Chapter 8

Inductive bounds, cubes, trees and matroids

The spectral gap of a MC can sometimes be bounded by a direct inductive argument. Given its conceptual simplicity, this inductive approach seems surprising powerful. To start with, however, we'll develop the tools in the context of the random walk on the n-dimensional cube. The simplicity of this example will bring the key ideas into sharp relief.

8.1 The cube

Suppose n is a positive integer (dimension) and $0 < p < 1/n$. We consider the random walk on $\Omega = \{0,1\}^n$ with transition probabilities given by

$$P(x,y) = \begin{cases} p & \text{if } |x-y| = 1; \\ 0 & \text{otherwise,} \end{cases}$$

where $|x-y|$ denotes Hamming distance between x and y. The MC (Ω, P) is ergodic with uniform stationary distribution. We already know two ways to upper bound the mixing time of this MC: coupling and canonical paths. A third is to give the state space a geometric interpretation and use isoperimetry. (Jerrum and Sinclair [45, §12.3] use the random walk on the cube as an illustration of the second and third of these approaches. Coupling is the subject of Exercise 8.4.) In this section we study a fourth. Why do we need another method? The advantage of this one is that it is robust, in the sense that applies to other MCs with inductively defined state spaces. This section and §8.3 is based on Jerrum and Son [47], and Jerrum, Son, Tetali and Vigoda [48].

A function $g : K \to \mathbb{R}$ defined on a convex set $K \subset \mathbb{R}^k$ is *convex* if $g(\alpha x + (1-\alpha)y) \le \alpha g(x) + (1-\alpha)g(y)$ for every $x, y \in K$ and $0 < \alpha < 1$. By expectation of a r.v. taking values in K we mean the obvious thing, namely, take expectations of the individual coordinates.

Lemma 8.1 (Jensen's inequality). *Let $K \subset \mathbb{R}^k$ be a compact convex set, X a r.v. taking values in K, and $g : K \to \mathbb{R}$ a convex real-valued function. Then $g(\mathbb{E}\, X) \le \mathbb{E}\, g(X)$.*

Exercise 8.2. Prove Jensen's inequality. Hint: Consider the graph $G(g) = \big\{(x, y) : x \in K \text{ and } y \ge g(x)\big\} \subset \mathbb{R}^{k+1}$ of g together with a supporting plane to $G(g)$ at the point $(\mathbb{E}\, X, g(\mathbb{E}\, X))$.

Suppose $\Omega = \Omega_0 \cup \Omega_1$ is a partition of the state space. (For the cube it is natural to take $\Omega_b = \{x = x_0 x_1 \dots x_{n-1} \in \Omega : x_0 = b\}$.) For π a probability distribution on Ω, we denote by $\pi_b : \Omega_b \to [0,1]$ the induced distribution $\pi/\pi(\Omega_b)$ on Ω_b. Let $\varphi : \Omega \to \mathbb{R}$ be any real-valued "test function" on Ω. (In previous chapters we used f for this purpose. The change to φ is just to avoid a notational clash later in this chapter.) Then (decomposition of variance)

$$\operatorname{Var}_\pi \varphi = \pi(\Omega_0) \operatorname{Var}_{\pi_0} \varphi + \pi(\Omega_1) \operatorname{Var}_{\pi_1} \varphi + \operatorname{Var}_\pi \bar{\varphi} \tag{8.1}$$

where

$$\operatorname{Var}_{\pi_b} \varphi = \sum_{x \in \Omega_b} \pi_b(x)(\varphi(x) - \mathbb{E}_{\pi_b} \varphi)^2,$$

$$\mathbb{E}_{\pi_b} \varphi = \sum_{x \in \Omega_b} \pi_b(x)\varphi(x)$$

and

$$\operatorname{Var}_\pi \bar{\varphi} = \pi(\Omega_0)\pi(\Omega_1)(\mathbb{E}_{\pi_0} \varphi - \mathbb{E}_{\pi_1} \varphi)^2.$$

The rationale for the notation $\operatorname{Var}_\pi \bar{\varphi}$ is that this "cross term" may be interpreted as the variance of the function $\bar{\varphi}$ that is constant $\mathbb{E}_{\pi_b} \varphi$ on Ω_b, for $b = 0, 1$. Also (decomposition of the Dirichlet form)

$$\mathcal{E}_P(\varphi, \varphi) = \pi(\Omega_0)\mathcal{E}_{P_0}(\varphi, \varphi) + \pi(\Omega_1)\mathcal{E}_{P_1}(\varphi, \varphi) + \mathcal{C}, \tag{8.2}$$

where

$$\mathcal{E}_{P_b}(\varphi, \varphi) = \frac{1}{2} \sum_{x, y \in \Omega_b} \pi_b(x) P(x, y)(\varphi(x) - \varphi(y))^2$$

and

$$\mathcal{C} = \sum_{x\in\Omega_0, y\in\Omega_1} \pi(x)P(x,y)(\varphi(x)-\varphi(y))^2.$$

In the definition of $\mathcal{C}$ we have assumed time reversibility of (Ω, P): the restriction of the sum to unordered pairs exactly accounts for the factor $\frac{1}{2}$ in the definition of the Dirichlet form.

Exercise 8.3. Verify (8.1) and (8.2). (One of these identities is actually trivial.)

All the above was for an arbitrary time-reversible MC with finite state space partitioned into two pieces. We now specialise to the uniform random walk on the n-dimensional Boolean cube. In this instance, π is the uniform distribution on Ω and π_b is the uniform distribution on Ω_b. Suppose, inductively, we had established Poincaré inequalities

$$\mathcal{E}_{P_b}(\varphi,\varphi) \geq \lambda_{n-1,p} \operatorname{Var}_{\pi_b} \varphi \tag{8.3}$$

for the subcubes. These will allow us to compare two of the three corresponding pairs of terms in (8.1) and (8.2). Thus we may obtain a Poincaré inequality for the n-dimensional cube provided we can relate the final pair of terms.

Consider the r.v. $(F_0, F_1) \in \mathbb{R}^2$ defined by the following trial: select $z \in \{0,1\}^{n-1}$ u.a.r.; then let $(F_0, F_1) = (\varphi(0z), \varphi(1z)) \in \mathbb{R}^2$. (Here bz denotes the element of Ω_b obtained by prefixing z by the bit b.) Then

$$\operatorname{Var}_\pi \bar\varphi = \pi(\Omega_0)\pi(\Omega_1)(\mathbb{E}_{\pi_0}\varphi - \mathbb{E}_{\pi_1}\varphi)^2 = \pi(\Omega_0)\pi(\Omega_1)(\mathbb{E}_z F_0 - \mathbb{E}_z F_1)^2$$

and

$$\mathcal{C} = \frac{p}{2}\mathbb{E}_z\left[(F_0-F_1)^2\right].$$

(Here we use $\mathbb{E}_z$ to denote expectations with respect to a uniformly selected $z \in \{0,1\}^{n-1}$.) But the function $\mathbb{R}^2 \to \mathbb{R}$ defined by $(\xi,\eta) \mapsto (\xi-\eta)^2$ is convex; so, by Lemma 8.1 (Jensen's Inequality),

$$\mathbb{E}_z\left[(F_0-F_1)^2\right] \geq (\mathbb{E}_z F_0 - \mathbb{E}_z F_1)^2$$

and hence

$$\mathcal{C} \geq \frac{p}{2\pi(\Omega_0)\pi(\Omega_1)} \operatorname{Var}_\pi \bar\varphi. \tag{8.4}$$

Substituting (8.3) and (8.4) into (8.2), and comparing with (8.1), we obtain

$$\lambda_{n,p} \geq \min\{\lambda_{n-1,p}, 2p\},$$

where we have used the fact that $\pi(\Omega_0) = \pi(\Omega_1) = \frac{1}{2}$. For the base case, $n = 1$, it is easy to check by direct calculation that $\lambda_{1,p} = 2p$. Thus, by a trivial induction, $\lambda_{n,p} \geq 2p$. This bound is tight, as can be seen by taking the function φ that is constant -1 on Ω_0 and constant 1 on Ω_1.

It follows from arguments in Chapter 5 — see Corollary 5.9, recalling $\varrho = \lambda^{-1}$ — that the mixing time of the random walk on the n-dimensional cube with transition probabilities $p = 1/n$ is $O\bigl(n(n + \log(1/\varepsilon))\bigr)$. (The first n is from the reciprocal of the Poincaré constant and the second from $\log(1/\pi(x_0))$.) Here we assume that periodicity is dealt with either by using the lazy version of the walk, or working in continuous time. The correct answer is $O\bigl(n \log(n/\varepsilon)\bigr)$, so no cigar... yet.

Exercise 8.4. Demonstrate that $O(n \log n)$ is the correct order of magnitude for the mixing time of the random walk on the cube. The upper bound can be obtained by coupling, the lower bound by a coupon collector argument. Warning: the lower bound may not be quite as simple as you expect!

It was suggested at the outset that the technique just applied in the context of the cube has a degree of robustness. "Twisted cubes" provide somewhat artificial confirmation of this claim. A *twisted cube* of dimension 1 is a complete graph on two vertices (i.e., two vertices joined by an edge); a twisted cube of dimension $n > 1$ is the union of two distinct twisted cubes (possibly different) of dimension $n - 1$, connected by an arbitrary perfect matching (of size 2^{n-1}). Observe that the inductive computation of $\lambda_{n,p}$ given in this section applies just as well to twisted cubes.

Exercise 8.5. For a twisted cube, what is the best upper bound on mixing time you can achieve by coupling?

8.2 Balanced Matroids

Twisted cubes in themselves aren't interesting, but there are more substantial examples where the ideas from §8.1 apply with no essential change. What do we need for the argument of §8.1? First, we need to be able to decompose the MC into two (or maybe more) smaller pieces "of the same kind". Second, we need the transitions that cross between the pieces to be such as to support a coupling of the r.v's F_0 and F_1, used in the derivation of (8.4).

A general class of random walks falling into this setting are random walks on the "bases-exchange graph" of a balanced matroid. The various technical terms appearing in that sentence will be explained presently. For the time being, let us merely note that this class includes, as a special case, a natural walk on spanning trees of a graph.

Let E be a finite ground set and $\mathcal{B} \subseteq 2^E$ a collection of subsets of E. We say that $\mathcal{B}$ forms the collection of *bases* of a *matroid* $M = (E, \mathcal{B})$ if the following two conditions hold:

1. All bases (sets in $\mathcal{B}$) have the same size, namely the *rank* of M.

2. 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 $X \cup \{f\} \setminus \{e\} \in \mathcal{B}$.

The above axioms for a matroid capture the notion of linear independence. Thus if $S = \{u_0, \dots, u_{m-1}\}$ is a set of n-vectors over a field K, then the maximal linearly independent subsets of S form the bases of a matroid with ground set S. The bases in this instance have size equal to the dimension of the vector space spanned by S, and they clearly satisfy the second or "exchange" axiom. A matroid that arises in this way is *vectorial*, and is said to be *representable over* K.

Several other equivalent axiomatisations of matroid are possible, each shedding different light on the notion of linear independence; the above choice turns out to be the most appropriate for our needs. For other possible axiomatisations, and more on matroid theory generally, consult Oxley [66] or Welsh [81].

The advantage of the abstract viewpoint provided by matroid theory is that it allows us to perceive and exploit formal linear independence in a variety of combinatorial situations. Most importantly, the spanning trees in an undirected graph $G = (V, E)$ form the bases of a matroid, the *cycle matroid of* G, with ground set E. A matroid that arises as the cycle matroid of some graph is called *graphic*.

Two absolutely central operations on matroids are contraction and deletion. An element $e \in E$ is said to be a *coloop* if it occurs in every basis. If $e \in E(M)$ is an element of the ground set of M then, provided e is not a coloop, the matroid $M \setminus e$ obtained by *deleting* e has ground set $E(M \setminus e) = E(M) \setminus \{e\}$ and bases $\mathcal{B}(M \setminus e) = \{X \subseteq E(M \setminus e) : X \in \mathcal{B}(M)\}$; and the matroid M/e obtained by *contracting* e has ground set $E(M/e) = E(M) \setminus \{e\}$ and bases $\mathcal{B}(M/e) = \{X \subseteq E(M/e) : X \cup \{e\} \in \mathcal{B}(M)\}$. Any matroid obtained from M by a series of contractions and deletions is a *minor* of M.

The matroid axioms given above suggest a very natural walk on the set of bases of a matroid M. The *bases-exchange graph* $G(M)$ of a matroid M

has vertex set $\mathcal{B}(M)$ and edge set

$$\big\{\{X,Y\} : X,Y \in \mathcal{B} \text{ and } |X \oplus Y| = 2\big\},$$

where $\oplus$ denotes symmetric difference. Note that the edges of the bases-exchange graph $G(M)$ correspond to the transformations guaranteed by the exchange axiom. Indeed, it is straightforward to check, using the exchange axiom, that the graph $G(M)$ is always connected. By simulating a random walk on $G(M)$ it is possible, in principle, to sample a basis (almost) u.a.r. from $\mathcal{B}(M)$. Although it has been conjectured that the random walk on $G(M)$ is rapidly mixing for all matroids M, the conjecture has never been proved and the circumstantial evidence in its favour seems slight. Nevertheless, there is an interesting class of matroids, the "balanced" matroids, for which rapid mixing has been established. The definition of balanced matroid is due to Feder and Mihail [32], as is the proof of rapid mixing. We follow their treatment quite closely, up to and including Lemma 8.8. We then deviate from their analysis, and instead use the methods of §8.1 in order to achieve a tighter bound on spectral gap.

For the rest of this section we usually drop explicit reference to the matroid M, and simply write $\mathcal{B}$ and E in place of $\mathcal{B}(M)$ and $E(M)$. Suppose a basis $X \in \mathcal{B}$ is chosen u.a.r. If $e \in E$, we let e stand (with a slight abuse of notation) for the event $e \in X$, and $\bar{e}$ for the event $e \notin X$. Furthermore, we denote conjunction of events by juxtaposition: thus $e\bar{f}$ denotes the event $e \in X \wedge f \notin X$, etc. The matroid M is said to possess the *negative correlation property* if the inequality $\Pr(ef) \le \Pr(e)\Pr(f)$ holds for all pairs of distinct elements $e, f \in E$. Another way of expressing negative correlation is by writing $\Pr(e \mid f) \le \Pr(e)$; in other words the knowledge that f is present in X makes the presence of e less likely.[1] Further, the matroid M is said to be *balanced* if all minors of M (including M itself) possess the negative correlation property. We shall see in §8.4 that graphic matroids, amongst others, are balanced. So the class is not without interest, even if it does not include all matroids.

If $E' \subseteq E$, then a *increasing property* over E' is a property of subsets of E' that is closed under the superset relation; equivalently, it is a property that may be expressed as a monotone Boolean formula in the indicator variables of the elements in E'. A *decreasing property* is defined analogously.

[1] We assume here that $\Pr(f) > 0$; an element f such that $\Pr(f) = 0$ is said to be a *loop*.

Lemma 8.6. *Suppose M is a balanced matroid. For every $e \in E(M)$ and every increasing property μ over $E(M) \setminus \{e\}$, the inequality* $\Pr(\mu e) \le \Pr(\mu)\Pr(e)$ *holds; in other words, μ is negatively correlated with e.*

Remark 8.7. The inequality $\Pr(ef) \le \Pr(e)\Pr(f)$ is a special case of Lemma 8.6.

Proof of Lemma 8.6. The proof is by induction on the size of the ground set. We may assume that $\Pr(\mu e) > 0$, otherwise the result is immediate. Conditional probabilities with respect to e and μe are thus well defined, and we may re-express our goal as $\Pr(\mu \mid e) \le \Pr(\mu)$. Further, we may assume that the rank r of M is at least 2, otherwise the result follows from the fact that μ is increasing.

From the identity

$$\sum_{f \neq e} \Pr(f \mid \mu e) = r - 1 = \sum_{f \neq e} \Pr(f \mid e),$$

and the assumption that $r \ge 2$, we deduce the existence of an element f satisfying $\Pr(f \mid \mu e) \ge \Pr(f \mid e) > 0$, and hence

$$\Pr(\mu \mid ef) \ge \Pr(\mu \mid e); \tag{8.5}$$

note that the conditional probability on the left is well defined. Two further inequalities that hold between conditional probabilities are

$$\Pr(f \mid e) \le \Pr(f) \tag{8.6}$$

and

$$\Pr(\mu \mid ef) \le \Pr(\mu \mid f); \tag{8.7}$$

the former comes simply from the negative correlation property, and the latter from applying the inductive hypothesis to the matroid M/f and the property derived from μ by forcing f to 1.

At this point we dispense with the degenerate case $\Pr(\bar{f} \mid e) = 0$. It follows from (8.6) that $\Pr(f) = 1$, and then from (8.7) that $\Pr(\mu \mid e) \le \Pr(\mu)$, as desired. So we may now assume $\Pr(\bar{f} \mid e) > 0$ and hence that probabilities conditional on the event $e\bar{f}$ are well defined. In particular,

$$\Pr(\mu \mid e\bar{f}) \le \Pr(\mu \mid \bar{f}), \tag{8.8}$$

as can be seen by applying the inductive hypothesis to the matroid $M \setminus f$ and the property derived from μ by forcing f to 0. Further, inequality (8.5) may be re-expressed as

$$\Pr(\mu \mid ef) \ge \Pr(\mu \mid e\bar{f}). \tag{8.9}$$

The inductive step is now achieved through a chain of inequalities based on (8.6)–(8.9):

$$\begin{aligned}
\Pr(\mu \mid e) &= \Pr(\mu \mid ef)\Pr(f \mid e) + \Pr(\mu \mid e\bar{f})\Pr(\bar{f} \mid e) \\
&= \Pr(\mu \mid ef)\Pr(f \mid e) + \Pr(\mu \mid e\bar{f})(1 - \Pr(f \mid e)) \\
&= \bigl[\Pr(\mu \mid ef) - \Pr(\mu \mid e\bar{f})\bigr]\Pr(f \mid e) + \Pr(\mu \mid e\bar{f}) \\
&\le \bigl[\Pr(\mu \mid ef) - \Pr(\mu \mid e\bar{f})\bigr]\Pr(f) + \Pr(\mu \mid e\bar{f}) && \text{by (8.6), (8.9)} \\
&= \Pr(\mu \mid ef)\Pr(f) + \Pr(\mu \mid e\bar{f})\Pr(\bar{f}) \\
&\le \Pr(\mu \mid f)\Pr(f) + \Pr(\mu \mid \bar{f})\Pr(\bar{f}) && \text{by (8.7), (8.8)} \\
&= \Pr(\mu).
\end{aligned}$$

□

Given $e \in E$, the set of bases $\mathcal{B}$ may be partitioned as $\mathcal{B} = \mathcal{B}_e \cup \mathcal{B}_{\bar{e}}$, where $\mathcal{B}_e = \{X \in \mathcal{B} : e \in X\}$ and $\mathcal{B}_{\bar{e}} = \{X \in \mathcal{B} : e \notin X\}$; observe that $\mathcal{B}_e$ and $\mathcal{B}_{\bar{e}}$ are isomorphic to $\mathcal{B}(M/e)$ and $\mathcal{B}(M \setminus e)$, respectively (assuming e is not a coloop). For $\mathcal{A} \subseteq \mathcal{B}_e$ (respectively, $\mathcal{A} \subseteq \mathcal{B}_{\bar{e}}$), let $\Gamma_e(\mathcal{A})$ denote the set of all vertices in $\mathcal{B}_{\bar{e}}$ (respectively, $\mathcal{B}_e$) that are adjacent to some vertex in $\mathcal{A}$. The bipartite subgraph of the bases-exchange graph induced by the bipartition $\mathcal{B} = \mathcal{B}_e \cup \mathcal{B}_{\bar{e}}$ satisfies a natural expansion property.

Lemma 8.8. *Suppose M is a balanced matroid, $e \in E(M)$, and that the partition $\mathcal{B} = \mathcal{B}_e \cup \mathcal{B}_{\bar{e}}$ is non-trivial. Then*

$$\begin{aligned}
\frac{|\Gamma_e(\mathcal{A})|}{|\mathcal{B}_{\bar{e}}|} &\ge \frac{|\mathcal{A}|}{|\mathcal{B}_e|}, \quad \textit{for all } \mathcal{A} \subseteq \mathcal{B}_e, \textit{ and} \\
\frac{|\Gamma_e(\mathcal{A})|}{|\mathcal{B}_e|} &\ge \frac{|\mathcal{A}|}{|\mathcal{B}_{\bar{e}}|}, \quad \textit{for all } \mathcal{A} \subseteq \mathcal{B}_{\bar{e}}.
\end{aligned}$$

Proof. For any $\mathcal{A} \subseteq \mathcal{B}_e$ let $\mu_{\mathcal{A}}$ denote the increasing property $\mu_{\mathcal{A}} = \bigvee_{X \in \mathcal{A}} \bigwedge_{f \in X \setminus \{e\}} f$. The collection of all bases in $\mathcal{B}_e$ satisfying $\mu_{\mathcal{A}}$ is precisely $\mathcal{A}$, while the collection of all bases in $\mathcal{B}_{\bar{e}}$ satisfying $\mu_{\mathcal{A}}$ is precisely $\Gamma_e(\mathcal{A})$. Hence the first part of the lemma is equivalent to the inequality $\Pr(\mu_{\mathcal{A}} \mid \bar{e}) \ge \Pr(\mu_{\mathcal{A}} \mid e)$, which follows from Lemma 8.6. Similarly, for any $\mathcal{A} \subseteq \mathcal{B}_{\bar{e}}$ let $\bar{\mu}_{\mathcal{A}}$ denote the decreasing property $\bar{\mu}_{\mathcal{A}} = \bigvee_{X \in \mathcal{A}} \bigwedge_{f \notin X \cup \{e\}} \bar{f}$. The set of all bases in $\mathcal{B}_{\bar{e}}$ satisfying $\bar{\mu}_{\mathcal{A}}$ is precisely $\mathcal{A}$, while the set of all bases in $\mathcal{B}_e$ satisfying $\bar{\mu}_{\mathcal{A}}$ is precisely $\Gamma_e(\mathcal{A})$. Hence the second part of the lemma is equivalent to the inequality $\Pr(\bar{\mu}_{\mathcal{A}} \mid e) \ge \Pr(\bar{\mu}_{\mathcal{A}} \mid \bar{e})$, which again follows from Lemma 8.6. □

8.3 Bases-exchange walk

Suppose M is a balanced matroid, and p satisfies $0 < p \le 1/rm$, where m is the size of the ground set of M and r its rank. Consider the MC (Ω, P) whose state space $\Omega = \mathcal{B}$ is the set of all bases in M, and whose transition probabilities P are given by

$$P(x,y) = \begin{cases} p & \text{if } (x,y) \text{ is an edge of the bases-exchange graph } G(M); \\ 0 & \text{otherwise,} \end{cases}$$

for all $x, y \in \Omega$ with $x \ne y$; loop probabilities are implicitly defined by complementation. Since the maximum degree of the bases-exchange graph of M is strictly less than rm, the transition probabilities are well defined. By the exchange property of matroids, (Ω, P) is irreducible, and since loop probabilities are non-zero it is also aperiodic. The transition probabilities are symmetric, so the stationary distribution is uniform. This MC is the *bases-exchange walk* associated with M.

We'll see that the expansion property formalised in Lemma 8.8 allows us to reuse the analysis of §8.1 almost exactly.

Remark 8.9. We can implement this random walk on $G(M)$ naturally as follows. The current state (basis) is X_0.

1. Choose e u.a.r. from E, and f u.a.r. from X_0.
2. If $Y = X_0 \cup \{e\} \setminus \{f\} \in \mathcal{B}$ then $X_1 = Y$; otherwise $X_1 = X_0$.

The new state is X_1.

Theorem 8.10. *Suppose M is a balanced matroid. The spectral gap of the bases-exchange walk associated with M is at least $\lambda \ge 2p$, where p is the uniform transition probability. For the above implementation, $p = 1/rm$.*

Corollary 8.11. *The mixing time of the bases-exchange walk on any balanced matroid of rank r on a ground set of size m is $O\big(rm(r \ln m + \ln \varepsilon^{-1})\big)$.*

Theorem 8.10 will follow fairly directly from Lemma 8.12 below. In order to make a connection with the argument of §8.1, we'll identify Ω_0 with $\mathcal{B}_{\bar{e}}$ and Ω_1 with $\mathcal{B}_e$. Recall that $\pi_b = \pi/\pi(\Omega_b)$, for $b = 0, 1$, is the induced distribution on Ω_b, in this case uniform.

Lemma 8.12. *The transitions from Ω_0 to Ω_1 support a fractional matching. Specifically, there is a function $w : \Omega_0 \times \Omega_1 \to \mathbb{R}^+$ such that (i) $\sum_{y \in \Omega_1} w(x,y) = \pi_0(x)$, for all $x \in \Omega_0$; (ii) $\sum_{x \in \Omega_0} w(x,y) = \pi_1(y)$, for all $y \in \Omega_1$; and (iii) $w(x,y) > 0$ entails $P(x,y) > 0$, for all $(x,y) \in \Omega_0 \times \Omega_1$.*

Proof (sketch). Follows from Lemma 8.8, using the the Max-flow, min-cut Theorem [79, Thm 7.1]. $\square$

Exercise 8.13. Prove Lemma 8.12. Start with the bipartite subgraph of the bases-exchange graph $G(M)$ induced by the vertex partition (Ω_0, Ω_1). Construct from it a flow network by adding a distinguished source s and sink t, arcs of capacity $\pi_0(x)$ from s to every node $x \in \Omega_0$, and arcs of capacity $\pi_1(y)$ from every node $y \in \Omega_1$ to t. All other arcs, corresponding to possible transitions from Ω_0 to Ω_1, have unbounded capacity. Use Lemma 8.8 to show that the network has a flow of value 1.

Remark 8.14. Note that

$$\sum_{(x,y)\in\Omega_0\times\Omega_1} w(x,y) = \sum_{x\in\Omega_0} \pi_0(x) = 1,$$

so $(\Omega_0 \times \Omega_1, w)$ is a probability space.

We are now ready to bound the spectral gap of the bases-exchange walk.

Proof of Theorem 8.10. Let $(F_0, F_1) \in \mathbb{R}^2$ be the r.v. defined on $(\Omega_0\times\Omega_1, w)$ as follows: select $(x, y) \in \Omega_0 \times \Omega_1$ according to distribution w and return $(F_0, F_1) = (f(x), f(y))$.

To carry out the programme of §8.1, need to compare the cross term of the variance

$$\operatorname{Var}_\pi \bar\varphi = \pi(\Omega_0)\pi(\Omega_1)(\mathbb{E}_{\pi_0}\varphi - \mathbb{E}_{\pi_1}\varphi)^2 = \pi(\Omega_0)\pi(\Omega_1)(\mathbb{E}_w F_0 - \mathbb{E}_w F_1)^2,$$

to the cross term $\mathcal{C}$ in the Dirichlet form. Without loss of generality, assume $\pi(\Omega_0) \ge \pi(\Omega_1)$. Now, $w(x, y) \le \pi_0(x) = \pi(x)/\pi(\Omega_0)$, for all $(x, y) \in \Omega_0 \times \Omega_1$, which implies $\pi(x)P(x, y) \ge p\,\pi(\Omega_0)w(x, y)$. (Note that we are using the fact that $w(x, y) = 0$ whenever $P(x, y) = 0$.) Thus

$$\begin{aligned}
\mathcal{C} &= \sum_{(x,y)\in\Omega_0\times\Omega_1} \pi(x)P(x,y)\big(\varphi(x)-\varphi(y)\big)^2 \\
&\ge p\,\pi(\Omega_0) \sum_{(x,y)\in\Omega_0\times\Omega_1} w(x,y)\big(\varphi(x)-\varphi(y)\big)^2 \\
&= p\,\pi(\Omega_0)\,\mathbb{E}_w\big[(F_0-F_1)^2\big] \\
&\ge p\,\pi(\Omega_0)(\mathbb{E}_w F_0 - \mathbb{E}_w F_1)^2 && \text{by Lemma 8.1} \\
&= \frac{p}{\pi(\Omega_1)}\,\pi(\Omega_0)\pi(\Omega_1)(\mathbb{E}_w F_0 - \mathbb{E}_w F_1)^2 \\
&\ge 2p\operatorname{Var}_\pi \bar\varphi.
\end{aligned}$$

We are now exactly in the situation of §8.1, when we were analysing the gap of the cube walk. In particular, denoting by $\lambda_{m,p}$ a lower bound on the spectral gap of the basis-exchange walk when the ground set of M has size m and the transition probabilities are all p, we have $\lambda_{m,p} \geq \min\{\lambda_{m-1,p}, 2p\}$, and hence $\lambda_{m,p} \geq 2p$. □

8.4 Examples of balanced matroids

A natural question now presents itself: how big is the class of balanced matroids?

A matroid that is representable over every field is called *regular*. The class of regular matroids is well studied is certainly wide enough to contain interesting examples; indeed, all graphic matroids are regular. The main result of this section is that all regular matroids are balanced. More precisely, we prove the equivalent result that all "orientable" matroids are balanced. The class of orientable matroids is known to be the same as the class of regular matroids [66, Corollary 13.4.6].[2]

In order to define the property of being orientable, we need some further matroid terminology. A *cycle* $C \subset E$ in a matroid $M = (E, \mathcal{B})$ is a minimal (under set inclusion) subset of elements that cannot be extended to a basis. A *cut* is a minimal set of elements whose complement does not contain a basis. Note that in the case of the cycle matroid of a graph, in which the bases are spanning trees, these terms are consistent with the usual graph-theoretic ones. Let $\mathcal{C} \subseteq 2^E$ denote the set of all cycles in M and $\mathcal{D} \subseteq 2^E$ the set of all cuts. We say that M is *orientable* if functions $\gamma : \mathcal{C} \times E \to \{-1, 0, +1\}$ and $\delta : \mathcal{D} \times E \to \{-1, 0, +1\}$ exist which satisfy the following three conditions, for all $C \in \mathcal{C}$ and $D \in \mathcal{D}$:

$$
\begin{aligned}
\gamma(C, g) &\neq 0 \text{ iff } g \in C, \\
\delta(D, g) &\neq 0 \text{ iff } g \in D, \text{ and} \\
\sum_{g \in E} \gamma(C, g)\delta(D, g) &= 0.
\end{aligned}
\tag{8.10}
$$

We work in this section towards the following result. In doing so, we'll follow Feder and Mihail [32] fairly closely.

Theorem 8.15. *Orientable (and hence regular) matroids are balanced.*

[2] When consulting this corollary, it is important to realise that Oxley applies the term "signable" to the class of matroids Feder and Mihail call "orientable," preferring to apply the latter term to a different and larger class. We follow Feder and Mihail's terminology.

In preparation for the proof of Theorem 8.15, we introduce some further notation and make some observations. A *near basis* of M is a set $N \subseteq E$ that can be augmented to a basis by the addition of a single element from the ground set. A *unicycle* of M is a set $U \subseteq E$ that can be reduced to a basis by the removal of a single element. A near basis N defines a unique cut D_N consisting of all elements of the ground set whose addition to N results in a basis. A unicycle U defines a unique cycle C_U consisting of all elements which whose removal from U results in a basis. Let e, f be distinct elements of the ground set E. We claim that

$$\gamma(C_U, e)\gamma(C_U, f) + \delta(D_N, e)\delta(D_N, f) = 0, \tag{8.11}$$

for all near-bases N and unicycles U that are related by $U = N \cup \{e, f\}$. To see this, note that the equation (8.10) simplifies in this situation to

$$\gamma(C_U, e)\delta(D_N, e) + \gamma(C_U, f)\delta(D_N, f) = 0, \tag{8.12}$$

since all terms in the sum are zero except from those obtained by setting $g = e$ and $g = f$. Now it may be that all four quantities in (8.12) are zero, in which case we are done. Otherwise, some quantity, say $\delta(D_N, e)$, is non-zero, in which case $D_N \cup \{e\} = C_U \setminus \{f\}$ is a basis and $\gamma(C_U, f)$ is non-zero also. Multiplying (8.12) through by $\gamma(C_U, f)\delta(D_N, e)$ yields

$$\gamma(C_U, e)\gamma(C_U, f)\delta(D_N, e)^2 + \gamma(C_U, f)^2\delta(D_N, e)\delta(D_N, f) = 0,$$

which simplifies to equation (8.11) as required, since the square factors are both one.

For distinct elements $e, f \in E$, define

$$\Delta_{ef} = \sum_N \delta(D_N, e)\delta(D_N, f) = -\sum_U \gamma(C_U, e)\gamma(C_U, f),$$

where the sums are over all near bases N and unicycles U. The equality of the two expressions above is a consequence of (8.11), and the bijection between non-zero terms in the two sums that is given by $N \mapsto N \cup \{e, f\} = U$. Select a distinguished element $e \in E$ and force $\gamma(C, e) = -1$ and $\delta(D, e) = 1$ for all cycles $C \ni e$ and cuts $D \ni e$. This can be done by flipping signs around cycles and cuts, without compromising the condition (8.10) for orientability, nor changing the value of Δ_{ef}. With this convention we have

$$\sum_{g \neq e} \gamma(C, g)\delta(D, g) = 1, \quad \text{provided } C \ni e \text{ and } D \ni e; \tag{8.13}$$

$$\gamma(C_U, f) = \delta(D_N, f), \quad \text{provided } U = N \cup \{e, f\}; \tag{8.14}$$

and

$$\Delta_{ef} = \sum_{U:e\in C_U} \gamma(C_U, f) = \sum_{N:e\in D_N} \delta(D_N, f), \tag{8.15}$$

where C, D, U and N denote, respectively, arbitrary cycles, cuts, unicycles and near bases satisfying the stated conditions. An intuitive reading of Δ_{ef} is as a measure of whether cycles containing e, f arising from unicycles tend to traverse e and f in the same or opposite directions; similarly for cuts arising from near bases.

We extend earlier notation in an obvious way, so that $\mathcal{B}_{ef}$ is the set of bases of M containing both e and f, and $\mathcal{B}_{\bar{e}f}$ is the set of bases excluding e but including f, etc.

Theorem 8.16. *The bases $\mathcal{B} = \mathcal{B}(M)$ of an orientable matroid M satisfy $|\mathcal{B}| \cdot |\mathcal{B}_{ef}| = |\mathcal{B}_e| \cdot |\mathcal{B}_f| - \Delta_{ef}^2$.*

Proof. We consider a pair of bases $(X, Y) \in \mathcal{B}_{\bar{e}} \times \mathcal{B}_{ef}$ to be adjacent to a pair $(X', Y') \in \mathcal{B}_e \times \mathcal{B}_{\bar{e}f}$ if (X', Y') can be obtained by an exchange involving e and a second element $g \neq e$:

$$X' = X \cup \{e\} \setminus \{g\} \tag{8.16}$$
$$Y' = Y \cup \{g\} \setminus \{e\}. \tag{8.17}$$

With each adjacent pair we associate a weight

$$\gamma(C_{X\cup\{e\}}, g)\delta(D_{Y\setminus\{e\}}, g). \tag{8.18}$$

Given a pair $(X, Y) \in \mathcal{B}_{\bar{e}} \times \mathcal{B}_{ef}$, the condition that an exchange involving g leads to a valid pair of bases (X', Y') via (8.16) and (8.17) is precisely that the weight (8.18) is non-zero. Note that whenever this occurs, $(X', Y') \in \mathcal{B}_e \times \mathcal{B}_{\bar{e}f}$. Thus

$$\begin{aligned} |\mathcal{B}_{\bar{e}}| \cdot |\mathcal{B}_{ef}| &= \sum_{(X,Y)\in\mathcal{B}_{\bar{e}}\times\mathcal{B}_{ef}} \left[\sum_{g\neq e} \gamma(C_{X\cup\{e\}}, g)\delta(D_{Y\setminus\{e\}}, g) \right] \\ &= W, \end{aligned} \tag{8.19}$$

where W is the total weight of adjacent pairs. Here we have used equation (8.13).

Now we perform a similar calculation, but in the other direction, starting at pairs $(X', Y') \in \mathcal{B}_e \times \mathcal{B}_{\bar{e}f}$. We apply a weight

$$\delta(D_{X'\setminus\{e\}}, g)\gamma(C_{Y'\cup\{e\}}, g) \tag{8.20}$$

to each adjacent pair, which is consistent, by (8.14), with the weight (8.18) applied earlier. Again, starting at (X', Y'), the condition for (X, Y), obtained by inverting the exchange given in (8.16) and (8.17), to be a valid pair of bases is that the weight (8.20) in non-zero. But now, even if (8.20) is non-zero, there remains the possibility that the new pair of bases (X, Y) is not a member of $\mathcal{B}_{\bar{e}} \times \mathcal{B}_{ef}$; this event will occur precisely when $g = f$. Thus

$$
\begin{aligned}
|\mathcal{B}_e| \cdot |\mathcal{B}_{\bar{e}f}| &= \sum_{(X',Y') \in \mathcal{B}_e \times \mathcal{B}_{\bar{e}f}} \Big[\sum_{g \neq e} \delta(D_{X' \setminus \{e\}}, g) \gamma(C_{Y' \cup \{e\}}, g) \Big] && (8.21)\\
&= \sum_{(X',Y') \in \mathcal{B}_e \times \mathcal{B}_{\bar{e}f}} \Big[\sum_{g \neq e, f} \delta(D_{X' \setminus \{e\}}, g) \gamma(C_{Y' \cup \{e\}}, g) \Big] \\
&\quad + \sum_{(X',Y') \in \mathcal{B}_e \times \mathcal{B}_{\bar{e}f}} \delta(D_{X' \setminus \{e\}}, f) \gamma(C_{Y' \cup \{e\}}, f) \\
&= W + \sum_{(X',Y') \in \mathcal{B}_e \times \mathcal{B}_{\bar{e}}} \delta(D_{X' \setminus \{e\}}, f) \gamma(C_{Y' \cup \{e\}}, f) && (8.22)\\
&= W + \sum_{X' \in \mathcal{B}_e} \delta(D_{X' \setminus \{e\}}, f) \sum_{Y' \in \mathcal{B}_{\bar{e}}} \gamma(C_{Y' \cup \{e\}}, f) \\
&= W + \Delta_{ef}^2. && (8.23)
\end{aligned}
$$

Here, step (8.21) is by (8.13); step (8.22) uses the observation that terms are non-zero only when $f \in Y'$; and (8.23) is from the definition (8.15) of Δ_{ef}.

Comparing (8.19) and (8.23) we obtain

$$|\mathcal{B}_e| \cdot |\mathcal{B}_{\bar{e}f}| = |\mathcal{B}_{\bar{e}}| \cdot |\mathcal{B}_{ef}| + \Delta_{ef}^2,$$

and the result now follows by adding $|\mathcal{B}_e| \cdot |\mathcal{B}_{ef}|$ to both sides. □

The main result of the section now follows easily.

Proof of Theorem 8.15. According to Theorem 8.16, all orientable matroids satisfy the negative correlation property. Moreover, it is easily checked that the class of orientable matroids is closed under contraction and deletion. □

Exercise 8.17. Proving that class of orientable matroids is the same as the class of regular matroids requires familiarity with matroid theory. However, the weaker claim that the cycle matroid of any graph is orientable is an exercise in straight combinatorics. Prove the claim.

Exercise 8.18. Another way to demonstrate that all graphic matroids are balanced is via the theory of electrical networks. Regard a graph $G = (V, E)$ as an electrical network, with vertices as terminals and edegs as unit resistors. The key facts are: (1) For any edge $e = \{u, v\}$, the effective between vertices u and v is equal to $\tau(G/e)/\tau(G)$, where $\tau(G)$ is the number of spanning trees in G, and $\tau(G/e)$ is the number of spanning trees in G that include the edge e. This result is essentially due to Kirchhoff; see Van Lint and Wilson [79, Thm. 34.3]. (2) If the resistance of some edge of a network is decreased, the effective resistance between any two terminals does not increase. This is "Rayleigh's Monotonicity Principle"; see Doyle and Snell [25].

Example 8.19. From the matroid-theoretic fact that graphic matroids are regular, or from Exercise 8.17, or indeed from Exercise 8.18, we know that graphic matroids are balanced. Let $G = (V, E)$ be a connected, undirected graph, and consider the following random walk on the spanning trees of G: Suppose the current state (tree) is $T \subseteq E$. Choose an edge e u.a.r. from E, and an edge f u.a.r. from T. If $T' = T \cup \{e\} \setminus \{f\}$ is a spanning tree then move to T', otherwise remain at T. The random walk just defined is the bases-exchange walk on a balanced matroid and, by Theorem 8.10, the spectral gap of this walk is $\Omega(1/mn)$, where $n = |V|$ and $m = |E|$. Thus, the mixing time of this natural random walk on spanning trees of a graph is just $O(mn^2 \log m)$. This is not a bad result, but we'll improve it further in the next chapter.

Remark 8.20. Regular matroids are always balanced, but not all balanced matroids are regular. The *uniform matroid* $U_{r,m}$ of rank r on a ground set E of size m has as its bases all subsets of E of size r. It is easy to check that all uniform matroids satisfy the negative correlation property and that the class of uniform matroids is closed under contraction and deletion; on the other hand, $U_{2,m}$ is not regular when $m \geq 4$. (Refer to Oxley [66, Theorem 13.1.1].)

Remark 8.21. Graphic matroids are always regular, but not all regular matroids are graphic. Let $G = (V, E)$ be an undirected graph. The *co-cycle matroid of* G again has ground set E but the bases are now complements (in E) of spanning trees. The relationship of the cycle and co-cycle matroids of G is a special case of a general one of *duality*. The co-graphic matroid of a non-planar graph is regular but not graphic.

Remark 8.22. The number of bases of a regular matroid may be computed exactly in polynomial time (in m) by an extension of Kirchhoff's Matrix-tree

Theorem. It can be shown that the bases of a regular matroid are in 1-1 correspondence with the non-singular $r \times r$ submatrices of an $r \times m$ unimodular matrix, and that the number of these can be computed using the Binet-Cauchy formula. Refer to Dyer and Frieze [26, §3.1] for a discussion of this topic. This approach gives alternative polynomial-time sampling procedure for bases of a regular matroid, not relying on Markov chain simulation. However, as we have seen, the class of balanced matroids is strictly larger than the class of regular matroids.

Remark 8.23. Following on from the previous remark, there exists a subclass of balanced matroids, the "sparse paving matroids", whose bases are hard to count exactly. A little less informally, the problem of counting bases of a sparse paving matroids is #P-hard. For a precise statement of this claim and a proof, refer to Jerrum [42].

Example 8.24. There exist non-balanced matroids. Let M be a matroid of rank r on ground set E. For any $0 < r' < r$,

$$\mathcal{B}' = \{X' : |X'| = r' \land \exists X \in \mathcal{B}(M).\, X' \subset X\}$$

is the collection of bases of a matroid M' on ground set E, the *truncation* of M to rank r'. The truncation of a graphic matroid may fail to be balanced. Consider the graph G with vertex set

$$\{u, v, y, z, 0, 1, 2, 3, 4\}$$

and edge set

$$\big\{\{u, v\}, \{y, z\}\big\} \cup \big\{\{u, i\} : 0 \le i \le 4\big\} \cup \big\{\{v, i\} : 0 \le i \le 4\big\}.$$

Let e denote the edge $\{u, v\}$ and f the edge $\{y, z\}$. Let $\mathcal{F}^6$ denote the set of forests in G with six edges, $\mathcal{F}^6_{ef}$ the number of such forests including edges e and f, etc. Then $\mathcal{F}^6_{ef} = 80$, $\mathcal{F}^6_{e\bar{f}} = 32$, $\mathcal{F}^6_{\bar{e}f} = 80$ and $\mathcal{F}^6_{\bar{e}\bar{f}} = 192$. Thus

$$\Pr(e \mid f) = 5/17 > 7/24 = \Pr(e),$$

contradicting negative correlation.