest in this article is in a class of inductively defined Markov chains, which we call "π–recursive". This class will be defined precisely in Section 3, but, roughly speaking, a π–recursive Markov chain is one that is either trivial or can be decomposed into two π–recursive (sub-) Markov chains, such that the transitions between the two chains support a matching of a certain kind.

The paradigmatic example of a π–recursive Markov chain is the Boolean cube, but the class extends well beyond this simple case to cover, for example, the "bases–exchange walk" (Alg. 6.1) on balanced matroids studied by Feder and Mihail [7]. The main results of this paper are lower bounds on the spectral gap (Theorem 4.4) and log–Sobolev constant (Theorem 5.4) of π–recursive Markov chains. Our approach yields a tight bound on spectral gap for the random walk on the Boolean cube (cf. Example 6.3). Furthermore, our bound on log–Sobolev constant for the cube is within a factor 4 of the spectral gap. Thus, owing to a result by Rothaus [19], our bound on the log–Sobolev constant is tight up to a constant factor.

Both spectral gap and log–Sobolev constant of the cube were already known by other methods [5]; however, the advantage of our "hands–on" approach is that it is robust and extends to π–recursive Markov chains in general. Since the log–Sobolev constant is more tightly related to mixing time, as measured by convergence to stationary in total variation distance, we obtain mixing times that are typically about $\widetilde{O}(k)$ shorter than those currently known. Here k is the roughly the "dimension" of the Markov chain, or the depth of recursion defining it.

As a specific example, consider the "natural" random walk on spanning trees of a graph G as defined by Broder [3]: Suppose the current spanning tree (state) is T; choose an edge $e \in E(G)$ and another $f \in T$ uniformly at random (u.a.r.); if $T' = T \cup \{e\} \setminus \{f\}$ is a spanning tree then T' is the new state, otherwise we remain at T. (This is a special case of the bases–exchange walk for matroids mentioned earlier.) We show that the mixing time of this random walk is $O(mn \log n)$, where n is the number of vertices of G and m the number of edges. As far as we are aware, the best previously known bound[1] was $O(mn^3 \log m)$, established by Feder and Mihail.

The rest of the paper is organised as follows. In Section 2 we give definitions of spectral gap and log–Sobolev constants and how they result in bounds on mixing time. Section 3 defines the above mentioned π–recursive Markov chains. In Section 4 lower bounds on the spectral gap of

*supported by EPSRC grant "Sharper Analysis of Randomized Algorithms: a Computational Approach" and the IST Programme of the EU under contract IST-1999-14036 (RAND-APX)

†supported by a DAAD grant

[1] Note that this is the best result for *general* m and n. For particular m and n there exist better results, e.g. [18]. Also note that the *modified* random walk by Feder and Mihail can mix faster.

such Markov chains are derived and in Section 5 lower bounds on the log–Sobolev constant. Section 6 contains applications of the results found in Section 4 and 5.

2 Spectral gap, log–Sobolev constants and mixing time

Unless stated otherwise, let throughout the paper $M = (X_0, X_1, X_2, \dots)$ be a time–reversible, irreducible and aperiodic Markov chain on finite state space Ω with transition kernel $P(x,y)$. Since M is finite, irreducible and aperiodic, it is also ergodic and has a unique stationary distribution π. Furthermore,

$$P^t(x,y) = \Pr(X_t = y \mid X_0 = x),$$

the probability that after t steps the chain M is at state y if it was initially at state x, converges toward the stationary distribution, i.e. $\lim_{t\to\infty} P^t(x,y) \to \pi(y)$ for all $x,y \in \Omega$. For the Markov chain Monte Carlo method to be of any practical significance, the Markov chain M has to converge quickly towards its stationary distribution. The earliest techniques for bounding the rate of convergence (or mixing time) were coupling [1] and conductance [14]. Since then, an abundance of work has been carried out and we refer readers to survey articles such as [16] and [12] and the references given therein. This might be particularly helpful to readers unfamiliar with this area since here, we will only mention concepts immediately relevant for this paper.

Given two probability distributions μ and ζ on state space Ω their *total variation distance* is

$$\|\mu - \zeta\|_{\mathrm{TV}} = \frac{1}{2}\sum_{x\in\Omega} |\mu(x) - \zeta(x)|\,.$$

There are various definitions of *mixing time* τ. Here let it be defined as

$$\tau = \max_{x\in\Omega} \min\left\{t > 0 : \|P^t(x,\cdot) - \pi\|_{\mathrm{TV}} \le e^{-1}\right\}.$$

Among the various techniques for bounding mixing time are spectral gap and logarithmic–Sobolev constants. The logarithmic Sobolev constant of a chain M is defined as

$$\alpha_M = \inf_f \left\{\frac{\mathcal{E}_M(f,f)}{\mathcal{L}_\pi(f)} : \mathcal{L}_\pi(f) \ne 0\right\},$$

where $f : \Omega \to \mathbb{R}$,

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

is the Dirichlet–form with respect to the Markov chain M with transition kernel P and stationary distribution π, and

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

is an entropy like quantity, with

$$\|f\|_{2,\pi} = \sqrt{\sum_{x\in\Omega} |f(x)|^2 \pi(x)}.$$

Formally, the log–Sobolev constant resembles the definition of the *spectral gap*

$$\lambda_M = \inf_f \left\{\frac{\mathcal{E}_M(f,f)}{\mathrm{Var}_\pi(f)} : \mathrm{Var}_\pi(f) \ne 0\right\},$$

$f : \Omega \to \mathbb{R}$, of the chain, where $\mathrm{Var}_\pi(f)$ is variance with respect to (w.r.t.) probability distribution π. But while spectral gap can be regarded as the difference between the first two largest eigenvalues of the transition kernel $P(x,y)$, there seems to be no analogous interpretation for log–Sobolev constants. The bounds on mixing time using spectral gap and log–Sobolev constants [5] are:

$$\tau \le \frac{2 + \log[1/\pi_*]}{2\lambda}$$

and

$$\tau \le \frac{4 + \log\log[1/\pi_*]}{4\alpha}, \tag{1}$$

where $\pi_* = \min_{x\in\Omega}\{\pi(x)\}$ (we dropped the subscripts of α and λ as they were inessential). Suppose that n measures input size, then typically $1/\pi_* = 2^{O(n)}$; thus using (1) gains a factor of $n/\log n$ whenever λ and α are of similar magnitude. It has been shown that $\alpha \le \lambda/2$ (see [19], [5]). However, in contrast to spectral gap, log–Sobolev constants are notoriously hard to compute. But in the few cases where both spectral gap and log–Sobolev constant are explicitly known the above bound has proven to be amazingly tight: for the symmetric two–point space, the Ornstein–Uhlenbeck process on $\mathbb{R}$ or the standard diffusion on the n–sphere $\alpha = \lambda/2$ (see [2], [8], [9]). A diffusion where $\alpha = \lambda/4$ is given in [15]. On the other hand, cases are known where $\alpha \ll \lambda$. Examples are random walks on generic regular r–graphs (see [5]) or complete graphs (see [11]). Very recent work [11] shows how quantities that are easier to compute, e.g. spectral gap or conductance, can be used to lower bound log–Sobolev constants. Here, we will directly compute the spectral gap and log–Sobolev constant of π–recursive Markov chains and show that our bound on the log–Sobolev constant is sometimes tight up to a constant factor. This is achieved by splitting the chain up into two sub–chains and applying induction to obtain the desired result. Other work where decomposition yields bounds on spectral gap and log–Sobolev constant is [4] and [17].

3 π–recursive Markov chains

Let $G = (V,E)$ be an undirected graph. The subgraph G_S of G induced by a subset $S \subseteq V$ has vertex set S and edges $E_S = \{\{x,y\} \in E \mid x,y \in S\}$.

Let $M = (\Omega, P)$ be a Markov chain. Recall that we stipulated that M be finite, time–reversible, irreducible and aperiodic. Hence, it converges towards a unique stationary distribution, say π. For $S \subseteq \Omega$ the *restriction chain* of M on S is the Markov chain $M_S = (S, P_S)$ with

$$P_S(x,y) = \begin{cases} P(x,y) & \text{if } x, y \in S,\ x \neq y \\ 1 - \sum_{\substack{y \in S \\ y \neq x}} P(x,y) & \text{if } x = y. \end{cases}$$

Notice that $M = M_\Omega$. Since M is time–reversible, it fulfils the *detailed balance* condition, which says that

$$\pi(x)P(x,y) = \pi(y)P(y,x) \text{ for all } x, y \in \Omega.$$

This condition is extremely useful because if a probability distribution μ meets detailed balance, then μ is a stationary distribution. Now observe that if M_S itself is irreducible, then its unique stationary distribution is π_S, where $\pi_S(x) = \pi(x)/\pi(S)$ for $x \in S$ and $\pi(S) = \sum_{x \in S} \pi(x)$; this is because π_S fulfils detailed balance (or equivalently M_S is time–reversible with respect to π_S).

Observe that M is equivalent to a random walk on an undirected graph $G = (V, E)$ with $V = \Omega$ and $E = \{\{x, y\} \mid P(x, y) > 0\}$ where for current vertex x the subsequent vertex y is chosen with probability $P(x, y)$. Call G the *underlying graph of* M. Let $A \neq \emptyset$, $\bar{A} \neq \emptyset$ be a bipartition of Ω. The set of *crossedges* $C(A, \bar{A}) \subseteq E$ is the set of edges with one endpoint in A and the other in $\bar{A}$. Any assignment $w : C(A, \bar{A}) \to \mathbb{R}^+$ satisfying

$$\sum_{y \in \bar{A}} w(x,y) = \pi_A(x) \quad \text{and} \quad \sum_{x \in A} w(x,y) = \pi_{\bar{A}}(y)$$

is a π*–matching.* The graph G is called π*–matchable* if there exists a bipartition A, $\bar{A}$ that allows a π–matching. The breaking up of G into the disjoint subgraphs G_A and $G_{\bar{A}}$ such that A, $\bar{A}$ allows a π–matching is referred to as 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. A Markov chain M is π*–recursive on* G if:

1. Ω is of size 1, or

2. G is π–matchable and if G is split into, say G_A and $G_{\bar{A}}$, then M_A is π_A–recursive on G_A and $M_{\bar{A}}$ is $\pi_{\bar{A}}$–recursive on $G_{\bar{A}}$.

A prototypical example is the random walk on the Boolean cube.

4 Spectral gap of π–recursive Markov chains

Our ultimate goal is to establish a lower bound on the log–Sobolev constant. However our approach is more easily understood in the context of spectral gap. So we consider this parameter first, on the one hand as a warm up exercise, on the other as a benchmark for our bound on the log–Sobolev constant.

Suppose that $M = (\Omega, P)$ is π–recursive on an underlying graph $G = (\Omega, E)$. The main result of this section is a lower bound on the spectral gap of M (Theorem 4.4). The idea of the proof is to find a value $\hat{\lambda}_M$ s.t. $\mathcal{E}_M(f, f) \geq \hat{\lambda}_M \mathrm{Var}_\pi(f)$ for all $f : \Omega \to \mathbb{R}$. Then, from the definition of spectral gap, $\lambda_M \geq \hat{\lambda}_M$ is an immediate consequence. We will find such a value $\hat{\lambda}_M$ by decomposition/induction.

The first ingredient is the observation that variance and Dirichlet form can be split up w.r.t. a partition of the state space:

Lemma 4.1 *Let A, $\bar{A}$ be a partition of Ω, $f : \Omega \to \mathbb{R}$ and π a probability distribution on Ω, 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

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

is a weighted crossterm,

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

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

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

Proof. Despite the fact that the lemma is standard knowledge now, we will include a proof for the sake of completeness. A proof of

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

can be found in [13]. Now, observe that

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

□

Similarly,

Lemma 4.2 *For a partition $A,\ \bar{A}$ of Ω, $f:\Omega\to\mathbb{R}$ and a probability distribution π on Ω*

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

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

Proof. As before, the lemma is standard and the proof included for the sake of completeness. 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 following lemma is the main stepping stone for the induction in Theorem 4.4.

Lemma 4.3 *Assume that $M=(\Omega,P)$ is a finite, time–reversible, irreducible and aperiodic Markov chain with stationary distribution π on a π–matchable graph $G=(\Omega,E)$. Then*

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

Proof. Since G allows a π–matching, there exists a bipartition $A,\ \bar{A}$ and 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:

$$\begin{aligned}
&\mathrm{ct}_{\mathrm{Var}_\pi}(f)\\
&=\pi(A)\pi(\bar{A})\Big(\sum_{\substack{x\in A\\ y\in\bar{A}}}\sqrt{w(x,y)}(f(x)-f(y))\sqrt{w(x,y)}\Big)^2.
\end{aligned}$$

Observe that $\sum_{\substack{x\in A\\ y\in\bar{A}}}w(x,y)=1$. Applying the Cauchy–Schwarz inequality yields:

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

Thus, to complete the proof it suffices to find a factor η such that

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

Observe that since M is time–reversible, we can always 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

$$\begin{aligned}
\pi(A)\pi(\bar{A})w(x,y)&\leq\pi(A)\pi(\bar{A})\frac{\pi(x)}{\pi(A)}\\
&\leq\frac{1}{2}\pi(x)\leq\frac{\pi(x)P(x,y)}{2\min_{\{x,y\}\in E}\{P(x,y)\}}.
\end{aligned}$$

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

The significance of Lemma 4.3 is that $2\min_{\{x,y\}\in E}\{P(x,y)\}$ is exactly the value $\hat{\lambda}_M$ we were looking for, i.e. it satisfies $\mathcal{E}_M(f,f)\geq\hat{\lambda}_M\mathrm{Var}_\pi(f)$ for all $f:\Omega\to\mathbb{R}$. This is made rigorous next.

Theorem 4.4 *Given a finite, time–reversible, irreducible and aperiodic Markov chain $M=(\Omega,P)$ with stationary distribution π on a graph $G=(\Omega,E)$ such that M is π–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 π_S. Since M is π–recursive on G, the chain M_S is π_S–recursive on G_S. Thus there is a split of G_S into, say G_A, $G_{\bar{A}}$. By Lemma 4.2: $\mathcal{E}_{M_S}(f,f) = \pi_S(A)\mathcal{E}_{M_A}(f,f) + \pi_S(\bar{A})\mathcal{E}_{M_{\bar{A}}}(f,f) + \mathrm{ct}_{\mathcal{E}_{M_S}}(f)$ which is by induction hypothesis greater than $2\min_{\{x,y\}}\{P(x,y)\} \times \big(\pi_S(A)\mathrm{Var}_{\pi_A}(f) + \pi_S(\bar{\mathrm{A}})\mathrm{Var}_{\pi_{\bar{A}}}(f)\big) + \mathrm{ct}_{\mathcal{E}_{M_S}}(f)$. By Lemma 4.3: $\mathrm{ct}_{\mathcal{E}_{M_S}}(f) \geq 2\min_{\{x,y\}\in E}\{P(x,y)\} \times \mathrm{ct}_{\mathrm{Var}_{\pi_S}}(f)$. Thus, applying Lemma 4.1 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 4.4 (and later Theorem 5.4) 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 δ is some small value much less than $\min_{\{x,y\}\in E}\{P(x,y)\}$. (The stationary distribution π 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 λ_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 4.3 and Theorem 4.4. Both continue to hold if $2\min_{\{x,y\}\in E}\{P(x,y)\}$ is replaced by

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

Thus, the tiny transition probability can be compensated for as long as the π–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 4.3 and Theorem 4.4 involving (2) 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 4.3 and Theorem 4.4 are sufficiently well suited.

5 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 π–recursive Markov chains. As before, we assume that $M = (\Omega, P)$ is π–recursive on the underlying graph $G = (\Omega, E)$ (Theorem 5.4). Again, we will identify a value $\hat{\alpha}_M$ such that if G_S is obtained by a series of splits: $\mathcal{E}_{M_S}(f,f) \geq \hat{\alpha}_M \mathcal{L}_{\pi_S}(f)$ for all $f : \Omega \to \mathbb{R}$. The approach is along the same lines as the proof of Theorem 4.4. 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 Ω.

Lemma 5.1

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

where

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

is a weighted crossterm,

$$\|f\|_{2,\pi_A} = \sqrt{\sum_{x\in A}|f(x)|^2\,\pi_A(x)}$$

and

$$\mathcal{L}_{\pi_A}(f) = \sum_{x\in A}|f(x)|^2\,\pi_A(x)\log\frac{|f(x)|^2}{\|f\|^2_{2,\pi_A}}$$

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

Proof. Again, the proof is standard and included merely for the sake of completeness. Recall that

$$\mathcal{L}_\pi(f) = \sum_{x\in\Omega}|f(x)|^2\,\pi(x)\log\frac{|f(x)|^2}{\|f\|^2_{2,\pi}}. \tag{3}$$

Splitting up (3) with respect to A, $\bar{A}$ yields

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

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

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

Then

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

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

Section 4 related the crossterms of Dirichlet form to variance. Now, we have to relate the crossterms of Dirichlet form and $\mathcal{L}_\pi(f)$.

Lemma 5.2 *Assume that $M = (\Omega, P)$ is a finite, time–reversible, irreducible and aperiodic Markov chain with stationary distribution π on a π–matchable graph $G = (\Omega, E)$. Then*

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

Proof. Let A, $\bar{A}$ be a split of G. Using $\ln x \leq x - 1$ we obtain

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

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

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

It is then enough to show that

$$\begin{aligned}\mathrm{ct}_{\mathcal{E}_M}(f) \geq \frac{1}{2}\min_{\{x,y\}\in E}\{P(x,y)\} \\ \times\,\pi(A)\pi(\bar{A})\frac{(\|f\|^2_{2,\pi_A} - \|f\|^2_{2,\pi_{\bar{A}}})^2}{\|f\|^2_{2,\pi}}.\end{aligned}$$

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

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

and the second one

$$\begin{aligned}\sum_{\substack{x\in A\\ y\in\bar{A}}}(f(x)-f(y))^2 w(x,y) \\ \geq \pi(A)\pi(\bar{A})\frac{(\|f\|^2_{2,\pi_A} - \|f\|^2_{2,\pi_{\bar{A}}})^2}{\|f\|^2_{2,\pi}}. \qquad (4)\end{aligned}$$

As regards the first inequality, observe the following. Since M is time–reversible w.r.t. π, we can choose A such that $\pi(\bar{A}) \leq \pi(A)$ without affecting the very first inequality which upper bounds $\mathrm{ct}_{\mathcal{L}_\pi}(f)$ and remember that M allows a π–matching between A and $\bar{A}$ such that $w(x,y) \leq \sum_{y\in\bar{A}} w(x,y) = \frac{\pi(x)}{\pi(A)}$. 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 (4) is valid. Set $F_A = \sum_{x\in A}|f(x)|^2\pi(x)$ and $F_{\bar{A}}$ analogously, then the RHS turns into $\frac{\pi(A)\pi(\bar{A})}{(F_A+F_{\bar{A}})}\left(\frac{F_A}{\pi(A)} - \frac{F_{\bar{A}}}{\pi(\bar{A})}\right)^2$. Observe that

$$\begin{aligned}\sum_{\substack{x\in A\\ y\in\bar{A}}}(f(x)-f(y))^2 w(x,y) \\ = \frac{F_A}{\pi(A)} - 2\sum_{\substack{x\in A\\ y\in\bar{A}}} f(x)f(y)w(x,y) + \frac{F_{\bar{A}}}{\pi(\bar{A})}.\end{aligned}$$

According to the Cauchy–Schwarz inequality

$$\sum_{\substack{x\in A\\ y\in\bar{A}}} f(x)\sqrt{w(x,y)}f(y)\sqrt{w(x,y)} \leq \sqrt{\frac{F_A F_{\bar{A}}}{\pi(A)\pi(\bar{A})}}$$

and so

$$\sum_{\substack{x\in A\\ y\in\bar{A}}}(f(x)-f(y))^2 w(x,y) \geq \left(\sqrt{\frac{F_A}{\pi(A)}} - \sqrt{\frac{F_{\bar{A}}}{\pi(\bar{A})}}\right)^2.$$

It therefore suffices to show that

$$\left(\sqrt{\frac{F_A}{\pi(A)}} - \sqrt{\frac{F_{\bar{A}}}{\pi(\bar{A})}}\right)^2 \geq \frac{\pi(A)\pi(\bar{A})}{(F_A+F_{\bar{A}})}\left(\frac{F_A}{\pi(A)} - \frac{F_{\bar{A}}}{\pi(\bar{A})}\right)^2$$

which is equivalent to

$$\left(\sqrt{\frac{F_A}{\pi(A)}} - \sqrt{\frac{F_{\bar{A}}}{\pi(\bar{A})}}\right)^2 \geq \pi(A)\pi(\bar{A})\left(\frac{F_A}{\pi(A)} - \frac{F_{\bar{A}}}{\pi(\bar{A})}\right)^2,$$

where $F_A + F_{\bar{A}}$ has been normalised to 1. Below, in Lemma 5.3 we will show that the last inequality is valid, which concludes the proof of Lemma 5.2. □

Lemma 5.3 *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. The claimed inequality is equivalent to

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

and hence to

$$y(1-y)\left(\sqrt{\frac{x}{y}} + \sqrt{\frac{1-x}{1-y}}\right)^2 \leq 1. \tag{5}$$

Since $x, y \in (0,1)$, inequality (5) 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 *Given a finite, time–reversible, irreducible and aperiodic Markov chain $M = (\Omega, P)$ with stationary distribution π on a graph $G = (\Omega, E)$ such that M is π–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 4.4. 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 π_S be a restriction chain on underlying graph $G_S = (S, E_S)$ obtained by a series of splits of G.

If $|S| = 1$, then $\pi_S(x) = 1$ for $x \in S$, in which case $\frac{1}{2}\sum_{x,y \in S}(f(x) - f(y))^2 \pi_S(x) P_S(x,y) \geq \frac{1}{2} \times \min_{\{x,y\}\in E}\{P(x,y)\} \sum_{x \in S} |f(x)|^2 \pi_S(x) \log \frac{|f(x)|^2}{\|f\|^2_{2,\pi_S}} = 0$ is trivially true.

Inductive step: As M_S is π_S–recursive, the underlying graph G_S can be split into two non–empty sets A and $\bar{A}$ such that: $\mathcal{E}_{M_S}(f,f) = \pi_S(A)\mathcal{E}_{M_A}(f,f) + \pi_S(\bar{A})\mathcal{E}_{M_{\bar{A}}}(f,f) + \mathrm{ct}_{\mathcal{E}_{M_S}}(f)$. By induction hypothesis this is greater than $\frac{1}{2}\min_{\{x,y\}\in E}\{P(x,y)\}\big(\pi_S(A)\mathcal{L}_{\pi_A}(f) + \pi_S(\bar{A})\mathcal{L}_{\pi_{\bar{A}}}(f)\big) + \mathrm{ct}_{\mathcal{E}_{M_S}}(f)$. By Lemma 5.2 this in turn is greater than $\frac{1}{2}\min_{\{x,y\}\in E}\{P(x,y)\}\big(\pi_S(A)\mathcal{L}_{\pi_A}(f) + \pi_S(\bar{A})\mathcal{L}_{\pi_{\bar{A}}}(f) + \mathrm{ct}_{\mathcal{L}_{\pi_S}}(f)\big)$ and finally using Lemma 5.1: $\mathcal{E}_{M_S}(f,f) \geq \frac{1}{2}\min_{\{x,y\}\in E}\{P(x,y)\}\mathcal{L}_{\pi_S}(f)$. □

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 λ stated in Theorem 4.4. (This follows from the known inequality $\alpha \leq \lambda/2$ [19], [5]). But the bound in Theorem 4.4 is tight, e.g in the case of the hypercube.

6 Application

In this section we will discuss some applications of Theorem 4.4 and Theorem 5.4. The applications we have in mind are random walks on bases–exchange walks on balanced matroids.

6.1 Balanced matroids

Matroids are algebraic structures that capture the concept of linear independence. Formally, a pair $\mathcal{M} = (S, \mathcal{I})$ with a finite ground set S and independent sets $\mathcal{I} \subseteq 2^S$ forms a *matroid* if

1. $\emptyset \in \mathcal{I}$,
2. if $A \in \mathcal{I}$ and $B \subseteq A$ then $B \in \mathcal{I}$, and
3. if $A, B \in \mathcal{I}$ and $|B| = |A| + 1$ then there exists $e \in B \setminus A$ such that $A \cup \{e\} \in \mathcal{I}$.

A matroid can be entirely defined by its set of maximal independent sets or *bases* $\mathcal{B} \subseteq 2^S$ (sometimes we will use $\mathcal{B}(\mathcal{M})$ to denote the bases of a matroid $\mathcal{M}$ or conversely $\mathcal{M}(\mathcal{B})$ to denote the matroid with bases $\mathcal{B}$). Bases possess the following two properties:

1. all bases are of the same size, namely the *rank* of $\mathcal{M}$ and
2. for every pair of bases $X, Y \in \mathcal{B}$ and every element $e \in X$ there exists an element $f \in Y$ s.t. $(X \setminus \{e\}) \cup \{f\} \in \mathcal{B}$.

The properties of matroid bases suggest a very natural random walk on $\mathcal{B}$. Define the *bases–exchange graph* $G(\mathcal{M}) = (\mathcal{B}, E)$ as the undirected graph with vertex set $\mathcal{B}$ and two vertices $X, Y \in G(\mathcal{M})$ are connected by an edge if $|X \oplus Y| = 2$, 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*. As an example take the following random walk:

Algorithm 6.1 *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).

Next, we introduce two important operations on matroids. For a ground set element $e \in S$ the matroid $\mathcal{M} \setminus e$ obtained by *deleting* e has ground set $S(\mathcal{M} \setminus e) = S \setminus \{e\}$ and set of bases $\mathcal{B}(\mathcal{M} \setminus e) = \{X \in \mathcal{B} \mid e \notin X\}$. The matroid $\mathcal{M}/e$ obtained by *contracting* e has ground set $S(\mathcal{M}/e) = S \setminus \{e\}$ and set of bases $\mathcal{B}(\mathcal{M}/e) = \{X \setminus \{e\} \mid X \in \mathcal{B} \wedge e \in X\}$. Any matroid obtained from $\mathcal{M}$ by a series of contractions and deletions is a *minor* of $\mathcal{M}$.

Let X be a basis chosen uniformly at random (u.a.r.) from $\mathcal{B}$ and $e \in S$ a ground set element. With abuse of notation, let e denote the event $e \in X$ and $\bar{e}$ the event $e \notin X$; juxtaposition expresses the conjunction of events, i.e. $\bar{e}f$ means $e \notin X \wedge f \in X$. A matroid is *negatively correlated* if the inequality $\Pr[ef] \leq \Pr[e]\Pr[f]$ holds for all pairs of distinct elements $e, f \in S$. We say that 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 the notion of balance and showed that Algorithm 6.1 on balanced matroids has mixing time polynomial in the rank n and size of ground set m, i.e., is rapidly mixing (cf. [7]).

6.1.1 Fractional matchings

In the same paper ([7]), Feder and Mihail discovered that vertices of the bases–exchange graph $G(\mathcal{M})$ of every balanced matroid have a natural bipartition into A, $\bar{A}$, such that the transitions between A and $\bar{A}$ support a *fractional matching*. Let A and $\bar{A}$ be the natural isomorphic copies of $\mathcal{B}(\mathcal{M} \setminus e)$ and $\mathcal{B}(\mathcal{M}/e)$ in $\mathcal{M}$ respectively for some ground set element e. A fractional matching is a function $\psi : A \times \bar{A} \to \mathbb{N}$ assigning integer weights to the cutset (the edges of $G(\mathcal{M})$ spanning A and $\bar{A}$) such that

$$\forall x \in A : \sum_{y \in \bar{A}} \psi(x, y) = \left|\bar{A}\right|$$

and

$$\forall y \in \bar{A} : \sum_{x \in A} \psi(x, y) = |A| \,.$$

Note that if π is uniform over A and $\bar{A}$, then the probability distribution

$$P_\psi = \left((x, y) \mapsto \frac{\psi(x, y)}{|A|\,|\bar{A}|}\right) : A \times \bar{A} \to [0, 1]$$

on the cutset is a π–matching.

6.2 Spectral gap and log–Sobolev constant of balanced matroids

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 $x \sim y$, i.e. x and y are connected by an edge, and $P(x, x) = 1 - \sum_{\{x,y\} \in E} P(x, y) > 0$ (so that the chain is aperiodic), 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 finite and irreducible; thus it has unique stationary distribution π. Since M fulfils the detailed balance condition (i.e. is time–reversible) w.r.t. the uniform distribution on $\mathcal{B}$, the stationary distribution π must be the uniform distribution on $\mathcal{B}$. 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 $P_\psi(\cdot, \cdot)$. 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[2].

Corollary 6.2 *Let $M = (\mathcal{B}, P)$ be a bases–exchange walk with constant transition probabilities, i.e. $P(x, y) = p$ if $x \neq y$, $x \sim y$ and $P(x, x) = 1 - \sum_{\{x,y\} \in E} P(x, y) > 0$, 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:*

$$\lambda_M \geq 2p \quad \textit{and} \quad \alpha_M \geq \frac{p}{2}.$$

Proof. Apply Theorems 4.4 and 5.4. ☐

We will give two example applications of Corollary 6.2:

Example 6.3 *Random walk on a hypercube:*

Consider the d–dimensional hypercube. The vertices can be denoted by d–tuples from $\{0, 1\}^d$ and edges connect vertices whose Hamming distance is 1, i.e. tuples that differ by exactly one entry. Consider the random walk M with transition probabilities $P(x, y) = 1/(d + 1)$ whenever x and y are connected by an edge. This random process has been studied quite extensively before (see [6] and the references given therein) and has uniform stationary distribution $\pi(x) = 1/2^d$. It is already known that the walk on the hypercube has spectral gap $\lambda = 2/(d+1)$ and log–Sobolev constant $\alpha = 1/(d + 1)$ (see [5] and [6] for a discussion). There are several ways to see that this Markov chain is π–recursive. The easiest way 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 π–matching is given by $w(x, y) = 2/2^d$. Using Theorem 4.4 and Theorem 5.4 we obtain the bounds $\lambda \geq 2/(d + 1)$

[2] Observe 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.

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 the factor 2. However, for the random walk on hypercubes it can be shown that Lemma 5.2 and Theorem 5.4 hold for $\mathrm{ct}_{\mathcal{E}_M}(f) \geq \mathrm{ct}_{\mathcal{L}_\pi}(f)/p$, which recovers the result given in [5] by Diaconis and Saloff–Coste.

The next example is more interesting.

Example 6.4 *Balanced matroids:*

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 6.1. 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 irreducible, aperiodic and time–reversible w.r.t. the uniform distribution on $\mathcal{B}$. If M is balanced, then applying Theorem 5.4 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 thus by (1) we obtain: $\tau \leq O(mn(\log n + \log\log m))$. This greatly improves on the mixing time for this walk given in [18]: $\tau \leq O(m^{3/2}n^2(\log n + \log\log m))$. Using a result by Heller [10] which says that for *simple* regular matroids, i.e. 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 $m = O(n^2)$. Thus for such matroids our bound on the log–Sobolev constant yields a mixing time of $O(n^3 \log n)$ 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[3].

Acknowledgement

We would like to thank the anonymous reviewer of this paper for an elegant and short version version of the proof of Lemma 5.3.

References

[1] 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*, number 986 in Springer Lecture Notes in Mathematics, pages 243–297. Springer, 1986.

[2] D. Bakry. L'hypercontractivité et son utilisation en théorie des semigroups. In *Ecole d'Été de Saint Flour 1992*, number 1581 in Lecture Notes in Mathematics. Springer, Berlin, 1994.

[3] A. Broder. Generating random spanning trees. In *Proceedings of the 30th Annual Symposium on Foundations of Computer Science*, pages 442–447. IEEE, 1989.

[4] 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.

[5] P. Diaconis and L. Saloff-Coste. Logarithmic sobolev inequalities for finite markov chains. *Ann. Appl. Probab.*, 6(3):695–750, 1996.

[6] P. Diaconis and D. Stroock. Geometric bounds for eigenvalues of markov chains. *Ann. Appl. Probab.*, 1(1):36–61, 1991.

[7] T. Feder and M. Mihail. Balanced matroids. In *Proceedings of the 24th Annual ACM Symposium on Theory of Computing*, pages 26–38. ACM Press, 1992.

[8] L. Gross. Logarithmic sobolev inequalities. *American Journal of Mathematics*, 97:1061–1083, 1976.

[9] L. Gross. *Logarithmic Sobolev inequalities and contractivity properties of semigroups*. Number 1563 in Lecture Notes in Mathematics. Springer, Berlin, 1993.

[10] I. Heller. On linear systems with integral valued solutions. *Pacific Journal of Mathematics*, 7:1351–1364, 1957.

[11] C. Houdré. Mixed and isoperimetric estimates on the log–sobolev constants of graphs and markov chains. *Combinatorica*, 21(4):489–513, 2001.

[12] 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.

[13] M. Jerrum. Counting, sampling and integrating. In *Algorithms and Complexity*. Birkhäuser, Basel, 2002.

[14] M. Jerrum and A. Sinclair. Approximating the permament. *SIAM Journal on Computing*, 18:1149–1178, 1989.

[15] A. Korzeniowski and D. Stroock. An example in the theory of hypercontractive semigroups. *Proceedings of the American Mathematical Society*, 94:87–90, 1985.

[16] L. Lovász. Random walks on graphs: a survey. In *Paul Erdös is Eighty (Volume 2)*, number 2 in Combinatorics, pages 1–46. Bolyai Society Mathematical Studies, Keszthely (Hungary), 1993.

[17] S. L. Lu and H.-T. Yau. Spectral gap and logarithmic sobolev inequality for kawasaki and glauber dynamics. *Communications in Mathemtical Physics*, 156(2):399–433, 1993.

[18] 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*, pages 704–711. ACM Press, 2001.

[19] O. Rothaus. Diffusion on compact riemannian manifolds and logarithmic sobolev inequalities. *Journal of Functional Analysis*, 42:102–109, 1981.

[3]However, if the size of the groundset is larger, say $m = \Omega(n^3)$, then Feder and Mihail's *modified* random walk on balanced matroids (see [7], Theorem 5.2) with $\tau \leq O(n^4 \log m)$ converges faster.