Chapter 1

Sampling and Counting

1.1 Introduction

The classical *Monte Carlo method* is an approach to estimating quantities that are hard to compute exactly. The quantity z of interest is expressed as the expectation $z = \mathbf{E}(Z)$ of a random variable (r.v.) Z over a probability space Ω, μ. It is assumed that some efficient procedure for sampling from Ω, μ is available. By taking the mean of some sufficiently large set of independent samples of Z, one may obtain an approximation to z. For example, suppose

$$S = \left\{(x, y) \in [0, 1]^2 : p_i(x, y) \leq 0, \text{ for all } i\right\}$$

is some region of the unit square defined by a system of polynomial inequalities $p_i(x, y) \leq 0$. Let Z be the r.v. defined by the following experiment or trial: choose a point (x, y) uniformly at random (u.a.r.) from $[0, 1]^2$; let $Z = 1$ if $p_i(x, y) \leq 0$ for all i, and $Z = 0$ otherwise. Then the area a of S is equal to $\mathbf{E}(Z)$, and an estimate of a may be obtained from the sample mean of a sufficiently long sequence of trials. In this example, the use of the Monte Carlo method is perhaps avoidable, at the expense of a more complex algorithm; for more essential uses, see, for example, Knuth's proposal for estimating the size of a tree by taking a random path from the root to a leaf, or Rasmussen's for estimating the permanent of a 0,1-matrix.

The main focus of this book is the *Markov chain* Monte Carlo (MCMC) method which is a development of the foregoing approach, which is sometimes applicable when Z cannot be sampled "directly". Z will often be the cardinality of some combinatorially defined set S. We design a Markov Chain $\mathfrak{M}$ with state space Ω (often S itself) whose steady state distribution is μ. Efficient sampling now rests on the rapid convergence of the chain to its steady state. These ideas will be made more explicit in Chapter 2 but for the moment we focus on the relationship between near uniform generation and approximate counting.

As a first example of the approach, we consider the problem of estimating the number of independent sets of a graph G with small maximum degree Δ. In Section 1.2 we show how sampling independent sets of G, generated independently and almost uniformly, can be used to obtain an estimate for their *number*. This step of the MCMC programme—how samples are used—is often (though not always) rather routine.

We then consider the reverse process i.e. we show how good estimates of the number of independent sets can be used to generate a near uniform sample. This illustrates a sort of equivalence between the problems of generation and counting. Section 1.4 discusses a formal framework within which this can be made precise.

1.2 Approximate counting, uniform sampling and their relationship

1.2.1 An example – Independent Sets

What do we mean precisely by (efficient) approximate counting and uniform sampling?

let $N = N(G)$ denote the number of independent sets of G. A *randomised approximation scheme* for N is a randomised algorithm that takes as input a graph G and an error bound $\varepsilon > 0$, and produces as output a number Y (a random variable) such that

$$\mathbf{Pr}\big((1-\varepsilon)N \le Y \le (1+\varepsilon)N\big) \ge \frac{3}{4}. \tag{1.1}$$

A randomised approximation scheme is said to be *fully polynomial* if it runs in time polynomial in n (the input length) and ε^{-1}. We shall abbreviate the rather unwieldy phrase "fully polynomial randomised approximation scheme" to FPRAS.

There is no significance in the constant $\frac{3}{4}$ appearing in the definition, beyond its lying strictly between $\frac{1}{2}$ and 1. Any success probability greater than $\frac{1}{2}$ may be boosted to $1-\delta$ for any desired $\delta > 0$ by performing a small number of trials and taking the median of the results; the number of trials required is $\mathrm{O}(\ln \delta^{-1})$. Indeed let $Y_1, Y_2, \ldots, Y_m$ be independent samples satisfying (1.1). Suppose that $\tilde{Y}$ is the median of $Y_1, Y_2, \ldots, Y_m$. Then

$$\mathbf{Pr}(\tilde{Y} \ge (1+\varepsilon)N) \le \mathbf{Pr}(|\{i :\ Y_i \ge (1+\varepsilon)N\}| \ge m/2) \le e^{-m/12}$$

using the Chernoff bounds. Similarly

$$\mathbf{Pr}(\tilde{Y} \ge (1-\varepsilon)N) \le e^{-m/8}.$$

Putting $m = \lceil 12 \ln(2/\delta) \rceil$ we get

$$\mathbf{Pr}\big((1-\varepsilon)N \le \tilde{Y} \le (1+\varepsilon)N\big) \ge 1-\delta. \tag{1.2}$$

For any two probability distributions π and π' on a countable set Ω, define the *total variation distance* between π and π' to be

$$D_{\mathrm{tv}}(\pi, \pi') := \max_{A \subseteq \Omega} |\pi(A) - \pi'(A)| = \frac{1}{2} \sum_{x \in \Omega} |\pi(x) - \pi'(x)|. \tag{1.3}$$

In our example $\Omega = \Omega(G)$ will be the set of independent sets of graph G and $\pi(I) = \frac{1}{|\Omega|}$ for each $I \in \Omega$ i.e. π is the uniform distribution over Ω. We will let π' be the distribution of the output of some randomised algorithm that generates a random independent subset of G.

A *good sampler* for Ω is a randomised algorithm that takes as input a graph G and a tolerance $\delta > 0$, and produces an independent set I (a random variable) such that the probability distribution of I is within variation distance δ of the uniform distribution on Ω. An almost uniform sampler is said to be *fully polynomial* if it runs in time polynomial in n (the input length) and $\log \delta^{-1}$.

From good sampling to approximate counting

Theorem 1.2.1 *Suppose we have a good sampler for the independent sets of a graph, which works for graphs G with maximum degree bounded by Δ and suppose that the sampler has time complexity $T(n, \delta)$, where n is the number of vertices in G, and δ the allowed deviation from uniformity in the sampling distribution. Then we may construct an FPRAS for the number of independent sets of a graph, which works for graphs G with maximum degree bounded by Δ, and which has time complexity*

$$O\left(\frac{m^2}{\varepsilon^2}\, T\Big(n, \frac{\varepsilon}{6m}\Big)\right),$$

where m is the number of edges in G, and ε the specified error bound.

Proof Let $G = G_m > G_{m-1} > \cdots > G_1 > G_0 = (V, \emptyset)$ be any sequence of graphs in which each graph G_{i-1} is obtained from the previous graph G_i by removing a single edge. We may express the quantity we wish to estimate as a product of ratios:

$$|\Omega(G)| = \frac{|\Omega(G_m)|}{|\Omega(G_{m-1})|} \times \frac{|\Omega(G_{m-1})|}{|\Omega(G_{m-2})|} \times \cdots \times \frac{|\Omega(G_1)|}{|\Omega(G_0)|} \times |\Omega(G_0)|, \tag{1.4}$$

where, it will be observed, $|\Omega(G_0)| = 2^n$. Our strategy is to estimate the ratio

$$\varrho_i = \frac{|\Omega(G_i)|}{|\Omega(G_{i-1})|}$$

for each i in the range $1 \leq i \leq m$, and by substituting these quantities into identity (1.4), obtain an estimate for the number of independent sets of G:

$$|\Omega(G)| = 2^n \varrho_1 \ldots \varrho_m. \tag{1.5}$$

To estimate the ratio ϱ_i we use the almost uniform sampler to obtain a sufficiently large sample of independent sets from $\Omega(G_{i-1})$ and compute the proportion of samples that lie in $\Omega(G_i)$.

The following lemma gives the basic probabilistic inequality we need.

Lemma 1.2.1 *For $i = 1, 2, \ldots, m$ let $0 \leq Z_i \leq 1$ be independent random variables on the probability space (Ω_i, π_i) where $\mathbf{E}(Z_i) = \mu_i$ and $\mu_{\min} = \min_i \mu_i > 0$.*

For $i = 1, 2, \ldots, m$ let $\hat{Z}_i$ denote the same random variable on the probability space $(\Omega_i, \hat{\pi}_i)$ where

$$d_{TV}(\pi_i, \hat{\pi}_i) \leq \delta = \frac{\varepsilon}{3m}\mu_{\min}.$$

For $i = 1, 2, \ldots, m$ let $\hat{\mu}_i = \mathbf{E}(\hat{Z}_i)$ and let $\hat{Z}_i^{(1)}, \ldots, \hat{Z}_i^{(s)}$ be a sequence of

$$s = \lceil 17m\mu_{\min}^{-2}\varepsilon^{-2} \rceil$$

independent copies of the random variable $\hat{Z}_i$ and let $\bar{Z}_i = s^{-1}\sum_{j=1}^{s} \hat{Z}_i^{(j)}$ be their mean. Let

$$W = \frac{\bar{Z}_1\bar{Z}_2\cdots\bar{Z}_m}{\mu_1\mu_2\cdots\mu_m}.$$

Then, for ϵ sufficiently small,

$$\mathbf{Pr}(|W - 1| \geq \varepsilon) \leq \frac{1}{4}.$$

Proof Note first that for $i = 1, 2, \ldots, m$,

$$|\hat{\mu}_i - \mu_i| \leq \delta \text{ and } \mathbf{Var}(\hat{Z}_i) \leq 1. \tag{1.6}$$

Let

$$\bar{W} = \frac{\bar{Z}_1\bar{Z}_2\cdots\bar{Z}_m}{\hat{\mu}_1\hat{\mu}_2\cdots\hat{\mu}_m}.$$

Now $\mathbf{E}(\bar{W}) = 1$ and (1.6) implies

$$\left(1 - \frac{\delta}{\mu_{\min}}\right)^m \leq \frac{W}{\bar{W}} \leq \left(1 + \frac{\delta}{\mu_{\min}}\right)^m.$$

So,

$$\left|\frac{W}{\bar{W}} - 1\right| \leq \frac{2\epsilon}{5}. \tag{1.7}$$

Furthermore

$$\begin{aligned}\mathbf{Var}(\bar{W}) &= \mathbf{E}\left(\prod_{i=1}^{m}\frac{\bar{Z}_i^2}{\hat{\mu}_i^2}\right)-1 \qquad (1.8)\\ &= \prod_{i=1}^{m}\left(1+\frac{\mathbf{Var}(\bar{Z}_i)}{\hat{\mu}_i^2}\right)-1\\ &= \prod_{i=1}^{m}\left(1+\frac{\mathbf{Var}(\hat{Z}_i)}{s\hat{\mu}_i^2}\right)-1\\ &\le \prod_{i=1}^{m}\left(1+\frac{1}{s\hat{\mu}_i^2}\right)-1\\ &\le \left(1+\frac{\epsilon^2}{17m}\right)^m-1\\ &\le \frac{\epsilon^2}{16}. \qquad (1.9)\end{aligned}$$

Thus by (1.7) and (1.9),

$$\mathbf{Pr}(|W-1|\ge\epsilon)\le\mathbf{Pr}(|\bar{W}-1|\ge\tfrac{\epsilon}{2})\le\frac{4}{\epsilon^2}\mathbf{Var}(\bar{W})\le\frac{1}{4}.$$

□

Suppose that the graphs G_i and G_{i-1} differ in the edge $\{u,v\}$, which is present in G_i but absent from G_{i-1}. Clearly, $\Omega(G_i)\subseteq\Omega(G_{i-1})$. Any independent set in $\Omega(G_{i-1})\setminus\Omega(G_i)$ contains u and v, and may be perturbed to an independent set in G_i by deleting vertex u. (To resolve ambiguity, let u be the smaller of the two vertices.) On the other hand, each independent set in G_i can be obtained in at most one way as the result of such a perturbation; hence $|\Omega(G_{i-1})\setminus\Omega(G_i)|\le|\Omega(G_i)|$ and

$$\frac{1}{2}\le\varrho_i\le 1. \qquad (1.10)$$

To avoid trivialities, assume $0<\varepsilon\le 1$ and $m\ge 1$. Let $Z_i\in\{0,1\}$ denote the random variable which results from choosing a random independent set from G_{i-1} and returning one if the resulting independent set is also independent in G_i and zero otherwise. Note that $\mu_i=\mathbf{E}(Z_i)=\varrho_i$ for $i=1,2,\ldots,m$. Let $\hat{Z}_i$ denote the random variable which results from running the postulated almost uniform sampler on the graph G_{i-1} and returning one if the resulting independent set is also independent in G_i and zero otherwise. We take $\delta=\frac{\epsilon}{6m}$ (in the sampler) and $s=\lceil 68m\epsilon^{-2}\rceil$. Let $Z_i^{(1)},\ldots,Z_i^{(s)}$ be a sequence of s independent copies of the random variable $\hat{Z}_i$. As our estimator for $|\Omega(G)|$, we use the random variable $Y=2^n\,\overline{Z}_1\overline{Z}_2\ldots\overline{Z}_m$. Applying Lemma 1.2.1 we see immediately that

$$\mathbf{Pr}\left(\left|\frac{Y}{|\Omega(G)|}-1\right|\ge\epsilon\right)\le\frac{1}{4}.$$

We use $s = O(m\epsilon^{-2})$ samples to estimate each ϱ_i and the time bound claimed in the theorem follows. □

From approximate counting to good sampling

Theorem 1.2.2 *Suppose that we have an FPRAS* APPROXCOUNT(G, ϵ, δ) *for the number of independent sets of a graph* $G = (V, E)$ *with maximum degree* Δ *and suppose that* APPROXCOUNT(G, ϵ, δ) *has time complexity* $T(n, \epsilon, \delta)$ *where* $n = |V|$*,* ϵ *is the required maximum relative error and* δ *is the allowed probability of failure. Then we can construct a good sampler* UGEN(G, δ) *for the independent sets of* G *with maximum degree* Δ *which has expected time complexity*

$$O\left(T\left(n, O\left(\frac{1}{n}\right), O\left(\frac{\delta}{n}\right)\right)\right). \tag{1.11}$$

Proof We will call our sampling procedure UGEN(G, δ): let

$$\delta_1 = \frac{\delta}{2n+1} \text{ and } \epsilon_1 = \frac{\log 2}{3n}.$$

UGEN(G, δ)
begin
  $N =$ APPROXCOUNT$(G, \epsilon_1, \delta_1)$
  Repeat until $I =$ UGENX$(G, \epsilon_1, \frac{1}{4N}) \neq \perp$
  Output I.
end

The precedure UGENX has an extra parameter ϕ which is needed to control the rate of some rejection sampling. We define UGENX recursively.

UGENX(G, ϵ_1, ϕ)
begin
  If $\phi > 1$ **then** output $I = \perp$ – failure.
  If $V = \emptyset$ **then** $I = \begin{cases} \emptyset & \text{probability } \phi \\ \perp & \text{probability } 1-\phi \end{cases}$
  else begin
    $v = \max V$ and X is the set of neighbours of v in G.
    $G_1 = G - v - X$ and $G_2 = G - v$
    $N_1 =$ APPROXCOUNT$(G_1, \epsilon_1, \delta_1)$ and $N_2 =$ APPROXCOUNT$(G_2, \epsilon_1, \delta_1)$

$$\textbf{Output } I = \begin{cases} v + \text{UGENX}\left(G_1, \epsilon_1, \phi\frac{N_1+N_2}{N_1}\right) & \text{probability } \frac{N_1}{N_1+N_2} \\ \text{UGENX}\left(G_2, \epsilon_1, \phi\frac{N_1+N_2}{N_2}\right) & \text{probability } \frac{N_2}{N_1+N_2} \end{cases}$$

end
end

For $I \in \Omega$ let p_I denote the probability that $Ugenx(G, \epsilon_1, \phi)$ generates I, conditional on all calls to APPROXCOUNT being successful. Then we will see that $\phi \leq p_I$ and at the bottom of the recursion, ϕ will have become ϕ/p_i and so I will be output with (conditional) probability $p_I \times \phi/p_I = \phi$ i.e. the conditional output is uniform.

Lemma 1.2.2 (a) *The probability that* APPROXCOUNT *gives a bad estimate during the execution of* UGEN *is at most* $(2n+1)\delta_1$.

(b *If* APPROXCOUNT *gives no bad estimates then* $\phi \leq 1$ *throughout the execution of* UGEN.

(c) *If* APPROXCOUNT *gives no bad estimates then the probability* UGEN *outputs* $\perp$ *is at most 2/3.*

(d) *If* APPROXCOUNT *gives no bad estimates then the output* I *is such that for any independent set* I_0 *of* G *we have* $\mathbf{Pr}(I = I_0) = \phi$.

(e) *Let* $\hat{\pi}$ *be the distribution of the output* I *of* UGEN *and let* π *denote the uniform distribution on* Ω. *Then* $D_{\mathrm{tv}}(\pi, \hat{p}) \leq \delta$.

Proof (a) This is clear from the fact that we call APPROXCOUNT at most $2n+1$ times during the execution of UGEN.

(b) If there is no bad estimate from APPROXCOUNT then we claim by induction on the depth of recursion d that whenever we invoke UGENX on a graph H, say, then we find the current value of ϕ, $\phi_d \leq \frac{(1+\epsilon_1)^d}{4(1-\epsilon_1)^{d+1}|\Omega(H)|}$. This is trivially true for $d = 0$ and assuming say that we recurse on H_1 we have in this call

$$\phi_{d+1} \leq \frac{(1+\epsilon_1)^d}{4(1-\epsilon_1)^{d+1}|\Omega(H)|}\frac{N_1+N_2}{N_1} \leq \frac{(1+\epsilon_1)^d}{4(1-\epsilon_1)^{d+1}|\Omega(H)|} \times \frac{(1+\epsilon_1)|\Omega(H)|}{(1-\epsilon_1)|\Omega(H_1)|}$$
$$= \frac{(1+\epsilon_1)^{d+1}}{4(1-\epsilon_1)^{d+2}|\Omega(H_1)|}$$

as required.

Thus throughout the execution of UGEN we have $\phi \leq \frac{(1+\epsilon_1)^n}{4(1-\epsilon_1)^{n+1}} < e^{n\epsilon_1}/2 < 1$.

(c) We prove by induction on $|V|$ that $\mathbf{Pr}(I = \perp) \leq 1 - \phi|\Omega(G)|$. This is clearly true if $V = \emptyset$. Otherwise

$$\mathbf{Pr}(I = \perp) \leq$$
$$\frac{N_1}{N_1+N_2}(1 - \phi\frac{N_1+N_2}{N_1}|\Omega(G_1)|) + \frac{N_2}{N_1+N_2}(1 - \phi\frac{N_1+N_2}{N_2}|\Omega(G_2)|)$$
$$= 1 - \phi|\Omega(G)|.$$

Thus $\mathbf{Pr}(I = \perp) \leq 2/3$ as required.

(d) This is clearly true if $V = \emptyset$. If $V \neq \emptyset$ and $v = \max V \in I_0$ then, by induction

$$\mathbf{Pr}(I = I_0) = \frac{N_1}{N_1 + N_2}\phi\frac{N_1 + N_2}{N_1} = \phi$$

and similarly $\mathbf{Pr}(I = I_0) = \phi$ if $v \notin I_0$.

(e) Let $\mathcal{E}$ denote the event that some output of APPROXCOUNT is bad in the iteration that produces output. Then for $A \subseteq \Omega$,

$$\begin{aligned}\hat{\pi}(A) - \pi(A) &\leq \mathbf{Pr}(I \in A \mid \bar{\mathcal{E}}) + \mathbf{Pr}(\mathcal{E}) - \pi(A)\\ &\leq \frac{|A|}{|\Omega|} + \delta - \frac{|A|}{|\Omega|}\\ &\leq \delta.\end{aligned}$$

□

We have therefore shown that by running UGENX for *constant* expected number of times, we will with probability at least $1 - \delta$ output a randomly chosen independent set. The expected running time of UGEN is clearly as given in (1.11) which is small enough to make it a good sampler.

Having dealt with a specific example we see how to put the above ideas into a formal framework. Before doing this we enumerate some basic facts about Markov Chains.

1.3 Markov Chains

Throughout $\mathbb{N} = \{0, 1, 2, \ldots\}$, $\mathbb{N}_+ = \mathbb{N} \setminus \{0\}$, $\mathbb{Q}_+ = \{q \in \mathbb{Q} : q > 0\}$, and $[n] = \{1, 2, \ldots, n\}$ for $n \in \mathbb{N}_+$.

A Markov chain $\mathcal{M}$ on the finite state space Ω, with transition matrix P is a sequence of random variables X_t, $t = 0, 1, 2, \ldots$, which satisfy

$$\mathbf{Pr}(X_t = \sigma \mid X_{t-1} = \omega, X_{t-2}, \ldots, X_0) = P(\omega, \sigma) \qquad (t = 1, 2, \ldots),$$

We sometimes write P_σ^ω. The value of X_t is referred to as the *state* of $\mathcal{M}$ at *time* t.

Consider the digraph $D_{\mathcal{M}} = (\Omega, A)$ where $A = \{(\sigma, \omega) \in \Omega \times \Omega : P(\sigma, \omega) > 0\}$. We will by and large be concerned with chains that satisfy the following assumptions:

M1 The digraph $D_{\mathcal{M}}$ is strongly connected.

M2 $\gcd\{|C| : C \text{ is a directed cycle of } D_{\mathcal{M}}\} = 1$

Under these assumptions, $\mathcal{M}$ is *ergodic* and therefore has a unique stationary distribution π i.e.

$$\lim_{t\to\infty} \mathbf{Pr}(X_t = \omega \mid X_0 = \sigma) = \pi(\omega) \tag{1.12}$$

i.e. the limit does not depend on the starting state X_0. Furthermore, π is the unique left eigen-vector of P with eigenvalue 1 i.e. satisfying

$$P^T\pi = \pi. \tag{1.13}$$

Another useful fact is that if τ_σ denotes the expected number of steps between successive visits to state σ then

$$\tau_\sigma = \frac{1}{\pi(\sigma)}. \tag{1.14}$$

In most cases of interest, $\mathcal{M}$ is *reversible*, i.e.

$$Q(\omega,\sigma) = \pi(\omega)P(\omega,\sigma) = \pi(\sigma)P(\sigma,\omega) \qquad (\forall\omega,\sigma\in\Omega). \tag{1.15}$$

The central role of reversible chains in applications rests on the fact that π can be deduced from (1.15). If $\mu : \Omega \longrightarrow \mathbb{R}$ satisfies (1.15), then it determines π up to normalization. Indeed, if (1.15) holds and $\sum_{\omega\in\Omega}\pi(\omega) = 1$ then

$$\sum_{\omega\in\Omega}\pi(\omega)P(\omega,\sigma) = \sum_{\omega\in\Omega}\pi(\sigma)P(\sigma,\omega) = \pi(\sigma)$$

which proves that π is a left eigenvector with eigenvalue 1.

In fact, we often *design* the chain to satisfy (1.15). Without reversibility, there is no apparent method of determining π, other than to explicitly construct the transition matrix, an exponential time (and space) computation in our setting.

As a canonical example of a reversible chain we have a *random walk* on a graph. A random walk on the undirected graph $G = (V, E)$ is a Markov chain with state space V associated with a particle that moves from vertex to vertex according to the following rule: the probability of a transition from vertex i, of degree d_i, to vertex j is $\frac{1}{d_i}$ if $\{i, j\} \in E$, and 0 otherwise. Its stationary distribution is given by

$$\pi(v) = \frac{d_v}{2|E|} \qquad\qquad v \in V. \tag{1.16}$$

To see this note that $Q(v, w) = Q(w, v)$ if v, w are not adjacent and otherwise

$$Q(v,w) = \frac{1}{2|E|} = Q(w,v),$$

verifying the *detailed balance* equations (1.15).

Note that if G is a regular graph then the steady state is uniform over V.

If G is bipartite then the walk as described is not ergodic. This is because all cycles are of even length. This is usually handled by adding d_v loops to vertex v for each vertex v. (Each loop counts as a single exit from v.) The net effect of this is to make the particle stay put with probability $\frac{1}{2}$ at each step. The steady state is unaffected. The chain is now *lazy*.

A chain is lazy if $P(\omega,\omega) \geq \frac{1}{2}$ for all $\omega \in \Omega$.

If $p_0(\omega) = \mathbf{Pr}(X_0 = \omega)$, then $p_t(\sigma) = \sum_\omega p_0(\omega)P^t(\omega,\sigma)$ is the distribution at time t. As a measure of convergence, the natural choice in this context is variation distance.

The *mixing time* of the chain is then

$$\tau(\varepsilon) = \max_{p_0} \min_t \{D_{\mathrm{tv}}(p_t,\pi) \leq \varepsilon\},$$

and it is easy to show that the maximum occurs when $X_0 = \omega_0$, with probability one, for some state ω_0. This is because $D_{\mathrm{tv}}(p_t,\pi)$ is a convex function of p_0 and so the maximum of $D_{\mathrm{tv}}(p_t,\pi)$ occurs at an extreme point of the set of probabilities p_0.

I think this should be moved to the next chapter

We now provide a simple lemma which indicates that variation distance $D_{\mathrm{tv}}(p_t,\pi)$ goes to zero exponentially. We define several related quantities: $p_t^{(i)}$ denotes the t-fold distribution, conditional on $X_0 = i$.

$$d_i(t) = D_{\mathrm{tv}}(p_t^{(i)},\pi),\ d(t) = \max_i d_i(t),\ \bar{d}(t) = \max_{i,j} D_{\mathrm{tv}}(p_t^{(i)},p_t^{(j)}).$$

Lemma 1.3.1 *For all $s,t \geq 0$,*

(a) $\bar{d}(s+t) \leq \bar{d}(s)\bar{d}(t)$.

(b) $d(s+t) \leq 2d(s)d(t)$.

(c) $d(s) \leq 2\bar{d}(s)$.

(d) $d(s) \leq d(t)$ *for* $s \leq t$.

Proof We will use the characterisation of variation distance as

$$D_{\mathrm{tv}}(\mu_1,\mu_2) = \min \mathbf{Pr}(X_1 \neq X_2) \tag{1.17}$$

where the minimum is taken over pairs of random variables X_1, X_2 such that X_i has distribution $\mu_i, i = 1,2$.

Fix states i_1, i_2 and times s,t and let Y^1, Y^2 denote the chains started at i_1, i_2 respectively. By (1.17) we can construct a joint distribution for (Y_s^1, Y_s^2) such that

$$\mathbf{Pr}(Y_s^1 \neq Y_s^2) = D_{\mathrm{tv}}(p_s^{(i_1)}, p_s^{(i_2)}) \leq \bar{d}(s).$$

Now for each pair j_1, j_2 we can use (1.17) to construct a joint distribution for (Y^1_{s+t}, Y^2_{s+t}) such that

$$\mathbf{Pr}(Y^1_{s+t} \neq Y^2_{s+t} \mid Y^1_s = j_1, Y^2_s = j_2) = D_{\mathrm{tv}}(p^{(j_1)}_t, p^{(j_2)}_t).$$

The RHS is 0 if $j_1 = j_2$ and otherwise at most $\bar{d}(t)$. So, unconditionally,

$$\mathbf{Pr}(Y^1_{s+t} \neq Y^2_{s+t}) \leq \bar{d}(s)\bar{d}(t)$$

and (1.17) establishes part (a) of the lemma.

For part (b), the same argument, with Y^2 now being the stationary chain shows

$$d(s+t) \leq d(s)\bar{d}(t) \tag{1.18}$$

and so (b) will follow from (c), which follows from the triangular inequality for variation distance. Finally note that (d) follows from (1.18). □

We will for the most part use carefully defined Markov chains as our good samplers. As an example, we now define a simple chain with state space Ω equal to the collection of independent sets of a graph G. The chain is ergodic and its steady state is uniform over Ω. So, running the chain for sufficiently long will produce a near uniformly chosen independent set, see (1.12). Unfortunately, this chain does not have a small enough mixing time for this to qualify as a good sampler, unless $\Delta(G) \leq 4$.

We define the chain as follows: suppose $X_t = I$. Then we choose a vertex v of G uniformly at random. If $v \in I$ then we put $X_{t+1} = I \setminus \{v\}$. If $v \notin I$ and $I \cup \{v\}$ is an indepedent set then we put $X_{t+1} = I \cup \{v\}$. Otherwise we let $X_{t+1} = X_t = I$. Thus the transition matrix can be described as follows: $n = |V|$ and I, J are independent sets of G.

$$P(I,J) = \begin{cases} \frac{1}{n} & |I \Delta J| = 1 \\ 0 & \text{otherwise} \end{cases}$$

Here $I \Delta J$ denotes the symmetric difference $(I \setminus J) \cup (J \setminus I)$.

The chain satisfies M1 and M2: In $D_{\mathcal{M}}$ every vertex can reach and is reachable from $\emptyset$, implying M1 holds. Also, $D_{\mathcal{M}}$ contains loops unless G has no edges. In both cases M2 holds trivially.

Note finally that $P(I,J) = P(J,I)$ and so (1.15) holds with $\pi(I) = \frac{1}{|\Omega|}$. Thus the chain is reversible and the steady state is uniform.

1.4 A formal computational framework

The sample spaces we have in mind are sets of combinatorial objects. However, in order to discuss the computational complexity of generation, it is necessary to consider a sequence of instances of increasing size. We therefore work within the following formal

framework. The models of computation are the Turing Machine (TM) for deterministic computations and the Probabilistic Turing Machine (PTM) for randomized computations. (A PTM is a TM with a source of uniform and independent random bits.) We must confine ourselves to some class of distributions which are "easily described", from a computational viewpoint, in large instances. We identify this below with a class of unnormalized measures which we call "weight functions".

Let Σ be a fixed alphabet of at least two symbols, and $W : \Sigma^* \times \Sigma^* \longrightarrow \mathbb{N}$ be such that, for some polynomial b, $W(\sigma, \omega) = 0$ unless $|\omega| \leq b(|\sigma|)$. Moreover $W(\sigma, \omega)$ must be computable in time polynomial in $|\sigma|$ whenever $W(\sigma, \omega) > 0$. (If the TM for W may ignore part of its input, this implies that W is *always* computable in polynomial time.) Let us call W a *weight function*. Here σ may be thought of as an encoding of an instance of some combinatorial problem, and the ω of interest are encodings of the structures we wish to generate.

Let $\Omega_\sigma = \{\omega : W(\sigma, \omega) > 0\}$. Then the sequence of discrete probability spaces determined by W is $(\Omega_\sigma, \pi_\sigma)$, where π_σ is the *density*

$$\pi_\sigma(\omega) = W(\sigma, \omega)/Z(\sigma), \quad \text{with} \quad Z(\sigma) = \sum_{\omega' \in \Omega_\sigma} W(\sigma, \omega')$$

being the corresponding *normalising function*. It is easy to see that the class of normalising functions so defined is essentially Valiant's class #P. The definition implies that, for some fixed $c \in \mathbb{N}$, $|\Omega_\sigma| \leq Z(\sigma) \leq 2^{|\sigma|^c}$. If $Z(\sigma) = 0$, then $\Omega_\sigma = \emptyset$ and π_σ is the unique (improper) measure on Ω_σ.

In our definition, two distinct weight functions may define the same sequence of spaces. Therefore let us say weight functions W_1, W_2 are *equivalent* if there exists $\kappa : \Sigma^* \longrightarrow \mathbb{Q}_+$ so that $W_2(\sigma, \omega) = \kappa(\sigma) W_1(\sigma, \omega)$ $(\forall \sigma, \omega \in \Sigma^*)$. Then there is a bijection between sequences of probability spaces $(\Omega_\sigma, \pi_\sigma)$ and equivalence classes of weight functions. Thus, if we write $\widetilde{W}$ for the equivalence class containing W, we may identify it with the sequence $(\Omega_\sigma, \pi_\sigma)$.

We insist that sample spaces are discrete, and weight functions are integer valued. Computationally, discrete spaces are essential. If we wish to work with continuous spaces, then approximations must be made to some predetermined number of bits. The same is true if we are interested in real-valued densities (as in some statistical applications). However, the effect of such approximations can be absorbed into the variation distance of the sampling procedure. The reader may still wonder why we require W to have codomain $\mathbb{N}$ rather than $\mathbb{Q}$, which would seem more natural. This is because we use unnormalised measures, and we wish to avoid the following technical difficulty. In a large sample space it is possible to specify polynomial size rationals for the unnormalised measure which result in exponential size rationals for the probabilities. An example is the set $[2^n]$, with the measure assigning probability proportional to $1/i$ to $i \in [2^n]$. In such spaces there is no possibility of *exact* sampling in sub-exponential expected time, and

we must accept approximations. We prefer not to deal with these anomalous spaces, but to insist that these approximations be made explicit. Thus, in this example we could use weights $\lfloor K/i \rfloor$ for some suitably large integer K.

A *fully polynomial approximate sampler* (which we shorten to *good sampler*) for $(\Omega_\sigma, \pi_\sigma)$ is a PTM which, on inputs σ and $\varepsilon \in \mathbb{Q}_+$ $(0 < \varepsilon \leq 1)$, outputs $\omega \in \Sigma^*$, according to a measure μ_σ satisfying $D_{\mathrm{tv}}(\mu_\sigma, \pi_\sigma) \leq \varepsilon$, in time bounded by a bivariate polynomial in $|\sigma|, \log \varepsilon^{-1}$. We allow $\omega \notin \Omega_\sigma$. If $\Omega_\sigma = \emptyset$, the algorithm does not terminate within its time bound. However, this can be detected, and we may construct a polynomial time algorithm which terminates either with a random ω or a proof that Ω_σ is empty.

Our real interest here is in combinatorial Markov chains, which we define as follows. Let $M : \Sigma^* \times \Sigma^* \times \Sigma^* \longrightarrow \mathbb{N}$ and define

$$\mathcal{R}_\sigma = \{(\omega, \omega') : \ M(\sigma, \omega, \omega') > 0\}, \quad \Omega_\sigma = \{\omega : \ \exists w' \text{ with } (\omega, \omega') \in \mathcal{R}_\sigma\}.$$

Let M have the following properties.

(a) There is a polynomial b such that $|\omega|, |\omega'| \leq b(|\sigma|)$ if $M(\sigma, \omega, \omega') > 0$, and M is computable in time polynomial in $|\sigma|$ whenever $M(\sigma, \omega, \omega') > 0$.

(b) There exist constants $K(\sigma) \in \mathbb{N}_+$, of polynomial size, such that

$$\sum_{\omega' \in \Sigma^*} M(\sigma, \omega, \omega') = K(\sigma) \qquad (\forall \omega \in \Omega_\sigma).$$

(c) The transitive closure of $\mathcal{R}_\sigma$ is $\Omega_\sigma \times \Omega_\sigma$, and for some ω, $(\omega, \omega) \in \mathcal{R}_\sigma$.

(d) Writing $M_\omega(\sigma, \omega') = M(\sigma, \omega, \omega')$ $(\omega \in \Sigma^*)$, it follows from (a) that M_ω is a weight function. We require that there is a good sampler for $\widetilde{M_\omega}$ $(\forall \omega)$.

We call M a *density matrix*, and associate with it a sequence of Markov chains $\mathcal{M}_\sigma = (\Omega_\sigma, P_\sigma)$, with transition matrices

$$P_\sigma(\omega_1, \omega_2) = M(\sigma, \omega_1, \omega_2)/K(\sigma) \qquad (\omega_1, \omega_2 \in \Omega_\sigma).$$

Properties (a) and (c) ensure that $\mathcal{M}_\sigma$ is finite and ergodic. Property (d) ensures that we can efficiently simulate $\mathcal{M}_\sigma$ to a close approximation for any given number of steps. Property (b) ensures that polynomial powers of the transition matrix cannot generate rationals of superpolynomial size, and hence the state probabilities at any polynomial time cannot be rationals of superpolynomial size. We include this property since we do not wish to preclude exact generation using Markov chains. In any case, this condition can always be satisfied to any desired approximation, and is usually satisfied naturally. There is little loss in restricting $K(\sigma)$ to be a power of 2. If any such $K(\sigma)$ exist, it is easy to show that there is a chain with the same stationary distribution and K a

power of 2, simply by increasing the "self-loop" probability on all states. Since we are interested in the stationary distribution, we can use this slightly slower chain. Thus we may insist on K being a power of 2 where convenient.

Density matrices M_1, M_2 are *equivalent* if there exists $\kappa : \Sigma^* \longrightarrow \mathbb{Q}_+$ such that $M_2(\sigma, \omega, \omega') = \kappa(\sigma) M_1(\sigma, \omega, \omega')$ for all $\sigma, \omega, \omega' \in \Sigma^*$. We can identify the equivalence class $\widetilde{M}$ with the sequence $\mathcal{M}_\sigma$. We say that $\mathcal{M}_\sigma$ is a *rapidly mixing Markov chain* if its mixing time $\tau_\sigma(\varepsilon)$ is bounded by a polynomial in $|\sigma|, \log \varepsilon^{-1}$.

If $\mathcal{M}_\sigma$ is a Markov chain sequence, let π_σ denote the stationary distribution of $\mathcal{M}_\sigma$. Then, if W is a weight function, $\mathcal{M}_\sigma$ is a *Monte Carlo Markov chain* (MCMC) for $\widetilde{W}$ if both $\widetilde{W}$, $\mathcal{M}_\sigma$ determine the same sequence of probability spaces $(\Omega_\sigma, \pi_\sigma)$. (This slight overloading of the MCMC abbreviation should not cause confusion.) The usual way to establish this is by reversibility, i.e. if $W(\sigma, \omega) M(\sigma, \omega, \omega') = W(\sigma, \omega') M(\sigma, \omega', \omega)$ for all $\sigma \in \Sigma^*$ and $\omega, \omega' \in \Omega_\sigma$. Clearly we have a good sampler for $\widetilde{W}$ if $\mathcal{M}_\sigma$ is a rapidly mixing Markov chain.

One of the main applications of sampling is to *approximate integration*. In our setting this means estimating $Z(\sigma)$ to some specified relative error. In the important case where W is a characteristic function, we call the approximate integration problem *approximate counting*. Specifically, a *fully polynomial randomized approximation scheme* (fpras) for $Z(\sigma)$ is a PTM which on input σ, ϵ outputs $\hat{Z}$ so that

$$\mathbf{Pr}(1/(1+\epsilon) \leq \hat{Z}/Z \leq 1+\epsilon) \geq \tfrac{3}{4},$$

and which runs in time polynomial in $|\sigma|$ and $1/\epsilon$.

The success probability can be increased to $1 - \delta$ by taking the median of $O(\log \delta)$ samples, see (1.2).

Let size : $\Sigma^* \longrightarrow \mathbb{N}$ be such that size(σ) is polynomially bounded in $|\sigma|$, and if size$(\sigma') <$ size(σ) then $|\sigma'|$ is polynomially bounded in $|\sigma|$. If size$(\sigma) = 0$, we call the problem a *base problem*. For the class of base problems, we assume the existence of a good sampler and a fpras for $Z(\sigma)$.

For all σ, let $\Xi(\sigma)$ be a polynomial time computable set such that

(a) size$(\xi) <$ size(σ) $(\forall \xi \in \Xi)$.

(b) There exist polynomial time computable constants $k_\xi(\sigma) \in \mathbb{Q}_+$ and injections $\phi_\xi(\sigma) : \Omega_\xi \longrightarrow \Omega_\sigma$ $(\forall \xi \in \Xi)$, such that

$$k_\xi W(\xi, \omega) \leq W(\sigma, \phi_\xi(\omega)) \qquad (\forall \omega \in \Omega_\xi).$$

Both $\phi_\xi(\omega)$ and ${\phi_\xi}^{-1}(\omega)$ must be computable in polynomial time, given $\omega \in \Omega_\xi$ and $\omega \in \Omega_\sigma$, respectively.

(c) For some $\zeta \in \Xi$, $Z(\sigma)/\left(k_\zeta(\sigma)Z(\zeta)\right)$ is polynomially bounded in $|\sigma|$.

If $\widetilde{W}$ satisfies these conditions, we call the problem *self-contractible*. Summing over $\omega \in \Omega_\xi$ and using the injectivity of ϕ_ξ shows that (b) implies $k_\xi Z(\xi) \leq Z(\sigma)\ \forall \xi \in \Xi$. Now, suppose we have a good sampler for $\widetilde{W}$. Then we may estimate $k_\xi Z(\xi)/Z(\sigma)$ by rejection sampling. We sample ω from $W(\sigma, \cdot)$, and accept with probability $k_\xi W(\xi, \phi^{-1}(\omega))/W(\sigma, \omega)$ if $\phi^{-1}(\omega) \neq \emptyset$. The overall acceptance probability is

$$\sum_{w \in \phi(\Omega_\xi)} \frac{k_\xi W(\xi, \phi^{-1}(\omega))}{W(\sigma, \omega)} \frac{W(\sigma, \omega)}{Z(\sigma)} = \frac{k_\xi Z(\xi)}{Z(\sigma)}.$$

Moreover, from (c) there is some $\zeta \in \Xi$ such that we can estimate this ratio to sufficient relative accuracy in polynomial time. Since $\text{size}(\zeta) < \text{size}(\sigma)$, we may repeat this process with ζ replacing σ. Then, letting $\sigma_0 = \sigma$, $\sigma_1 = \zeta$, ..., we may iterate until $\text{size}(\sigma_r) = 0$. Now $|\sigma_i|$ is polynomially bounded in $|\sigma|$ for all $i = 0, 1, \ldots, r$. For σ_r we can approximate $Z(\sigma_r)$ in polynomial time. Then we may multiply estimates together to approximate

$$Z(\sigma_r) \prod_{i=1}^{r} \frac{Z(\sigma_{i-1})}{k_{\sigma_i}(\sigma_{i-1}) Z(\sigma_i)} = \frac{Z(\sigma)}{\prod_{i=1}^{r} k_{\sigma_i}(\sigma_{i-1})}$$

to the required relative error, and hence $Z(\sigma)$. A converse result may be obtained under rather stronger conditions. Suppose that the base problems are such that $Z(\sigma)$ may be determined *exactly* and $\Omega(\sigma)$ can be sampled *perfectly*. Suppose that (b) and (c) are strengthened to

(b)$'$ There exist polynomial time computable constants $k_\xi(\sigma) \in \mathbb{Q}_+$ and injections $\phi_\xi(\sigma) : \Omega_\xi \longrightarrow \Omega_\sigma$ $(\forall \xi \in \Xi)$, such that

$$k_\xi W(\xi, \omega) = W(\sigma, \phi_\xi(\omega)) \qquad (\forall \omega \in \Omega_\xi).$$

Both $\phi_\xi(\omega)$ $(\omega \in \Omega_\xi)$ and ${\phi_\xi}^{-1}(\omega)$ $(\omega \in \Omega_\sigma)$ must be computable in polynomial time.

(c)$'$ The sets $\phi_\xi(\Omega_\xi)$ form a partition of Ω_σ.

Let us call such a problem *self-partitionable*. Clearly (b)$'$ implies (b). Also, from (b)$'$ and (c)$'$, since

$$\begin{aligned} \sum_{\xi \in \Xi} k_\xi Z(\xi) = \sum_{\xi \in \Xi} \sum_{\omega \in \Omega_\xi} k_\xi W(\xi, \omega) &= \sum_{\xi \in \Xi} \sum_{\omega \in \Omega_\xi} W(\sigma, \phi_\xi(\omega)) \\ = \sum_{\omega \in \Omega_\sigma} W(\sigma, \omega) &= Z(\sigma), \end{aligned} \tag{1.19}$$

and the polynomial size of Ξ now implies (c). We sketch the generation procedure, skipping details. Suppose we can estimate $Z(\sigma)$ by $\hat{Z}(\sigma)$ within relative error ϵ to high enough probability. We branch to $\xi \in \Xi$ with probability $k_\xi \hat{Z}(\xi)/(1+\epsilon)\hat{Z}(\sigma)$. If the total of these probabilities over $\Xi(\sigma)$ is more than 1 we "fail", i.e. we abandon this whole sampling "trial". If the total is less than 1, as we would expect, then we fail with the (small) unassigned probability. Otherwise we repeat, getting $\sigma = \sigma_0, \sigma_1, \ldots, \sigma_r$ until we reach a base case, and then we generate ω' from $W(\sigma_r, .)$. Then ω is determined from $\omega_{i-1} = \phi_{\sigma_i}(\omega_i)$ $(i = 1, \ldots, r)$, with $\omega_0 = \omega$, $\omega_r = \omega'$. Then under the assumption that our approximations $\hat{Z}(\sigma_i)$ are always within bounds, the probability that ω is generated is

$$\frac{k_{\sigma_1}\hat{Z}(\sigma_1)}{(1+\epsilon)\hat{Z}(\sigma_0)}\frac{k_{\sigma_2}\hat{Z}(\sigma_2)}{(1+\epsilon)\hat{Z}(\sigma_1)}\cdots\frac{k_{\sigma_r}Z(\sigma_r)}{(1+\epsilon)\hat{Z}(\sigma_{r-1})}\frac{W(\sigma_r,\omega_r)}{Z(\sigma_r)} = \frac{W(\sigma,\omega)}{(1+\epsilon)^r\hat{Z}(\sigma)},$$

after an easy induction. This is equivalent to the desired weight function. Provided that ϵ is sufficiently small, the failure probability and the variation distance can be kept small on a single trial. Then we may output an arbitrary ω if we fail after some large enough number of trials. Hence the overall variation distance is small. The running time of the algorithm will depend polyonomially on the *logarithm* of the error ϵ, since it is linked to the failure probability of the approximation algorithm. It follows that for self-partitionable problems, approximate integration and good sampling are equivalent. It is easy to see that self-reducible problems are self-partitionable, but the converse is not necessarily true. An example is the volume approximation problem.

We can show that approximate integration implies good sampling under rather weaker conditions than self-partitionability. We do not develop this here, however, since we have no example of a problem satisfying these conditions which is not self-partitionable. In any case, the usual direction in applications is to go from sampling to integration.

Chapter 2

Bounding the Mixing Time

2.1 Spectral Gap

Let P be the transition matrix of an ergodic, reversible Markov chain on state space Ω, Let π be its stationary distribution. Let $N = |\Omega|$ and assume w.l.o.g. that $\Omega = \{0, 1, \ldots, N-1\}$. Let the eigenvalues of P be $1 = \lambda_0 > \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{N-1}$. They are all real valued. Let $\lambda_{\max} = \max\{|\lambda_i| : i > 0\}$. The fact that $\lambda_{\max} < 1$ is a classical result of the theory of non-negative matrices. The *spectral gap* $1 - \lambda_{\max}$ determines the mixing rate of the chain in an essential way. The larger it is, the more rapidly does the chain mix. For $U \subseteq \Omega$ let

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

Theorem 2.1.1 *For all $U \subseteq \Omega$ and $t \geq 0$,*

$$\Delta_U \leq \frac{\lambda_{\max}^t}{\min_{i \in U} \pi(i)}.$$

Proof Let $D^{1/2}$ be the diagonal $\Omega \times \Omega$ matrix with diagonal entries $\sqrt{\pi(\omega)}$, $\omega \in \Omega$ and let $D^{-1/2}$ be its inverse. Then the reversibility of of the chain (1.15) implies that the matrix $S = D^{1/2} P D^{-1/2}$ is symmetric. It has the same eigenvalues as P and its symmetry means that these are all real. We can select an orthonormal basis of column vectors $\mathbf{e}^{(i)}$, $i \in \Omega$ for $\mathbb{R}^\Omega$ consisting of left eigenvectors of S where $\mathbf{e}^{(i)}$ has associated eigenvalue λ_i and $\mathbf{e}^{(0)} = \pi^T D^{-1/2}$. S has the *spectral decomposition*

$$S = \sum_{i=0}^{N-1} \lambda_i \mathbf{e}^{(i)} \mathbf{e}^{(i)^T} = \sum_{i=0}^{N-1} \lambda_i E^{(i)},$$

where $E^{(i)} = \mathbf{e}^{(i)}\mathbf{e}^{(i)^T}$. Note that $E^{(i)}E^{(j)} = 0$ for $i \neq j$ and ${E^{(i)}}^2 = E^{(i)}$. It follows that for any $t = 0, 1, 2, \ldots$, $S^t = \sum_{i=0}^{N-1} \lambda_i^t E^{(i)}$. Hence

$$\begin{aligned} P^t = D^{-1/2}S^tD^{1/2} &= \sum_{i=0}^{N-1} \lambda_i^t (D^{-1/2}\mathbf{e}^{(i)}))(\mathbf{e}^{(i)^T}D^{1/2}) \\ &= \mathbf{1}_N\pi^T + \sum_{i=1}^{N-1} \lambda_i^t (D^{-1/2}\mathbf{e}^{(i)}))(\mathbf{e}^{(i)^T}D^{1/2}), \end{aligned}$$

where $\mathbf{1}_N$ is the N-vector all of whose components are 1. In component form, we get with the help of the Cauchy-Schwartz inequality:

$$\begin{aligned} |P^t(j,k) - \pi_k| &= \left| \sqrt{\frac{\pi_k}{\pi_j}} \sum_{i=1}^{N-1} \lambda_i^t \mathbf{e}_j^{(i)} \mathbf{e}_k^{(i)} \right| \\ &\leq \sqrt{\frac{\pi_k}{\pi_j}} \lambda_{\max}^t \left(\sum_{i=0}^{N-1} {\mathbf{e}_j^{(i)}}^2 \right)^{1/2} \left(\sum_{i=0}^{N-1} {\mathbf{e}_k^{(i)}}^2 \right)^{1/2} \\ &= \sqrt{\frac{\pi_k}{\pi_j}} \lambda_{\max}^t. \end{aligned} \tag{2.1}$$

The theorem follows by substitution of the above inequality in the definition of Δ_U. □

In terms of mixing time we have

Corollary 2.1.1

$$\tau(\varepsilon) \leq \left\lceil \frac{\log \varepsilon\pi_{\min}}{\log \lambda_{\max}} \right\rceil.$$

Proof For $A \subseteq \Omega$ we have

$$p_t(A) - \pi(A) \leq \frac{\lambda_{\max}^t}{\pi_{\min}} \pi(A) \leq \frac{\lambda_{\max}^t}{\pi_{\min}}.$$

□

As an example we consider random walk $\mathcal{W}_n$ on the unit hypercube. Here the graph is the n-cube $Q_n = (X_n = \{0,1\}^n, E_n)$ where $x, y \in X_n$ are adjacent in Q_n if their Hamming distance is one i.e. if $|\{i \in [n] : \ x_i \neq y_i\}| = 1$. We add n self loops to each vertex to make the chain lazy.

If G is a d-regular graph without loops and A_G is its adjacency matrix then the probability transition matrix P_G of a random walk on G satisfies $P_G = d^{-1}A_G$.

For graphs $G_i = (V_i, E_i), i = 1, 2$ we can define their product $G = G_1 \times G_2 = (V, E)$ where $V = V_1 \times V_2$ and $E = \{((v_1, v_2), (w_1, w_2)) : \ v_1 = w_1 \text{ and } (v_2, w_2) \in E_2 \text{ or } v_2 = w_2 \text{ and } (v_1, w_1) \in E_1\}$. Then

$$Q_n = K_2 \times K_2 \times \cdots \times K_2 \qquad (n \text{ fold product}). \tag{2.2}$$

Theorem 2.1.2 *If $\mu_i, i = 1, 2, \ldots, m$ and $\nu_i, i = 1, 2, \ldots, n$ are the eigenvalues of matrices A_{G_1}, A_{G_2} respectively, then the eigenvalues of A_G are $\{\mu_i + \nu_j : 1 \le i \le m, 1 \le j \le n\}$.*

Proof A_G can be obtained from A_{G_1} by replacing each 1 by the $|V_2|$ identity matrix I_2, the off-diagonal 0's by the $|V_2| \times |V_2|$ matrix of 0's and replacing each diagonal entry by A_{G_2}. So if $p_G(\lambda) = \det(\lambda I - A_G)$ then

$$p_G(\lambda) = \det p_{G_1}(\lambda I_2 - A_{G_2}).$$

This follows from the following: Suppose the $mn \times mn$ matrix A is decomposed into an $m \times m$ matrix of $n \times n$ blocks $A_{i,j}$. Suppose also that the $A_{i,j}$ commute among themselves. Then

$$\det A = \det \left(\sum_{\sigma} (-1)^{\mathrm{sign}(\sigma)} \prod_{i=1}^{m} A_{i,\sigma(i)} \right),$$

i.e. one can produce an $m \times m$ matrix by a "determinant" calculation and then take its determinant. Needs a proof

So

$$p_G(\lambda) = \det \prod_{i=1}^{n} (\lambda I_2 - A_{G_2} - \mu_i I_2) = \prod_{i=1}^{n} p_{G_2}(\lambda - \mu_i) = \prod_{i=1}^{n} \prod_{j=1}^{n} (\lambda - \mu_i - \nu_j).$$

□

The eigenvalues of K_2 are $\{1, -1\}$ and applying (2.2) we see that the eigenvalues of Q_n are $\{0, \pm 1, \pm 2, \ldots, \pm n\}$ (ignoring multiplicities). To get the eigenvalues for our random walk we (i) divide by n and then (ii) replace each eigenvalue λ by $\frac{1+\lambda}{2}$ to account for adding loops. Thus the second eigenvalue of the walk is $1 - \frac{1}{2n}$.

Applying Corollary 2.1.1 we obtain $\tau(\varepsilon) \le \log(\varepsilon^{-1}) + O(n^2)$. This is a poor estimate, due to our use of the Cauchy-Schwartz inequality in the proof of Theorem 2.1.1. We get an easier and better estimate by using *coupling*.

2.1.1 Decomposition Theorem

2.2 Conductance

The conductance Φ of $\mathcal{M}$ is defined by

$$\Phi = \min\{\Phi_S : S \subseteq \Omega,\ 0 < \pi(S) \le 1/2\}$$

where if $Q(\omega, \sigma) = \pi(\omega) P(\omega, \sigma)$ and $\bar{S} = \Omega \setminus S$,

$$\Phi_S = \pi(S)^{-1} Q(S, \bar{S}).$$

Thus Φ_S is the steady state probability of moving from S to $\bar{S}$ in one step of the chain, conditional on being in S.

Clearly $\Phi \leq \frac{1}{2}$ if $\mathcal{M}$ is lazy.

Note that

$$\Phi_S \pi(S) = Q(S, \bar{S}) = Q(\bar{S}, S) = \Phi_{\bar{S}} \pi(\bar{S}). \tag{2.3}$$

Indeed,

$$Q(S, \bar{S}) = Q(\Omega, \bar{S}) - Q(\bar{S}, \bar{S}) = \pi(\bar{S}) - Q(\bar{S}, \bar{S}) = Q(\bar{S}, S).$$

Let $\pi_{\min} = \min\{\pi(\omega) : \omega \in \Omega\} > 0$ and $\pi_{\max} = \max\{\pi(\omega) : \omega \in \Omega\}$.

2.2.1 Reversible Chains

In this section we show how conductance gives us an estimate of the spectral gap of a reversible chain.

Lemma 2.2.1 *If $\mathcal{M}$ is lazy and ergodic then all eigenvalues are positive.*

Proof $Q = 2P - I \geq 0$ is stochastic and has eigenvalues $\mu_i = 2\lambda_i - 1$, $i = 0, 1, \ldots N-1$. The result follows from $\mu_i > -1$, $i = 0, 1, \ldots N-1$. □

For $y \in \mathbb{R}^N$ let

$$\mathcal{E}(y, y) = \sum_{i<j} \pi_i P_{i,j} (y_i - y_j)^2.$$

Lemma 2.2.2 *If $\mathcal{M}$ is reversible then*

$$1 - \lambda_1 = \min_{\pi^T y = 0} \frac{\mathcal{E}(y, y)}{\sum_i \pi_i y_i^2}.$$

Proof Let $D, S, \mathbf{e}^{(0)}$ be as in Section 2.1. Then by the Rayleigh principle,

$$\lambda_1 = \max_{\pi^T D^{-1/2} x = 0} \frac{x^T D^{1/2} P D^{-1/2} x}{x^T x}.$$

Thus

$$\begin{aligned} 1 - \lambda_1 &= \min_{\pi^T D^{-1/2} x = 0} \frac{x^T D^{1/2} (I - P) D^{-1/2} x}{x^T x} \\ &= \min_{\pi^T y = 0} \frac{y^T D (I - P) y}{y^T D y}. \end{aligned} \tag{2.4}$$

Now

$$\begin{aligned} y^T D(I-P)y &= -\sum_{i\neq j} y_i y_j \pi_i P_{i,j} + \sum_i \pi_i (1-P_{i,i}) y_i^2 \\ &= -\sum_{i\neq j} y_i y_j \pi_i P_{i,j} + \sum_{i\neq j} \pi_i P_{i,j} \frac{y_i^2+y_j^2}{2} \\ &= \sum_{i<j} \pi_i P_{i,j} (y_i - y_j)^2 \\ &= \mathcal{E}(y,y), \end{aligned}$$

and the lemma follows from (2.4). □

Theorem 2.2.1 *If $\mathcal{M}$ is a reversible chain then*

$$1-\lambda_1 \geq \frac{\Phi^2}{2}.$$

Proof Assume now that $\pi^T y = 0$, $y_1 \geq y_2 \geq \cdots \geq y_N$ and that

$$\pi_1 + \pi_2 + \cdots + \pi_{r-1} \leq \frac{1}{2} < \pi_1 + \pi_2 + \cdots + \pi_r.$$

Let $z_i = y_i - y_r$ for $i = 1, 2, \ldots, n$. Then

$$z_1 \geq z_2 \geq \cdots \geq z_r = 0 \geq z_{r+1} \geq \cdots \geq z_N,$$

and

$$\begin{aligned} \frac{\mathcal{E}(y,y)}{\sum_i \pi_i y_i^2} &= \frac{\mathcal{E}(z,z)}{-y_r^2 + \sum_i \pi_i z_i^2} \\ &\geq \frac{\mathcal{E}(z,z)}{\sum_i \pi_i z_i^2}. \qquad (2.5) \\ &= \frac{\left(\sum_{i<j} \pi_i P_{i,j} (z_i - z_j)^2\right)\left(\sum_{i<j} \pi_i P_{i,j} (|z_i| + |z_j|)^2\right)}{\left(\sum_i \pi_i z_i^2\right)\left(\sum_{i<j} \pi_i P_{i,j} (|z_i| + |z_j|)^2\right)} \\ &= \frac{A}{B}, \qquad \text{say.} \end{aligned}$$

Now,

$$\begin{aligned} A &\geq \left(\sum_{i<j} \pi_i P_{i,j} |z_i - z_j| (|z_i| + |z_j|)\right)^2 \qquad \text{by Cauchy-Schwartz} \\ &\geq \left(\sum_{i<j} \pi_i P_{i,j} \sum_{k=i}^{j-1} |z_{k+1}^2 - z_k^2|\right)^2. \qquad (2.6) \end{aligned}$$

We verify (2.6) later. Also,

$$\sum_{i<j} \pi_i P_{i,j} (|z_i| + |z_j|)^2 \le 2 \sum_{i<j} \pi_i P_{i,j} (z_i^2 + z_j^2) \le 2 \sum_i \pi_i z_i^2.$$

So,

$$\frac{\mathcal{E}(y,y)}{\sum_i \pi_i y_i^2} \ge \frac{A}{B} \ge \frac{\left(\sum_{i<j} \pi_i P_{i,j} \sum_{k=i}^{j-1} |z_{k+1}^2 - z_k^2| \right)^2}{2 \left(\sum_i \pi_i z_i^2 \right)^2}.$$

Now let $S_k = \{1, 2, \ldots, k\}$ and $C_k = \{(i,j) : i \le k < j\}$. Then

$$\begin{aligned}
\sum_{i<j} \pi_i P_{i,j} \sum_{k=i}^{j-1} |z_{k+1}^2 - z_k^2| &= \sum_{k=1}^{N-1} |z_{k+1}^2 - z_k^2| \sum_{(i,j) \in C_k} \pi_i P_{i,j} \\
&\ge \Phi \left(\sum_{k=1}^{r-1} (z_k^2 - z_{k+1}^2) \pi(S_k) + \sum_{k=r}^{N-1} (z_{k+1}^2 - z_k^2)(1 - \pi(S_k)) \right) \\
&= \Phi \left(\sum_{k=1}^{N-1} (z_k^2 - z_{k+1}^2) \pi(S_k) + (z_N^2 - z_r^2) \right) \\
&= \Phi \left(\sum_{k=1}^{N} \pi_k z_k^2 \right)
\end{aligned}$$

since $z_r = 0$.

Thus if $\pi^T y = 0$ then

$$\frac{\mathcal{E}(y,y)}{\sum_i \pi_i y_i^2} \ge \frac{\Phi^2}{2}$$

and Theorem 2.2.1 follows.

Proof of (2.6)

We show that if $i < j$ then

$$|z_i - z_j|(|z_i| + z_j|) \ge \sum_{k=i}^{j-1} |z_{k+1}^2 - z_k^2|. \tag{2.7}$$

If $r \notin \{i, i+1, \ldots, j\}$ i.e. if z_i, z_j have the same sign then LHS(2.7)=RHS(2.7)=$|z_i^2 - z_j^2|$. Otherwise LHS(2.7)=$(|z_i| + |z_j|)^2$ and RHS(2.7)=$z_i^2 + z_j^2$. □

In terms of mixing time we obtain from Corollary 2.1.1,

Corollary 2.2.1 *If $\mathcal{M}$ is a lazy ergodic chain then*

$$\tau(\varepsilon) \le \left\lceil \frac{2 |\log \varepsilon \pi_{\min}|}{\Phi^2} \right\rceil.$$

Proof Lemma 2.2.1 implies that $\lambda_1 = \lambda_{\max}$ and then

$$\frac{1}{\log \lambda_{\max}^{-1}} \leq \frac{1}{\log(1-\Phi^2/2)^{-1}} \leq \frac{2}{\Phi^2}.$$

□

Now consider the conductance of a random walk on a graph $G = (V, E)$. For $S, T \subseteq V$ let $E(S,T) = \{(v,w) \in E : \ v \in S, w \in T\}$ and $e(S,T) = |E(S,T)$. Then, by definition,

$$\Phi_S = \frac{\displaystyle\sum_{(v,w)\in E(S,\bar{S})} \frac{d_v}{2|E|}\frac{1}{d_v}}{\displaystyle\sum_{v\in S}\frac{d_v}{2|E|}} = \frac{e(S,\bar{S})}{\displaystyle\sum_{v\in S} d_v}.$$

In particular when G is an r-regular graph

$$\Phi = r^{-1} \min_{|S|\leq \frac{1}{2}|V|} \frac{e(S,\bar{S})}{|S|}. \tag{2.8}$$

The minimand above is referred to as the *expansion* of G. This graphs with good expansion (*expander graphs*) have large conductance and random walks on them mix rapidly.

As an example consider the n-cube Q_n. For $S \subseteq X_n$ let $in(S)$ denote the number of edges of Q_n which are wholly contained in S.

Lemma 2.2.3 *If $\emptyset \neq S \subseteq X_n$ then $in(S) \leq \frac{1}{2}|S|\log_2|S|$.*

Proof We prove this by induction on n. It is trivial for $n = 1$. For $n > 1$ let $S_i = \{x \in S : \ x_n = i\}$ for $i = 1, 2$. Then

$$in(S) \leq in(S_0) + in(S_1) + \min\{|S_0|, |S_1|\}$$

since the term $\min\{|S_0|, |S_1|\}$ bounds the number of edges which are contained in S and join S_0, S_1. The lemma follows from the inequality

$$x\log_2 x + y\log_2 y + 2y \leq (x+y)\log_2(x+y)$$

for all $x \geq y \geq 0$. The proof is left as a simple exercise in calculus. □

By summing the degrees at each vertex of S we see that

$$e(S,\bar{S}) + 2in(S) = n|S|.$$

By the above lemma we have

$$e(S,\bar{S}) \geq n|S| - \tfrac{1}{2}|S|\log_2|S| \geq |S|$$

assuming $|S| \le 2^{n-1}$. It follows from (2.8) that $\Phi \ge \frac{1}{n}$. Adding the self-loops to delay the walk will halve the conductance – the denominator $\sum_{v\in S} d_v$ doubles without changing the numerator in the definition of Φ_S. This gives us the estimate of $\frac{1}{8n^2}$ for the spectral gap, which is off by a factor of n – see Section 2.1.

We finish this section by proving a sort of converse to Theorem 2.2.1.

Theorem 2.2.2 *If $\mathcal{M}$ is a reversible chain then*

$$1 - \lambda_1 \le 2\Phi$$

Proof We use Lemma 2.2.2. Let S be a set of states which minimises Φ_S and define y by $y_j = \frac{1}{\pi(S)}$ if $j \in S$ and $y_j = -\frac{1}{\pi(S)}$ if $j \in \bar{S}$. It is easy to check that $\pi^T y = 0$. Then

$$\mathcal{E}(y,y) = \left(\frac{1}{\pi(S)} + \frac{1}{\pi(\bar{S})}\right)^2 Q(S,\bar{S}) \text{ and } \sum \pi_i y_i^2 = \frac{1}{\pi(S)} + \frac{1}{\pi(\bar{S})}.$$

Thus

$$1 - \lambda_{\max} \le \Phi_S \pi(S)\left(\frac{1}{\pi(S)} + \frac{1}{\pi(\bar{S})}\right) \le 2\Phi_S = 2\Phi.$$

□

2.2.2 General Chains

Theorem 2.2.3 *Suppose that $\mathcal{M}$ is lazy and*

$$\pi_{\max} \le \frac{\Phi^2}{20}.$$

Then

$$|p_t(\omega) - \pi(\omega)| \le \pi_{\min}^{-1/2}\left(1 - \tfrac{1}{2}\Phi^2\right)^t.$$

Proof For $0 \le x \le 1$ let

$$h_t(x) = \max\left\{\sum_{\omega\in\Omega}(p^t(\omega) - \pi(\omega))\xi(\omega) :\ \xi \in [0,1]^\Omega, \sum_{\omega\in\Omega}\xi(\omega)\pi(\omega) = x\right\}.$$

By putting $\xi = 1_{\omega\in S}$ in the above definition we see that

$$p_t(S) - \pi(S) \le h_t(\pi(S))$$

for all $S \subseteq \Omega$.

So, in particular $p_t(\omega) - \pi(\omega) \leq h_t(\pi(\omega))$ ($x = \pi(\omega)$) and $\pi(\omega) - p_t(\omega) \leq h_t(1 - \pi(\omega))$ ($x = 1 - \pi(\omega)$) and so

$$|p_t(\omega) - \pi(\omega)| \leq \min\{h_t(\pi(\omega)), h_t(1 - \pi(\omega))\}. \tag{2.9}$$

Order the elements $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$ so that

$$\frac{p_t(\omega_1)}{\pi(\omega_1)} \geq \frac{p_t(\omega_2)}{\pi(\omega_2)} \geq \cdots \geq \frac{p_t(\omega_N)}{\pi(\omega_N)},$$

and let $\theta_k = \sum_{i=1}^{k} \pi(\omega_i)$. Find the index k such that $\theta_{k-1} \leq x < \theta_k$. Then

$$h_t(x) = \sum_{i=1}^{k-1}(p_t(\omega_i) - \pi(\omega_i)) + \frac{x - \theta_{k-1}}{\pi(\omega_k)}(p_t(\omega_k) - \pi(\omega_k)).$$

This is because putting

$$\xi(\omega_i) = \begin{cases} 1 & i < k \\ \frac{x-\theta_{k-1}}{\pi(\omega_k)} & i = k \\ 0 & i > k \end{cases}$$

yields an optimal basic feasible solution to the linear program in the definition of $h_t(x)$.

It follows that $h_t(x)$ is a concave piece-wise linear function on the interval $[0, 1]$ with breakpoints at $0 = \theta_0 < \theta_1 < \cdots < \theta_N = 1$. Trivially, $h_t(0) = h_t(1) = 0$ and $0 \leq h_t(x) \leq 1$ for all t and x.

Now let

$$C = \max\left\{ \frac{h_0(x)}{\min\{\sqrt{x}, \sqrt{1-x}\}} : \ 0 < x < 1 \right\}.$$

If $a, b, c, d \geq 0$ then the function $f(\xi) = (a + b\xi)/\sqrt{c + d\xi}$ is monotone on $[0, 1]$ and so the value of x defining C must occur at one of the breakpoints of h_t. It follows easily that

$$\begin{aligned} C &\leq \max_{S \subseteq \Omega} \frac{|\pi_0(S) - \pi(S)|}{\min\left\{\sqrt{\pi(S)}, \sqrt{1 - \pi(S)}\right\}} \\ &= \max_{\substack{S \subseteq \Omega \\ \pi(S) \leq 1/2}} \frac{|\pi_0(S) - \pi(S)|}{\sqrt{\pi(S)}} \\ &\leq \frac{1}{\sqrt{\pi_{\min}}}. \end{aligned} \tag{2.10}$$

(The second equation comes from considering $\Omega \setminus S$ when $\pi(S) \geq 1/2$.)

We now prove that for $t \geq 1$ and $x \in \{\theta_0, \theta_1, \ldots, \theta_N\}$,

$$h_t(x) \leq \tfrac{1}{2}(h_{t-1}(x - 2\Phi \min\{x, 1-x\}) + h_{t-1}(x + 2\Phi \min\{x, 1-x\})) \tag{2.11}$$

Fix k and let $u_i = \sum_{j\le k} p_{i,j}$ where $p_{i,j} = P(\omega_i, \omega_j)$. Clearly

$$1 \ge u_i \ge p_{i,i} \ge \tfrac{1}{2} \ (i \le k) \ \text{ and } \ 0 \le u_i \le 1 - p_{i,i} \le \tfrac{1}{2} \ (i > k).$$

Now

$$\pi(\omega_j) = \sum_{i=1}^{N} \pi(\omega_i) p_{i,j} \text{ and } \sum_{i=1}^{N} p_{t-1}(\omega_i) p_{i,j} = p_t(\omega_j)$$

and so if $x = \theta_k$

$$\begin{aligned} h_t(x) &= \sum_{j=1}^{k} (p_t(\omega_j) - \pi(\omega_j)) = \sum_{j=1}^{k} \sum_{i=1}^{N} p_{i,j}(p_{t-1}(\omega_i) - \pi(\omega_i)) \\ &= \sum_{i=1}^{N} (p_{t-1}(\omega_i) - \pi(\omega_i)) u_i. \end{aligned} \tag{2.12}$$

Moreover, $0 \le u_i \le 1$ and

$$\sum_{i=1}^{N} \pi(\omega_i) u_i = \sum_{i=1}^{N} \sum_{j=1}^{k} \pi(\omega_i) p_{i,j} = \sum_{j=1}^{k} \sum_{i=1}^{N} \pi(\omega_i) p_{i,j} = \sum_{j=1}^{k} \pi(\omega_j) = x. \tag{2.13}$$

Now let

$$u_i' = \begin{cases} 2u_i - 1 & i \le k \\ 0 & i > k \end{cases} \text{ and } u_i'' = \begin{cases} 1 & i \le k \\ 2u_i & i > k \end{cases}$$

Then $0 \le u_i', u_i'' \le 1$ and $u_i' + u_i'' = 2u_i$. Let $x' = \sum_{i=1}^{N} \pi(\omega_i) u_i'$ and $x'' = \sum_{i=1}^{N} \pi(\omega_i) u_i''$. Then (2.13) implies $x' + x'' = 2x$ and so by (2.12)

$$\begin{aligned} h_t(x) &= \tfrac{1}{2} \sum_{i=1}^{N} (p_{t-1}(\omega_i) - \pi(\omega_i)) u_i' + \tfrac{1}{2} \sum_{i=1}^{N} (p_{t-1}(\omega_i) - \pi(\omega_i)) u_i'' \\ &\le \tfrac{1}{2} h_{t-1}(x') + \tfrac{1}{2} h_{t-1}(x''). \end{aligned}$$

Furthermore,

$$\begin{aligned} x - x' &= \sum_{i=1}^{N} \pi(\omega_i)(u_i - u_i') = \sum_{i=1}^{k} \pi(\omega_i)(1 - u_i) + \sum_{i=k+1}^{N} \pi(\omega_i) u_i \\ &= \sum_{i=1}^{k} \pi(\omega_i) \left(1 - \sum_{j=1}^{k} p_{i,j} \right) + \sum_{i=k+1}^{N} \pi(\omega_i) \sum_{j=1}^{k} p_{i,j} \\ &= \sum_{i=1}^{k} \sum_{j=k+1}^{N} \pi(\omega_i) p_{i,j} + \sum_{i=k+1}^{N} \pi(\omega_i) \sum_{j=1}^{k} p_{i,j} \\ &\ge 2\Phi \min\{x, 1 - x\}, \end{aligned}$$

using (2.3) and the definition of Φ. So, $x' \le x - 2\Phi \min\{x, 1-x\}$ and similarly $x'' \ge x + 2\Phi \min\{x, 1-x\}$. (2.11) now follows from the concavity of h_{t-1}.

Now consider an x, k such that $\theta_{k-1} < x < \theta_k < \frac{1}{2}$. Let $x = \alpha\theta_{k-1} + (1-\alpha)\theta_k$. Then

$$\begin{aligned}
h_t(x) &= \alpha h_t(\theta_{k-1}) + (1-\alpha)h_t(\theta_k) \\
&\le \tfrac{1}{2}(\alpha(h_{t-1}(\theta_{k-1}(1-2\Phi)) + h_{t-1}(\theta_{k-1}(1+2\Phi))) \\
&\qquad + (1-\alpha)(h_{t-1}(\theta_k(1-2\Phi)) + h_{t-1}(\theta_k(1+2\Phi)))) \\
&= \tfrac{1}{2}(\alpha(h_{t-1}(\theta_{k-1}(1-2\Phi))) + (1-\alpha)(h_{t-1}(\theta_k(1-2\Phi)))) \\
&\qquad + \tfrac{1}{2}(\alpha(h_{t-1}(\theta_{k-1}(1+2\Phi))) + (1-\alpha)(h_{t-1}(\theta_k(1+2\Phi)))) \\
&\le \tfrac{1}{2}(h_{t-1}(x(1-2\Phi)) + h_{t-1}(x(1+2\Phi)))
\end{aligned}$$

from the concavity of h_{t-1}. Thus (2.11) holds for such an x. A similar argument shows that (2.11) holds for an x, k such that $\frac{1}{2} < \theta_{k-1} < x < \theta_k$. So let ℓ be such that $\theta_{\ell-1} \le \frac{1}{2} < \theta_\ell$ and suppose $\theta_{\ell-1} < x < \theta_\ell$. For such x we can only prove

$$h_t(x) \le \tfrac{1}{2}(h_{t-1}(x - x\gamma_x\Phi) + h_{t-1}(x + x\gamma_x\Phi)) \tag{2.14}$$

where $\gamma_x \ge 2 - 4\pi_{\max}$.

$$\begin{aligned}
h_t(x) &= \alpha h_t(\theta_{\ell-1}) + (1-\alpha)h_t(\theta_\ell) \\
&\le \tfrac{1}{2}(\alpha(h_{t-1}(\theta_{\ell-1}(1-2\Phi)) + h_{t-1}(\theta_{\ell-1}(1+2\Phi))) \\
&\qquad + (1-\alpha)((h_{t-1}(\theta_\ell - 2\Phi(1-\theta_\ell))) + h_{t-1}(\theta_\ell + 2\Phi(1-\theta_\ell)))) \\
&\le \tfrac{1}{2}(h_{t-1}(x - 2\Phi(x - (1-\alpha)(2\theta_\ell - 1))) + h_{t-1}(x + 2\Phi(x - (1-\alpha)(2\theta_\ell - 1))))
\end{aligned}$$

Thus (2.14) holds with

$$\begin{aligned}
\gamma_x &= 2 - \frac{(1-\alpha)(2\theta_\ell - 1)}{x} \\
&\ge 2 - \frac{2\theta_\ell - 1}{\theta_{\ell-1}} \\
&\ge 2 - 4(\theta_\ell - \theta_{\ell-1})
\end{aligned}$$

and (2.14) follows.

Combining (2.11) and (2.14) we get that for $0 \le x \le 1$

$$h_t(x) \le \tfrac{1}{2}(h_{t-1}(x - \gamma_x\Phi \min\{x, 1-x\}) + h_{t-1}(x + \gamma_x\Phi \min\{x, 1-x\})) \tag{2.15}$$

We now prove inductively that

$$h_t(x) \le C \min\left\{\sqrt{x}, \sqrt{1-x}\right\} \left(1 - \frac{\Phi^2}{2}\right)^t. \tag{2.16}$$

For $t = 0$ (2.16) follows trivially from the definition of C. Let $t \geq 1$ and suppose for example $0 \leq x \leq \frac{1}{2}$. Then (2.15) implies

$$\begin{aligned} h_t(x) &\leq \tfrac{1}{2}C\left(1-\tfrac{1}{2}\Phi^2\right)^{t-1}\left(\sqrt{x-\gamma_x\Phi x}+\sqrt{x+\gamma_x\Phi x}\right) \\ &= C\left(1-\tfrac{1}{2}\Phi^2\right)^{t-1}\sqrt{x}\left(\sqrt{1-\gamma_x\Phi}+\sqrt{1+\gamma_x\Phi}\right)/2. \end{aligned}$$

The last factor can be estimated by

$$\begin{aligned} \tfrac{1}{2}(\sqrt{1-\gamma_x\Phi}+\sqrt{1+\gamma_x\Phi}) &= \tfrac{1}{2}\left(\sum_{r=0}^{\infty}(-1)^r\binom{\frac{1}{2}}{r}(\gamma_x\Phi)^r+\sum_{r=0}^{\infty}\binom{\frac{1}{2}}{r}(\gamma_x\Phi)^r\right) \\ &= \sum_{r=0}^{\infty}\binom{\frac{1}{2}}{2r}(\gamma_x\Phi)^r = 1-\tfrac{1}{8}(\gamma_x\Phi)^2-\tfrac{5}{128}(\gamma_x\Phi)^4-\cdot \\ &\leq 1-\tfrac{1}{2}\Phi^2. \end{aligned}$$

This completes the induction for $x \leq \frac{1}{2}$. For $x > \frac{1}{2}$ we put $x = 1-y$ and define $\hat{h}_t(y) = h_t(1-y)$. Then (2.15) gives

$$\hat{h}_t(y) \leq \tfrac{1}{2}(\hat{h}_{t-1}(y-\gamma_x\Phi y)+\hat{h}_{t-1}(y+\gamma_x\Phi y))$$

from which we obtain

$$\hat{h}_t(y) \leq C\sqrt{y}\left(1-\tfrac{1}{2}\Phi^2\right)^t$$

as before. □

Suppose now that we define the following "distance" M between measures $\hat{\pi}$ and π on space Ω.

$$M(\hat{\pi},\pi) = \max_{\emptyset\neq A\subseteq\Omega}\frac{|\hat{\pi}(A)-\pi(A)|}{\sqrt{\pi(A)}}. \tag{2.17}$$

Corollary 2.2.1 *Let a lazy ergodic Markov chain with steady state π be started with distribution π_0 and let π_t denote the distribution after t steps. If $\pi_{\max} \leq \frac{\Phi^2}{20}$ then*

$$M(\pi_t,\pi) \leq M(\pi_0,\pi)\left(1-\tfrac{1}{2}\Phi^2\right)^t.$$

Proof Fix $S \subseteq \Omega$. It follows from (2.10) and (2.16) that

$$|\pi_t(S)-\pi(S)| \leq \sqrt{\pi(S)}M(\pi_0,\pi)\left(1-\frac{\Phi^2}{2}\right)^t.$$

□

2.2.3 Path Congestion

Supose that for each pair $(x, y) \in \Omega \times \Omega$ we have a *canonical path* γ_{xy} from x to y in the digraph $D_{\mathcal{M}} = (\Omega, A)$ (defined in Section 1.3). Let

$$\bar{\rho} = \max_{e \in A} \frac{1}{Q(e)} \sum_{\gamma_{xy} \ni e} \pi(x)\pi(y)|\gamma_{xy}|$$

where if $e = (\sigma, \tau)$ then $Q(e) = \pi(\sigma)P(\sigma, \tau)$ and $|\gamma_{xy}|$ is the number of arcs in γ_{xy}.

Theorem 2.2.4 *Assume that $\mathcal{M}$ is reversible. Then*

$$1 - \lambda_1 \geq \frac{1}{\bar{\rho}}.$$

Proof We use Lemma 2.2.2. Assume $\sum_i \pi_i y_i = 0$. Then

$$\begin{aligned} 2\sum_{i=1}^{N} \pi_i y_i^2 &= \sum_{i=1}^{N}\sum_{j=1}^{N} \pi_i \pi_j (y_i - y_j)^2 \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N} \pi_i \pi_j \left(\sum_{e \in \gamma_{ij}} (y_{e^+} - y_{e^-}) \right)^2 \end{aligned}$$

where edge $e = (e^-, e^+)$

$$\leq \sum_{i=1}^{N}\sum_{j=1}^{N} \pi_i \pi_j |\gamma_{ij}| \sum_{e \in \gamma_{ij}} (y_{e^+} - y_{e^-})^2$$

by Cauchy-Schwartz

$$\begin{aligned} &= \sum_{e \in A} (y_{e^+} - y_{e^-})^2 \sum_{\gamma_{xy} \ni e} \pi_x \pi_y |\gamma_{xy}| \\ &\leq \sum_{e \in A} (y_{e^+} - y_{e^-})^2 Q(e) \bar{\rho} \\ &= 2\bar{\rho}\mathcal{E}(y, y). \end{aligned}$$

□

This theorem often gives stronger bounds on the spectral gap than Theorem 2.2.1. We apply it now to our example of a random walk $\mathcal{W}_n$ on the cube.

Let $x = (x_0, x_1, \ldots, x_{n-1})$ and $y = (y_0, y_1, \ldots, y_{n-1})$ be arbitrary members of X_n. The canonical path γ_{xy} from x to y is composed of n edges, 0 to $n-1$, where edge i is simply

$$\big((y_0, \ldots, y_{i-1}, x_i, x_{i+1}, \ldots x_{n-1}),\ (y_0, \ldots, y_{i-1}, y_i, x_{i+1}, \ldots x_{n-1})\big),$$

i.e., we change the ith component from x_i to y_i. Note that some of the edges may be loops (if $x_i = y_i$). To compute $\bar{\varrho}$, fix attention on a particular (oriented) edge

$$t = (w, w') = \big((w_0, \ldots, w_i, \ldots w_{n-1}), (w_0, \ldots, w'_i, \ldots w_{n-1})\big),$$

and consider the number of canonical paths γ_{xy} that include t. The number of possible choices for x is 2^i, as the final $n - i$ positions are determined by $x_j = w_j$, for $j \geq i$; and by a similar argument the number of possible choices for y is 2^{n-i-1}. Thus the total number of canonical paths using a particular edge t is 2^{n-1}; furthermore, $Q(w, w') = \pi(w)P(w, w') \geq 2^{-n}(2n)^{-1}$, and the length of every canonical path is exactly n. Plugging all these bounds into the definition of $\bar{\rho}$ yields $\bar{\rho} \leq n^2$. Thus, by Theorem 2.2.4, the mixing time of $\mathcal{W}_n$ is $\tau(\varepsilon) \leq n^2(n \ln q + \ln \varepsilon^{-1})$.

2.2.4 Comparison Theorems

2.2.5 Decomposition Theorem

2.3 Coupling

A *coupling* $\mathcal{C}(\mathcal{M})$ for $\mathcal{M}$ is a stochastic process (X_t, Y_t) on $\Omega \times \Omega$ such that each of X_t, Y_t is marginally a copy of $\mathcal{M}$,

$$\begin{aligned} \mathbf{Pr}(X_t = \sigma_1 \mid X_{t-1} = \omega_1) &= P(\omega_1, \sigma_1), \\ \mathbf{Pr}(Y_t = \sigma_2 \mid Y_{t-1} = \omega_2) &= P(\omega_2, \sigma_2), \end{aligned} \qquad (\forall t > 0). \tag{2.18}$$

The following simple but powerful inequality then follows easily from these definitions.

Lemma 2.3.1 (Coupling Lemma) *Let X_t, Y_t be a coupling for $\mathcal{M}$ such that Y_0 has the stationary distribution π. Then, if X_t has distribution p_t,*

$$D_{\mathrm{tv}}(p_t, \pi) \leq \mathbf{Pr}(X_t \neq Y_t). \tag{2.19}$$

Proof Suppose $A_t \subseteq \Omega$ maximizes in (1.3). Then, since Y_t has distribution π,

$$\begin{aligned} D_{\mathrm{tv}}(p_t, \pi) &= \mathbf{Pr}(X_t \in A_t) - \mathbf{Pr}(Y_t \in A_t) \\ &\leq \mathbf{Pr}(X_t \in A_t, Y_t \notin A_t) \\ &\leq \mathbf{Pr}(X_t \neq Y_t). \end{aligned}$$

□

It is important to remember that the Markov chain Y_t is simply a proof construct, and X_t the chain we actually observe. We also require that $X_t = Y_t$ implies $X_{t+1} = Y_{t+1}$,

since this makes the right side of (2.19) nonincreasing. Then the earliest epoch T at which $X_T = Y_T$ is called *coalescence*, making T a random variable. A *successful coupling* is such that $\lim_{t\longrightarrow\infty} \mathbf{Pr}(X_t \neq Y_t) = 0$. Clearly we are only interested in successful couplings.

As an example consider our random walk on the cube Q_n. We can define a coupling as follows: Given (X_t, Y_t) we

(a) Choose i uniformly at random from $[n]$.

(b) Put $X_{t+1,j} = X_{t,j}$ and $Y_{t+1,j} = Y_{t,j}$ for $j \neq i$.

(c) If $X_{t,i} = Y_{t,i}$ then

$$X_{t+1,i} = Y_{t+1,i} = \begin{cases} X_{t,i} & \text{prob } \frac{1}{2} \\ 1 - X_{t,i} & \text{prob } \frac{1}{2} \end{cases}$$

(d) otherwise

$$(X_{t+1,i}, Y_{t+1,i}) = \begin{cases} (X_{t,i}, 1 - Y_{t,i}) & \text{prob } \frac{1}{2} \\ (1 - X_{t,i}, Y_{t,i}) & \text{prob } \frac{1}{2} \end{cases}$$

It should hopefully be clear that this is a coupling i.e. the marginals are correct and $X_t = Y_t$ implies $X_{t+1} = Y_{t+1}$.

Now let $I_t = \{j : \ i \text{ is chosen in (a) of steps } 1, 2, \ldots, t$. Then $I_t = [n]$ implies that $X_\tau = Y_\tau$ for $\tau \geq t$. So

$$\begin{aligned} \mathbf{Pr}(X_t \neq Y_t) &\leq \mathbf{Pr}(I_t \neq [n]) \\ &= \mathbf{Pr}(\bar{I}_t \neq \emptyset) \\ &\leq \mathbf{E}(|\bar{I}_t|) \\ &= n\left(1 - \tfrac{1}{n}\right)^t. \end{aligned}$$

So if $t = n(\log n + \log \epsilon^{-1})$ we have $d_{TV}(p_t, \pi) \leq \epsilon$.

A coupling is a *Markovian coupling* if the process $\mathcal{C}(\mathcal{M})$ is a Markov chain on $\Omega \times \Omega$. There always exists a *maximal coupling*, which gives equality in (2.19). This maximal coupling is in general non-Markovian, and is seemingly not constructible without knowing p_t $(t = 1, 2, \ldots)$. But coupling has little algorithmic value if we already know p_t. More generally, it seems difficult to prove mixing properties of non-Markovian couplings in our setting. Therefore we restrict attention to Markovian couplings, at the (probable) cost of sacrificing equality in (2.19).

Let $\mathcal{C}(\mathcal{M})$ be a Markovian coupling, with Q its transition matrix, i.e. the probability of a joint transition from (ω_1, ω_2) to (σ_1, σ_2) is $Q^{\omega_1\omega_2}_{\sigma_1\sigma_2}$. The precise conditions required of

Q are then

$$Q^{\omega\,\omega}_{\sigma_1\sigma_2} \neq 0 \qquad \text{implies} \quad \sigma_1 = \sigma_2 \qquad (\forall \omega \in \Omega), \tag{2.20}$$

$$\sum_{\sigma_2 \in \Omega} Q^{\omega_1\omega_2}_{\sigma_1\sigma_2} = P^{\omega_1}_{\sigma_1} \quad (\forall \omega_2 \in \Omega), \quad \sum_{\sigma_1 \in \Omega} Q^{\omega_1\omega_2}_{\sigma_1\sigma_2} = P^{\omega_2}_{\sigma_2} \quad (\forall \omega_1 \in \Omega). \tag{2.21}$$

Here (2.20) implies equality after coalescence, and (2.21) implies the marginals are copies of $\mathcal{M}$. Our goal is to design Q so that $\mathbf{Pr}(X_t \neq Y_t)$ quickly becomes small. We need only specify Q to satisfy (2.21) for $\omega_1 \neq \omega_2$. The other entries are completely determined by (2.20) and (2.21).

In general, to prove rapid mixing using coupling, it is usual to map $\mathcal{C}(\mathcal{M})$ to a process on $\mathbb{N}$ by defining a function $\psi : \Omega \times \Omega \longrightarrow \mathbb{N}$ such that $\psi(\omega_1, \omega_2) = 0$ implies $\omega_1 = \omega_2$. We call this a *proximity function*. Then $\mathbf{Pr}(X_t \neq Y_t) \leq \mathbf{E}(\psi(X_t, Y_t))$, by Markov's inequality, and we need only show that $\mathbf{E}(\psi(X_t, Y_t))$ converges quickly to zero.

2.4 Path coupling

A major difficulty with coupling is that we are obliged to specify it, and show improvement in the proximity function, for every pair of states. The idea of *path coupling*, where applicable, can be a major saving in this respect. We describe the approach below.

As a simple example of this approach consider a Markov chain where $\Omega \subseteq S^m$ for some set S and positive integer m. Suppose also that if $\omega, \sigma \in \Omega$ and $h(\omega, \sigma) = d$ (Hamming distance) then there exists a sequence $\omega = x_0, x_1, \ldots, x_d = \sigma$ of members of Ω such that (i) $\{x_0, x_1, \ldots, x_d\} \subseteq \Omega$, (ii) $h(x_i, x_{i+1}) = 1$, $i = 0, 1, \ldots, d-1$ and (iii) $P(x_i, x_{i+1}) > 0$.

Now suppose we define a coupling of the chains (X_t, Y_t) *only* for the case where $h(X_t, Y_t) = 1$. Suppose then that

$$\mathbf{E}(h(X_{t+1}, Y_{t+1}) \mid h(X_t, Y_t) = 1) \leq \beta \tag{2.22}$$

for some $\beta < 1$. Then

$$\mathbf{E}(h(X_{t+1}, Y_{t+1})) \leq \beta h(X_t, Y_t), \tag{2.23}$$

in all cases. It then follows that

$$d_{TV}(p_t, \pi) \leq \mathbf{Pr}(X_t \neq Y_t) \leq n\beta^t.$$

Equation (2.23) is shown by choosing a sequence $X_t = Z_0, Z_1, \ldots, Z_d = Y_t, d = h(X_t, Y_t)$ $Z_0, Z_1, \ldots, Z_d$ satisfy (i),(ii),(iii) above. Then we can couple Z_i and Z_{i+1}, $1 \leq i < d$ so that $X_{t+1} = Z'_0, Z'_1, \ldots, Z'_d = Y_{t+1}$ and (i) $\mathbf{Pr}(Z'_i = \sigma \mid Z_i = \omega) = P(\omega, \sigma)$ and (ii)

$\mathbf{E}(h(Z_i', Z_{i+1}')) \le \beta$. Therefore

$$\mathbf{E}(h(X_{t+1}, Y_{t+1})) \le \sum_{i=1}^{d} \mathbf{E}(h(Z_i', Z_{i+1}')) \le \beta d$$

and (2.23) follows.

As an example, let $G = (V, E)$ be a graph with maximum degree Δ and let $k \ge 2\Delta + 1$ be an integer. Let Ω_k be the set of proper k- vertex colourings of G i.e. $\{c : V \to [k]\}$ such that $(v, w) \in E$ implies $c(v) \ne c(w)$. We describe a chain which provides a good sampler for the uniform distribution over Ω_k. We let $\Omega = V^k$ be all k-colourings, including improper ones and describe a chain on Ω for which only proper colourings have a positive steady state probability.

To describe a general step of the chain asume $X_t \in \Omega$. Then

Step 1 Choose w uniformly from V and x uniformly from $[k]$.

Step 2 Let $X_{t+1}(v) = X_t(v)$ for $v \in V \setminus \{w\}$.

Step 3 If no neighbour of w in G has colour x then put $X_{t+1}(w) = x$, otherwise put $X_{t+1}(w) = x$.

Note that $P(\omega, \sigma) = P(\sigma, \omega) = \frac{1}{nk}$ for two proper colourings which can be obtained from each other by a single move of the chain. It follows from (1.15) that the steady state is uniform over Ω_k.

We first describe a coupling which is extremely simple but needs $k > 3\Delta$ in order for (2.22) to be satisfied. Let $h(X_t, Y_t) = 1$ and let v_0 be the unique vertex of V such that $X_t(v) \ne Y_t(v)$. In our coupling we choose w, x as in Step 1 and try to colour w with x in both chains.

We claim that

$$\mathbf{E}(h(X_{t+1}, Y_{t+1}) \le 1 - \frac{1}{n}\left(1 - \frac{\Delta}{k}\right) + \frac{\Delta}{n}\frac{2}{k} = 1 - \frac{k - 3\Delta}{kn}. \tag{2.24}$$

and so we can take $\beta \le 1 - \frac{1}{kn}$ in (2.23) if $k > 3\Delta$.

The term $\frac{1}{n}\left(1 - \frac{\Delta}{k}\right)$ in (2.24) lower bounds the probability that $w = v_0$ and that x is not used in the neighbourhood of v_0. In which case we will have $X_{t+1} = Y_{t+1}$. Next let $c_X \ne c_Y$ be the colours of v_0 in X_t, Y_t respectively. The term $\frac{\Delta}{n}\frac{2}{k}$ in (2.24) is an upper bound for the probability that w is in the neighbourhood of v_0 and $x \in \{c_X, c_Y\}$ and in which case we might have $h(X_{t+1}, Y_{t+1}) = 2$. In all other cases we find that $h(X_{t+1}, Y_{t+1}) = h(X_t, Y_t) = 1$.

A better coupling gives the desired result. We proceed as above except for the case where w is a neighbour of v_0 and $x \in \{c_X, c_Y\}$. In this case with probability $\frac{1}{2}$ we try to colour w with c_X in X_t and colour w with c_Y in Y_t, and fail in both cases. With probability $\frac{1}{2}$ we try to colour w with c_Y in X_t and colour w with c_X in Y_t, in which case the hamming distance may increase by one. Thus for this coupling we have

$$\mathbf{E}(h(X_{t+1}, Y_{t+1}) \leq 1 - \frac{1}{n}\left(1 - \frac{\Delta}{k}\right) + \frac{1}{2}\frac{\Delta}{n}\frac{2}{k} = 1 - \frac{k - 2\Delta}{kn}$$

and we can take $\beta \leq 1 - \frac{1}{kn}$ in (2.23) if $k > 2\Delta$.

We now give a more general framework for the definition of path coupling. Recall that a *quasi-metric* satisfies the conditions for a metric except possibly the symmetry condition. Any metric is a quasi-metric, but a simple example of a quasi-metric which is not a metric is directed edge distance in a digraph.

Suppose we have a relation $S \subseteq \Omega \times \Omega$ such that S has transitive closure $\Omega \times \Omega$, and suppose that we have a proximity function defined for all pairs in S, i.e. $\psi : S \longrightarrow \mathbb{N}$. Then we may lift ψ to a quasi-metric $\phi(\omega, \omega')$ on Ω as follows. For each pair $(\omega, \omega') \in \Omega \times \Omega$, consider the set $\mathcal{P}(\omega, \omega')$ of all sequences

$$\omega = \omega_1, \omega_2, \ldots, \omega_{r-1}, \omega_r = \omega' \quad \text{with} \quad (\omega_i, \omega_{i+1}) \in S \quad (i = 1, \ldots, r-1). \tag{2.25}$$

Then we set

$$\phi(\omega, \omega') \;=\; \min_{\mathcal{P}(\omega,\omega')} \sum_{i=1}^{r-1} \psi(\omega_i, \omega_{i+1}). \tag{2.26}$$

It is easy to prove that ϕ is a quasi-metric. We call a sequence minimizing (2.26) *geodesic*. We now show that, without any real loss, we may define the (Markovian) coupling only on pairs in S. Such a coupling is a called a path coupling. We give a detailed development below. Clearly $S = \Omega \times \Omega$ is always a relation whose transitive closure is $\Omega \times \Omega$, but path coupling is only useful when we can define a suitable S which is "much smaller" than $\Omega \times \Omega$. A relation of particular interest is $\mathcal{R}_\sigma$ from Section 1.4, but this is not always the best choice.

As in Section 2.3, we use σ (or σ_i) to denote a state obtained by performing a single transition of the chain from the state ω (or ω_i). Let P^ω_σ denote the probability of a transition from state ω to state σ in the Markov chain, and let $Q^{\omega\omega'}_{\sigma\sigma'}$ denote the probability of a joint transition from (ω, ω') to (σ, σ'), where $(\omega, \omega') \in S$, as specified by the path coupling. Since this coupling has the correct marginals, we have

$$\sum_{\sigma' \in \Omega} Q^{\omega\omega'}_{\sigma\sigma'} = P^\omega_\sigma, \qquad \sum_{\sigma \in \Omega} Q^{\omega\omega'}_{\sigma\sigma'} = P^{\omega'}_{\sigma'} \qquad (\forall (\omega, \omega') \in S). \tag{2.27}$$

We extend this to all pairs $(\omega, \omega') \in \Omega \times \Omega$, as follows. For each pair, fix a sequence $(\omega_1, \omega_2, \ldots, \omega_r) \in \mathcal{P}(\omega, \omega')$. We do not assume the sequence is geodesic here, or indeed

the existence of any proximity function, but this is our eventual purpose. The implied global coupling $\bar{Q}^{\omega_1\omega_r}_{\sigma_1\sigma_r}$ is then defined along this sequence by successively conditioning on the previous choice. Using (2.27), this can be written explicitly as

$$\bar{Q}^{\omega_1\omega_r}_{\sigma_1\sigma_r} = \sum_{\sigma_2\in\Omega}\sum_{\sigma_3\in\Omega}\cdots\sum_{\sigma_{r-1}\in\Omega} Q^{\omega_1\omega_2}_{\sigma_1\sigma_2}\frac{Q^{\omega_2\omega_3}_{\sigma_2\sigma_3}}{P^{\omega_2}_{\sigma_2}}\cdots\frac{Q^{\omega_{r-1}\omega_r}_{\sigma_{r-1}\sigma_r}}{P^{\omega_{r-1}}_{\sigma_{r-1}}}. \tag{2.28}$$

Summing (2.28) over σ_r or σ_1, and again applying (2.27), causes the right side to successively simplify, giving

$$\sum_{\sigma_r\in\Omega}\bar{Q}^{\omega_1\omega_r}_{\sigma_1\sigma_r} = P^{\omega_1}_{\sigma_1}\quad(\forall\omega_r\in\Omega),\qquad \sum_{\sigma_1\in\Omega}\bar{Q}^{\omega_1\omega_r}_{\sigma_1\sigma_r} = P^{\omega_r}_{\sigma_r}\quad(\forall\omega_1\in\Omega). \tag{2.29}$$

Hence the global coupling satisfies (2.21), as we would anticipate from the properties of conditional probabilities.

Now suppose the global coupling is determined by geodesic sequences. We bound the expected value of $\phi(\sigma_1,\sigma_r)$. This is

$$\begin{aligned}
\mathbf{E}(\phi(\sigma_1,\sigma_r)) &= \sum_{\sigma_1}\cdots\sum_{\sigma_r}\phi(\sigma_1,\sigma_r)\frac{Q^{\omega_1\omega_2}_{\sigma_1\sigma_2}Q^{\omega_2\omega_3}_{\sigma_2\sigma_3}\cdots Q^{\omega_{r-1}\omega_r}_{\sigma_{r-1}\sigma_r}}{P^{\omega_2}_{\sigma_2}\cdots P^{\omega_{r-1}}_{\sigma_{r-1}}}\\
&\le \sum_{\sigma_1}\cdots\sum_{\sigma_r}\sum_{i=1}^{r-1}\phi(\sigma_i,\sigma_{i+1})\frac{Q^{\omega_1\omega_2}_{\sigma_1\sigma_2}Q^{\omega_2\omega_3}_{\sigma_2\sigma_3}\cdots Q^{\omega_{r-1}\omega_r}_{\sigma_{r-1}\sigma_r}}{P^{\omega_2}_{\sigma_2}\cdots P^{\omega_{r-1}}_{\sigma_{r-1}}}\\
&= \sum_{i=1}^{r-1}\sum_{\sigma_1}\cdots\sum_{\sigma_r}\phi(\sigma_i,\sigma_{i+1})\frac{Q^{\omega_1\omega_2}_{\sigma_1\sigma_2}Q^{\omega_2\omega_3}_{\sigma_2\sigma_3}\cdots Q^{\omega_{r-1}\omega_r}_{\sigma_{r-1}\sigma_r}}{P^{\omega_2}_{\sigma_2}\cdots P^{\omega_{r-1}}_{\sigma_{r-1}}}\\
&= \sum_{i=1}^{r-1}\sum_{\sigma_i}\sum_{\sigma_{i+1}}\phi(\sigma_i,\sigma_{i+1})Q^{\omega_i\omega_{i+1}}_{\sigma_i\sigma_{i+1}},
\end{aligned} \tag{2.30}$$

where we have used the triangle inequality for a quasi-metric and the same observation as that leading from (2.28) to (2.29).

Suppose we can find $\beta\le 1$, such that, for all $(\omega,\omega')\in S$,

$$\mathbf{E}(\phi(\sigma,\sigma')) \;=\; \sum_{\sigma}\sum_{\sigma'}\phi(\sigma,\sigma')\,Q^{\omega\omega'}_{\sigma\sigma'} \;\le\; \beta\,\phi(\omega,\omega'). \tag{2.31}$$

Then, from (2.30), (2.31) and (2.26) we have

$$\mathbf{E}(\phi(\sigma_1,\sigma_r)) \;\le\; \sum_{i=1}^{r-1}\beta\,\phi(\omega_i,\omega_{i+1}) \;=\; \beta\sum_{i=1}^{r-1}\phi(\omega_i,\omega_{i+1}) \;=\; \beta\,\phi(\omega_1,\omega_r). \tag{2.32}$$

Thus we can show (2.31) for every pair, merely by showing that this holds for all pairs in S. To apply path coupling to a particular problem, we must find a relation S and

proximity function ψ so that this is possible. In particular we need $\phi(\omega, \omega')$ for $(\omega, \omega') \in S$ to be easily deducible from ψ.

Suppose that Ω has *diameter* D, i.e. $\phi(\omega, \omega') \leq D$ for all $\omega, \omega' \in \Omega$. Then, $\mathbf{Pr}(X_t \neq Y_t) \leq \beta^t D$ and so if $\beta < 1$ we have, since $\log \beta^{-1} \geq 1 - \beta$,

$$D_{\mathrm{tv}}(p_t, \pi) \leq \varepsilon \quad \text{for} \quad t \geq \log(D\varepsilon^{-1})/(1-\beta). \tag{2.33}$$

This bound is polynomial even when D is exponential in the problem size. It is also possible to prove a bound when $\beta = 1$, provided we know the quasi-metric cannot "get stuck". Specifically, we need an $\alpha > 0$ (inversely polynomial in the problem size) such that, in the above notation,

$$\mathbf{Pr}(\phi(\sigma, \sigma') \neq \phi(\omega, \omega')) \geq \alpha \qquad (\forall \omega, \omega' \in \Omega). \tag{2.34}$$

Observe that it is not sufficient simply to establish (2.34) for pairs in S. However, the structure of the path coupling can usually help in proving it. In this case, we can show that

$$D_{\mathrm{tv}}(p_t, \pi) \leq \varepsilon \quad \text{for} \quad t \geq \lceil eD^2/\alpha \rceil \lceil \ln(\varepsilon^{-1}) \rceil. \tag{2.35}$$

This is most easily shown using a martingale argument. Here we need D to be polynomial in the problem size.

Consider a sequence $(\omega_0, \omega_0'), (\omega_1, \omega_1') \ldots, (\omega_t, \omega_t')$ and define the random time $T^{\omega,\omega'} = \min\{t : \phi(\omega_t, \omega_t') = 0\}$, assuming that $\omega_0 = \omega, \omega_0' = \omega'$. We prove that

$$\mathbf{E}(T^{\omega,\omega'}) \leq D^2/\alpha. \tag{2.36}$$

Let

$$Z(t) = \phi(\omega_t, \omega_t')^2 - 2D\phi(\omega_t, \omega_t') - \alpha t$$

and let

$$\delta(t) = \phi(\omega_{t+1}, \omega_{t+1}') - \phi(\omega_t, \omega_t').$$

Then

$$\mathbf{E}(Z(t+1) \mid Z(0), Z(1), \ldots, Z(t)) - Z(t) = \\ 2(\phi(\omega_t, \omega_t') - D)\mathbf{E}(\delta(t) \mid \omega_t, \omega_t') + (\mathbf{E}(\delta(t)^2 \mid \omega_t, \omega_t') - \alpha) \geq 0.$$

Hence $Z(t)$ is a submartingale. The stopping time $T^{\omega,\omega'}$ has finite expectation and $|Z(t+1) - Z(t)| \leq D^2$. We can therefore apply the Optional Stopping Theorem for submartingales to obtain

$$\mathbf{E}(Z(T^{\omega,\omega'})) \geq Z(0).$$

This implies

$$-\alpha \mathbf{E}(T^{\omega,\omega'}) \geq \delta(0)^2 - 2D\delta(0)$$

and (2.36) follows.

So for any ω, ω'

$$\mathbf{Pr}(T^{\omega,\omega'} \geq eD^2/\alpha) \leq e^{-1}$$

and by considering k consecutive time intervals of length k we obtain

$$\mathbf{Pr}(T^{\omega,\omega'} \geq keD^2/\alpha) \leq e^{-k}$$

and (2.35) follows.

2.5 Hitting Time Lemmas

For a finite Markov chain $\mathcal{M}$ let $\mathbf{Pr}_i, \mathbf{E}_i$ denote probability and expectation, given that $X_0 = i$.

For a set $A \subseteq \Omega$ let

$$T_A = \min\{t \geq 0 : \ X_t \in A\}.$$

Then for $i \neq j$ the *hitting time*

$$H_{i,j} = \mathbf{E}_i(T_j)$$

is the expected number of steps needed to get from state i to state j.

The *commute time*

$$C_{i,j} = H_{i,j} + H_{j,i}.$$

Lemma 2.5.1 *Assume $X_0 = i$ and S is a stopping time with $X_S = i$. Let j be an arbitrary state. Then*

$$\mathbf{E}_i(\textit{number of visits to state } j \textit{ before time } S) = \pi_j \mathbf{E}_i(S).$$

Proof Consider the renewal process whose inter-renewal time is distributed as S. The reward-renewal theorem states that the asymptotic proportion of time spent in state j is given by

$$\mathbf{E}_i(\text{number of visits to } j \text{ before time } S)/\mathbf{E}_i(S).$$

This also equal to π_j, by the ergodic theorem. □

Lemma 2.5.2

$$\mathbf{E}_j(\textit{number of visits to } j \textit{ before } T_i) = \pi_j C_{i,j}.$$

Proof Let S be the time of the first return to i after the first visit to j. Apply Lemma 2.5.1. □

The *cover time* $C(\mathcal{M})$ of $\mathcal{M}$ is $\max_i C_i(\mathcal{M})$ where $C_i(\mathcal{M}) = \mathbf{E}_i(\max_j T_j)$ is the expected time to visit all states starting at i.

Let $\mathcal{M}_G$ denote a random walk on the connected graph $G = (V, E)$. Here $|V| = n$ and $|E| = m$.

Lemma 2.5.3 *For Markov chain $\mathcal{M}_G$ and $e = \{u, v\} \in E$, $C_{u,v} \leq 2m$.*

Proof The random walk on G induces a Markov chain on $A = \{(x, y) : \{x, y\}\}$ the set of oriented edges obtainable by replacing each edge $\{x, y\} \in E$ by a pair of oppositely oriented edges. It is can be easily checked that the all 1's vector satisfies (1.13) and hence the steady state of the induced walk is uniform. It follows from (1.14) the expected time between traversals of (v, u) is $\frac{1}{2m}$. So conditional on entering u from v the expected time to visit v and subsequently visit u is at most $\frac{1}{2m}$. Conditioning on initially traversing (v, u) is irrelevant to the time to subsequently visit v and then u and the lemma follows. □

We can use this to obtain a bound on the cover time of $\mathcal{M}_G$.

Lemma 2.5.4

$$C(\mathcal{M}_G) \leq 2m(n-1).$$

Proof Let T be any spanning tree of G and let $v_0, v_1, \ldots, v_{2n-2} = v_0$ be a traversal of G which crosses each edge of T in each direction. Now consider the expected time for the random walk, started at v_0, to make journeys from v_0 to v_1, then from v_1 onto v_2 and so on until $v_0, v_1, \ldots, v_{2n-2}$ have been visited. This journey visits every vertex of G and so its expected length is an upper bound on the cover time, from v_0. Thus

$$C_{v_0}(\mathcal{M}_G) \leq \sum_{i=0}^{2n-3} H_{v_i,v_{i+1}} = \sum_{\{u,v\}\in T} C_{u,v}.$$

The result now follows from Lemma 2.5.3. □

2.6 Optimal Stopping Rules

Lovász and Winkler, see for example [**?**] have been studying *optimal stopping rules*. We need a little of that theory here. For us a stopping rule is a function $\rho : \Omega^* :\rightarrow [0, 1]$ where $\Omega^* = \{(X_0, X_1, \ldots, X_t) : t \geq 0\}$ is the set of possible sequences of states generated by our Markov chain. $\rho(X_0, X_1, \ldots, X_t)$ is the probability that we stop the chain at time t. If X_0 is chosen with probability distribution σ and τ is the distribution of the state where we stop then we say that ρ is a stopping rule from σ to τ and write $\rho \in SR(\sigma, \tau)$. We are naturally mainly interested in the case where $\tau = \pi$.

For a stopping rule ρ let T_ρ be the random number of steps taken until we stop. Let

$$H(\sigma, \tau) = \inf \{\mathbf{E}(T_\rho) : \ \rho \in SR(\sigma, \tau)\}$$

denote the minimum expected number of steps in a stopping rule from σ to τ. If σ is concentrated on a single state s then we write $H(s,\tau)$.

For a stopping rule ρ and $j \in \Omega$ let $x_j = x_j(\rho)$ be the expected number of exits from i before stopping i.e. the expected number of times that the chain leaves i.

Lemma 2.6.1 *If $\rho \in SR(\sigma,\tau)$ then*

$$x_j + \tau_j = \sum_{i\in\Omega} x_i P(i,j) + \sigma_j.$$

Proof Let $T = T_\rho$ and consider the identity

$$\sum_{t=0}^{T-1} 1_{X_t=j} + 1_{T<\infty, X_T=j} = 1_{X_0=j} + \sum_{t=1}^{T} 1_{X_t=j}$$

where both sides count the number of times that $X_t = j$.

Taking expectations we have

$$\begin{aligned} x_j + \tau_j = \sigma_j + \sum_{t=1}^{T} \mathbf{Pr}(X_t = j) &= \sigma_j + \sum_{t=0}^{T-1}\sum_{i\in\Omega} \mathbf{Pr}(X_t = i)P(i,j) \\ &= \sigma_j + \sum_{i\in\Omega}\sum_{t=0}^{T-1} 1_{X_t=i}P(i,j) = \sigma_j + \sum_{i\in\Omega} x_i P(i,j). \end{aligned}$$

□

Corollary 2.6.1 *Let $\rho_1, \rho_2 \in SR(\sigma,\tau)$. Then for all $i \in \Omega$*

$$x_i(\rho_1) - x_i(\rho_2) = D\pi_i$$

where $D = \mathbf{E}(T_{\rho_1} - T_{\rho_2})$.

Proof It follows from Lemma 2.6.1 that $\xi = x(\rho_1) - x(\rho_2)$ satisfies

$$\xi_j = \sum_{i\in\Omega} \xi_i P(i,j).$$

Therefore $\xi = A\pi$ for some $A \geq 0$. Now for $k = 1, 2$

$$T_{\rho_k} = \sum_{j\in\Omega} x_j(\rho_k)$$

and the result follows. □

A state j is a halting state for rule ρ if $x_j(\rho) = 0$. This implies that if the chain ever enters state j then it stops. Using Corollary 2.6.1 we can prove the following remarkable theorem:

Theorem 2.6.1 *A stopping rule $\rho \in SR(\sigma, \tau)$ has a minimum mean expected stopping time iff there is a halting state.*

Proof If there exists j such that $x_j = 0$ then Corollary 2.6.1 implies that for $\rho' \in SR(\sigma, \tau)$

$$\mathbf{E}(T_{\rho'} - T_\rho) = \frac{x_j(\rho')}{\pi_j} \geq 0$$

implying that $\mathbf{E}(T_\rho)$ is minimal. It only remains to show that there exists a stopping rule in $SR(\sigma, \tau)$ which has at least one halting state.

The rule we define has a particular format. We define a nested sequence of sets $S_i = \{v_i, v_{i+1}, \ldots, v_n\}$ where $\Omega = \{v_1, v_2, \ldots, v_n\}$. For each i we will define $q^{(i)}$ by

$$q_j^{(i)} = \mathbf{Pr}(v_j \text{ is the first vertex of } S_i \text{ visited}).$$

In particular $q^{(1)} = \sigma$. We choose $S_1, S_2, \ldots, S_n$ so that we can write

$$\tau = \alpha_1 q^{(1)} + \alpha_2 q^{(2)} + \cdots + \alpha_n q^{(n)} \tag{2.37}$$

where $\alpha \geq 0$ and $\alpha_1 + \alpha_2 + \cdots + \alpha_n = 1$. Our stopping rule ρ is then:

(i) Choose i with probability α_i.

(ii) Choose X_0 with probability σ and then run the chain until S_i is reached and then stop.

It should be clear that $\rho \in SR(\sigma, \tau)$. If $S_1, S_2, \ldots, S_n$ can be constructed so that (2.37) holds then we are done: v_n is a halting state.

Assume inductively that we have found $S_1, S_2, \ldots, S_i$ and $\alpha_1, \alpha_2, \ldots, \alpha_{i-1} \geq 0$ such that

$$\tau^{(i-1)} = \tau - (\alpha_1 q^{(1)} + \alpha_2 q^{(2)} + \cdots + \alpha_{i-1} q^{(i-1)}) \geq 0 \tag{2.38}$$

and

$$\alpha_1 + \alpha_2 + \cdots + \alpha_{i-1} \leq 1.$$

Putting $S_1 = \Omega$ does this for $i = 1$ and then for general i let

$$\alpha_i = \min_{j \in S_i} \frac{\tau_j^{(i)}}{q_j^{(i)}}$$

and let v_i be a state of S_i which achieves the minimum. Clearly $\alpha_i \geq 0$ and

$$\tau^{(i)} = \tau^{(i-1)} - \alpha_i q^{(i)} \geq 0 \tag{2.39}$$

from the definition of α_i.

Furthermore

$$\sum_{j=1}^{i} \alpha_j = \sum_{j=1}^{i} \alpha_j \sum_{k=1}^{n} q_k^{(j)} \leq \sum_{k=1}^{n} \tau_k = 1 \tag{2.40}$$

completing the induction.

Finally note that when $i = n$ the construction yields equality in (2.39) and then (2.37) holds and we obtain equality in (2.40). □

We now relate optimal stopping rules and mixing time. Let

$$T_{\text{mix}} = \max_{s \in \Omega} H(s, \pi).$$

Theorem 2.6.2

$$\tau(\epsilon) \leq 8 T_{\text{mix}} \log_2(1/\epsilon).$$

Proof Let $s \in \Omega$ and let ρ be an optimal stopping rule from s to π. Consider a modification: Follow ρ until it stops after $T = T_\rho$ steps and then generate $\xi \in \{0, 1, \ldots, t-1\}$ uniformly and independently of the previous walk, and then walk ξ more steps. Let the walk be $v_1, v_2, \ldots, v_{T+\xi}$. Then let $\eta = T + \xi \pmod t$ and note that η is uniformly distributed over $\{0, 1, \ldots, t-1\}$. Then for $i \in \Omega$

$$\mathbf{Pr}(v_\eta = i) \geq \mathbf{Pr}(v_{T+\xi} = i) - \mathbf{Pr}(v_{T+\xi} = i, v_\eta \neq i) \geq \pi_i - \mathbf{Pr}(v_{T+\xi} = i, T + \xi \geq t)$$

since $v_{T+\xi}$ is in the stationary distribution and $T + \xi < t$ implies $\eta = T + \xi$.

Hence, for every $A \subseteq \Omega$,

$$\pi(A) - \mathbf{Pr}(v_\eta \in A) \leq \mathbf{Pr}(v_{T+\xi} \in A, T + \xi \geq t) \leq \mathbf{Pr}(T + \xi \geq t).$$

Now for any fixed value of T, $\mathbf{Pr}(T + \xi \geq t) \leq \frac{T}{t}$ and so

$$\mathbf{Pr}(T + \xi \geq t) \leq \frac{\mathbf{E}(T)}{t} = \frac{H(s, \pi)}{t}$$

and

$$\pi(A) - \mathbf{Pr}(v_\eta \in A) \leq \frac{H(s, \pi)}{t}.$$

It follows from Lemma 1.3.1(d) that

$$d(t) \leq \frac{T_{\text{mix}}}{t}$$

and so

$$d(4T_{\text{mix}}) \leq \frac{1}{4}.$$

Applying Lemma 1.3.1(b) we see that

$$d(8T_{\text{mix}} \log_2 \epsilon^{-1}) \leq \epsilon.$$

□

We can now prove a refinement of the usual conductance bound on the mixing time (Corollary 2.2.1) due to Kannan and Lovász [**?**]. Thus for $0 \leq x \leq \frac{1}{2}$ let

$$\Phi(x) = \min_{\substack{S \subseteq \Omega \\ \pi(S) \leq x}} \frac{Q(S, \bar{S})}{\pi(S)\pi(\bar{S})}$$

and let $\Phi(x) = \Phi(\frac{1}{2})$ for $\frac{1}{2} < x \leq 1$. Note that $\Phi(x) \leq 2$.

Theorem 2.6.3 *If* $0 \leq \xi \leq 1$ *then*

$$T_{\text{mix}} \leq \frac{30}{\xi^2} + 30 \int_{x=\pi_\xi}^{1} \frac{dx}{x\Phi(x)^2}$$

where $\pi_\xi = \inf \{y : \exists S \text{ such that } \pi(S) \leq y \text{ and } \Phi(S) < \xi\}$.

Proof Let $s \in \Omega$ and ρ be an optimal stopping rule from s to π. Let $y_i = x_i/\pi_i$, $i = 1, 2, \ldots, n$ be the *scaled exit frequencies* of ρ. Now order the states so that $y_1 \leq y_2 \leq \cdots \leq y_n$. We first claim that with this ordering

$$y_1 = 0 \text{ and } n = s. \tag{2.41}$$

The first assertion comes from Theorem 2.6.1. For the second we use Lemma 2.6.1 and write, for $j \in \Omega$,

$$\sum_{i \in \Omega} \pi_i P(i, j) y_i - \pi_j y_j = \pi_j - 1_{j=s}.$$

Putting $j = n$ we obtain

$$\pi_n - 1_{n=s} \leq \sum_{i \in \Omega} \pi_i P(i, n) y_n - \pi_n y_n = 0$$

and (2.41) follows.

Now fix $1 \leq k < m \leq n$ and let $A = \{1, 2, \ldots, k\}$, $B = \{k+1, k+2, \ldots, m-1\}$ and $C = \{m, m+1, \ldots, n\}$. We show next that

$$y_m - y_k \leq \frac{\pi(A)}{Q(C, A)}. \tag{2.42}$$

We start with the identity

$$\sum_{i=1}^{k} \sum_{j=k+1}^{n} y_j Q(j, i) - \sum_{i=1}^{k} \sum_{j=k+1}^{n} y_i Q(i, j) = \pi(A). \tag{2.43}$$

The left hand side counts the expected number of steps from $V \setminus A$ to A less the expected number of steps from A to $V \setminus A$, when following an optimal rule. Since we do not start in A ($s = n$) and stop in A with probability $\pi(A)$, (2.43) follows.

Now we estimate the left hand side of (2.43) as follows:

$$\sum_{i=1}^{k}\sum_{j=k+1}^{n} y_j Q(j,i) \geq \sum_{i=1}^{k}\sum_{j=m+1}^{n} y_m Q(j,i) + \sum_{i=1}^{k}\sum_{j=k+1}^{m-1} y_k Q(j,i)$$
$$= y_m Q(C,A) + y_k Q(B,A)$$

and

$$\sum_{i=1}^{k}\sum_{j=k+1}^{n} y_i Q(i,j) \leq \sum_{i=1}^{k}\sum_{j=k+1}^{n} y_k Q(i,j) = y_k Q(A, B\cup C) = y_k Q(B\cup C, A).$$

Substituting into (2.43) we get

$$y_m Q(C,A) + y_k Q(B,A) - y_k Q(B\cup C, A) = (y_m - y_k)Q(C,A) \leq \pi(A)$$

which proves (2.42).

We now observe that since $y_1 = 0$,

$$H(s,\pi) = \sum_{i=1}^{n} \pi_i y_i = \sum_{j=1}^{n-1} (y_{j+1} - y_j)\pi_{>j} \tag{2.44}$$

where $\pi_{>j} = \sum_{r=j+1}^{n} \pi_j$.

We now define a sequence $1 = m_0 < m_1 < \cdots m_k < m_{k+1}$ so that if $T_i = \{1, 2, \ldots, m_i\}$, $\bar{T}_i = \Omega \setminus T_i$ and $a_i = \pi(T_i)$ then

$$a_{i+1} - \pi_{m_{i+1}} < a_i \left(1 + \frac{\Phi(a_i)}{4}\right) \leq a_{i+1} \tag{2.45}$$

and

$$a_k \leq \frac{1}{2} < a_{k+1}. \tag{2.46}$$

This definition can be justified as follows: Given m_i with $a_i \leq \frac{1}{2}$ we let m_{i+1} be the first integer such that (2.45) holds. Since $a_n = 1$ and $a_i \left(1 + \frac{\Phi(a_i)}{4}\right) \leq \frac{3}{2} a_i$, such an m_{i+1} exists. k exists for the same reason.

We bound a portion of the sum in the right hand side of (2.44) by

$$\sum_{j=m_i}^{m_{i+1}-1} (y_{j+1} - y_j)\pi_{>j} \leq (1 - a_i)(y_{m_{i+1}} - y_{m_i}) \leq \frac{a_i(1-a_i)}{Q(\bar{T}_{i+1} \cup \{m_{i+1}\}, T_i)} \tag{2.47}$$

where the second inequality follows from (2.43). Now,

$$\begin{aligned} Q(\bar{T}_{i+1} \cup \{m_{i+1}\}, T_i) &= Q(T_i, \bar{T}_i) - Q(\bar{T}_i \setminus (\bar{T}_{i+1} \cup \{m_{i+1}\}, T_i)) \geq \\ &Q(T_i, \bar{T}_i) - \pi(\bar{T}_i \setminus (\bar{T}_{i+1} \cup \{m_{i+1}\})) \geq \\ &\Phi(a_i)a_i(1-a_i) - a_{i+1} + \pi_{m_{i+1}} + a_i > \Phi(a_i)a_i(1-a_i)/2. \end{aligned}$$

Hence we obtain from (2.47) that

$$\sum_{j=m_i}^{m_{i+1}-1} (y_{j+1} - y_j)\pi_{>j} \leq \frac{2}{\Phi(a_i)}. \tag{2.48}$$

Now define i_0 by $\Phi(a_i) \geq \xi$ iff $i \leq i_0$. It follows from (2.45) that

$$i_0 \leq \frac{\ln 2}{\ln\left(1 + \frac{\xi}{4}\right)} \leq \frac{5}{\xi}.$$

So from (2.48) we see that

$$\sum_{j=1}^{m_{i_0+1}-1} (y_{j+1} - y_j)\pi_{>j} \leq \sum_{i=1}^{i_0} \frac{2}{\Phi(a_i)} \leq \frac{10}{\xi^2}. \tag{2.49}$$

In general we have

$$\begin{aligned} \int_{a_i}^{a_{i+1}} \frac{dx}{x\Phi(x)^2} \geq \frac{1}{\Phi(a_i)^2} \int_{a_i}^{a_{i+1}} \frac{dx}{x} &= \frac{1}{\Phi(a_i)^2} \ln(a_{i+1}/a_i) \\ &\geq \frac{1}{\Phi(a_i)^2} \ln\left(1 + \frac{\Phi(a_i)}{4}\right) \geq \frac{1}{5\Phi(a_i)} \end{aligned} \tag{2.50}$$

since $\Phi(a_i) \leq 2$.

So from (2.48), (2.49) and (2.50) we have

$$\sum_{j=1}^{m_{k+1}-1} (y_{j+1} - y_j)\pi_{>j} \leq \frac{10}{\xi^2} + 10 \int_{\pi_\xi}^{1} \frac{dx}{x\Phi(x)^2}. \tag{2.51}$$

The estimate for the other half of the sum on the right hand sise of (2.44) is similar. We define a sequence $n_0 = n > n_1 > \cdots > n_r$ and sets $S_i = \{n_i, n_i + 1, \ldots, n\}$, $\bar{S}_i = \Omega \setminus S_i$ and $b_i = \pi(S_i)$ for $i = 1, 2, \ldots, r+1$ such that

$$b_{i+1} - \pi_{n_{i+1}} < b_i \left(1 + \frac{\Phi(b_i)}{4}\right) \leq b_{i+1}$$

and

$$b_r \leq \frac{1}{2} < b_{r+1}.$$

As before we consider the partial sum

$$\sum_{j=n_{i+1}}^{n_i-1}(y_{j+1}-y_j)\pi_{>j}\le(b_{i+1}-\pi_{n_{i+1}})(y_{n_i}-y_{n_{i+1}})\le\frac{(b_{i+1}-\pi_{n_{i+1}})(1-b_{i+1}+\pi_{n_{i+1}})}{Q(S_i,\bar{S}_{i+1}\cup\{n_{i+1}\})}$$

where the second inequality follows from (2.42).

Now

$$Q(S_i,\bar{S}_{i+1}\cup\{n_{i+1}\})=Q(\bar{S}_{i+1}\cup\{n_{i+1}\},S_i)=$$
$$Q(\bar{S}_i,S_i)-Q(\bar{S}_i\setminus(\bar{S}_{i+1}\cup\{n_{i+1}\}),S_i)\ge Q(\bar{S}_i,S_i)-\pi(\bar{S}_{i+1}\setminus\bar{S}_i)+\pi_{n_{i+1}}\ge$$
$$\Phi(b_i)b_i(1-b_i)-b_{i+1}+\pi_{n_{i+1}}+b_i>\Phi(b_i)b_i(1-b_i)/2.$$

Hence

$$\sum_{j=n_{i+1}}^{n_i-1}(y_{j+1}-y_j)\pi_{>j}\le\frac{2(b_{i+1}-\pi_{n_{i+1}})}{b_i}\frac{1}{\Phi(b_i)}\le\frac{4}{\Phi(b_i)}$$

since $b_{i+1}-\pi_{n_{i+1}}\le b_i\left(1+\frac{\Phi(b_i)}{4}\right)\le 2b_i$. So as before we get

$$\sum_{j=n_{r+1}}^{n-1}(y_{j+1}-y_j)\pi_{>j}\le\frac{20}{\xi^2}+20\int_{\pi_\xi}^1\frac{dx}{x\Phi(x)^2}$$

and combined with (2.49) and (2.44) we have the theorem. □

Of paticular interest to us is the case where for some $A=A(n)<B=B(n)$, $\Phi(x)$ satisfies

$$\Phi(x)\ge\min\left\{A\log\frac{1}{x(1-x)},B\right\}\tag{2.52}$$

for $x\le 1/2$.

Theorem 2.6.4 *If (2.52) holds then the mixing time*

$$\tau(\epsilon)\le cA^{-2}$$

for some absolute constant $c>0$.

Proof It follows from (2.52) that for $\xi\le B$ we have $\pi_\xi\ge e^{-\xi/A}$. ($\pi_\xi<e^{-\xi/A}$ implies that $\exists S:\ x=\pi(S)<e^{-\xi/A}$ and $\Phi(x)<\xi$, which implies that $\min\{A\log\frac{1}{x(1-x)},B\}<\xi$, contradiction). Define x_0 by $A\log\frac{1}{x_0(1-x_0)}=B$. Then for $\xi\le B$ we have by Theorems 2.6.2 and 2.6.3 that

$$\begin{aligned}\tau(\epsilon)&=O\left(\frac{1}{\xi^2}+\frac{1}{B^2}\int_{e^{-\xi/A}}^{x_0}\frac{dx}{x}+\frac{1}{A^2}\int_{x_0}^{1/2}\frac{dx}{x(\log x)^2}+\frac{1}{A^2(\log 4)^2}\int_{1/2}^1\frac{dx}{x}\right)\\&=O\left(\frac{1}{\xi^2}+\frac{1}{B^2}\left(\log x_0+\frac{\xi}{A}\right)+\frac{1}{A^2}\left(\frac{1}{\log x_0}+\frac{1}{\log 2}\right)+\frac{1}{A^2}\right)\\&=O(A^{-2})\end{aligned}$$

where we use $\log x_0 = \Theta(B/A)$ and take $\xi = (AB^2/2)^{1/3}$ and absorb terms of order $(AB)^{-1}$ or B^{-2}. □

2.7 Coupling from the Past

Chapter 3

Matchings and related structures

A problem that has played a historically important role in the development both of complexity theory and algorithm design is that of evaluating the permanent function. The *permanent* of an $n \times n$ integer matrix $A = (a_{ij} : 0 \leq i, j \leq n-1)$ is defined by

$$\operatorname{per} A = \sum_{\pi} \prod_{i=0}^{n-1} a_{i,\pi(i)} \,,$$

where the sum is over all permutations π of $[n] = \{0, \ldots, n-1\}$. Evaluating the permanent of a 0,1-matrix is complete for the class #P; thus, we cannot expect to find an algorithm that solves the problem exactly in polynomial time. Interest has therefore centred on finding computationally feasible approximation algorithms. In contrast, as is well known, the superficially related *determinant* of an $n \times n$ matrix can be evaluated in $\mathrm{O}(n^3)$ arithmetic operations using Gaussian elimination.

A *matching* in a graph $G = (V, E)$ is any subset $A \subseteq E$ of edges that are pairwise vertex disjoint. A matching is said to be *perfect* if it covers every vertex; clearly a perfect matching, which can exist only if $|V|$ is even, has size $|V|/2$. Specialised to the case when A is a 0,1-matrix, $\operatorname{per} A$ is equal to the number of perfect matchings in the bipartite graph $G = (V_1, V_2, E)$, where $V_1 = V_2 = [n]$, and $(i, j) \in E$ iff $a_{ij} = 1$.

In the light of the above connection, a promising approach to computing an approximation of the permanent of A, at least when A is a 0,1-matrix, is through sampling perfect matchings in the related bipartite graph G. We shall immediately generalise the situation to that of sampling a (weighted) matching in a general graph. In the next section we attack that sampling problem through Markov chain simulation; then in subsequent sections we shall apply the methods we develop there to related problems, including the approximation of the permanent.

3.1 Weighted matchings (the monomer-dimer model)

Let G be a graph, not necessarily bipartite, with an even number $2n = |V|$ of vertices. The assumption that the number of vertices in G is even is inessential and is made for notational convenience. To each matching M, a *weight* $w(M) = \lambda^{|M|}$ is assigned, where λ is a positive real parameter. The generating (or partition) function of matchings in G is

$$Z(\lambda) \equiv Z_G(\lambda) = \sum_M w(M) = \sum_{k=0}^{n} m_k \lambda^k, \tag{3.1}$$

where $m_k \equiv m_k(G)$ is the number of k-matchings in G. In statistical physics, a matching is termed a "monomer-dimer configuration": the edges in M are the "dimers" and the unmatched (uncovered) vertices are "monomers". Thus $m_k(G)$ counts the number of monomer-dimer configurations with k dimers. The weight parameter λ reflects the contribution of a dimer to the energy of the system.

Our main goal in this section is the development of an algorithm for approximating Z_G at an arbitrary point $\lambda \geq 0$. The running time of the algorithm is $\mathrm{poly}(n, \epsilon, \max\{\lambda, 1\})$, where ϵ, as usual, controls the relative error that will be tolerated in the output. Thus the algorithm will meet the specification of an FPRAS for Z_G, provided λ is specified in unary notation. Our approach is to simulate a suitable Markov chain $\mathcal{M}_{\mathrm{match}}(\lambda)$, parameterised on the the graph G and edge weight λ. The state space, Ω, is the set of all matchings in G, and the transitions are constructed so that the chain is ergodic with stationary distribution π_λ given by

$$\pi_\lambda(M) = \frac{\lambda^{|M|}}{Z(\lambda)}. \tag{3.2}$$

(Since G is fixed from now on, we drop the subscript from Z.) In other words, the stationary probability of each matching (monomer-dimer configuration) is proportional to its weight in the partition function (3.1). The Markov chain $\mathcal{M}_{\mathrm{match}}(\lambda)$, if simulated for sufficiently many steps, provides a method of sampling matchings from the distribution π_λ.

It is not hard to construct a Markov chain $\mathcal{M}_{\mathrm{match}}(\lambda)$ with the right asymptotic properties. Consider the chain in which transitions from any matching M are made according to the following rule:

1. with probability $\frac{1}{2}$ let $M' = M$; otherwise,

2. select an edge $e = \{u, v\} \in E$ u.a.r. and set

$$M' = \begin{cases} M - e & \text{if } e \in M; \\ M + e & \text{if both } u \text{ and } v \text{ are unmatched in } M; \\ M + e - e' & \text{if exactly one of } u \text{ and } v \text{ is matched in } M \\ & \qquad \text{and } e' \text{ is the matching edge;} \\ M & \text{otherwise;} \end{cases}$$

3. go to M' with probability $\min\{1, \pi_\lambda(M')/\pi_\lambda(M)\}$.

It is helpful to view this chain as follows. There is an underlying graph defined on the set of matchings Ω in which the neighbours of matching M are all matchings M' that differ from M via one of the following local perturbations: an edge is removed from M (a $\downarrow$-*transition*); an edge is added to M (a $\uparrow$-*transition*); or a new edge is exchanged with an edge in M (a $\leftrightarrow$-*transition*). Transitions from M are made by first selecting a neighbour M' u.a.r., and then actually making, or *accepting* the transition with probability $\max\{1, \pi_\lambda(M')/\pi_\lambda(M)\}$. Note that the ratio appearing in this expression is easy to compute: it is just λ^{-1}, λ or 1 respectively, according to the type of the transition.

As the reader may easily verify, this acceptance probability is constructed so that the transition probabilities $P(M, M')$ satisfy the detailed balance condition

$$Q(M, M') = \pi_\lambda(M)P(M, M') = \pi_\lambda(M')P(M', M), \quad \text{for all } M, M' \in \Omega,$$

i.e., $\mathcal{M}_{\text{match}}(\lambda)$ is reversible. This fact, together with the observation that $\mathcal{M}_{\text{match}}(\lambda)$ is irreducible (i.e., all states communicate, for example via the empty matching) and aperiodic (by step 1, the self-loop probabilities $P(M, M)$ are all non-zero), ensures that $\mathcal{M}_{\text{match}}(\lambda)$ is ergodic with stationary distribution π_λ, as required. The device of performing random walk on a connected graph with acceptance probabilities of this form is well known in Monte Carlo physics under the name of the "Metropolis process". Clearly, it can be used to achieve any desired stationary distribution π for which the ratio $\pi(u)/\pi(v)$ for neighbours u, v can be computed easily. It is also the standard mechanism used in combinatorial optimisation by simulated annealing.

Having constructed a family of Markov chains with stationary distribution π_λ, our next task is to explain how samples from this distribution can be used to obtain a reliable statistical estimate of $Z(\lambda)$ at a specified point $\lambda = \widehat{\lambda} \geq 0$. Our strategy is to express $Z(\widehat{\lambda})$ as the product

$$Z(\widehat{\lambda}) = \frac{Z(\lambda_r)}{Z(\lambda_{r-1})} \times \frac{Z(\lambda_{r-1})}{Z(\lambda_{r-2})} \times \cdots \frac{Z(\lambda_2)}{Z(\lambda_1)} \times \frac{Z(\lambda_1)}{Z(\lambda_0)} \times Z(\lambda_0), \tag{3.3}$$

where $0 = \lambda_0 < \lambda_1 < \lambda_2 < \cdots < \lambda_{r-1} < \lambda_r = \widehat{\lambda}$ is a suitably chosen sequence of values. Note that $Z(\lambda_0) = Z(0) = 1$. We will then estimate each factor $Z(\lambda_i)/Z(\lambda_{i-1})$ in this

product by sampling from the distribution π_{λ_i}. This approach is analogous to that used in the context of independent sets in the proof of Theorem 1.2.1; refer in particular to equation (1.4). For reasons that will become clear shortly, we will use the sequence of values $\lambda_1 = (2|E|)^{-1}$ and $\lambda_i = (1 + \frac{1}{n})^{i-1}\lambda_1$ for $1 \leq i < r$. The length r of the sequence is taken to be minimal such that $(1 + \frac{1}{n})^{r-1}\lambda_1 \geq \widehat{\lambda}$, so we have the bound

$$r \leq \lceil 2n\big(\ln \widehat{\lambda} + \ln(2|E|)\big) \rceil + 1. \tag{3.4}$$

To estimate the ratio $Z(\lambda_i)/Z(\lambda_{i-1})$, we will express it, or rather its reciprocal, as the expectation of a suitable random variable. Specifically, define the random variable $Z_i(M) = \big(\frac{\lambda_{i-1}}{\lambda_i}\big)^{|M|}$, where M is a matching chosen from the distribution π_{λ_i}. Then we have

$$\mathbf{E}(Z_i) = \sum_M \left(\frac{\lambda_{i-1}}{\lambda_i}\right)^{|M|} \frac{\lambda_i^{|M|}}{Z(\lambda_i)} = \frac{1}{Z(\lambda_i)} \sum_M \lambda_{i-1}^{|M|} = \frac{Z(\lambda_{i-1})}{Z(\lambda_i)}.$$

Thus the ratio $\rho_i = Z(\lambda_{i-1})/Z(\lambda_i)$ can be estimated by sampling matchings from the distribution π_{λ_i} and computing the sample mean of Z_i. Following (3.3), our estimator of $Z(\widehat{\lambda})$ will be the product of the reciprocals of these estimated ratios. Summarising this discussion, our algorithm can be written down as follows:

Step 1 Compute the sequence $\lambda_1 = (2|E|)^{-1}$ and $\lambda_i = \big(1 + \frac{1}{n}\big)^{i-1} \lambda_1$ for $1 \leq i < r$, where r is the least integer such that $\big(1 + \frac{1}{n}\big)^{r-1} \lambda_1 \geq \widehat{\lambda}$. Set $\lambda_0 = 0$ and $\lambda_r = \widehat{\lambda}$.

Step 2 For each value $\lambda = \lambda_1, \lambda_2, \ldots, \lambda_r$ in turn, compute an estimate X_i of the ratio ρ_i as follows:

- by performing S independent simulations of the Markov chain $\mathcal{M}_{\text{match}}(\lambda_i)$, each of length T_i, obtain an independent sample of size S from (close to) the distribution π_{λ_i};
- let X_i be the sample mean of the quantity $\big(\frac{\lambda_{i-1}}{\lambda_i}\big)^{|M|}$.

Step 3 Output the product $Y = \prod_{i=1}^{r} X_i^{-1}$.

Figure 3.1: Algorithm MatchSample

To complete the description of the algorithm, we need to specify the sample size S in Step 2, and the number of simulation steps T_i required for each sample. Our goal is to show that, with suitable values for these quantities, Algorithm MatchSample is an FPRAS for $Z(\lambda)$.

The issue of the sample size S is straightforward. Now $\mathrm{e}^{-1} \leq Z_i \leq 1$ and so using Lemma 1.2.1 of Chapter 1 we see

Proposition 3.1.1 *In Algorithm* MatchSample*, suppose the sample size S in Step 2 is $S = \lceil 17\mathrm{e}^2\epsilon^{-2}r \rceil$, and that the simulation length T_i is large enough that the variation distance of $\mathcal{M}_{\mathrm{match}}(\lambda_i)$ from its stationary distribution π_{λ_i} is at most $\epsilon/(3\mathrm{e}r)$. Then the output random variable Y satisfies*

$$\mathbf{Pr}\big((1-\epsilon)Z(\widehat{\lambda}) \le Y \le (1+\epsilon)Z(\widehat{\lambda})\big) \ge \tfrac{3}{4}.$$

Since r is a relatively small quantity (essentially linear in n: see (3.4)), this result means that a modest sample size at each stage suffices to ensure a good final estimate Y, provided of course that the samples come from a distribution that is close enough to π_{λ_i}.

It is in determining the number of simulation steps, T_i, required to achieve this that the meat of the analysis lies: of course, this is tantamount to investigating the mixing time of the Markov chain $\mathcal{M}_{\mathrm{match}}(\lambda_i)$. Our main task in this section will be to show:

Proposition 3.1.2 *The mixing time of the Markov chain $\mathcal{M}_{\mathrm{match}}(\lambda)$ satisfies*

$$\tau_X(\epsilon) \le 4|E|n\lambda'\big(n(\ln n + \ln\lambda') + \ln\epsilon^{-1}\big),$$

where $\lambda' = \max\{\lambda, 1\}$.

The proof of this result will make use of the full power of the machinery introduced in Section 2.2.3 of Chapter 2. Note that Proposition 3.1.2 is a very strong statement: it says that we can sample from (close to) the complex distribution π_λ over the exponentially large space of matchings in G, by performing a Markov chain simulation of length only a low-degree polynomial in the size of G.[1]

According to Proposition 3.1.1, we require a variation distance of $\epsilon/(3\mathrm{e}r)$, so Proposition 3.1.2 tells us that it suffices to take

$$T_i = \big\lceil 4|E|n\lambda_i'\big(n(\ln n + \ln\lambda_i') + \ln(3\mathrm{e}r/\epsilon)\big)\big\rceil. \tag{3.5}$$

This concludes our specification of the Algorithm MatchSample.

Before proceeding to prove the above statements, let us convince ourselves that together they imply that Algorithm MatchSample is an FPRAS for $Z(\lambda)$. First of all, Proposition 3.1.1 ensures that the output of Algorithm MatchSample satisfies the requirements of an FPRAS for Z. It remains only to verify that the running time is bounded by a polynomial in n, $\widehat{\lambda}'$ and ϵ^{-1}. Evidently the running time is dominated by the number of Markov chain simulations steps, which is $\sum_{i=1}^{r} ST_i$; since T_i increases with i, this is at most rST_r. Substituting the upper bound for r from (3.4), and values

[1] Incidentally, we should point out that Proposition 3.1.2 immediately tells us that we can sample monomer-dimer configurations from the canonical distribution π_λ, in time polynomial in n and λ'. This is in itself an interesting result, and allows estimation of the expectation of many quantities associated with monomer-dimer configurations.

for S from Proposition 3.1.1 and T_r from (3.5), we see that the overall running time of Algorithm MatchSample is bounded by[2]

$$O\big(n^4|E|\widehat{\lambda}'(\ln n\widehat{\lambda}')^3\epsilon^{-2}\big),$$

which grows only polynomially with n, $\widehat{\lambda}'$ and ϵ^{-1}. We have therefore proved

Theorem 3.1.1 *Algorithm* MatchSample *is an FPRAS for the partition function of an arbitrary monomer-dimer system.*

We turn now to the question of proving Proposition 3.1.2. Our strategy will be to carefully choose a collection of canonical paths $\Gamma = \{\gamma_{XY} : X, Y \in \Omega\}$ in the Markov chain $\mathcal{M}_{\text{match}}(\lambda)$ for which the "bottleneck" measure $\bar{\rho}(\Gamma)$ of Section 2.2.3 is small. We can then appeal to Theorem 2.2.4 and Corollary 2.1.1 to bound the mixing time. Specifically, we shall show that our paths satisfy

$$\bar{\rho}(\Gamma) \leq 4|E|n\lambda'. \tag{3.6}$$

Since the number of matchings in G is certainly bounded above by $(2n)!$, the stationary probability $\pi_\lambda(X)$ of any matching X is bounded below by $\pi_\lambda(X) \geq 1/(2n)!\lambda'^n$. Using (3.6) and the fact that $\ln n! \leq n \ln n$, the bound on the mixing time in Proposition 3.1.2 can now be read off from Theorem 2.2.4 and Corollary 2.1.1.

It remains for us to find a set of canonical paths Γ satisfying (3.6). For a pair of matchings X, Y in G, we define the canonical path γ_{XY} as follows. Consider the symmetric difference $X \oplus Y$. A moment's reflection should convince the reader that this consists of a disjoint collection of paths in G (some of which may be closed cycles), each of which has edges that belong alternately to X and to Y. Now suppose that we have fixed some arbitrary ordering on all simple paths in G, and designated in each of them a so-called "start vertex", which is arbitrary if the path is a closed cycle but must be an endpoint otherwise. This ordering induces a unique ordering $P_1, P_2, \ldots, P_m$ on the paths appearing in $X \oplus Y$. The canonical path from X to Y involves "unwinding" each of the P_i in turn as follows. There are two cases to consider:

1. *P_i is not a cycle.* Let P_i consist of the sequence $(v_0, v_1, \ldots, v_l)$ of vertices, with v_0 the start vertex. If $(v_0, v_1) \in Y$, perform a sequence of $\leftrightarrow$-transitions replacing (v_{2j+1}, v_{2j+2}) by (v_{2j}, v_{2j+1}) for $j = 0, 1, \ldots$, and finish with a single $\uparrow$-transition if l is odd. If on the other hand $(v_0, v_1) \in X$, begin with a $\downarrow$-transition removing (v_0, v_1) and proceed as before for the reduced path $(v_1, \ldots, v_l)$.

[2]In deriving the O-expression, we have assumed w.l.o.g. that $T_r = O\big(|E|n^2\widehat{\lambda}' \ln n\widehat{\lambda}'\big)$. This follows from (3.5) with the additional assumption that $\ln \epsilon^{-1} = O(n \ln n)$. This latter assumption is justified since the problem can always be solved exactly by exhaustive enumeration in time $O(n(2n)!)$, which is $O(\epsilon^{-2})$ if $\ln \epsilon^{-1}$ exceeds the above bound.

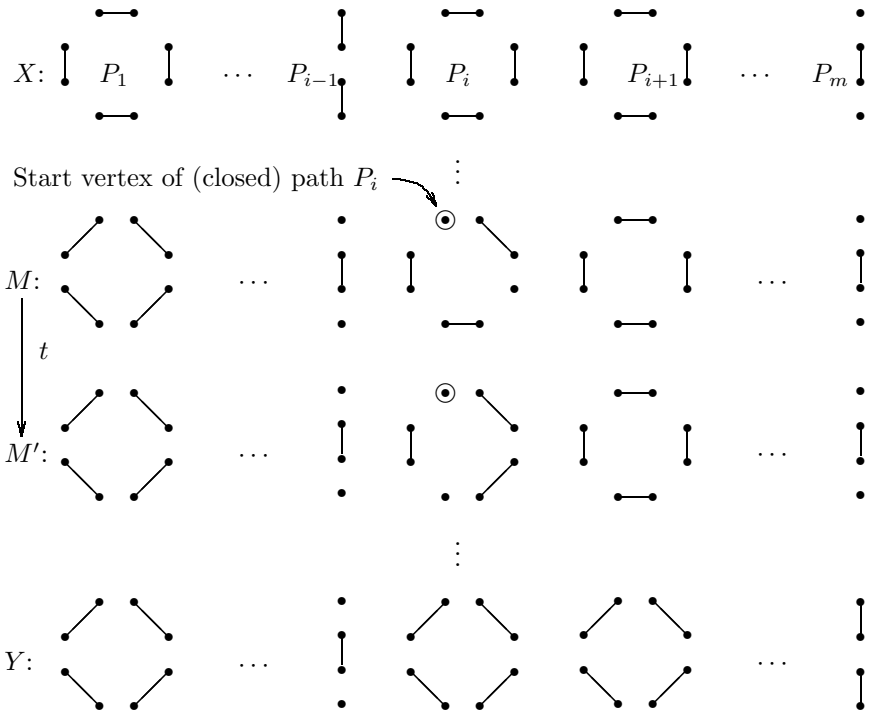

Figure 3.2: A transition t in the canonical path from X to Y

2. *P_i is a cycle.* Let P_i consist of the sequence $(v_0, v_1, \ldots, v_{2l+1})$ of vertices, where $l \geq 1$, v_0 is the start vertex, and $(v_{2j}, v_{2j+1}) \in X$ for $0 \leq j \leq l$, the remaining edges belonging to Y. Then the unwinding begins with a $\downarrow$-transition to remove (v_0, v_1). We are left with an open path O with endpoints v_0, v_1, one of which must be the start vertex of O. Suppose v_k, $k \in \{0, 1\}$, is *not* the start vertex. Then we unwind O as in (i) above but treating v_k as the start vertex. This trick serves to distinguish paths from cycles, as will prove convenient shortly.

This concludes our definition of the family of canonical paths Γ. Figure 3.2 will help the reader picture a typical transition t on a canonical path from X to Y. The path P_i (which happens to be a cycle) is the one currently being unwound; the paths $P_1, \ldots, P_{i-1}$ to the left have already been processed, while the ones $P_{i+1}, \ldots, P_m$ are yet to be dealt with.

We now proceed to bound the "bottleneck" measure $\bar{\rho}(\Gamma)$ for these paths. Let ϵ be an arbitrary edge in the Markov chain, i.e., a transition from M to $M' \neq M$, and let $\mathrm{cp}(\epsilon) = \{(X, Y) : \gamma_{XY} \ni \epsilon\}$ denote the set of all canonical paths that use ϵ. (We use the notation ϵ in place of e here to avoid confusion with edges of G.) We will obtain a bound on the total weight of all paths that pass through ϵ by defining an injective mapping $\eta_\epsilon : \mathrm{cp}(\epsilon) \to \Omega$. What we would like to do is to set $\eta_\epsilon(X, Y) = X \oplus Y \oplus (M \cup M')$; the intuition for this is that $\eta_\epsilon(X, Y)$ should agree with X on paths that have already

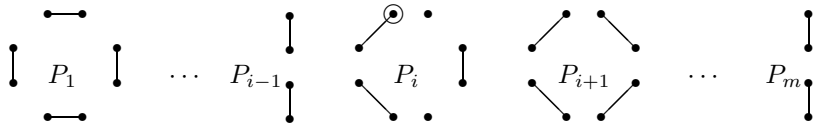

Figure 3.3: The corresponding encoding $\eta_t(X, Y)$

been unwound, and with Y on paths that have not yet been unwound. However, there is a minor complication concerning the path that we are currently processing: in order to ensure that $\eta_\epsilon(X, Y)$ is indeed a matching, we may — as we shall see — have to remove from it the edge of X adjacent to the start vertex of the path currently being unwound: we shall call this edge e_{XYt}. This leads us to the following definition of the mapping η_ϵ:

$$\eta_\epsilon(X,Y) = \begin{cases} X \oplus Y \oplus (M \cup M') - e_{XYt}, & \text{if } \epsilon \text{ is a } \leftrightarrow\text{-transition} \\ & \text{and the current path is a cycle;} \\ X \oplus Y \oplus (M \cup M'), & \text{otherwise.} \end{cases}$$

Figure 3.5 illustrates the encoding $\eta_t(X, Y)$ that would result from the transition t on the canonical path sketched in Figure 3.2.

Let us check that $\eta_\epsilon(X, Y)$ is always a matching. To see this, consider the set of edges $A = X \oplus Y \oplus (M \cup M')$, and suppose that some vertex, u say, has degree two in A. (Since $A \subseteq X \cup Y$, no vertex degree can exceed two.) Then A contains edges $\{u, v_1\}, \{u, v_2\}$ for distinct vertices v_1, v_2, and since $A \subseteq X \cup Y$, one of these edges must belong to X and the other to Y. Hence both edges belong to $X \oplus Y$, which means that neither can belong to $M \cup M'$. Following the form of $M \cup M'$ along the canonical path, however, it is clear that there can be at most one such vertex u; moreover, this happens precisely when the current path is a cycle, u is its start vertex, and ϵ is a $\leftrightarrow$-transition. Our definition of η_ϵ removes one of the edges adjacent to u in this case, so all vertices in $\eta_\epsilon(X, Y)$ have degree at most one, i.e., $\eta_\epsilon(X, Y)$ is indeed a matching.

We now have to check that η_ϵ is injective. It is immediate from the definition of η_ϵ that the symmetric difference $X \oplus Y$ can be recovered from $\eta_\epsilon(X, Y)$ using the relation

$$X \oplus Y = \begin{cases} \eta_\epsilon(X,Y) \oplus (M \cup M') + e_{XYt}, & \text{if } \epsilon \text{ is a } \leftrightarrow\text{-transition} \\ & \text{and the current path is a cycle;} \\ \eta_\epsilon(X,Y) \oplus (M \cup M'), & \text{otherwise.} \end{cases}$$

Note that, once we have formed the set $\eta_\epsilon(X, Y) \oplus (M \cup M')$, it will be apparent whether the current path is a cycle from the sense of unwinding. (Note that e_{XYt} is the unique edge that forms a cycle when added to the path.) Given $X \oplus Y$, we can at once infer the sequence of paths $P_1, P_2, \ldots, P_m$ that have to be unwound along the canonical path from X to Y, and the transition t tells us which of these, P_i say, is the path currently being unwound. The partition of $X \oplus Y$ into X and Y is now straightforward: X has the same parity as $\eta_\epsilon(X, Y)$ on paths $P_1, \ldots, P_{i-1}$, and the same parity as M on

paths $P_{i+1}, \ldots, P_m$. Finally, the reconstruction of X and Y is completed by noting that $X \cap Y = M - (X \oplus Y)$, which is immediate from the definition of the paths. Hence X and Y can be uniquely recovered from $\eta_\epsilon(X, Y)$, so η_ϵ is injective.

We are almost done. What we now require in addition is that η_ϵ be "weight-preserving," in the sense that $Q(\epsilon)\pi_\lambda(\eta_\epsilon(X, Y)) \approx \pi_\lambda(X)\pi_\lambda(Y)$. More precisely, we will show in a moment that

$$\pi_\lambda(X)\pi_\lambda(Y) \leq 2|E|\lambda'^2 Q(\epsilon)\pi_\lambda(\eta_\epsilon(X, Y)). \tag{3.7}$$

First, let us see why we need a bound of this form in order to estimate $\bar{\rho}$. We have

$$\begin{aligned}
\frac{1}{Q(\epsilon)} \sum_{\gamma_{XY} \ni \epsilon} \pi_\lambda(X)\pi_\lambda(Y)|\gamma_{XY}| &\leq 2|E|\lambda'^2 \sum_{\gamma_{XY} \ni \epsilon} \pi_\lambda(\eta_\epsilon(X, Y))\,|\gamma_{XY}| \\
&\leq 4|E|n\lambda'^2 \sum_{\gamma_{XY} \ni \epsilon} \pi_\lambda(\eta_\epsilon(X, Y)) \\
&\leq 4|E|n\lambda'^2\,,
\end{aligned} \tag{3.8}$$

where the second inequality follows from the fact that the length of any canonical path is bounded by $2n$, and the last inequality from the facts that η_ϵ is injective and π_λ is a probability distribution.

It remains for us to prove inequality (3.7). Before we do so, it is helpful to notice that $Q(\epsilon) = (2|E|)^{-1} \min\{\pi_\lambda(M), \pi_\lambda(M')\}$, as may easily be verified from the definition of $\mathcal{M}_{\text{match}}(\lambda)$. We now distinguish four cases:

1. *ϵ is a $\downarrow$-transition.* Suppose $M' = M - e$. Then $\eta_\epsilon(X, Y) = X \oplus Y \oplus M$, so, viewed as multisets, $M \cup \eta_\epsilon(X, Y)$ and $X \cup Y$ are identical. Hence we have

$$\begin{aligned}
\pi_\lambda(X)\pi_\lambda(Y) &= \pi_\lambda(M)\pi_\lambda(\eta_\epsilon(X, Y)) \\
&= \frac{2|E|Q(\epsilon)}{\min\{\pi_\lambda(M), \pi_\lambda(M')\}} \times \pi_\lambda(M)\pi_\lambda(\eta_\epsilon(X, Y)) \\
&= 2|E|Q(\epsilon) \max\{1, \pi_\lambda(M)/\pi_\lambda(M')\}\pi_\lambda(\eta_\epsilon(X, Y)) \\
&\leq 2|E|\lambda' Q(\epsilon)\pi_\lambda(\eta_\epsilon(X, Y)),
\end{aligned}$$

   from which (3.7) follows.

2. *ϵ is a $\uparrow$-transition.* This is handled by a symmetrical argument to (i) above, with the roles of M and M' interchanged.

3. *ϵ is a $\leftrightarrow$-transition and the current path is a cycle.* Suppose $M' = M + e - e'$, and consider the multiset $M \cup \eta_\epsilon(X, Y)$. Then $\eta_\epsilon(X, Y) = X \oplus Y \oplus (M + e) - e_{XYt}$, so the multiset $M \cup \eta_\epsilon(X, Y)$ differs from $X \cup Y$ only in that e and e_{XYt} are missing from it. Thus we have

$$\begin{aligned}
\pi_\lambda(X)\pi_\lambda(Y) &\leq \lambda'^2 \pi_\lambda(M)\pi_\lambda(\eta_\epsilon(X, Y)) \\
&= 2|E|\lambda'^2 Q(\epsilon)\pi_\lambda(\eta_\epsilon(X, Y)),
\end{aligned}$$

since in this case $\pi_\lambda(M) = \pi_\lambda(M')$, and so $Q(\epsilon) = (2|E|)^{-1}\pi_\lambda(M)$. Thus (3.7) is again satisfied.

4. *ϵ is a $\leftrightarrow$-transition and the current path is not a cycle.* This is identical with (iii) above, except that the edge e_{XYt} does not appear in the analysis. Accordingly, the bound is
$$\pi_\lambda(X)\pi_\lambda(Y) \leq 2|E|\lambda' Q(\epsilon)\pi_\lambda(\eta_\epsilon(X,Y)).$$

This concludes our proof of (3.7). We may now deduce from (3.8), that $\bar{\rho}(\Gamma) \leq 4|E|n\lambda'^2$. However, one additional observation will allow us to improve the bound to $\bar{\rho}(\Gamma) \leq 4|E|n\lambda'$, which is what we claimed in (3.6). Looking at the above case analysis we see that, in all cases except case (iii), (3.7), and hence (3.8), actually hold with λ'^2 replaced by λ'. But in case (iii) we can argue that $\eta_\epsilon(X,Y)$ must have such a restricted form that $\sum_{\gamma_{XY} \ni \epsilon} \pi_\lambda(\eta_\epsilon(X,Y))$ is bounded above by λ'^{-1}. Using this fact in the final inequality in (3.8), we get the improved upper bound of $4|E|n\lambda'$ in this case, and hence in all cases. This will complete our verification of the bound (3.6) on $\bar{\rho}(\Gamma)$.

To justify the above claim, note that $\eta_\epsilon(X,Y)$ has at least two unmatched vertices, namely the start vertex of the current cycle and the vertex that is common to both e and e'. Moreover, in $\eta_\epsilon(X,Y) \oplus M$ these vertices are linked by an alternating path that starts and ends with an edge of M. So we may associate with each matching $\eta_\epsilon(X,Y)$ another matching, say $\eta'_\epsilon(X,Y)$, obtained by augmenting $\eta_\epsilon(X,Y)$ along this path. But this operation is uniquely reversible, so all matchings $\eta'_\epsilon(X,Y)$ created in this way are distinct. Moreover, $\pi_\lambda(\eta_\epsilon(X,Y)) = \lambda\pi_\lambda(\eta_\epsilon(X,Y))$. Hence we have $\sum \pi_\lambda(\eta_\epsilon(X,Y)) = \lambda^{-1}\sum \pi_\lambda(\eta'_\epsilon(X,Y)) \leq \lambda^{-1}$, so $\sum \pi_\lambda(\eta_\epsilon(X,Y)) \leq \lambda'^{-1}$ as claimed.

3.2 Perfect Matchings

The question of whether there exists an FPRAS for the permanent of an arbitrary 0,1-matrix has recently been positivley resolved by Jerrum, Sinclair and Vigoda [**?**]. Thus there is an FPRAS for the number of perfect matchings in an arbitrary bipartite graph. Their result does not seem to carry over for arbitrary graphs and so we first concentrate on seeing how to use the methods and results of the previous section to construct an FPRAS that covers many cases, even a majority in some sense. To state the result precisely, we will use the perfect matching formulation. Let $G = (V,E)$ be a graph with $|V| = 2n$. A special role will be played in the result by the number of *near-perfect* matchings in G, i.e., matchings with exactly two unmatched vertices. Following the notation of the previous section, let us write $m_k = m_k(G)$ for the number of k-matchings in G. Then the number of perfect matchings is m_n, and the number of near-perfect matchings is m_{n-1}.

Theorem 3.2.1 *There exists a randomized approximation scheme for the number of perfect matchings m_n whose running time is polynomial in n, ϵ^{-1} and the ratio m_{n-1}/m_n.*

Note that this algorithm is not in general an FPRAS, since there exist $2n$-vertex graphs G for which the ratio m_{n-1}/m_n is exponential in n. However, it turns out that these examples are atypical in the sense that the probability that a randomly selected G on $2n$ vertices violates the inequality $m_{n-1}/m_n \leq 4n$ tends to 0 as $n \to \infty$. Thus the above algorithm constitutes an FPRAS for almost all graphs; moreover, the condition that the ratio m_{n-1}/m_n be bounded by a specified polynomial in n can be tested for an arbitrary graph in polynomial time. It is also known that *every* sufficiently dense graph (specifically, those in which every vertex has degree at least $\frac{1}{2}n$) satisfies $m_{n-1}/m_n = \mathrm{O}(n^2)$. Moreover, it has been shown ratio m_{n-1}/m_n is guaranteed to be small for a wide class of homogeneous graphs G, including the important case of geometric lattice graphs in any number of dimensions.

Our approximation algorithm for the number of perfect matchings follows quite painlessly from our results about the matchings problem derived in the previous section. Note that m_n is precisely the leading coefficient of the partition function $Z_G(\lambda)$ of the monomer-dimer system associated with G (see (3.1)). In the previous section, we saw how to sample matchings in G from the distribution

$$\pi_\lambda(M) = \frac{\lambda^{|M|}}{Z_G(\lambda)} = \frac{\lambda^{|M|}}{\sum_{k=0}^{n} m_k \lambda^k} \tag{3.9}$$

for any desired $\lambda > 0$, in time polynomial in n and $\lambda' = \max\{\lambda, 1\}$, by Monte Carlo simulation of the Markov chain $\mathcal{M}_{\mathrm{match}}(\lambda)$. We also saw how this fact can be used to compute $Z_G(\lambda)$ to good accuracy in time polynomial in n and λ'. Suppose then that we have computed a good estimate $\widehat{Z}_G(\lambda)$ of $Z_G(\lambda)$. Then we can get a good estimator for m_n by sampling matchings from the distribution π_λ and computing the proportion, X, of the sample that are perfect matchings; since $\mathbf{E}X = m_n\lambda^n/Z_G(\lambda)$, our estimator is $Y = X\lambda^{-n}\widehat{Z}_G(\lambda)$.

The sample size required to ensure a good estimate depends on the variance of a single sample, or more precisely on the quantity $(\mathbf{E}X)^{-1}$. Clearly, by making λ large enough, we can make this quantity, and hence the sample size, small: this corresponds to placing very large weight on the perfect matchings, so that their proportion can be estimated well by random sampling. How large does λ have to be? This analysis is eased by a beautiful fact. A sequence $a_1, a_2, \ldots, a_n$ of positive reals is *log-concave* if $a_{k-1}a_{k+1} \leq a_k^2$ for $k = 1, 2, \ldots, n-1$.

Lemma 3.2.1 *The sequence $m_0, m_1, \ldots, m_n$ is log-concave.*

Proof Let $M_k = M_k(G)$ be the set of k-matchings of G. Thus $m_k = |M_k(G)|$. We need to show that $m_{k-1}m_{k+1} \leq m_k^2$ and so we can assume that $m_{k+1} > 0$. Let $A = M_{k+1} \times M_{k-1}$ and $B = M_k \times M_k$.

If M, M' are matchings then we know that $M \oplus M$ consists of paths and cycles. Let a path of $M \oplus M$ be an M-path if it contains more M-edges than M'-edges and an M' path if the reverse is true. For any pair $(M, M') \in A$ the number of M paths exceeds the number of M' paths by exactly two. We partition A into disjoint classes $A_r, r = 1, 2, \ldots, k$ where

$$A_r = \{(M, M') \in A : \ M \oplus M' \text{ contains } r+1 \ M - \text{paths and } r-1 \ M' - \text{paths}\}.$$

Similarly the sets

$$B_r = \{(M, M') \in B : \ M \oplus M' \text{ contains } r \ M - \text{paths and } r \ M' - \text{paths}\}.$$

partition B. The lemma will follow from the fact that $|A_r| \leq |B_r|$ for each $r > 0$.

Let us call a pair $(L, L') \in B_r$ *reachable* from $(M, M') \in A_r$ iff $L \oplus L' = M \oplus M'$ and $L = M \oplus P$ for some M-path P of $M \oplus M'$. Clearly the number of elements of B_r reachable from a given $(M, M') \in A_r$ is $r + 1$. Conversely, any given element of B_r is reachable from precisely r elements of A_r. Hence if $|A_r| > 0$ we have $|B_r|/|A_r| = (r+1)/r > 1$. □

As a consequence, it follows that that $m_k/m_n \leq (m_{n-1}/m_n)^{n-k}$. This means that, if we take $\lambda \geq m_{n-1}/m_n$, we get

$$\mathbf{E}X = \frac{m_n \lambda^n}{Z_G(\lambda)} = \frac{m_n \lambda^n}{\sum_{k=0}^{n} m_k \lambda^k} \geq \frac{1}{n+1}, \tag{3.10}$$

which implies that the sample size required grows only linearly with n. Thus it is enough to take λ about as large as the ratio m_{n-1}/m_n. Of course we do not have a priori knowledge of this ratio and so we run the algorithm with $\lambda = 2, 4, 8, \ldots$ until we find that at least $1/n$ proportion of the matchings produced are perfect. Since the time required to generate a single sample grows linearly with λ (see Proposition 3.1.2), the running time of the overall algorithm is polynomial in n, ϵ^{-1} and the ratio m_{n-1}/m_n, as claimed.

We conclude thsi section by mentioning some extensions. First of all, it is not hard to see, again using the log-concavity property, that the above technique can be extended to approximate the entire sequence (m_k), or equivalently all the coefficients of the monomer-dimer partition function. The running time per coefficient is no worse than for m_n.

A Special Case: Vertex Transitive graphs

In this section we discuss a class of graphs for which we can prove that m_{n-1}/m_n is polynomially bounded. A graph G is *vertex transitive* if for every pair of vertices u, v there is an automorphism g of G such that $g(u) = v$. As an example consider the *discrete torus* $T_{d,L}$ in d dimensions. Here we take $V = \{0, 1, \ldots, L-1\}$ for some $L > 0$ and two vertices x, y are adjacent if there exists an index j such that $x_i = y_i$, $i \neq j$ and $|x_j - y_j| = 1 \bmod L$.

Theorem 3.2.2 *In any vertex transitive graph G, the number of near-perfect matchings in G exceeds the number of perfect matchings by a factor at most n^3.*

Proof Let G be any graph with transitive automorphism group. Denote by $\mathcal{M}$ the set of all perfect matchings in G, by $\mathcal{N}(u,v)$ the set of near-perfect matchings that leave vertices u and v uncovered, and by $\mathcal{N} = \bigcup_{u,v} \mathcal{N}(u,v)$ the set of all near-perfect matchings. Let $\mu = |\mathcal{M}|$, $\nu = \max_{u,v} |\mathcal{N}(u,v)|$, and select vertices u_0 and v_0 satisfying $|\mathcal{N}(u_0,v_0)| = \nu$. We assume, contrary to the statement of the theorem, that $\mu < \nu/2n$, and obtain a contradiction.

Let u, v be any pair of non-adjacent vertices. We show that there exists a vertex u' with $\mathrm{dist}(u',v) < \mathrm{dist}(u,v)$ satisfying $|\mathcal{N}(u',v)| \geq |\mathcal{N}(u,v)| - \nu/2n$. By induction on distance (starting with the base case $|\mathcal{N}(u_0,v_0)| = \nu$) it follows that there exists a pair of adjacent vertices (u,v) satisfying $|\mathcal{N}(u,v)| \geq \nu/2$ and hence that $\mu = |\mathcal{M}| \geq \nu/2$, contradicting our initial assumption.

So suppose u and v are non-adjacent, and let u' be any vertex adjacent to u satisfying $\mathrm{dist}(u',v) < \mathrm{dist}(u,v)$. Let g be any automorphism of G mapping u_0 to u' and let v' be the image of v_0 under g. If $v' = v$ we are done, since then $|\mathcal{N}(u',v)| = |\mathcal{N}(u_0,v_0)| = \nu$. So assume the contrary. Define a mapping

$$f : \mathcal{N}(u,v) \times \mathcal{N}(u',v') \to \mathcal{M} \times \mathcal{N}(v,v') \cup \mathcal{N}(u,v') \times \mathcal{N}(u',v)$$

as follows. Let $N \in \mathcal{N}(u,v)$ be a "red" and $N' \in \mathcal{N}(u',v')$ a "blue" near-perfect matching. $N \oplus N'$ consists of two alternating paths together with a number of cycles. We distinguish two cases.

1. There is a blue-blue path from u to v and red-red path from u' to v', *or* there is a blue-red path from u to v' and a red-blue path from u' to v. In this case add a red edge $\{u,u'\}$ and exchange the colours along the (u',v') or (u',v) path, as appropriate. This operation yields a perfect matching in $\mathcal{M}$, and a near-perfect matching in $\mathcal{N}(v,v')$.

2. There is a blue-red path from u to u' and a blue-red path from v to v'. In this case, exchange the colours along the (u,u') path to yield a blue near-perfect matching in $\mathcal{N}(u,v')$ and a red near-perfect matching in $\mathcal{N}(u',v)$.

The operations described above are reversible, so the mapping f is injective. Thus

$$|\mathcal{N}(u,v)| \times |\mathcal{N}(u',v')| \leq |\mathcal{M}| \times |\mathcal{N}(v,v')| + |\mathcal{N}(u,v')| \times |\mathcal{N}(u',v)|.$$

Since $|\mathcal{N}(u',v')| = |\mathcal{N}(u_0,v_0)| = \nu$ and $|\mathcal{M}| = \mu < \nu/2n$ by assumption, the required inequality $|\mathcal{N}(u',v)| \geq |\mathcal{N}(u,v)| - \nu/2n$ follows. □

3.3 The Permanent

Let $G = (V_1, V_2, E)$ be a bipartite graph on $n + n$ vertices. Let $\mathcal{M}_k = \mathcal{M}_k(G)$ be the set of k-matchings of G. There is a rapidly mixing Markov chain on state space $\Omega = \mathcal{M}_{n-1} \cup \mathcal{M}_n$ similar to that described in Section 3.1 that has a uniform steady state distribution. Of course if m_{n-1}/m_n is too large then it will tend to only generate near-perfect matchings within a reasonable time limit. The idea from [**?**] is to modify this chain so that the steady state is still uniform over perfect matchings and the near-perfect matchings have sufficiently smaller weight.

Theorem 3.3.1 *There exists a fully-polynomial randomized approximation scheme for the permanent of an arbitrary $n \times n$ 0-1 matrix A.*

It will actually prove technically convenient to introduce edge weights also. Thus for each edge $(y, z) \in E$, we introduce a positive weight $\lambda(y, z)$, which we call its *activity*. We extend the notion of activities to matchings M (of any cardinality) by $\lambda(M) = \prod_{(i,j)\in M} \lambda(i, j)$. Similarly, for a set of matchings $\mathcal{S}$ we define $\lambda(\mathcal{S}) = \sum_{M\in\mathcal{S}} \lambda(M)$. For our purposes, the advantage of edge weights is that they allow us to work with the complete graph $K_{n,n}$ on $n + n$ vertices, rather than with an arbitrary graph $G = (V_1, V_2, E)$: we can do this by setting $\lambda(e) = 1$ for $e \in E$, and $\lambda(e) \leq 1/n!$ for $e \notin E$. This ensures that the "bogus" matchings have little effect, as will be described shortly.

Let $\mathcal{M}$ denote the set of perfect matchings of $K_{n,n}$ and for $u \in V_1$ and $v \in V_2$ we let $\mathcal{M}(u, v)$ denote the set of near perfect matchings of $K_{n,n}$ that leave only u, v isolated. We are now ready to specify the desired stationary distribution of our Markov chain. This will be the distribution π over Ω defined by $\pi(M) \propto \Lambda(M)$, where

$$\Lambda(M) = \begin{cases} \lambda(M)w(u, v) & \text{if } M \in \mathcal{M}(u, v) \text{ for some } u, v; \\ \lambda(M) & \text{if } M \in \mathcal{M}, \end{cases}$$

and $w : V_1 \times V_2 \to \mathbb{R}^+$ is the weight function for *holes* to be specified shortly.

To construct a Markov chain having π as its stationary distribution, we use the original chain of [Bro86, JS89] augmented with a Metropolis acceptance rule for the transitions. Thus transitions from a matching M are defined as follows:

1. Choose an edge $e = (u, v)$ uniformly at random.
2. (i) If $M \in \mathcal{M}$ and $e \in M$, let $M' = M \setminus \{e\} \in \mathcal{M}(u, v)$;
   (ii) if $M \in \mathcal{M}(u, v)$, let $M' = M \cup \{e\} \in \mathcal{M}$;
   (iii) if $M \in \mathcal{M}(u, z)$ where $z \neq v$ and $(y, v) \in M$, let $M' = M \cup \{e\} \setminus \{(y, v)\} \in \mathcal{M}(y, z)$;

(iv) if $M \in \mathcal{M}(y, v)$ where $y \neq u$ and $(u, z) \in M$, let $M' = M \cup \{e\} \setminus \{(u, z)\} \in \mathcal{M}(y, z)$.

3. With probability $\min\{1, \Lambda(M')/\Lambda(M)\}$ go to M'; otherwise, stay at M.

The Metropolis rule in the final step ensures that this Markov chain is reversible with $\pi(M) \propto \Lambda(M)$ as its stationary distribution. Finally, to make the chain lazy we add a self-loop probability of $1/2$ to every state; i.e., on every step, with probability $1/2$ we make a transition as above and otherwise do nothing.

Next we need to specify the weight function w. Ideally we would like to take $w = w^*$, where

$$w^*(u, v) = \frac{\lambda(\mathcal{M})}{\lambda(\mathcal{M}(u, v))} \tag{3.11}$$

for each pair of holes u, v.

We will not be able to determine w^* exactly but will content ourselves with weights w satisfying

$$w^*(y, z)/2 \leq w(y, z) \leq 2w^*(y, z), \tag{3.12}$$

with very high probability.

The main technical result of this paper is the following theorem, which says that, provided the weight function w satisfies condition (3.12), the Markov chain is rapidly mixing. We present the theorem as it applies to an arbitrary bipartite graph, hence let $m = |E|$. Since we are working with $K_{n,n}$, for our purposes $m = n^2$.

Theorem 3.3.2 *Assuming the weight function w satisfies inequality (3.12) for all $(y, z) \in V_1 \times V_2$, then the mixing time of the Markov chain MC is bounded above by $\tau(\delta) = O(m^6 n^8 (n \log n + \log \delta^{-1}))$, provided the initial state is a perfect matching of maximum activity.*

Initially we have to take $\lambda(e) = 1$ for all $e \in V_1 \times V_2$ and $w(u, v) = n$ for all $u \in V_1, v \in V_2$. This a natural way of starting with parameters for which (3.12) holds. By a sequence of iterations to be described we are able to maintain (3.12) and at the same time reduce $\lambda(e)$ to at most $1/n!$ for all $e \notin E$.

At this point we see that if $\mathcal{M}'$ denotes the perfect matchings of G and $\mathcal{M}''$ denotes the perfect matchings of $K_{n,n}$ which are not in G then since $\Lambda(\mathcal{M}') = |\mathcal{M}'|$, $\Lambda(\mathcal{M}'') \leq 1$ and $\Lambda(\mathcal{M}(u, v)) \leq 2\Lambda(\mathcal{M})$ for all u, v, we see that

$$\frac{\Lambda(\mathcal{M}')}{\Lambda(\Omega)} \geq \frac{|\mathcal{M}'| - 1}{2(n^2 + 1)}. \tag{3.13}$$

It follows from this and Theorem 3.3.2 that we can in polynomial time generate a near uniform perfect matching of G. We repeat this process a number of times until we find

an edge $e_1 = (u_1, v_1)$ whiich is in at least a fraction $1/(2n^2)$ of the matchings in $\mathcal{M}'$. We then estimate this proportion to within accuracy $\epsilon/(2n)$ with probability at least δ/n. Let us call our estimate ρ_1. We then apply the same strategy to $G - \{u_1, v_1\}$ and so on to obtain estimated proportions $\rho_1, \rho_2, \ldots, \rho_n$. Our final estimate of the number of perfect matchings in G is then $\rho_1^{-1}\rho_2^{-1}\cdots\rho_n^{-1}$.

Now let us see how to go about reducing $\lambda(e)$ for $e \notin E$. Assuming that (3.12) holds we get sufficient samples from our chain so that we can estimate all of the $w^*(y, z)$ to within a factor $4/3$ say. Let these estimates be denoted by $w'(y, z)$. If now there is an edge $e \notin E$ such that $\lambda(e) > 1/n!$ then we replace $\lambda(e)$ by $3\lambda(e)/4$. The effect of this is to change any value of w^* by at most $4/3$. If we replace our old w values by the corresponding w' values then (3.12) will still hold, since now $w'(y, z)/w^*(y, z) \in [9/16, 16/9]$ for all y, z and we repeat our attempts at reducing the λ's where necessary. Thus to prove Theorem 3.3.1, it is sufficient to prove Theorem 3.3.2.

3.3.1 Proof of Theorem 3.3.2

Theorem 3.3.3 *For an ergodic, reversible Markov chain with self-loop probabilities $P(y, y) \geq 1/2$ for all states y, and any initial state $x \in \Omega$,*

$$\tau_x(\delta) \leq \frac{2}{\Phi^2}\left(\ln \pi(x)^{-1} + \ln \delta^{-1}\right).$$

We bound the conductance by defining canonical paths $\gamma_{I,F}$ from all $I \in \Omega$ to all $F \in \mathcal{M}$. By upper bounding the maximum number of paths through any particular transition we will obtain a lower bound on the conductance. Using the fact that perfect matchings are likely under the stationary distribution, it will be sufficient to only consider a portion of particular canonical paths. Denote the set of all canonical paths by $\Gamma = \{\gamma_{I,F} : (I, F) \in \Omega \times \mathcal{M}\}$. Certain transitions on a canonical path will be deemed *chargeable*. For each transition t denote by

$$\mathrm{cp}(t) = \{(I, F) : \gamma_{I,F} \text{ contains } t \text{ as a chargeable transition}\}.$$

The canonical paths are defined by superimposing I and F. If $I \in \mathcal{M}$, then $I \oplus F$ consists of a collection of alternating cycles. We assume that the cycles are ordered in some canonical fashion; for example, having ordered the vertices, we may take as the first cycle the one that contains the least vertex in the order, as the second cycle the one that contains the least vertex amongst those remaining, and so on. Furthermore we assume that each cycle has a distinguished start vertex (e.g., the least in the order).

The canonical path $\gamma_{I,F}$ from $I \in \mathcal{M}$ to F is obtained by unwinding these cycles in the canonical order. A cycle $v_0 \sim v_1 \sim \ldots \sim v_{2k} = v_0$, where we assume w.l.o.g. that the edge (v_0, v_1) belongs to I, is unwound by: (i) removing the edge (v_0, v_1), (ii) successively,

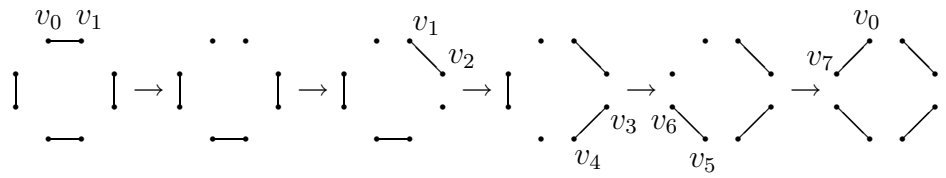

Figure 3.4: Unwinding a cycle with $k = 4$.

for each $1 \le i \le k-1$, exchanging the edge (v_{2i}, v_{2i+1}) with (v_{2i-1}, v_{2i}), and (iii) adding the edge (v_{2k-1}, v_{2k}). (Refer to figure 3.4.) All transitions on the path $\gamma_{I,F}$ are deemed chargeable. A canonical path joining two perfect matchings, as just described, will be termed "type A."

If $I \in \mathcal{M}(y,z)$ for some $(y,z) \in V_1 \times V_2$, then $I \oplus F$ consists of a collection of alternating cycles together with a single alternating path from y to z. The canonical path $\gamma_{I,F}$ from I to F is obtained by unwinding the path and then unwinding the cycles in some canonical order. In this case, only the transitions involved in the unwinding of the path are deemed chargeable. The alternating path $y = v_0 \sim \ldots \sim v_{2k+1} = z$ is unwound by: (i) successively, for each $1 \le i \le k$, exchanging the edge (v_{2i-1}, v_{2i}) with (v_{2i-2}, v_{2i-1}), and (ii) adding the edge (v_{2k}, v_{2k+1}). A canonical path joining a near-perfect to a perfect matching will be termed "type B."

We define a notion of congestion of Γ that accounts only for the chargeable transitions:

$$\varrho(\Gamma) := \max_{t \in T} \left\{ \frac{1}{Q(t)} \sum_{(I,F) \in \mathrm{cp}(t)} \pi(I)\pi(F) \right\}. \tag{3.14}$$

Our main task will be to derive an upper bound on $\varrho(\Gamma)$, which we state in the next lemma. From this, it will be a straightforward matter to obtain a lower bound on the conductance Φ (see Lemma 3.3.2 below) and hence, via Theorem 3.3.3, a bound on the mixing time. In the following lemma recall that $m = |E|$, where for our purposes $m = n^2$.

Lemma 3.3.1 *Assuming the weight function w satisfies inequality (3.12) for all $(y,z) \in V_1 \times V_2$, then $\varrho(\Gamma) \le 16m$.*

In preparation for proving Lemma 3.3.1, we establish some combinatorial inequalities concerning weighted near-perfect matchings that will be used in the proof.

Lemma 3.3.2 *Let G be as above, and let $u, y \in V_1$, $v, z \in V_2$.*

1. *$\lambda(u,v)\lambda(\mathcal{M}(u,v)) \le \lambda(\mathcal{M})$, for all vertices u, v with $u \sim v$;*

2. $\lambda(u,v)\lambda(\mathcal{M}(u,z))\lambda(\mathcal{M}(y,v)) \leq \lambda(\mathcal{M})\lambda(\mathcal{M}(y,z))$, *for all distinct vertices* u,v,y,z *with* $u \sim v$.

Proof The mapping from $\mathcal{M}(u,v)$ to $\mathcal{M}$ defined by $M \mapsto M \cup \{(u,v)\}$ is injective, and preserves activities modulo a factor $\lambda(u,v)$; this dispenses with (i). For (ii), suppose $M_{u,z} \in \mathcal{M}(u,z)$ and $M_{y,v} \in \mathcal{M}(y,v)$, and consider the superposition of $M_{u,z}$, $M_{y,v}$ and the single edge (u,v). Observe that $M_{u,z} \oplus M_{y,v} \oplus \{(u,v)\}$ decomposes into a collection of cycles together with an odd-length path O joining y and z.[3] Let $O = \{e_0, e_1, \ldots, e_{2k}\}$ be an enumeration of the edges of this path, starting at y and working towards z. Denote by O_0 the $k+1$ even edges, and by O_1 the k odd edges. Finally define a mapping from $\mathcal{M}(u,z) \times \mathcal{M}(y,v)$ to $\mathcal{M} \times \mathcal{M}(y,z)$ by $(M_{u,z}, M_{y,v}) \mapsto (M, M_{y,z})$, where $M := M_{u,z} \cup O_0 \setminus O_1$ and $M_{y,z} := M_{y,v} \cup O_1 \setminus O_0$. Note that this mapping is injective, since we may uniquely recover $(M_{u,z}, M_{y,v})$ from $(M, M_{y,z})$. (To see this, observe that $M \oplus M_{y,z}$ decomposes into a number of cycles, together with a single odd-length path joining y and z. This path is exactly the path O considered in the forward map. There is only one way to apportion edges from $O \setminus \{(u,v)\}$ between $M_{u,z}$ and $M_{y,v}$.) Moreover, the mapping preserves activities modulo a factor $\lambda(u,v)$. □

Corollary 3.3.1 *Let* G *be as above, and let* $u,y \in V_1$, $v,z \in V_2$. *Then, provided in each case that the left hand side of the inequality is defined,*

1. $w^*(u,v) \geq \lambda(u,v)$, *for all vertices* u,v *with* $u \sim v$;
2. $w^*(u,z)w^*(y,v) \geq \lambda(u,v)w^*(y,z)$, *for all distinct vertices* u,v,y,z *with* $u \sim v$;
3. $w^*(u,z)w^*(y,v) \geq \lambda(u,v)\lambda(y,z)$, *for all distinct vertices* u,v,y,z *with* $u \sim v$ *and* $y \sim z$.

Proof Inequalities (i) and (ii) follow from the correspondingly labelled inequalities in Lemma 3.3.2, and the definition of w^*. Inequality (iii) is implied by (i) and (ii). □

Armed with Corollary 3.3.1, we can now turn to the proof of our main lemma.

Proof [Proof of Lemma 3.3.1] Note from the Metropolis rule that for any pair of states M, M' such that the probability of transition from M to M' is non-zero, we have $Q(M,M') = \min\{\pi(M), \pi(M')\}/2m$. We will show that for any transition $t = (M,M')$ and any pair of states $I, F \in \mathrm{cp}(t)$, we can define an encoding $\eta_t(I,F) \in \Omega$ such that $\eta_t : \mathrm{cp}(t) \to \Omega$ is an injection (i.e., (I,F) can be recovered uniquely from $\eta_t(I,F)$), and

$$\pi(I)\pi(F) \leq 8\min\{\pi(M), \pi(M')\}\pi(\eta_t(I,F)) = 16m\,Q(t)\pi(\eta_t(I,F)). \tag{3.15}$$

[3] It is at this point that we rely crucially on the bipartiteness of G. If G is non-bipartite, we may end up with an even-length path and an odd-length cycle, and the proof cannot proceed.

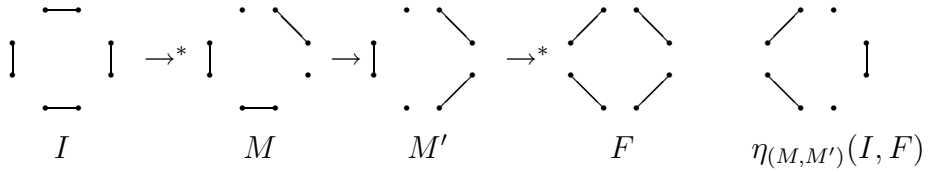

Figure 3.5: A canonical path through transition $M \to M'$ and its encoding.

Summing inequality (3.15) over $(I, F) \in \mathrm{cp}(t)$, we get

$$\frac{1}{Q(t)} \sum_{(I,F)\in \mathrm{cp}(t)} \pi(I)\pi(F) \le 16m \sum_{(I,F)\in \mathrm{cp}(t)} \pi(\eta_t(I,F)) \le 16m,$$

where we have used the fact that η_t is an injection. This immediately yields the claimed bound on $\varrho(\Gamma)$.

We now proceed to define the encoding η_t and show that it has the above properties. For a transition $t = (M, M')$ which is involved in stage (ii) of unwinding a cycle, the encoding is

$$\eta_t(I, F) = I \oplus F \oplus (M \cup M') \setminus \{(v_0, v_1)\}.$$

(Refer to figure 3.5, where just a single alternating cycle is viewed in isolation.) Otherwise, the encoding is

$$\eta_t(I, F) = I \oplus F \oplus (M \cup M').$$

It is not hard to check that $C = \eta_t(I, F)$ is always a matching in Ω (this is the reason that the edge (v_0, v_1) is removed in the first case above), and that η_t is an injection. To see this for the first case, note that $I \oplus F$ may be recovered from the identity $I \oplus F = (C \cup \{(v_0, v_1)\}) \oplus (M \cup M')$; the apportioning of edges between I and F can then be deduced from the canonical ordering of the cycles and the position of the transition t. The remaining edges, namely those in the intersection $I \cap F$, are determined by $I \cap F = M \cap M' \cap C$. The second case is similar, but without the need to reinstate the edge (v_0, v_1).

It therefore remains only to verify inequality (3.15) for our encoding η_t.

Consider first the case where $I \in \mathcal{M}$ and $t = (M, M')$ is the initial transition in the unwinding of an alternating cycle in a type A canonical path, where $M = M' \cup \{(v_0, v_1)\}$. Since $I, F, C, M \in \mathcal{M}$ and $M' \in \mathcal{M}(v_0, v_1)$, inequality (3.15) simplifies to

$$\lambda(I)\lambda(F) \le 8 \min\{\lambda(M), \lambda(M')w(v_0, v_1)\}\lambda(C).$$

The inequality in this form can be seen to follow from the identity

$$\lambda(I)\lambda(F) = \lambda(M)\lambda(C) = \lambda(M')\lambda(v_0, v_1)\lambda(C),$$

using of inequality (i) of Corollary 3.3.1, and inequality (3.12). The situation is symmetric for the final transition in unwinding an alternating cycle.

Staying with the type A path, i.e., with the case $I \in \mathcal{M}$, suppose the transition $t = (M, M')$ is traversed in stage (ii) of unwinding an alternating cycle, i.e., exchanging edge (v_{2i}, v_{2i+1}) with (v_{2i-1}, v_{2i}). In this case we have $I, F \in \mathcal{M}$ while $M \in \mathcal{M}(v_0, v_{2i-1}), M' \in \mathcal{M}(v_0, v_{2i+1})$ and $C \in \mathcal{M}(v_{2i}, v_1)$. Since

$$\begin{aligned}\lambda(I)\lambda(F) &= \lambda(M)\lambda(C)\lambda(v_{2i}, v_{2i-1})\lambda(v_0, v_1) \\ &= \lambda(M')\lambda(C)\lambda(v_{2i}, v_{2i+1})\lambda(v_0, v_1),\end{aligned}$$

inequality (3.15) simplifies to

$$1 \leq 8 \min\left\{\frac{w(v_0, v_{2i-1})}{\lambda(v_{2i}, v_{2i-1})}, \frac{w(v_0, v_{2i+1})}{\lambda(v_{2i}, v_{2i+1})}\right\} \frac{w(v_{2i}, v_1)}{\lambda(v_0, v_1)}.$$

This inequality can be verified by reference to Corollary 3.3.1: specifically, it follows from inequality (iii) in the general case $i \neq 1$, and by two applications of inequality (i) in the special case $i = 1$.

We now turn to the type B canonical paths. Suppose $I \in \mathcal{M}(y, z)$, and consider a transition $t = (M, M')$ from stage (i) of the unwinding of an alternating path, i.e., exchanging edge (v_{2i}, v_{2i-1}) with (v_{2i-2}, v_{2i-1}). Observe that $F \in \mathcal{M}, M \in \mathcal{M}(v_{2i-2}, z), M' \in \mathcal{M}(v_{2i}, z)$ and $C \in \mathcal{M}(y, v_{2i-1})$. Moreover, $\lambda(I)\lambda(F) = \lambda(M)\lambda(C)\lambda(v_{2i-2}, v_{2i-1}) = \lambda(M')\lambda(C)\lambda(v_{2i}, v_{2i-1})$. In inequality (3.15) we are left with

$$w(y, z) \leq 8 \min\left\{\frac{w(v_{2i-2}, z)}{\lambda(v_{2i-2}, v_{2i-1})}, \frac{w(v_{2i}, z)}{\lambda(v_{2i}, v_{2i-1})}\right\} w(y, v_{2i-1}),$$

which holds by inequality (ii) of Corollary 3.3.1. Note that the factor $8 = 2^3$ is determined by this case, since we need to apply inequality (3.12) three times.

The final case is the last transition $t = (M, M')$ in unwinding an alternating path, where $M' = M \cup (z', z)$. Note that $I, C \in \mathcal{M}(y, z)$, $F, M' \in \mathcal{M}, M \in \mathcal{M}(z', z)$ and $\lambda(I)\lambda(F) = \lambda(M')\lambda(C) = \lambda(M)\lambda(z', z)\lambda(C)$. (Here we have written z' for v_{2k}.) Plugging these into inequality (3.15) leaves us with

$$1 \leq 8 \min\left\{\frac{w(z', z)}{\lambda(z', z)}, 1\right\},$$

which follows from inequality (i) of Corollary 3.3.1.

We have thus shown that the encoding η_t satisfies inequality (3.15) in all cases. This completes the proof of the lemma. □

Based on the upper bound on congestion, we can derive the following lower bound on the conductance.

Corollary 3.3.2 *Assuming the weight function w satisfies inequality (3.12) for all $(y,z) \in V_1 \times V_2$, then $\Phi \geq 1/100\varrho^3 n^4 \geq 1/10^6 m^3 n^4$.*

Proof Set $\alpha = 1/10\varrho n^2$. Let $\mathcal{S}, \overline{\mathcal{S}}$ be a partition of the state-space. (Note that we do not assume that $\pi(\mathcal{S}) \leq \pi(\overline{\mathcal{S}})$.) We distinguish two cases, depending on whether or not the perfect matchings $\mathcal{M}$ are fairly evenly distributed between $\mathcal{S}$ and $\overline{\mathcal{S}}$. If the distribution is fairly even, then we can show $\Phi(\mathcal{S})$ is large by considering type A canonical paths, and otherwise by using the type B paths.

Case I. $\pi(\mathcal{S} \cap \mathcal{M})/\pi(\mathcal{S}) \geq \alpha$ and $\pi(\overline{\mathcal{S}} \cap \mathcal{M})/\pi(\overline{\mathcal{S}}) \geq \alpha$. Just looking at canonical paths of type A we have a total flow of $\pi(\mathcal{S} \cap \mathcal{M})\pi(\overline{\mathcal{S}} \cap \mathcal{M}) \geq \alpha^2 \pi(\mathcal{S})\pi(\overline{\mathcal{S}})$ across the cut. Thus $\varrho\, Q(\mathcal{S}, \overline{\mathcal{S}}) \geq \alpha^2 \pi(\mathcal{S})\pi(\overline{\mathcal{S}})$, and $\Phi(\mathcal{S}) \geq \alpha^2/\varrho = 1/100\varrho^3 n^4$.

Case II. Otherwise (say) $\pi(\mathcal{M} \cap \mathcal{S})/\pi(\mathcal{S}) < \alpha$. Note the following estimates:

$$\begin{aligned} \pi(\mathcal{M}) &\geq \frac{1}{4n^2+1} \geq \frac{1}{5n^2}; \\ \pi(\mathcal{S} \cap \mathcal{M}) &< \alpha\pi(\mathcal{S}) < \alpha; \\ \pi(\mathcal{S} \setminus \mathcal{M}) &= \pi(\mathcal{S}) - \pi(\mathcal{S} \cap \mathcal{M}) > (1-\alpha)\pi(\mathcal{S}).\alpha \geq 1/7n^2. \end{aligned}$$

Consider the cut $\mathcal{S} \setminus \mathcal{M} : \overline{\mathcal{S}} \cup \mathcal{M}$. The weight of canonical paths (all chargeable as they cross the cut) is $\pi(\mathcal{S} \setminus \mathcal{M})\pi(\mathcal{M}) \geq (1-\alpha)\pi(\mathcal{S})/5n^2 \geq \pi(\mathcal{S})/6n^2$. Hence $\varrho\, Q(\mathcal{S} \setminus \mathcal{M}, \overline{\mathcal{S}} \cup M) \geq \pi(\mathcal{S})/6n^2$. Noting $Q(\mathcal{S} \setminus \mathcal{M}, \mathcal{S} \cap M) \leq \pi(\mathcal{S} \cap M) \leq \alpha\pi(\mathcal{S})$ we have

$$\begin{aligned} Q(\mathcal{S}, \overline{\mathcal{S}}) &\geq Q(\mathcal{S} \setminus \mathcal{M}, \overline{\mathcal{S}}) \\ &= Q(\mathcal{S} \setminus \mathcal{M}, \overline{\mathcal{S}} \cup M) - Q(\mathcal{S} \setminus \mathcal{M}, \mathcal{S} \cap M) \\ &\geq (1/6\varrho n^2 - \alpha)\pi(\mathcal{S}) \\ &\geq \pi(\mathcal{S})/15\varrho n^2 \\ &\geq \pi(\mathcal{S})\pi(\overline{\mathcal{S}})/15\varrho n^2. \end{aligned}$$

Rearranging, $\Phi(\mathcal{S}) = Q(\mathcal{S}, \overline{\mathcal{S}})/\pi(\mathcal{S})\pi(\overline{\mathcal{S}}) \geq 1/15\varrho n^2$.

Clearly, it is Case I that dominates, giving the claimed bound on Φ. □

Theorem 3.3.2 of the previous section, now follows immediately.

Proof [Proof of Theorem 3.3.2] The condition on the starting state ensures $\log(\pi(X_0)^{-1}) = O(n \log n)$, where X_0 in the initial state. The lemma now follows from (2.1) and Theorem 2.2.1. □

initialize $\lambda(e) \leftarrow a_{\max}$ for all $e \in V_1 \times V_2$
initialize $w(u,v) \leftarrow na_{\max}$ for all $(u,v) \in V_1 \times V_2$
while $\exists e$ with $\lambda(e) > a(e)$ do
    take S samples from MC with parameters λ, w,
        each after a simulation of T steps
    use the sample to obtain estimates $w'(u,v)$ satisfying
        $3w^*(u,v)/4 \leq w(u,v) \leq 4w^*(u,v)/3$ with high probability $\forall u,v$
    set $\lambda(e) \leftarrow \max\{3\lambda(e)/4, a(e)\}$ and $w(u,v) \leftarrow w'(u,v)$ $\forall u,v$
output final weights $w(u,v)$

Figure 3.6: The algorithm for non-negative entries.

3.3.2 Arbitrary weights

Our algorithm easily extends to compute the permanent of a matrix A with non-negative entries. Let $a_{\max} = \max_{i,j} a(i,j)$ and $a_{\min} = \min_{i,j} a(i,j)$. Assuming $\mathrm{per}(A) > 0$, then it is clear that $\mathrm{per}(A) \geq (a_{\min})^n$. Rounding zero entries $a(i,j)$ to $(a_{\min})^n/n!$, the algorithm follows as described in figure 3.6.

The running time of this algorithm is polynomial in n and $\log(a_{\max}/a_{\min})$. For completeness, we provide a strongly polynomial time algorithm, i.e., one whose running time is polynomial in n and independent of $a_{\max}$ and $a_{\min}$, assuming arithmetic operations are treated as unit cost. The algorithm of Linial, Samorodnitsky and Wigderson [**?**] converts, in strongly polynomial time, the original matrix A into a nearly doubly stochastic matrix B such that $1 \geq \mathrm{per}(B) \geq \exp(-n - o(n))$ and $\mathrm{per}(B) = \alpha\,\mathrm{per}(A)$ where α is an easily computable function. Thus it suffices to consider the computation of $\mathrm{per}(B)$. In which case we can afford to round up any entries smaller than say n^{-2n} to n^{-2n}. The analysis for the 0,1-case now applies with the same running time.

3.3.3 Problems reducible to matchings

There are a number of problems reducible to counting and generating matchings. Here we discuss counting graphs with a fixed degree sequence, counting Hamilton cycles in *dense* graphs and counting Euler orientations.

Counting Graphs with a Fixed Degree Sequence

Let $\mathcal{G}(\mathbf{d})$ denote the set of all labelled graphs with vertex set $\{1, 2, \ldots, n\}$ and degree sequence $\mathbf{d} = (d_1, \ldots, d_n)$. It costs us very little in terms of complexity of discussion if we extend the discussion to allow the exclusion of a set of edges X. Thus let $\mathcal{G}(\mathbf{d}, X)$

denote the set of all graphs $G \in \mathcal{G}(\mathbf{d})$ for which the edge set of G is disjoint from X and so $\mathcal{G}(\mathbf{d}) = \mathcal{G}(\mathbf{d}, \emptyset)$. Our goal is to provide a fully polynomial almost uniform generator for $\mathcal{G}(\mathbf{d}, X)$, subject to appropriate conditions on $\mathbf{d}$ and X.

Define $\mathcal{G}'(\mathbf{d}, X)$ to be $\bigcup_{\mathbf{d}'} \mathcal{G}(\mathbf{d}', X)$, where the union ranges over vectors $\mathbf{d}' \in \mathbf{N}^n$ which satisfy $\mathbf{d}' \leq \mathbf{d}$ and $\sum_{i=1}^{n} |d_i - d_i'| \leq 2$. Call a class of degree sequences/excluded pairs *P-stable* if there exists a polynomial p such that $|\mathcal{G}'(\mathbf{d}, X)|/|\mathcal{G}(\mathbf{d}, X)| \leq p(n)$ for every sequence $\mathbf{d} = (d_1, \ldots, d_n)$ in the class. Informally, a degree sequence/excluded pairs $\mathbf{d}, X$ is P-stable if $|\mathcal{G}(\mathbf{d}, X)|$ does not change radically when $\mathbf{d}$ is slightly perturbed. Although the class of *all* graphical degree sequences (with $X = \emptyset$) is not P-stable, there are natural subclasses which are. We shall return to this issue in the next section.

Our aim is to construct a fully polynomial almost uniform generator for $\mathcal{G}(\mathbf{d}, X)$, which is valid for all sequences $\mathbf{d}$ within a specified P-stable class.

Theorem 3.3.4 *There is a good sampler and an FPRAS for $\mathcal{G}(\mathbf{d}, X)$ provided the pair $\mathbf{d}, X$ are drawn from some P-stable class.*

Proof For given degree sequence $\mathbf{d} = (d_1, \ldots, d_n)$ and excluded set X, let $\Gamma = \Gamma(\mathbf{d}, X)$ be the undirected graph with vertex set

$$V(\Gamma) = \{v_{ik} : 1 \leq i \leq n \text{ and } 1 \leq k \leq d_i\} \cup \{u_{ij} : 1 \leq i, j \leq n,\ i \neq j \text{ and } (i,j) \notin X\}$$

and edge set

$$\begin{aligned} E(\Gamma) = &\{(v_{ik}, u_{ij}) : 1 \leq i, j \leq n,\ 1 \leq k \leq d_i,\ i \neq j \text{ and } (i,j) \notin X\} \\ &\cup \{(u_{ij}, u_{ji}) : 1 \leq i, j \leq n,\ i \neq j \text{ and } (i,j) \notin X\}. \end{aligned}$$

The intention is to set up a correspondence between perfect matchings M in Γ and elements of $\mathcal{G}(\mathbf{d}, X)$. Informally, Γ contains an edge (u_{ij}, u_{ji}) corresponding to each potential edge (i,j) in a graph $G \in \mathcal{G}(\mathbf{d}, X)$; the *presence* of the edge (u_{ij}, u_{ji}) in M models the *absence* of the edge (i,j) in G. Additionally there are n *clusters* of vertices of the form $\{v_{ik} : 1 \leq k \leq d_i\}$ which, together with their incident edges, enforce the degree contraints at each vertex i in G.

Let ϕ be the function, from matchings in Γ to (undirected) graphs on vertex set $\{1, \ldots, n\}$, which maps the matching $M \subseteq E(\Gamma)$ to the graph with edge set

$$\{(i,j) : i \neq j \text{ and } (u_{ij}, u_{ji}) \notin M\}.$$

It is a straightforward task to verify that $\phi(\mathcal{M}) = \mathcal{G}(\mathbf{d}, X)$ and, moreover, that each graph in $\mathcal{G}(\mathbf{d}, X)$ is the image of precisely $\prod_{i=1}^{n} d_i!$ elements of $\mathcal{M}$. In particular, when $(u_{ij}, u_{ji}) \notin M$ there exist edges $(v_{is}, u_{ij}), (v_{jt}, u_{ji}) \in M$ and this enforces the degree constraints.

Thus, to generate elements of $\mathcal{G}(\mathbf{d}, X)$ almost uniformly, it is enough to generate perfect matchings in $\Gamma(\mathbf{d})$ almost uniformly. By Theorem 3.2.1, this will be possible provided

$|\mathcal{N}(\Gamma(\mathbf{d}))|/|\mathcal{M}(\Gamma(\mathbf{d}))| \leq q(m)$, where $m = |V(\Gamma)|$ and q is some fixed polynomial. (The polynomial q will depend on the polynomial p in the definition of P-stable.)

Call a matching M of Γ *normalised* iff either (i) M is a perfect matching, or (ii) M is a near-perfect matching whose unmatched vertices are both cluster vertices. Denote the set of all normalised matchings of Γ by $\mathcal{N}'(\Gamma)$. Let $M \in \mathcal{N}(\Gamma)$ be a matching in which the vertex u_{ij} is unmatched. By adding the edge (u_{ij}, u_{ji}) to M, and removing the edge from M which was previously incident at u_{ji}, we succeed in moving one unmatched vertex into the set of cluster vertices. Two such operations are sufficient to normalise any near-perfect matching. (If vertices u_{ij} and u_{ji} are both unmatched, then M can be normalised by adding the single edge (u_{ij}, u_{ji}).)

The normalising operation maps at most n^2 distinct matchings onto a single normalised matching; hence $|\mathcal{N}(\Gamma)| \leq n^2|\mathcal{N}'(\Gamma)|$. It is straightforward to check that $\phi(\mathcal{N}'(\Gamma)) = \mathcal{G}'(\mathbf{d}, X)$ and that each element of $\mathcal{G}'(\mathbf{d}, X)$ is the image of at most $\prod_{i=1}^{n} d_i!$ elements of $\mathcal{N}'(\Gamma)$. Putting these facts together we have

$$\frac{|\mathcal{N}(\Gamma)|}{|\mathcal{M}(\Gamma)|} \leq \frac{n^2|\mathcal{N}'(\Gamma)|}{|\mathcal{M}(\Gamma)|} \leq \frac{n^2|\mathcal{G}'(\mathbf{d}, X)|}{|\mathcal{G}(\mathbf{d}, X)|} \leq n^2 p(n).$$

The proof is completed by appealing to Theorem 3.2.1; the degree of the polynomial q in the statement of that theorem can be taken as $\lceil \frac{1}{2} \deg p \rceil + 1$. □

The proof of Theorem 3.3.4 hinged on the fact that a polynomial bound on the ratio $|\mathcal{G}'(\mathbf{d}, X)|/|\mathcal{G}(\mathbf{d}, X)|$ implies a polynomial bound on the ratio $|\mathcal{N}(\Gamma(\mathbf{d}))|/|\mathcal{M}(\Gamma(\mathbf{d}))|$; in fact the reverse implication also holds. P-stability is therefore not merely a sufficient, but also a necessary condition for our reduction to be applicable.

In the next section, we will develop a simple numerical condition on degree sequences/exclud pairs which is sufficient to guarantee P-stability.

A criterion for P-stability

For a graphical degree sequence $\mathbf{d} = (d_1, \ldots, d_n)$, define $d_{\max} = \max_i d_i$ and $e(\mathbf{d}) = \frac{1}{2}\sum_{i=1}^{n} d_i$. Note that $e(\mathbf{d})$ is integral. Next let $x_{\max}$ denote the maximum degree in the graph induced by X. We now derive a very useful sufficient condition for $\mathbf{d}$ to be P-stable in terms of the quantities $d_{\max}$, $e(\mathbf{d})$ and $x_{\max}$.

Theorem 3.3.5 *The class of all pairs $\mathbf{d}, X$ which satisfy $e(\mathbf{d}) > d_{\max}(d_{\max} + x_{\max} - 1)$ is P-stable.*

Proof Let $\mathbf{d}, X$ satisfy the above condition. We will show how to associate with each graph $G = (v, E) \in \mathcal{G}'(\mathbf{d}, X)$ a graph $\bar{G} \in \mathcal{G}(\mathbf{d}, X)$ which is "close to" G, in the sense that G can be transformed into $\bar{G}$ via a simple edge exchange operation. The

result will then follow from the observation that no graph in $\mathcal{G}(\mathbf{d}, X)$ can be close to too many graphs in $\mathcal{G}'(\mathbf{d}, X)$.

If $G \in \mathcal{G}(\mathbf{d}, X)$ we simply set $\bar{G} = G$, so assume that $G \in \mathcal{G}'(\mathbf{d}, X) - \mathcal{G}(\mathbf{d}, X)$. We describe the operation first in the case that G has degree sequence $\mathbf{d}'$ of the form

$$d'_i = \begin{cases} d_i - 1 & \text{for } i \in \{k, l\}; \\ d_i & \text{otherwise.} \end{cases}$$

If $(k, l) \notin E \cup X$ then we just add this edge to G to form $\bar{G}$. If on the other hand the edge $(k, l) \in E \cup X$ we look for a pair x, y of vertices such that $(x, y) \in E$

(i) x, y, k, l are all distinct;

(ii) (x, k) and (y, l) are not edges of G.

(iii) $(x, k), (y, l) \notin X$.

Then the graph $\bar{G}$ is formed by adding to G the edges (x, k) and (y, l) and deleting the edge (x, y). We claim that such a pair x, y can always be found. To see this, note that there are $2e(\mathbf{d}') = 2(e(\mathbf{d}) - 1)$ candidates for the ordered pair x, y among endpoints of edges of G, some of which are excluded by requirements (i) – (iii). Elementary counting reveals that the number of candidates excluded by (i) is at most $2(d'_k + d'_l) - 2 \leq 2(2d_{\max} - 3)$. Similarly, the number excluded by (ii) is at most $2(d_{\max} - 2)(d_{\max} - 1)$ and the number excluded by (iii) is at most $2x_{\max}d_{\max}$. It follows that a suitable pair x, y can be found provided that

$$2(e(\mathbf{d})-1) > 2(2d_{\max}-3)+2(d_{\max}-2)(d_{\max}-1)+2x_{\max}d_{\max} = 2d_{\max}(d_{\max}+x_{\max}-1)-2,$$

which is equivalent to the condition on $\mathbf{d}$ stipulated in the theorem.

It remains to describe $\bar{G}$ when the degree sequence of G is

$$d'_i = \begin{cases} d_i - 2 & \text{for } i = k; \\ d_i & \text{otherwise.} \end{cases}$$

In this case, we seek an edge (x, y) of G for which (i) $x, y \neq k$, and (ii) $(x, k), (y, k) \notin E \cup X$. The graph $\bar{G}$ is then obtained from G by adding the edges (x, k) and (y, k) and deleting (x, y). Using similar reasoning to the above, the reader may easily verify that a suitable edge (x, y) always exists under the stated condition on $\mathbf{d}$.

Now for any graph $H \in \mathcal{G}(\mathbf{d}, X)$, define the set

$$\mathcal{K}(H) = \{G \in \mathcal{G}'(\mathbf{d}, X) : \bar{G} = H\}.$$

Note that the sets $\mathcal{K}(H)$ partition $\mathcal{G}'(\mathbf{d}, X)$. It is a straightforward task to verify that each element of $\mathcal{K}(H)$ can be coded by a unique tuple (x, y, k, l), and hence that $|\mathcal{K}(H)| \leq n^4$. We therefore conclude that $|\mathcal{G}'(\mathbf{d}, X)| \,/\, |\mathcal{G}(\mathbf{d}, X)| \leq n^4$, so this class of degree sequences is indeed P-stable. □

For the remainder of this discussion we consider $X = \emptyset$ and talk about degree sequences only. Informally, Theorem 3.3.5 says that a degree sequence belongs to a P-stable class provided its maximum and average degrees do not differ by too much. Let us mention two important types of degree sequence which satisfy this condition:

(i) Let $\mathbf{d} = (k, k, \ldots, k)$, i.e., $\mathbf{d}$ is k-regular, with $k \leq n/2$. Then setting $e(\mathbf{d}) = nk/2$ and $d_{\max} = k$, we see at once that the hypothesis of Theorem 3.3.5 is satisfied.

(ii) Suppose all degrees in $\mathbf{d}$ lie in the range $[1, \sqrt{n/2}\,]$. Then setting $e(\mathbf{d}) \geq \frac{1}{2}(d_{\max} + n - 1)$ and $d_{\max} \leq \sqrt{n/2}$, the hypothesis of Theorem 3.3.5 is again seen to hold.

Note also that there is an obvious bijection between the sets $\mathcal{G}(\mathbf{d})$ and $\mathcal{G}(\overline{\mathbf{d}})$, where $\overline{\mathbf{d}}$ is the complement of $\mathbf{d}$, i.e., $\overline{d}_i = n - 1 - d_i$. Hence for the purposes of generation and counting $\mathbf{d}$ and $\overline{\mathbf{d}}$ are equivalent.

Remark 3.3.1 The class of *all* graphical degree sequences is not P-stable. To see this, consider the family of degree sequences on $2k$ vertices of the form

$$\mathbf{d}^{(k)} = (1, 2, \ldots, k-1, k, k, k+1, \ldots, 2k-1)$$

for $k = 3, 4, \ldots$. Observe that the selection of a graph with any given degree sequence $\mathbf{d} = (d_1, \ldots, d_n)$ can be viewed recursively as follows:

1. Select a set of d_n neighbours for vertex n.

2. Recursively select a graph with degree sequence $\mathbf{d}' = (d'_1, \ldots, d'_{n-1})$, where $\mathbf{d}'$ is obtained from $\mathbf{d}$ by deleting the final component and decrementing by 1 the components which correspond to the d_n chosen neighbours for vertex n.

Adopting this view, it is clear that there is a unique graph with degree sequence $\mathbf{d}^{(k)}$.

Now modify $\mathbf{d}^{(k)}$ by decrementing by 1 the final two components. Note that graphs on the modified degree sequence are members of $\mathcal{G}'(\mathbf{d}^{(k)})$. Applying the recursive selection

procedure we therefore have

$$\begin{aligned}
|\mathcal{G}'(\mathbf{d}^{(k)})| &\geq |\mathcal{G}(1,2,\ldots,k,k,k+1,\ldots,2k-3,2k-3,2k-2)| \\
&\geq |\mathcal{G}(1,1,2,\ldots,k-1,k-1,k,\ldots,2k-4,2k-4)| \\
&\geq |\mathcal{G}(1,1,1,2,\ldots,k-2,k-2,k-1,\ldots,2k-5)| \\
&\geq 3\times|\mathcal{G}(1,1,1,2,\ldots,k-3,k-3,k-2,\ldots,2k-7)| \\
&\geq 3^2\times|\mathcal{G}(1,1,1,2,\ldots,k-4,k-4,k-3,\ldots,2k-9)| \\
&\vdots \\
&\geq 3^{k-3}|\mathcal{G}(1,1,1,1)| \\
&= 3^{k-2}.
\end{aligned}$$

(The factor of 3 arises at each stage from the freedom to choose one of three degree-one vertices to be adjacent to the vertex of largest degree.) The ratio $|\mathcal{G}'(\mathbf{d}^{(k)})|/|\mathcal{G}(\mathbf{d}^{(k)})|$ is exponential in k, and hence in $n=2k$, the number of vertices.

Remark 3.3.2 When considering bipartite graphs we can dispense with worrying about P-stability. Suppose we consider bipartite graphs with $m+n$ vertices. Then we change the definition of Γ so that it is bipartite, allowing the generation of a (near) random perfect matching without qualification.

$$\begin{aligned}
V(\Gamma) =&\{v_{ik} : 1\leq i\leq m \text{ and } 1\leq k\leq d_i\}\cup\{w_{jk} : 1\leq j\leq n \text{ and } 1\leq k\leq d_j\} \\
&\cup\{u_{ij},u'_{ij} : 1\leq i\leq m,\ 1\leq j\leq n, (i,j)\notin X\}
\end{aligned}$$

and edge set

$$\begin{aligned}
E(\Gamma) =& \{(v_{ik},u_{ij}) : 1\leq i,\leq m,\ 1\leq j\leq n\, 1\leq k\leq d_i,\ \text{ and } (i,j)\notin X\} \\
&\cup\{(w_{jk},u'_{ij}) : 1\leq i\leq m,\ 1\leq j\leq n\, 1\leq k\leq d_i,\ \text{ and } (i,j)\notin X\} \\
&\cup\{(u_{ij},u'_{ij}) : 1\leq i,j\leq n,\ \text{ and } (i,j)\notin X\}.
\end{aligned}$$

Digraphs

Consider a directed graph $\overrightarrow{G}=(\overrightarrow{V},\overrightarrow{E})$, where the in-degree (out-degree, respectively) of a vertex $v\in\overrightarrow{V}$ is denoted by $d_-(v)$ ($d_+(v)$). A *0,1-flow* is defined as a subset of edges $\overrightarrow{E'}\subset\overrightarrow{E}$ such that in the resulting subgraph $(\overrightarrow{V},\overrightarrow{E'})$, $d_-(v)=d_+(v)$ for all $v\in\overrightarrow{V}$. Counting the number 0,1 flows is reducible to computing the 0,1 permanent of an undirected bipartite graph $G=(V,E)$ as follows.

The graph $G = (V, E)$ consists of:

$$V = \left\{ \begin{array}{ll} h_{i,j}, m_{i,j}, t_{i,j} & \text{for all } \overrightarrow{v_i v_j} \in \overrightarrow{E}, \\ u_i^1, \ldots, u_i^{d_-(v_i)} & \text{for all } v_i \in \overrightarrow{V} \end{array} \right\},$$

$$E = \left\{ \begin{array}{ll} (h_{i,j}, m_{i,j}), (m_{i,j}, t_{i,j}) & \text{for all } \overrightarrow{v_i v_j} \in \overrightarrow{E}, \\ (u_i^k, h_{i,j}) & \text{for all } i, j, k \text{ where } u_i^k, h_{i,j} \in \overrightarrow{V}, \\ (u_i^k, t_{j,i}) & \text{for all } i, j, k \text{ where } u_i^k, t_{j,i} \in \overrightarrow{V} \end{array} \right\}$$

A $0,1$-flow $\overrightarrow{E'}$ is mapped to a perfect matching M in the following manner. For each $\overrightarrow{v_i v_j} \in \overrightarrow{E'}$ add the edge $(h_{i,j}, m_{i,j})$ to M, while for each $\overrightarrow{v_i v_j} \in \overrightarrow{E} \setminus \overrightarrow{E'}$ add the edge $(m_{i,j}, t_{i,j})$ to M. Now for $v_i \in \overrightarrow{V}$, observe that the set of vertices $\{h_{i,j}\}_j \cup \{t_{j',i}\}_{j'}$, consists of exactly $d_-(v_i)$ unmatched vertices. There are $d_-(v_i)!$ ways of pairing these unmatched vertices with the set of vertices $\{u_i^k\}_k$. Thus the flow E' corresponds to $\prod_{v \in \overrightarrow{V}} d_-(v)!$ perfect matchings of G, and it is clear that the mapping is a bijection. This implies the following corollary.

Corollary 3.3.3 *For an arbitrary directed graph $\overrightarrow{G}$, there exists an fpras for counting the number of 0,1 flows.*

Suppose the directed graph $\overrightarrow{G}$ has a fixed source s and sink t. After adding a simple gadget from t to s we can estimate the number of maximum $0,1$ flows from s to t by estimating the number of $0,1$ flows in the resulting graph.

Hamilton Cycles in Dense Graphs

Let $G = (V, E)$ be a graph, where $V = \{v_1, v_2, \ldots, v_n\}$. Denote the degree of vertex v_i by d_i, for $i = 1, 2, \ldots, n$. We will say that G is *dense* if $\min_i d_i \geq (\frac{1}{2} + \alpha)n$, where $0 < \alpha \leq \frac{1}{2}$ is a fixed constant. Under these circumstances it is known that G must contain a Hamilton cycle. Moreover, the proof of this fact is easily modified to give a simple polynomial-time algorithm for constructing such a Hamilton cycle. This algorithm, which uses edges whose existence is guaranteed by the pigeonhole principle to "patch together" disjoint cycles, provides the required easy decision procedure.

We consider here the natural but more difficult problems of counting the *number* of Hamilton paths and cycles in such graphs. These problems are in fact #P-complete, so exact counting is presumably intractable. More positively, our main results in Sections 3.3.3 and 3.3.3 establish the existence of *fpras's* for these counting problems when $\alpha > 0$. We may observe that if the degree condition is relaxed to $\min_i d_i \geq (\frac{1}{2} - \alpha_n)n$ with $\alpha_n = \Omega(n^{\kappa - 1})$ for any fixed $\kappa > 0$, then the question of the existence of any Hamilton path or cycle becomes NP-Complete and so approximate counting is NP-hard. This is

true even if we insist on G being k-connected for any $k = o(n)$. Start with an arbitrary graph G and add a clique C of size $m = n^{1/\kappa}$ and an independent set I of size $m - 1$ and then join every vertex in C to every other vertex, to produce a graph Γ. Then G has a Hamilton path if and only if Γ has a Hamilton cycle. Also Γ contains a Hamilton path if and only if G contains two vertex disjoint paths that cover all its vertices.

Thus our results establish quite precisely the difficulty of the counting problem except in the region where α is close to zero. Section 5 extends the positive results of the earlier sections to cover self-avoiding paths and cycles of all lengths.

The natural approach given previous successes in this area is to try to find a rapidly mixing Markov chain with state space the set of Hamilton cycles of a given dense graph, and possibly its Hamilton paths as well. Earlier attempts with this approach have proved fruitless. Somewhat surprisingly, the key lies in the fact that in dense graphs, Hamilton cycles form a substantial fraction of the set of 2-factors, a *2-factor* being defined as a set of vertex-disjoint cycles which together contain all vertices of G. This is not obvious a priori and the main technical difficulty in the approach lies in obtaining a good upper bound on the ratio of 2-factors to Hamilton cycles in a dense graph. A direct attack — relating the number of 2-factors with k cycles to the number with $k+1$ cycles — appears unworkable. Instead, we introduce a weight function on 2-factors that allows us to argue about the distribution of total weight as a function of the number of cycles. By a rather delicate analysis, we are able to show that the Hamilton cycles carry sufficient weight for our purpose. In summary we prove

Theorem 3.3.6 *If G is dense then there are fpras's for*

(a) *approximating its number of Hamilton cycles,*

(b) *approximating its number of Hamilton paths,*

(c) *approximating its number of cycles of all sizes,*

(d) *approximating its number of paths of all sizes.*

Outline Approach

Our approach to constructing an fpras for Hamilton cycles in a dense graph G is via a randomized reduction to sampling and estimating *2-factors* in G. Using the results of Section 3.3.3 we prove

Theorem 3.3.7 *There exist both a good sampler and an FPRAS for the set of 2-factors in a dense graph.*

Proof The set of 2-factors in a graph $G = (V, E)$ is equal to $\mathcal{G}(\mathbf{d}, X)$, where $\mathbf{d} = (2, 2, \ldots, 2)$, and $X = V^{(2)} - E$ is the complementary edge set to E. The result now follows from Theorems 3.3.4 and 3.3.5, since, for a dense G and n sufficiently large, $d_{\max} = 2$, $x_{\max} < \frac{1}{2}n - 1$, and $d_{\max}(d_{\max} + x_{\max} - 1) < n = e(\mathbf{d})$. □

Given Theorem 3.3.7, the reduction from Hamilton cycles to perfect matchings is easy to describe. We estimate first the number of 2-factors in G, and then the proportion of 2-factors which are Hamilton cycles. Both counting and sampling phases run in polynomial time, by Theorem 3.3.7, provided only that G is dense. For the sampling phase to also produce an accurate estimate of the number of Hamilton cycles, it is necessary that the ratio of 2-factors to Hamilton cycles in G not be too large, i.e. bounded by a polynomial in n. This will be established in Section 3.3.3.

Many 2-factors are Hamiltonian

Let n be a natural number and $\beta = 10/\alpha^2$. Let $k_0 = \lfloor \beta \ln n \rfloor$, and for $1 \leq k \leq n$, define $g(k) = n^\beta k!(\beta \ln n)^{-k}$, and

$$f(k) = \begin{cases} g(k), & \text{if } k \leq k_0; \\ g(k_0), & \text{otherwise.} \end{cases}$$

Lemma 3.3.3 *Let f be the function defined above. Then*

1. *f is non-increasing and satisfies*
$$\min\{f(k-1), f(k-2)\} = f(k-1) \geq (\beta \ln n)k^{-1}f(k);$$
2. *$f(k) \geq 1$, for all k.*

Proof Observe that g is unimodal, and that k_0 is the value of k minimizing $g(k)$; it follows that f is non-increasing. When $k \leq k_0$, we have $f(k-1) = g(k-1) = (\beta \ln n)k^{-1}g(k) = (\beta \ln n)k^{-1}f(k)$; otherwise, $f(k-1) = g(k_0) = f(k) \geq (\beta \ln n)k^{-1}f(k)$. In either case, the inequality in part 1 of the lemma holds.

Part 2 of the lemma follows from the chain of inequalities

$$\frac{1}{f(k)} \leq \frac{1}{g(k_0)} \leq \frac{(\beta \ln n)^{k_0}}{n^\beta k_0!} \leq n^{-\beta} \sum_{k=0}^{\infty} \frac{(\beta \ln n)^k}{k!} = n^{-\beta} \exp(\beta \ln n) = 1.$$

□

Lemma 3.3.4 *Suppose α is constant greater than 0. Let $G = (V, E)$ be an undirected graph of order n and minimum degree $(\frac{1}{2} + \alpha)n$. Then the number of 2-factors in G exceeds the number of Hamilton cycles by at most a polynomial (in n) factor, the degree of the polynomial depending only on α.*

Proof For $1 \leq k \leq \lfloor n/3 \rfloor$, let Φ_k be the set of all 2-factors in G containing exactly k cycles, and let $\Phi = \cup_k \Phi_k$ be the set of all 2-factors. Define

$$\Psi = \big\{(F, F') : F \in \Phi_k, F' \in \Phi_{k'}, k' < k, \text{ and } F \oplus F' \cong C_6\big\},$$

where $\oplus$ denotes symmetric difference, and C_6 is the cycle on 6 vertices. Observe that (Φ, Ψ) is an acyclic directed graph; let us agree to call its component parts *nodes* and *arcs* to avoid confusion with the vertices and edges of G. Observe also that if $(F, F') \in \Psi$ is an arc, then F' can be obtained from F by deleting three edges and adding three others, and that this operation can decrease the number of cycles by at most two. Thus every arc $(F, F') \in \Psi$ is directed from a node F in some Φ_k to a node F' in Φ_{k-1} or Φ_{k-2}.

Our proof strategy is to define a positive weight function on the arc set Ψ such that the total weight of arcs leaving each node (2-factor) $F \in \Phi \setminus \Phi_1$ is at least one greater than the total weight of arcs entering F. This will imply that the total weight of arcs entering Φ_1 is an upper bound on the number of non-Hamilton 2-factors in G, and that the maximum total weight of arcs entering a single node in Φ_1 is an upper bound on the ratio $|\Phi \setminus \Phi_1|/|\Phi_1|$.

The weight function $w : \Psi \to \mathbb{R}^+$ we employ is defined as follows. For any arc (F, F') with $F' \in \Phi_k$: if the 2-factor F' is obtained from F by coalescing two cycles of lengths l_1 and l_2 into a single cycle of length $l_1 + l_2$, then $w(F, F') = (l_1^{-1} + l_2^{-1})f(k)$; if F' results from coalescing three cycles of length l_1, l_2 and l_3 into a single one of length $l_1 + l_2 + l_3$, then $w(F, F') = (l_1^{-1} + l_2^{-1} + l_3^{-1})f(k)$.

Let $F \in \Phi_k$ be a 2-factor with $k > 1$ cycles $\gamma_1, \gamma_2, \ldots, \gamma_k$, of lengths $n_1, n_2, \ldots, n_k$. We proceed to bound from below the total weight of arcs *leaving* F. For this purpose imagine that the cycles $\gamma_1, \gamma_2, \ldots, \gamma_k$ are oriented in some way, so that we can speak of each oriented edge (u, u') in some cycle γ_i as being "forward" or "backward". Since we are interested in obtaining a *lower* bound, it is enough to consider only arcs (F, F^+) from F of a certain kind: namely, those for which the 6-cycle $\gamma = F \oplus F^+$ is of the form $\gamma = (x, x', y, y', z, z')$, where $(x, x') \in F$ is a forward cycle edge, $(y, y') \in F$ is a forward edge in a cycle distinct from the first, and $(z, z') \in F$ is a backward cycle edge. The edge (z, z') may be in the same cycle as either (x, x') or (y, y'), or in a third cycle. Observe that (x', y), (y', z) and (z', x) must necessarily be edges of F^+. It is routine to check that any cycle $\gamma = (x, x', y, y', z, z')$ satisfying the above constraints does correspond to a valid arc from F. The fact that (z, z') is oriented in the opposite sense to (x, x') and (y, y') plays a crucial role in ensuring that the number of cycles decreases in the passage to F^+ when only two cycles involved.

First, we estimate the number of cycles γ for which (x, x') is contained in a particular cycle γ_i of F. We might say that γ is *rooted* at γ_i. Assume, for a moment, that the vertices x, x', y, y' have already been chosen. There are at least $(\frac{1}{2} + \alpha)n - 5$ ways to extend the path (x, x', y, y'), first to z and then to z', which are consistent with the rules given above; let Z' be the set of all vertices z' so reachable. Denote by $G(x)$ the set

of vertices adjacent to x. The number of ways of completing the path (x, x', y, y') to a valid 6-cycle is at least

$$\begin{aligned} |G(x) \cap Z'| &\geq |G(x)| + |Z'| - n \\ &\geq (\tfrac{1}{2} + \alpha)n + [(\tfrac{1}{2} + \alpha)n - 5] - n \\ &= 2\alpha n - 5 \\ &\geq \alpha n, \end{aligned}$$

for n sufficiently large. A lower bound on the number of 6-cycles γ rooted at γ_i now follows easily: there are n_i choices for (x, x'); then at least $(\frac{1}{2} + \alpha)n - n_i$ choices for (y, y'); and finally — as we have just argued — at least αn ways to complete the cycle. Thus the total number of 6-cycles rooted at γ_i is at least $\alpha n n_i[(\frac{1}{2} + \alpha)n - n_i]$.

We are now poised to bound the total weight of arcs leaving F. Each arc (F, F^+) defined by a cycle γ rooted at γ_i has weight at least $n_i^{-1} \min\{f(k-1), f(k-2)\}$, which, by Lemma 3.3.3, is bounded below by $(\beta \ln n)(kn_i)^{-1} f(k)$. Thus the total weight of arcs leaving F is bounded as follows:

$$\sum_{F^+:(F,F^+)\in\Psi} w(F, F^+) \geq \sum_{i=1}^{k} \alpha n n_i[(\tfrac{1}{2} + \alpha)n - n_i] \frac{(\beta \ln n) f(k)}{k n_i} \tag{3.16}$$

$$\begin{aligned} &= \alpha n^2 [(\tfrac{1}{2} + \alpha)k - 1] \frac{(\beta \ln n) f(k)}{k} \\ &\geq \alpha^2 \beta \, f(k) n^2 \ln n \\ &\geq 10 f(k) n^2 \ln n, \end{aligned} \tag{3.17}$$

where we have used the fact that $k \geq 2$. Note that the presence of a unique backward edge, namely (z, z'), ensures that each cycle γ has a distinguishable root, and hence that the arcs (F, F^+) were not overcounted in summation (3.16).

We now turn to the corresponding *upper* bound on the total weight of arcs $(F^-, F) \in \Psi$ *entering* F. It is straightforward to verify that the cycle $\gamma = (x, x', y, y', z, z') = F^- \oplus F$ must contain three edges — (x, x'), (y, y') and (z, z') — from a single cycle γ_i of F, the remaining edges coming from F^-. The labeling of vertices in γ can be made canonical in the following way: assume an ordering on vertices in V, and assign label x to the smallest vertex. The condition $(x, x') \in F$ uniquely identifies vertex x', and the labeling of the other vertices in the cycle γ follows.

Removing the three edges (x, x'), (y, y') and (z, z') from γ_i leaves a triple of simple paths of lengths (say) $a - 1$, $b - 1$ and $c - 1$: these lengths correspond (respectively) to the segment containing x, the segment containing x', and the remaining segment. Going round the cycle γ_i, starting at x' and ending at x, the vertices x, x', y, y', z, z' may appear in one of eight possible sequences:

$$x', y', y, z', z, x;$$

$$
\begin{array}{l}
x', z, z', y, y', x; \\
x', z, z', y', y, x; \\
x', z', z, y, y', x; \\
x', y', y, z, z', x; \\
x', y, y', z', z, x; \\
x', z', z, y', y, x; \\
x', y, y', z, z', x.
\end{array}
$$

For a given triple of lengths (a, b, c), each of the above sequences corresponds to at most n_i possible choices for the edges (x, x'), (y, y') and (z, z'), yielding a maximum of $8n_i$ in total. To see this, observe that the edge (x, x') may be chosen in n_i ways (minimality of x fixes the orientation of the edge), and that the choice of (x, x') combined with the information provided by the sequence completely determines the triple of edges.

The eight sequences divide into five possible cases, as the first four sequences lead to equivalent outcomes (covered by case 1 below). Taken in order, the five cases are:

1. For at most $4n_i$ of the choices for the edges (x, x'), (y, y') and (z, z'), $\gamma_i \oplus \gamma$ is a single cycle;

2. for at most n_i choices, $\gamma_i \oplus \gamma$ is a pair of cycles of lengths a and $b + c$;

3. for at most n_i choices, $\gamma_i \oplus \gamma$ is a pair of cycles of lengths b and $a + c$;

4. for at most n_i choices, $\gamma_i \oplus \gamma$ is a pair of cycles of lengths c and $a + b$;

5. for at most n_i choices, $\gamma_i \oplus \gamma$ is a triple of cycles of lengths a, b and c.

The first case does not yield an arc (F^-, F), since the number of cycles does not decrease when passing from $F^- = F \oplus \gamma$ to F, but the other four cases do have to be reckoned with.

The total weight of arcs entering F can be bounded above as follows:

$$\begin{aligned}\sum_{F^-:(F^-,F)\in\Psi} w(F^-,F) &\le \sum_{i=1}^{k} n_i f(k) \sum_{\substack{a,b,c\ge 1\\ a+b+c=n_i}} \left[\left(\frac{1}{a}+\frac{1}{b}+\frac{1}{c}\right)+\right.\\ &\qquad \left.\left(\frac{1}{a}+\frac{1}{b+c}\right)+\left(\frac{1}{b}+\frac{1}{a+c}\right)+\left(\frac{1}{c}+\frac{1}{a+b}\right)\right]\\ &= \sum_{i=1}^{k} n_i f(k) \sum_{\substack{a,b,c\ge 1\\ a+b+c=n_i}} \left[\frac{6}{a}+\frac{3}{b+c}\right]\\ &\le \sum_{i=1}^{k} n_i f(k) n \sum_{a=1}^{n_i-1} \left[\frac{6}{a}+\frac{3}{n_i-a}\right]\\ &\le 9f(k)n^2 H_n \qquad (3.18)\end{aligned}$$

where $H_n = \sum_{i=1}^{n} i^{-1} \le \ln n + 1$ is the nth harmonic number. Combining inequalities (3.17) and (3.18), we have

$$\begin{aligned}\sum_{F^+:(F,F^+)\in\Psi} w(F,F^+) \;-\; \sum_{F^-:(F^-,F)\in\Psi} w(F^-,F) &\ge 10f(k)n^2\ln n - 9f(k)n^2 H_n\\ &\ge f(k)n^2(\ln n - 9)\\ &\ge n^2(\ln n - 9),\end{aligned}$$

where the final inequality is by Lemma 3.3.3. Thus the total weight of arcs leaving F exceeds the total weight of arcs entering by at least 1, provided n is sufficiently large. The number of non-Hamilton 2-factors $|\Phi \setminus \Phi_1|$ is bounded above by the total weight of arcs entering Φ_1, which in turn is bounded — see inequality (3.18) — by $|\Phi_1| \times 9f(1)n^2 H_n = |\Phi_1| \times O(n^{2+\beta})$. This establishes the lemma. □

Counting the number of cycles of all sizes

We will first consider approximating the total number of cycles in graphs with minimum degree $(\frac{1}{2}+\alpha)n$.

We first note that if we add a loop to each vertex and extend the definition of 2-factor to include loops as cycles of length one, then the argument of Section 3.3.3 may be extended to this case (note that we still forbid cycles of length two i.e. double edges). Thus there exists both a fully polynomial randomized approximation scheme and a fully polynomial almost uniform sampler for the set of *partial* 2-factors in a dense graph. Let a partial 2-factor be *cyclic* if it consists of a single cycle of length at least three and a collection of loops. Clearly the number of cyclic partial 2-factors is the same as the number of cycles.

The procedure for approximating the number of cycles of all sizes is as follows: we estimate first the number of partial 2-factors in G, and then the number of cyclic partial 2-factors by standard sampling methods as a proportion of the number of partial 2-factors. To produce an accurate estimate in polynomial time it is only necessary to show that the ratio of partial 2-factors to cyclic partial 2-factors is not too large. Let

$$\mathcal{F}_\ell = \{\text{partial 2-factors with } \ell \text{ loops}\}, \quad \text{and} \quad f_\ell = |\mathcal{F}_\ell|.$$

For a given $F \in \mathcal{F}_\ell$ let $L = \{\text{loops of } F\}$, which we will now identify with the corresponding set of vertices. For $v \in L$ let d_v denote the number of neighbours of v in L and $D = \sum_{v\in L} d_v$.

If $v \in L$ then there are at least $2\alpha n - 2d_v$ ways of adding v to a cycle C of F by deleting an edge (a, b) of C and adding edges $(a, v), (v, b)$. Indeed we go round each cycle C of F; if the successor b of a vertex a neighbouring v is also a neighbour of v, then it forms an (a, b, v) triangle. The number of such triangles is at least $2\alpha n - 2d_v$.

So in total there are at least

$$\sum_{v\in L}(2\alpha n - 2d_v) = 2\ell\alpha n - 2D \tag{3.19}$$

$$\geq 2\ell(\alpha n - (\ell - 1)) \tag{3.20}$$

such augmentations.

Suppose first that $\ell \leq \ell_1 = \lfloor \alpha n/2 \rfloor$. Then (3.20) gives at least $\ell\alpha n$ augmentations of $F \in \mathcal{F}_\ell$ to an $F' \in \mathcal{F}_{\ell-1}$. Each $F' \in \mathcal{F}_{\ell-1}$ arises in at most n ways and so

$$\frac{f_{\ell-1}}{f_\ell} \geq \alpha\ell.$$

Putting $\ell_0 = \lceil 2/\alpha \rceil$ we see that

$$f_{\ell_1} + f_{\ell_1-1} + \cdots + f_{\ell_0+1} \leq f_{\ell_0} \leq f_{\ell_0} + f_{\ell_0-1} + \cdots + f_0. \tag{3.21}$$

Suppose next that $\ell > \ell_1$. Note first that since a graph with r vertices and s edges contains at least $r - s + 1$ distinct cycles, we see that L contains at least

$$\frac{D}{2} - \ell + 1 \tag{3.22}$$

distinct cycles.

Adding a cycle C contained in L to F and removing $|C|$ loops gives us a 2-factor in $\mathcal{F}_{\ell'}$ where $\ell' < \ell$. From (3.19) and (3.22) we see that there are at least

$$\left(\frac{2\ell\alpha n - 2D}{4}\right)^+ + \left(\frac{D}{2} - \ell\right)^+ \geq \ell\left(\frac{\alpha n}{2} - 1\right) \tag{3.23}$$

$$\geq \frac{\ell\alpha n}{3} \tag{3.24}$$

augmentations of either sort from F. Each $F' \in \mathcal{F}_{<\ell}$ arises in at most $n+n$ ways (accounting for both ways of reducing L) and so

$$\begin{aligned} f_\ell &\le \frac{6}{\alpha\ell}(f_{\ell-1} + f_{\ell-2} + \cdots + f_0) \\ &\le \theta(f_{\ell-1} + f_{\ell-2} + \cdots + f_0), \end{aligned}$$

where $\theta = 12/(\alpha^2 n)$, assuming $\ell > \ell_1$.

Thus

$$\frac{f_\ell + f_{\ell-1} + \cdots + f_0}{f_{\ell-1} + f_{\ell-2} + \cdots + f_0} \le 1 + \theta$$

and so

$$f_\ell + f_{\ell-1} + \cdots + f_0 \le (1+\theta)^{\ell-\ell_1}\Sigma_1, \tag{3.25}$$

where $\Sigma_1 = f_{\ell_1} + f_{\ell_1-1} + \cdots + f_0$. We weaken (3.25) to

$$\begin{aligned} f_{\ell_1+k} &\le (1+\theta)^k \Sigma_1 \\ &\le e^{12\alpha^{-2}}\Sigma_1. \end{aligned} \tag{3.26}$$

It follows from (3.21) and (3.26) that

$$\frac{f_0 + f_1 + \cdots + f_n}{f_0 + f_1 + \cdots + f_{\ell_0}} \le 2 + 2ne^{12\alpha^{-2}}. \tag{3.27}$$

Now take an $F \in \mathcal{F}_\ell$ where $\ell \le \ell_0$ and fix its set of loops L. The number of partial 2-factors with this same L is at most a polynomial factor, $p(n)$ say, of the number of cycles of size $n-\ell$ through $V \setminus L$, by the results of Section 3. (It is clear that because ℓ is small here, the required degree conditions are satisfied.) Thus, by (3.27), the ratio of partial 2-factors to cyclic partial 2-factors is $O(np(n))$ and we have proved the existence of an fpras for the number of cycles.

Paths and Hamilton Paths

We obtain an fpras for counting the number of Hamilton paths in the following way. We add a vertex v_0 and join it by an edge to every vertex of G. Call this new graph G^*. The number of Hamilton cycles in G^* is equal to the number of Hamilton paths in G. Since G^* is dense we can approximate the latter quantity by approximating the former.

Similarly, to estimate the number of paths of all lengths, we compute an estimate c^* for the number of cycles in G^* and an estimate ρ^* for the proportion ρ of cycles which contain v_0. Since the number of cycles containing v_0 is the number of paths in G, this provides an estimate $\rho^* c^*$ for the number of paths. Also, this will give us an fpras provided ρ is not too small. But clearly $\rho \ge 3/4$ and we are done.

Eulerian Orientations

A graph $G = (V, E)$ is Eulerian if it is connected and all of its vertex degrees $d(v)$, $v \in V$ are even. An orientation σ of G defines a digraph $D_\sigma = (V, A_\sigma)$ where the unoriented edge $\{u, v\} \in E$ is replaced either by (u, v) or (v, u) in A_σ. Thus there are precisely $2^{|E|}$ distinct orientations of G.

Let $d^+_\sigma(v), d^-_\sigma(v)$ denote the outdegree, indegree of vertex v under orientation σ. σ is an Eulerian orientation if $d^+_\sigma(v) = d^-_\sigma(v)$ for all $v \in V$.

Theorem 3.3.8 *There is a good sampler and an FPRAS for the Euler orientations of an Eulerian graph G.*

Proof We reduce the problem of sampling/counting Euler orientations to that of sampling/counting perfect matchings in an associated bipartite graph G'. The graph G' has vertex bipartition $V' = V_1 \cup V_2$ where

$$V_1 = \bigcup_{v \in V} X_v \text{ and } X_v = \{x_{v,e} : \ v \in e \in E\} \text{ for } v \in V.$$

$$V_2 = \{w_e : \ e \in E\} \cup \bigcup_{v \in V} Y_v \text{ and } Y_v = \{y_{v,i} : \ 1 \le i \le d(v)/2\} \text{ for } v \in V.$$

The edge set E' of G' is defined by

$$E' = \{\{x_{u,e}, w_e\} : \ u \in e \in E\} \cup \bigcup_{v \in V} X_v \times Y_v.$$

Let $n' = 2m = |V_1| = |V_2|$ where $m = |E|$. Let now $\mathcal{M}'$ denote the set of perfect matchings of G'. Let $\mathcal{P}_0$ denote the set of Eulerian orientations of G.

Lemma 3.3.5 *$\mathcal{M}'$ can be partitioned into $\mathcal{M}'_\sigma : \sigma \in \mathcal{P}_0$ so that $|\mathcal{M}'_\sigma| = \prod_{v \in V}(d(v)/2)!$ for all $\sigma \in \mathcal{P}_0$.*

It follows immediately that sampling/counting for $\mathcal{P}_0$ can be reduced to sampling/counting for $\mathcal{M}'$.

Proof of Lemma 3.3.5 Given a perfect matching M of G' we let $X^M_v = \{x_{v,e} \in X_v : \ \{x_{v,e}, w_e\} \in M\}$. Then define an orientation σ as follows: If $e = \{u, v\} \in E$ then put (u, v) into A_σ if $x_{u,e} \in X^M_u$ and put (v, u) into A_σ if $x_{v,e} \in X^M_v$. Exactly one of these is true, in order that w_e is covered by M. The orientation is Eulerian because $d^+_\sigma(v) = |X^M_v| = d(v)/2$.

Furthermore, there are $\nu = \prod_{v \in V}(d(v)/2)!$ different matchings for each fixed collection $X^M_v : v \in V$, all giving the same orientation. The construction is reversible i.e. given $\sigma \in \mathcal{P}_0$ we let $X^M_v = \{x_{v,e} : \ e \text{ is oriented away from } v\}$ for $v \in V$. □

Chapter 4

Computing the volume of a convex body

The mathematical study of areas and volumes is as old as civilization itself, and has been conducted for both intellectual and practical reasons. As far back as 2000 B.C., the Egyptians[1] had methods for approximating the areas of fields (for taxation purposes) and the volumes of granaries. The exact study of areas and volumes began with Euclid[2] and was carried to a high art form by Archimedes[3]. The modern study of this subject began with the great astronomer Johann Kepler's treatise[4] *Nova stereometria doliorum vinariorum*, which was written to help wine merchants measure the capacity of their barrels.

We consider here the problem of computing the volume of a convex body in $\mathbb{R}^n$, where n is assumed to be relatively large.

[1]The Rhind Papyrus (copied ca. 1650 BC by a scribe who claimed it derives from the "middle kingdom" about 2000 - 1800 BC) consists of a list of problems and solutions, 20 of which relate to areas of fields and volumes of granaries.

[2]The exact study of volumes of pyramids, cones, spheres and regular solids may be found in Euclid's Elements (ca. 300 BC).

[3]Archimedes (ca. 240 BC) developed the method of exhaustion (found in Euclid) into a powerful technique for comparing volumes and areas of solids and surfaces. Manuscripts:

1. Measurement of the Circle. (Proves $3\frac{10}{71} < \pi < 3\frac{1}{7}$).
2. Quadrature of the Parabola
3. On the Sphere and Cylinder
4. On Spirals
5. On Conoids and Spheroids

[4]The application of modern infinitesimal ideas begins with Kepler's *Nova stereometria doliorum vinariorum* (New solid geometry of wine barrels), 1615.

4.1 The oracle model

A convex body $K \subseteq \mathbb{R}^n$ could be be given in a number of ways. For example K could be a polyhedron and we are given a list of its faces, as we would be in the domain of Linear Programming. We could also be given a set of points in $\mathbb{R}^n$ and told that K is its convex hull.

In general, however, K may not be a polyhedron, and it might be difficult (or even impossible) to give a compact description of it. For example, if $K = \{(y, z) \in \mathbb{R}^{m+1} : v(y) \geq z\}$, where $v(y) = \max\{cx : Ax = y, x \geq 0\}$ is the value function of a linear program (A is an $m \times n$ matrix.)

We want a way of defining convex sets which can handle all these cases. This can be achieved by taking an "operational" approach to defining K i.e. we assume that information about K can be found by asking an oracle. We assume that we have access to a *strong membership* oracle. Given $x \in \mathbb{R}^n$ we can "ask" the oracle whether or not $x \in K$. The oracle is assumed to answer immediately. Thus the work that the oracle does is hidden from us, but in most cases of interest it would be a polynomial time computation. For example, if K is a polyhedron given by its facets, all the oracle needs to do is check whether or not x is on the right side of each defining hyperplane.

With such an oracle, we will need to be given a litle more information. For $x \in \mathbb{R}^n$ and $r > 0$ we let $B(x, r)$ denote a ball of radius r with centre x and let $B = B(0, 1)$. We assume that there exists $d \in \mathbb{R}$ such that

$$B \subseteq K \subseteq dB. \tag{4.1}$$

In this case we say that the oracle is *well-guaranteed*.

Without a similar such guarantee, one could not be certain of finding even a single point of K in finite time.

4.2 Sampling from a convex body

We discuss generating random points in a convex body K by the use of random walks. Here $\delta > 0$ is some parameter to be defined later.

Ball Walk: $\mathcal{BW}$
Let $v_0 = 0$ and generate $v_1, v_2, \ldots, v_k, \ldots$ as follows: With probability $1/2$ put $v_{k+1} = v_k$. Otherwise, choose y randomly from $B(v_k, \delta)$. If $y \in K$ put $v_{k+1} = y$, otherwise $v_{k+1} = v_k$.

A step of $\mathcal{BW}$ where $y \notin K$ is called an *improper* step and the other steps are called *proper* steps.

One immediate problem with this Markov chain is that the state space K is far from

finite and we have only been discussing finite Markov chains. Also, in practice we can only compute the coordinates of points $x \in K$ to finite precision. We therefore let $\eta = 2^{-\lceil 10n \log_2(\epsilon^{-1}dn)\rceil} \leq (\epsilon^{-1}dn)^{-10n}$ for some $0 < \epsilon < 1$ and let $\mathcal{L}$ be the lattice $\eta\mathbb{Z}^n$. Our random walk will therefore be on $K_\mathcal{L}$ where for convex set S we let $S_\mathcal{L} = K \cap \mathcal{L}$ and we re-define $\mathcal{BW}$. For $x \in \mathcal{L}$ and $r > 0$ we let $B_\mathcal{L}(x, r) = \{y \in \mathcal{L} : |y - x| \leq r\}$:

Ball Walk: $\mathcal{BW}$
Let v_0 be chosen with distribution P and generate $v_1, v_2, \ldots, v_k, \ldots$ as follows: With probability $1/2$ put $v_{k+1} = v_k$. Otherwise, choose y randomly from $B_\mathcal{L}(v_k, \delta)$. If $y \in K$ put $v_{k+1} = y$, otherwise $v_{k+1} = v_k$.

The steady state distribution Q of $\mathcal{BW}$ is uniform over $K_\mathcal{L}$: if $x, y \in K_\mathcal{L}$ and $|x - y| \leq \delta$ then

$$\mathbf{Pr}(v_{k+1} = y \mid v_k = x) = \frac{1}{2|B_\mathcal{L}(0, \delta)|}.$$

Thus the uniform distribution satisfies the detailed balance equations (1.15).

Speedy Walk: $\mathcal{SW}$
If we ignore the improper steps of $\mathcal{BW}$ then we have a sequence $v_0' = v_0, v_1', v_2', \ldots$ which define a new Markov chain called the *speedy walk*.

The steady state distribution of $\mathcal{SW}$ is not uniform. For $x \in K_\mathcal{L}$ let its *local conductance* $\ell_\mathcal{L}(x)$ be defined by

$$\ell_\mathcal{L}(x) = \frac{|K_\mathcal{L} \cap B_\mathcal{L}(x, \delta)|}{|B_\mathcal{L}(x, \delta)|}.$$

The steady state distribution $Q_\mathcal{L}$ of $\mathcal{SW}$ is given by

$$Q_\mathcal{L}(x) = \frac{\ell_\mathcal{L}(x)}{\ell_\mathcal{L}(K_\mathcal{L})}.$$

This is just the usual degree formula (1.16). The *average local conductance* is

$$\lambda_\mathcal{L} = \frac{\ell_\mathcal{L}(K_\mathcal{L})}{|K_\mathcal{L}|}.$$

Theorem 4.2.1 *If $K \supseteq B$ then*

$$\lambda_\mathcal{L} \geq 1 - \delta\sqrt{n}.$$

Proof (Deferred to Section 4.2.1) □

It will be useful to note that $U_\mathcal{L}$ is close in distribution to $Q_\mathcal{L}$.

Lemma 4.2.1

$$D_{\mathrm{tv}}(Q_\mathcal{L}, U_\mathcal{L}) \leq \frac{1 - \lambda_\mathcal{L}^2}{\lambda_\mathcal{L}}.$$

Proof Let $A \subseteq K_{\mathcal{L}}$. Then

$$|Q_{\mathcal{L}}(A) - U_{\mathcal{L}}(A)| = \left|\sum_{x\in A}\left(\frac{\ell_{\mathcal{L}}(x)}{\ell_{\mathcal{L}}(K_{\mathcal{L}})} - \frac{1}{|K_{\mathcal{L}}|}\right)\right| \leq$$
$$\sum_{x\in A}\left(\frac{\ell_{\mathcal{L}}(x)}{\ell_{\mathcal{L}}(K_{\mathcal{L}})} - \frac{\ell_{\mathcal{L}}(x)}{|K_{\mathcal{L}}|}\right) + \sum_{x\in A}\left(\frac{1}{|K_{\mathcal{L}}|} - \frac{\ell_{\mathcal{L}}(x)}{|K_{\mathcal{L}}|}\right)$$
$$\leq \frac{|A|(|K_{\mathcal{L}}| - \ell_{\mathcal{L}}(K_{\mathcal{L}}))}{\ell_{\mathcal{L}}(K_{\mathcal{L}})|K_{\mathcal{L}}|} + \frac{|K_{\mathcal{L}}| - \ell_{\mathcal{L}}(K_{\mathcal{L}})}{|K_{\mathcal{L}}|} \leq \frac{1-\ell_{\mathcal{L}}}{\ell_{\mathcal{L}}} + 1 - \ell_{\mathcal{L}}.$$

□

Our choice of $\delta = o(1/\sqrt{n})$ is such that $\lambda_{\mathcal{L}} = 1 - o(1)$ for all bodies discussed in this and later sections.

Theorem 4.2.2 *Let $K \subseteq dB$ be a convex body, $d > 32\delta$. Then the mixing time $\tau_{\mathcal{SW}}(\epsilon)$ of the speedy walk satisfies*

$$\tau_{\mathcal{SW}}(\epsilon) \leq \kappa n d^2 \delta^{-2} \log(1/\epsilon)$$

for some absolute constant $\kappa > 0$.

Proof (Deferred to Section 4.5.3) □

We will refer later to two sampling algorithms. First we need the following nesting of convex sets: Let K be a convex body where $B \subseteq K \subseteq dB$. Let $K^{(i)} = (2^{i/n}B) \cap K$ for $0 \leq i \leq m = n\log_2 d$. Then $K^{(0)} = B$ and $K^{(m)} = K$. In general quantities superscripted by i will refer to $K^{(i)}$ e.g. $\lambda_{\mathcal{L}}^{(i)}$ denotes the average local conductance of $K^{(i)}$.

Algorithm NESTED SAMPLING:
begin
Choose u_0 uniformly from $B_{\mathcal{L}}$.
For $i = 1$ **to** m **do**
$v \leftarrow u_{i-1}$
begin
A: Carry out a t-step speedy walk on $K_{\mathcal{L}}^{(i)}$ starting at v and ending at w.
If $u = \frac{2n}{2n-1}w \in K_{\mathcal{L}}^{(i)}$ **then** $u_i \leftarrow u$ and go to the next i.
Otherwise $v \leftarrow w$, **goto** A.
end
end

Algorithm ORDINARY SAMPLING:

begin
Run Algorithm NESTED SAMPLING to obtain $v_1 = u_m$.
For $i = 1$ **to** p **do**
$w \leftarrow v_{i-1}$
begin
A: Carry out a t-step speedy walk on $K_{\mathcal{L}}$ starting at w and ending at x.
If $v = \frac{2n}{2n-1}x \in K_{\mathcal{L}}$ **then** $v_i \leftarrow v$ and go to the next i.
Otherwise $w \leftarrow x$, **goto** A.
end
end

Let $U_{\mathcal{L}}$ denote the uniform distribution on $K_{\mathcal{L}}$ and let $U_{\mathcal{L}}^{(i)}$ denote the uniform distribution on $K_{\mathcal{L}}^{(i)}$ for $i = 1, 2, \ldots, m$.

Theorem 4.2.3 *Let* $0 < \alpha < \frac{1}{50}$ *and*

$$\delta \le \frac{\alpha}{\sqrt{n}} \text{ and } t = \left\lceil \kappa n d^2 \delta^{-2} \ln(10/\alpha) \right\rceil.$$

Let $U^{(i)}$ *denote the distribution of* u_i *in* NESTED SAMPLING, given $u_j, j \neq i$ *and let* $V^{(i)}$ *denote the distribution of* v_i *in* ORDINARY SAMPLING, given $v_j, j \neq i$. *Then*

(a) $D_{\mathrm{tv}}(U_{\mathcal{L}}^{(i)}, U^{(i)}) \le 4\alpha$ *for* $i = 1, 2, \ldots, m$ *and conditional on an event* $\mathcal{G}$ *of probability at least* $1 - 5m\alpha$, *the expected number of oracle calls in* NESTED SAMPLING *is at most* $10mt$.

(b) $D_{\mathrm{tv}}(U_{\mathcal{L}}, V^{(i)}) \le 4\alpha$ *for* $i = 1, 2, \ldots, p$ *and conditional on an event* $\mathcal{G}$ *of probability at least* $1 - 5m\alpha$, *the expected number of oracle calls in* ORDINARY SAMPLING *is at most* $10(m + p)t$.

Proof Deferred to Section 4.5.1 □

4.3 Volume Algorithm

$0 < \epsilon < 1$ be given. Let

$$\delta = \frac{1}{\sqrt{8n\ln(n/\epsilon)}},\ p = \left\lceil \frac{120m}{\epsilon^2} \right\rceil,\ \epsilon_0 = \left\lceil \frac{\epsilon}{12800^2 mnp} \right\rceil \text{ and } t = \left\lceil \kappa n \left(\frac{d}{\delta}\right)^2 \ln \frac{10}{\epsilon_0} \right\rceil. \tag{4.2}$$

We write

$$|K_{\mathcal{L}}| = |K_{\mathcal{L}}^{(0)}| \prod_{i=1}^{m} \frac{|K_{\mathcal{L}} i|}{K_{\mathcal{L}}^{(i-1)}|}.$$

We need to be sure that an estimate of $|K_{\mathcal{L}}|$ yields a good estimate of $\mathrm{vol}(K)$.

Theorem 4.3.1 *Let $S \supseteq \alpha B$ be a convex set in $\mathbb{R}^n$ where $\alpha \geq \delta$. Then*

$$|vol(S) - \eta^n |S_{\mathcal{L}}|| \leq \frac{\epsilon}{10} vol(S).$$

Proof (Deferred to Section 4.2.1) □

We note that $K^{(i)} \subseteq 2^{1/n} K^{(i-1)}$ and so

$$\frac{1}{2} \leq \rho_i = \frac{|K_{\mathcal{L}}^{(i-1)}|}{|K^{(i)}|} \leq 1. \tag{4.3}$$

We use sampling to estimate the ratios ρ_i, $i = 1, 2, \ldots, m$. Since $|K_{\mathcal{L}}^{(0)}| = |B_{\mathcal{L}}|$ can be computed to arbitrary accuracy, we see that this will give us an estimate of $|K_{\mathcal{L}}|$ and hence of vol(K).

Algorithm VOLUME COMPUTATION
Run Algorithm NESTED SAMPLE p times with t as in (4.2) and $\alpha = \epsilon_0$ to obtain $u_{i,r}$, $i = 0, 1, \ldots, m$, $r = 1, 2, \ldots, p$.

Now for $r = 1, 2, \ldots, p$ define

$$a_{i,r} = \begin{cases} 1 & u_{i,r} \in K_{\mathcal{L}}^{(i-1)} \\ 0 & \text{otherwise} \end{cases}$$

and let $b_i = a_{i,1} + \cdots + a_{i,p}$ for $i = 1, 2, \ldots, m$.

Then put

$$\zeta = \eta^n \frac{|B_{\mathcal{L}}| p^m}{b_1 b_2 \cdots b_m}.$$

Theorem 4.3.2 *Assume that $B \subseteq K \subseteq dB$. Then*

$$\mathbf{Pr}(\zeta \in [(1-\epsilon)vol(K), (1+\epsilon)vol(K)]) \geq \frac{3}{4} - o(1).$$

Furthermore, conditional on an event $\mathcal{G}$ of probability $1 - O(\frac{\epsilon}{n})$ the expected number of oracle calls for Algorithm VOLUME COMPUTATION *is $O(n^4 d^2 \ln d (\ln n/\epsilon)^2 \epsilon^{-2})$.*

Proof Let $\beta_i = p\rho_i$ and

$$X = \sum_{i=1}^{m+1} \ln\left(\frac{b_i}{\beta_i}\right).$$

We show

$$\mathbf{Pr}(|X| \geq \epsilon/2) \leq \frac{1}{4} + o(1). \tag{4.4}$$

Included in this calculation is the assumption that $\mathcal{G}$ occurs. Now, $\mathbf{Pr}(\mathcal{G}) = 1 - O(\frac{\epsilon}{n})$ and so what we actually prove is

$$\mathbf{Pr}(|X| < \epsilon/2) \geq \mathbf{Pr}(|X| < \epsilon/2 \mid \mathcal{G})\mathbf{Pr}(\mathcal{G}) \geq \frac{3}{4}\left(1 - O\left(\frac{\epsilon}{n}\right)\right).$$

Now $e^X\zeta = \eta^n|K_{\mathcal{L}}|$ and so $|X| < \epsilon/2$ implies

$$\mathrm{vol}(K) \in \left[\left(1 - \frac{\epsilon}{10}\right)e^{-\epsilon/2}\zeta, \left(1 + \frac{\epsilon}{10}\right)e^{\epsilon/2}\zeta\right].$$

Hence for $i = 1, 2, \ldots, m$, b_i has the binomial distribution $B(p, \alpha_i)$ where

$$\alpha_i = \mathbf{Pr}(u_i \in K_{\mathcal{L}}^{(i-1)}) = \frac{|K_{\mathcal{L}}^{(i-1)}|}{|K_{\mathcal{L}}^{(i)}|} + \epsilon_1 = \rho_i + \epsilon_1$$

where $|\epsilon_1| \leq \epsilon_0$.

It follows from (4.3) that $\alpha_i \geq 1/3, 1 \leq i \leq m$.

Applying the Chernoff bounds we obtain that with probability $1 - o(1)$

$$b_i \geq \frac{\beta_i}{\sqrt{2}} \qquad i = 1, 2, \ldots, m. \tag{4.5}$$

Set

$$A = \sum_{i=1}^{m} \frac{b_i - \beta_i}{\beta_i},\ C = \sum_{i=1}^{m} \left(\frac{b_i - \beta_i}{\beta_i}\right)^2 \text{ and } D = \sum_{i<j} \frac{(b_i - \beta_i)(b_j - \beta_j)}{\beta_i\beta_j}.$$

Using the formula for the variance of the binomial distribution we get

$$\mathbf{E}(C) = \sum_{i=1}^{m} \frac{1}{\beta_i}\left(1 - \frac{\beta_i}{p}\right) + O(m\epsilon_0) \leq \frac{\epsilon^2}{150}.$$

Now $a_{i,r}, a_{j,s}$ are independent for $r \neq s$ and Theorem 4.2.3 implies $|\mathbf{Pr}(a_{i,r} = 1 \mid a_{j,r} = 1) - \alpha_i| \leq \epsilon_0$ for arbitrary i, j, r and so

$$\mathbf{E}\left(\frac{(b_i - \beta_i)(b_j - \beta_j)}{\beta_i\beta_j}\right) \leq \frac{4p^2\epsilon_0}{\beta_i\beta_j} \leq 40\epsilon_0$$

and hence

$$\mathbf{E}(D) < 20m^2\epsilon_0 \leq \frac{\epsilon^2}{640}.$$

Claim 4.3.1 *Whenever (4.5) holds, $C < \epsilon^2/30$ and $D < \epsilon^2/64$ then we have $|X| < \epsilon/2$.*

Proof of Claim Since $A^2 = C + 2D$ we see that $|A| < \epsilon/\sqrt{15}$. If $X \geq 0$ then using the inequality $\ln x \leq x - 1$, we obtain $X \leq A < \epsilon/2$. If $X < 0$ then using (4.5) and the inequality $\ln x \geq x - 1 - (x-1)^2$ (for $x \geq 1/\sqrt{2}$) we get $X \geq A - C \geq -\epsilon/2$.
End of proof of claim

By Markov's inequality, the probability that either $C > \epsilon^2/30$ or $D > \epsilon^2/64$ is at most .25 and (4.4) follows.

We now consider the expected number of steps in the algorithm. We first remark that the algorithm requires

$$t(m+1)p = O(n^4 d^2 \ln d (\ln n/\epsilon)^2 \epsilon^{-4}) \text{ proper steps.}$$

Given $\mathcal{G}$, the expected total number of steps, proper and improper is say $\leq 5Ctmp$ for some $C > 0$. We stop the algorithm if fewer than $Ctmp$ proper steps have been made after $50Ctmp$ steps in total. Then we succeed in producing an estimate with probability at least $\frac{9}{10} - o(1)$.

We have not said anything about the size of d. Using the Ellipsoid Algorithm one can in $O^*(n^4)$ steps[5] find an affine transformation $K' = AK + b$ of K such that $B \subseteq K' \subseteq O(n^{3/2})B$. Here A is an $n \times n$ matrix and $\text{vol}(K') = \det(A)\text{vol}(K)$. So, applying the above theorem we obtain an $O^*(n^7)$ algorithm. In Section 4.4 we show how to reduce d to $O^*(n^{1/2})$ and obtain an $O^*(n^5)$ algorithm.

4.4 Putting a body into isotropic position

For a convex body K and real function f we let

$$\mathbf{E}_K(f) = \int_K f(x)dx.$$

The definition extends naturally to vectors of functions.

A body K is in *isotropic position* if its *centre of gravity*

$$b = b(K) = \mathbf{E}_K(x) = 0 \qquad \text{and} \qquad \mathbf{E}_K((v^T x)^2) = 1 \text{ for all } |v| = 1.$$

reference?

If a body is in isotropic position then it contains B and most of its volume is contained in $O(\sqrt{n})B$. This makes it a useful concept for volume approximation. It is known that for any convex body K there is an affine transformation T such that TK is in isotropic position. We only manage to obtain θ-isotropic position in this section, i.e.

$$|b(K)| \leq \theta \qquad \text{and} \qquad 1 - \theta \leq \mathbf{E}_K((v^T x)^2) \leq 1 + \theta \text{ for all } |v| = 1.$$

[5] O^* notation ignores all factors other than powers of n.

Theorem 4.4.1 *If K is in θ-nearly isotropic position then*

(a) $(1-2\theta)B \subseteq K \subseteq (1+2\theta)(n+1)B$.

(b) $vol(K \cap d_{\epsilon,\theta}B) \geq (1-\epsilon)vol(K)$ *where* $d_{\epsilon,\theta} = \left(\frac{(1+\theta)n}{\epsilon}\right)^{1/2}$.

Proof (a) .
(b) Let x be chosen randomly from K. By assumption, $\mathbf{E}(|x|^2) \leq (1+\theta)n$. Therefore, for any $d > 0$,

Can't find a simple proof of (a)

$$1 - \frac{\mathrm{vol}(K \cap dB)}{\mathrm{vol}(K)} = \mathbf{Pr}(|x| > d) \leq \frac{\mathbf{E}(|x|^2)}{d^2} \leq \frac{(1+\theta)n}{d^2}$$

and (b) follows. □

Our aim is to desribe an algorithm for finding an affine transformation A such that AK is in θ-nearly isotropic position for some small θ. **Algorithm** ISOTROPY 1
Suppose $0 < \theta, \gamma \leq 1/4$ and let

The number of samples needed has been reduced to $O(n(\log n)^3)$. This (mercifully) obviates the need for the stuff in Section 4.4.1

$$\epsilon_1 = \frac{\gamma^2\theta^2}{32(n+1)^4} \text{ and } m_1 = \left\lceil \frac{80n^2}{\theta^2\gamma^2} \right\rceil.$$

use Algorithm ORDINARY SAMPLE with $p = m_1$ and $\alpha = \epsilon_1$ to obtain $y^{(1)}, y^{(2)}, \ldots, y^{(m_1)} \in K_{\mathcal{L}}$. Compute the vector

$$\bar{y} = \frac{1}{m_1}\sum_{i=1}^{m_1} y^{(i)}$$

and the matrix

$$Y = \frac{1}{m_1}\sum_{i=1}^{m_1}(y^{(i)} - \bar{y})(y^{(i)} - \bar{y})^T.$$

If Y is singular, declare failure and repeat. Otherwise, output $K' = Y^{-1/2}(K - \bar{y})$.

Theorem 4.4.2 *With probability $\geq 1-\gamma$, the body K' is in θ-nearly isotropic position.*

Proof (Deferred to Section 4.6.2) □

The rest of this section describes how to obtain $y^{(1)}, y^{(2)}, \ldots, y^{(m_1)}$ with expected number $O^*(n^5)$ of oracle calls. Having done this, the $O^*(n^5)$ volume algorithm is immediate. It follows from Theorem 4.4.1 that $B \subseteq (1-2\theta)^{-1}(K' \cap d_{\epsilon,\theta}B) \subseteq (1-2\theta)^{-1}d_{\epsilon,\theta}B$. Applying the algorithm of Section 4.3 we obtain a good approximation to $(1-2\theta)^{-n}\mathrm{vol}(K' \cap d_{\epsilon,\theta}B)$ which by Theorem 4.4.1 is a good approximation to $(1-2\theta)^{-n}\mathrm{vol}(K') = (1-2\theta)^{-n}\det(Y^{-1/2})\mathrm{vol}(K)$. The expected number of oracle calls for the volume computation is $O(n^4 d_{\epsilon,\theta}^2 \ln d_{\epsilon,\theta}(\ln n/\epsilon)^2\epsilon^{-4}) = O^*(n^5)$.

Let us assume for the moment that

$$B \subseteq K \subseteq 10nB. \tag{4.6}$$

We will need to compute an approximation to $b(K)$.

Algorithm BARYCENTRE
Let $0 < \phi, \gamma < 1$ be given and let

$$m_2 = \left\lceil \frac{8n}{\phi\gamma} \right\rceil \text{ and } \epsilon_2 = \frac{\phi\gamma}{20(n+1)^2}.$$

Apply Algorithm ORDINARY SAMPLE with $p = m_2$ and $\alpha = \epsilon_2$ to obtain $z^{(1)}, z^{(2)}, \ldots, z^{(m_2)}$ and compute their centre of gravity g.

Theorem 4.4.3

(a) *With probability* $\geq 1 - \gamma$, $g - b(K) \in \phi(K - b(K))$.

(b) *Assume that* K *is in isotropic position. Then with probability* $\geq 1-\gamma$, $|g-b(K)| \leq \phi$.

Proof (Deferred to Section 4.6.1). □

We should mentin what the algorithm does

We may assume therefore that we have carried out Algorithm BARYCENTRE and that

$$b(K) \in -\frac{1}{10}K. \tag{4.7}$$

The algorithm for generating $y^{(1)}, y^{(2)}, \ldots, y^{(m_1)}$ in Algorithm ISOTROPY is then

(Step 1) Apply Algorithm ORDINARY SAMPLE with $p = 1$ and $\alpha = \gamma/6$ to get a single point u. This needs $O^*(n^5)$ oracle calls, in expectation.

(Step 2) Rescale K so that

$$\frac{1}{\sqrt{n}}B \subseteq K \subseteq 10\sqrt{n}B, \tag{4.8}$$

which implies

$$\mathbf{E}_K(|x|^2) \leq 100n. \tag{4.9}$$

Apply Algorithm LOCAL CONDUCTANCE (Section 4.4.1 below) to increase the local conductance of K to at least $1 - 100\theta \geq .999$. This algorithm needs $O^*(n^5)$ oracle calls, in expectation.

(Step 3) Apply Algorithm ORDINARY SAMPLE to generate $y^{(1)}, y^{(2)}, \ldots, y^{(m_1)}$. The expected number of oracle calls is $O^*(n^5)$, because we have reduced the diameter to $O(\sqrt{n})$ having used Algorithm LOCAL CONDUCTANCE to get the local conductance up to .999.

This completes the description for getting K into θ-nearly isotropic position when (4.8) is satisfied. We now show how to eliminate this assumption. Assume that $K \subseteq dB$.

Algorithm ISOTROPY 2
begin
$p = \lceil \log_2 d \rceil$, $K_0 = K$.
For $i = 0$ **to** p **do**
begin
$K_i' = K_i \cap 10nB$.
Use Algorithm ISOTROPY 1 to find with probabilty $\geq 1 - \frac{\gamma}{\log_2 d}$,
a map T_i which takes K_i' into θ-nearly isotropic position.
$K_{i+1} = T_i K_i$.
end
Output K_p.
end

Theorem 4.4.4 *With probability $\geq 1-\gamma$, the body K_p produced by Algorithm* ISOTROPY 2 *is in θ-nearly isotropic position.*

Proof Define $d_i = \max\left\{\frac{d}{2^i}, 10n\right\}$. It suffices to prove by induction on i that if all iterations are successful (which happens with probability at least $1-\gamma$) then

$$K_i \subseteq d_i B. \tag{4.10}$$

In which case, since $d_{p-1} = 10n$, $K_p = T_p T_{p-1} \cdots T_0 K$ is in θ-nearly isotropic position.

The case $i = 0$ is trivial and so let $i > 0$. Let $v \in K_i$ be the image under T_{i-1} of $u \in K_{i-1}$. If $u \in K_{i-1}'$ then v lies in $T_{i-1}K_{i-1}'$ which is in θ-nearly isotropic position and so by Theorem 4.4.1 $|v| \leq (1+2\theta)(n+1) \leq 2n$. So assume that $u \in K_{i-1} \setminus K_{i-1}'$. Let q be the point where the line segnment $[0,u]$ meets the boundary of K_{i-1}'. Let $x = T_{i-1}(0)$ and $y = T_{i-1}(q)$ where $|x|, |y| \leq 2n$. Now $u = \tau q$ where $\tau = |u|/(10n) > 1$. Since T_{i-1} is affine, it follows that $v = \tau y + (1-\tau)x$ and hence

$$|v| \leq \tau|y| + |1-\tau|\,|x| < 4n\tau < \frac{|u|}{2} \leq \frac{d}{2^i},$$

which proves (4.10). □

4.4.1 Improving Local Conductance

We assume that (4.7), (4.8) hold. We now define a *flat step*. Its aim is to improve local conductance. Suppose we carry out $\mathcal{SW}$ and we do an improper step from $u \in K$ to

$v \notin K$. We find by binary search a point $u' \in [u, v]$ such that $u' \in (2^{1/n}K) \setminus K$. We assume now that we have a *separation oracle* for K. It returns a hyperpalne H through u' such that K is contained in one of the open half-spaces of H. Let h be the unit normal of H whhich is directed away from the origin. Let

$$U_h = \left(1 - \frac{1}{2n}\right)(I + hh^T). \tag{4.11}$$

A flat step replaces K by $U_h K$ if H is at distance $< 1/2$ from the origin.

Algorithm LOCAL CONDUCTANCE
Let $0 < \phi, \gamma < 1$ be given and let

$$\delta_0 = \frac{\{\theta, \gamma\}}{24\sqrt{n}},\ d_0 = \sqrt{\frac{2n}{\theta}},\ M = \left\lceil \frac{32}{\theta} n \log n \right\rceil,\ T = \left\lceil \kappa n^2 d^2 \delta^{-2} \ln(\delta_0 \sqrt{n}) \right\rceil.$$

Step 1 Select an integer N uniformly at random from $\{0, 1, \ldots, M-1\}$.

Step 2 Using ORDINARY SAMPLE generate a point $u \in K_{\mathcal{L}}$ whose distribution is closer than $\lambda/6$ in variation distance to $U_{\mathcal{L}}$.

Step 3 Let $K_0 = K$.

for $i = 0$ **to** $N - 1$ **do**
begin
Starting at u execute $\mathcal{SW}$ on $K_i' = K_i \cap d_0 B$ until either
(i) T proper steps are made, (ii) a flat step is made.
In case (i) $K_{i+1} = K_i$ and in case (ii) $K_{i+1} = U_h K_i$.
end

Step 4 **Output** K_N.

Theorem 4.4.5 *Assume that K satisfies (4.7), (4.8) and (4.9). Then Algorithm* LOCAL CONDUCTANCE *produces a convex body K_N which also satisfies (4.7), (4.8) and (4.9). The expectation of the average local conductance of K_N is at least $1 - 100\theta$. With probability $\geq 1 - \gamma$, the number of calls to the oracle is at most $3MT = O^*(n^5)$.*

Proof [Deferred to Section 4.6.3]

4.5 Deferred proofs of Section 4.3

4.5.1 Proof of Theorem 4.2.3

(a) It follows from Theorem 4.2.2 that given v, the conditional distribution $W^{(i)}$ of w is within variation distance α of $Q_{\mathcal{L}}^{(i)}$. We must examine the distribution of $u = \frac{2n}{2n-1} w$.

Let $c = \frac{2n-1}{2n}$.

$$Q_{\mathcal{L}}^{(i)}(cK_{\mathcal{L}}^{(i)}) = \sum_{x \in cK_{\mathcal{L}}^{(i)}} \frac{|(x+\delta B_{\mathcal{L}}) \cap K_{\mathcal{L}}^{(i)}|}{|x+\delta B_{\mathcal{L}}|} \geq \sum_{x \in cK_{\mathcal{L}}^{(i)}} \frac{|(x+c\delta B_{\mathcal{L}}) \cap cK_{\mathcal{L}}^{(i)}|}{|x+c\delta B_{\mathcal{L}}|} \times \frac{|x+c\delta B_{\mathcal{L}}|}{|x+\delta B_{\mathcal{L}}|}$$

$$\geq c^n(1-c^{-1}\alpha)|cK_{\mathcal{L}}^{(i)}| \geq c^{2n}(1-c^{-1}\alpha)Q_{\mathcal{L}}^{(i)}(K_{\mathcal{L}}^{(i)}) \geq \frac{8}{15}Q_{\mathcal{L}}^{(i)}(K_{\mathcal{L}}^{(i)})$$

The first inequality in the above comes from Theorem 4.2.1. It follows that

$$W^{(i)}(cK_{\mathcal{L}}^{(i)}) \geq Q_{\mathcal{L}}^{(i)}(cK_{\mathcal{L}}^{(i)}) - \alpha \geq \frac{8}{15}Q_{\mathcal{L}}^{(i)}(K_{\mathcal{L}}^{(i)}) - \alpha \geq \frac{8}{15} - \frac{23}{15}\alpha \geq \frac{1}{2}$$

and so the expected number of executions of A to find $u \in K_{\mathcal{L}}^{(i)}$ is at most 2.

The distribution of the first point u such that $u \in cK_{\mathcal{L}}^{(i)}$ is proportional to the restriction of $W^{(i)}$ to $cK_{\mathcal{L}}^{(i)}$. Therefore, for $S \subseteq K_{\mathcal{L}}^{(i)}$,

$$V^{(i)}(S) - U_{\mathcal{L}}^{(i)}(S) = \frac{W^{(i)}(cS)}{W^{(i)}(cK_{\mathcal{L}}^{(i)})} - U_{\mathcal{L}}^{(i)}(S) \leq \frac{Q_{\mathcal{L}}^{(i)}(cS)+\alpha}{Q_{\mathcal{L}}^{(i)}(cK_{\mathcal{L}}^{(i)})-\alpha} - U_{\mathcal{L}}^{(i)}(S).$$

Here

$$Q_{\mathcal{L}}^{(i)}(cS) \leq \frac{|cS|}{\ell_{\mathcal{L}}^{(i)}(K_{\mathcal{L}}^{(i)})} \text{ and } Q_{\mathcal{L}}^{(i)}(cK_{\mathcal{L}}^{(i)}) = \frac{\ell_{\mathcal{L}}^{(i)}(cK_{\mathcal{L}}^{(i)})}{\ell_{\mathcal{L}}^{(i)}(K_{\mathcal{L}}^{(i)})} \geq \frac{(1-\alpha c^{-1})|cK_{\mathcal{L}}^{(i)}|}{\ell_{\mathcal{L}}^{(i)}(K_{\mathcal{L}}^{(i)})} > \frac{1}{2}.$$

Hence,

$$V^{(i)}(S) - U_{\mathcal{L}}^{(i)}(S) \leq \frac{|cS| + \alpha\ell_{\mathcal{L}}^{(i)}(K_{\mathcal{L}}^{(i)})}{(1-\alpha)|cK_{\mathcal{L}}^{(i)}| - \alpha\ell_{\mathcal{L}}^{(i)}(K_{\mathcal{L}}^{(i)})} - \frac{|cS|}{|cK_{\mathcal{L}}^{(i)}|} < 4\alpha.$$

To complete the proof of (a) we need to discuss the average number of actual steps for each speedy step. Consider a walk in $K_{\mathcal{L}}^{(i+1)}$ starting from u_i. Let $W^{(i)}$ be the distribution of u_i. Let

$$B^{(i)} = \left\{ x \in K_{\mathcal{L}}^{(i)} : \frac{W^{(i)}(x)}{Q_{\mathcal{L}}^{(i+1)}(x)} > 3 \right\}.$$

Now for $x \in K_{\mathcal{L}}^{(i)}$ we have

$$Q_{\mathcal{L}}^{(i+1)}(x) = \frac{\ell_{\mathcal{L}}^{(i+1)}(x)}{\ell_{\mathcal{L}}^{(i+1)}(K_{\mathcal{L}}^{(i+1)})} \geq \frac{\ell_{\mathcal{L}}^{(i)}(x)}{\ell_{\mathcal{L}}^{(i+1)}(K_{\mathcal{L}}^{(i+1)})} \geq \frac{2}{5}Q_{\mathcal{L}}^{(i)}(x) \tag{4.12}$$

where the last inequality comes from (4.3) and Lemma 4.2.1. Hence

$$B^{(i)} \subseteq \hat{B}^{(i)} = \left\{ x \in K_{\mathcal{L}}^{(i)} : \frac{W^{(i)}(x)}{Q_{\mathcal{L}}^{(i)}(x)} > \frac{6}{5} \right\}.$$

So,

$$\frac{1}{5}W^{(i)}(\hat{B}^{(i)}) \leq |W^{(i)}(\hat{B}^{(i)}) - Q_{\mathcal{L}}^{(i)}(\hat{B}^{(i)})| \leq \alpha. \tag{4.13}$$

We show next that if Z denotes the number of oracle calls per speedy step in the walk from u_i then

$$\mathbf{E}(Z \mid v_i \notin B^{(i)}) \leq \frac{4}{\ell_{\mathcal{L}}^{(i)}} \leq 5. \tag{4.14}$$

Let $u_i = x_0, x_1, \ldots,$ denote the sequence of points reached by proper steps. Let W_k denote the distribution of x_k conditional on $u_i \notin B^{(i)}$. Then for $x \in K_{\mathcal{L}}^{(i+1)} \setminus B^{(i)}$

$$W_0(x) = \frac{P^{(i)}(x)}{P^{(i)}(K_{\mathcal{L}}^{(i+1)} \setminus B^{(i)})} \leq \frac{3Q_{\mathcal{L}}^{(i+1)}(x)}{1 - 5\epsilon_0} \leq 4Q_{\mathcal{L}}^{(i+1)}(x).$$

It follows by induction that

$$W_k(x) \leq 4Q_{\mathcal{L}}^{(i+1)}(x) \qquad \text{for } k = 1, 2, \ldots$$

Given $\ell_{\mathcal{L}}^{(i+1)}(x_k)$, the expected number of steps (proper and improper) between x_k and x_{k+1} is $1/\ell_{\mathcal{L}}^{(i+1)}(x_k)$. The expected number conditional only on $v_i \notin B^{(i)}$ is thus

$$\sum_{x \in K_{\mathcal{L}}^{(i+1)}} \frac{W_k(x)}{\ell_{\mathcal{L}}^{(i+1)}(x)} \leq 4 \sum_{x \in K_{\mathcal{L}}^{(i+1)}} \frac{Q_{\mathcal{L}}^{(i+1)}(x)}{\ell_{\mathcal{L}}^{(i+1)}(x)} = \frac{4}{\lambda_{\mathcal{L}}^{(i)}}$$

proving (4.14).

We therefore define $\mathcal{G} = \left\{u_i \notin B^{(i)} : \ i = 1, 2, \ldots, m\right\}$. The proof of (b) is similar to (a). □

4.5.2 Proof of Theorem 4.2.1

We start by considering a *continuous local conductance*. For $x \in K$ we let

$$\ell(x) = \frac{\mathrm{vol}(B(x, \delta) \cap K)}{\mathrm{vol}(B(x, \delta))}$$

and the *average continuous local conductance*

$$\lambda = \frac{\lambda(K)}{\mathrm{vol}(K)}.$$

Lemma 4.5.1 *Let L be a measurable subset of the surface of a convex set K in $\mathbb{R}^n$ and let S be the set of pairs (x, y) with $x \in K, y \notin K, |x - y| \leq \delta$, and such that the line segment xy meets L. Then the $(2n)$-dimensional volume of S is at most*

$$\delta vol_{n-1}(L)\frac{c_{n-1}}{(n+1)c_n}vol_n(\delta B) \tag{4.15}$$

where c_l denotes the volume of the unit ball in $\mathbb{R}^l$.

Proof It suffices to prove the assertion for the case when L is "infinitesimally small". In this case, the measure of S is maximised when the surface of K is a hyperplane H in a larger neighbourhood of L. Then the measure of S is independent of K and is given by

$$\mathrm{vol}_{n-1}(L)\int_{\alpha=0}^{\delta}\int_{\theta=0}^{\pi}\frac{\alpha\sin\theta}{n+1}\left(\left(\frac{\delta}{\alpha}\right)^{n-1}-1\right)(n-1)c_{n-1}(\alpha\cos\theta)^{n-2}\alpha d\alpha d\theta$$

Fix the distance α of x from L and the angle θ between the line joining x and L and the hyperplane H. The volume of the corresponding y's is $\mathrm{vol}_{n-1}(L)\frac{\alpha\sin\theta}{n+1}\left(\left(\frac{\delta}{\alpha}\right)^{n-1}-1\right)$ – the volume of the part of a cone on the y-side of H. Now multiply by $(n-1)c_{n-1}(\alpha\cos\theta)^{n-2}$ – the $(n-2)$-volume swept out by x the surface of an $n-1$ dimensional ball of radius $\alpha\cos\theta$. Then integrate over α, θ. □

Corollary 4.5.1 *Let K and L be as in Lemma 4.5.1. Choose x uniformly from K and choose u uniformly from δB. The probability that $[x, x + u]$ intersects L is at most*

$$\frac{\delta vol_{n-1}(L)}{2\sqrt{n}vol(K)}.$$

Proof Divide (4.15) by $\mathrm{vol}(K) \times \mathrm{vol}(\delta B)$ and use the fact that $c_n/c_{n-1} > 2/\sqrt{n}$ for $n > 2$. □

The average local conductance λ thus satisfies

$$\lambda \geq 1 - \frac{\delta}{2\sqrt{n}}\frac{\mathrm{vol}_{n-1}(\partial K)}{\mathrm{vol}(K)}.$$

If $K \supseteq B$ then $\mathrm{vol}(K) \geq \mathrm{vol}_{n-1}(\partial K)/n$ and so

$$\lambda \geq 1 - \frac{\delta\sqrt{n}}{2}. \tag{4.16}$$

We now need to relate λ to $\lambda_{\mathcal{L}}$. We first prove Theorem 4.3.1.

Proof of Theorem 4.3.1.

Let $S_{\mathcal{L}}^I = \{x \in S_{\mathcal{L}} : C(x) \subseteq S\}$ and let $S_{\mathcal{L}}^B = S_{\mathcal{L}} \setminus S_{\mathcal{L}}^I$ denote the interior and border Needs fixing!

points of $S_{\mathcal{L}}$ respectively. $S \supseteq \alpha B$ implies that

$$(1+2\alpha^{-1}\eta\sqrt{n})^{-1}\bigcup_{x\in S_{\mathcal{L}}} C(x) \subseteq S \subseteq (1+\alpha^{-1}\eta\sqrt{n})\bigcup_{x\in S^I_{\mathcal{L}}} C(x). \tag{4.17}$$

The theorem follows easily from this. □

Note that (4.17) implies

$$\begin{aligned} |S^B_{\mathcal{L}}| &\leq \eta^{-n}\mathrm{vol}(S)((1+2\alpha^{-1}\eta\sqrt{n})^n - (1+\alpha^{-1}\eta\sqrt{n})^{-n}) \\ &\leq 4\alpha^{-1}\eta^{-(n-1)}\sqrt{n}\mathrm{vol}(S). \end{aligned} \tag{4.18}$$

Lemma 4.5.2

(a) *$x \in K$ implies that $\ell(x) \geq \left(\frac{\delta}{10d}\right)^n$.*

(b) *$x, x' \in K$ and $|x - x'| \leq \eta\sqrt{n}$ implies $|\ell(x) - \ell(x')| \leq \frac{\epsilon}{n}\ell(x)$.*

(c) *$x \in K_{\mathcal{L}}$ implies $|\ell(x) - \ell_{\mathcal{L}}(x)| \leq \frac{\epsilon}{n}\ell(x)$.*

Note: the estimates $\frac{\epsilon}{n}$ are much larger than we will actually prove.

Proof
(a) Consider the finite cone C with point x and base the intersection of B with the hyperplane through the origin O which is perpendicular to the line L joining x to O. C contains a ball of radius $\frac{\delta}{10}$ with centre on L, at distance $\frac{\delta}{2}$ from x.

(b)

$$\mathrm{vol}(B(x,\delta) \setminus B(x',\delta)) \leq (1-(1-\delta^{-1}\eta\sqrt{n})^n)\mathrm{vol}(B(x,\delta) \leq 2\delta^{-1}n^{3/2}\eta\delta^n.$$

Now use (a).

(c) This follows from (b) and Theorem 4.3.1, taking account of the note prior to the proof. □

Theorem 4.2.1 follows.

4.5.3 Proof of Theorem 4.2.2

Geometric lemmas related to local condutance

The following classic theorem is basic to the study of convexity:

Theorem 4.5.1 Brunn-Minkowski Theorem *Let K_1, K_2 be convex bodies in $\mathbb{R}^n$. Then*

$$vol(K_1 + K_2)^{1/n} \geq vol(K_1)^{1/n} + vol(K_2)^{1/n}.$$

Corollary 4.5.2 *Let K_1, K_2 be convex bodies. Then the function $f(x) = vol((x + K_1) \cap K_2)^{1/n}$ is concave.*

Proof This follows from Theorem 4.5.1 and

$$(\lambda x + (1-\lambda)y + K_1) \cap K_2 \supseteq \lambda((x + K_1) \cap K_2) + (1-\lambda)((y + K_1) \cap K_2).$$

□

In the remainder of this subsection x, y are members of K and $|x - y| < \delta/\sqrt{n}$. Let

$$C = (x + \delta B) \cap (y + \delta B),\ M_x = (x + \delta B) \setminus C \text{ and } M_y = (y + \delta B) \setminus C,$$

$$R_x = M_x \cap (x - y + C) \text{ and } R_y = M_y \cap (y - x + C).$$

Let C' be obtained by blowing up C from its centre $(x + y)/2$ by a factor $\rho = 1 + \frac{4}{4n-1}$.

Lemma 4.5.3

$$M_x \setminus R_x \subseteq C'.$$

Proof Assume w.l.o.g. that $x = -y$ and let $z = \mu x + w \in M_x \setminus R_x$ where w is orthogonal to x. It can be seen that $0 < \mu < 2$. Now $|\rho^{-1}z - x| \leq |\rho^{-1}z - y|$ and so it is enough to show that $|\rho^{-1}z - y| \leq \delta$. This follows by straightforward calculation. □ More needed

Lemma 4.5.4

$$vol(K \cap (M_x \setminus R_x)) \leq 3vol(K \cap C).$$

Proof By Lemma 4.5.3, blowing up C by a factor ρ covers both $K \cap C$ and $K \cap (M_x \setminus R_x)$. Hence

$$\text{vol}(K \cap (C \cup (M_x \setminus R_x))) \leq \text{vol}(K \cap C') \leq \left(1 + \frac{4}{4n-1}\right)^n \text{vol}(K \cap C) \leq 3\text{vol}(K \cap C)$$

and the lemma follows. □

We say that a real-valued function $f(x)$ defined on the convex set $K \subseteq \mathbb{R}^n$ is *log-concave* if $\log f(x)$ is concave on K. This clearly entails $f(x) > 0$ on K.

In particular:

$$\text{If } f(x)^\alpha \text{ is concave, for some } \alpha > 0, \text{ then } f \text{ is itself log-concave.} \tag{4.19}$$

Lemma 4.5.5

$$vol(K \cap C)^2 \geq vol(K \cap R_x)vol(K \cap R_y).$$

Proof Corollary 4.5.2 and (4.19) imply that the function $g(u) = \mathrm{vol}(K \cap (u + C))$ is log-concave. Therefore

$$g(0)^2 \geq g(x-y)g(y-x) = \mathrm{vol}(((x-y)+C) \cap K)\mathrm{vol}(((y-x)+C) \cap K) \\ \geq \mathrm{vol}(K \cap R_x)\mathrm{vol}(K \cap R_y).$$

□

Lemma 4.5.6

$$vol(K \cap C) \geq \frac{1}{5} \min \{vol(K \cap (x + \delta B)), vol(K \cap (y + \delta B))\}.$$

Proof We have

$$\mathrm{vol}(K \cap R_x) = \mathrm{vol}(K \cap M_x) - \mathrm{vol}(K \cap (M_x \setminus R_x)) \\ \geq \mathrm{vol}(K \cap (x + \delta B)) - \mathrm{vol}(K \cap C) - 3\mathrm{vol}(K \cap C)$$

by Lemma 4.5.4. We also get a symmetric lower bound for $\mathrm{vol}(K \cap R_y)$. Then by Lemma 4.5.5 we have

$$\mathrm{vol}(K \cap C) \geq \min \{\mathrm{vol}(K \cap R_x), \mathrm{vol}(K \cap R_y)\} \\ \geq \min \{\mathrm{vol}(K \cap (x + \delta B)), \mathrm{vol}(K \cap (y + \delta B))\} - 4\mathrm{vol}(K \cap C).$$

The lemma follows. □

Lemma 4.5.7 *Suppose S_1, S_2 is a partition of K into two measurable sets where $x \in S_1$, $y \in S_2$ and $|x - y| \leq \delta/\sqrt{n}$. Then*

$$\frac{vol((x + \delta B) \cap S_2)}{vol(\delta B)} + \frac{vol((y + \delta B) \cap S_1)}{vol(\delta B)} \geq \frac{1}{5} \min \{\ell(x), \ell(y)\}. \tag{4.20}$$

Proof The LHS Λ of (4.20) is at least

$$\frac{1}{\mathrm{vol}(\delta B)}(\mathrm{vol}(S_1 \cap C) + \mathrm{vol}(S_2 \cap C)) = \frac{\mathrm{vol}(K \cap C)}{\mathrm{vol}(\delta B)}.$$

Thus by Lemma 4.5.6,

$$\Lambda \geq \frac{1}{5\mathrm{vol}(\delta B)} \min \{\mathrm{vol}(K \cap (x + \delta B)), \mathrm{vol}(K \cap (y + \delta B))\} = \frac{1}{5} \min \{\ell(x), \ell(y)\}.$$

□

Geometric lemmas for the main argument

Now, for any $a \in \mathbb{R}^n$, $|a| = 1$, consider the set of hyperlanes $H(s) = \{ax = s\}$ orthogonal to a, and half-spaces $H^+(s) = \{ax \leq s\}$, $H^-(s) = \{ax \geq s\}$ they define. If K is any convex body, let $K(s) = K \cap H(s)$, $K^+(s) = K \cap H^+(s)$, $K^-(s) = K \cap H^-(s)$. (We call $K(s)$ a "cross section" of K in "direction" a.) Let $\beta_1 = \inf_s\{K(s) \neq \emptyset\}$, $\beta_2 = \sup_s\{K(s) \neq \emptyset\}$. Then $w = \beta_2 - \beta_1$ is the *width* of K in direction a, and we will write $w = W(K, a)$. Note that

Lemma 4.5.8 *diameter* $K = \max_a W(K, a)$.

Proof

$$\begin{aligned}\text{diameter } K &= \max\{|x - y| : x, y \in K\} = \max\{|z| : z \in K - K\} \\ &= \max_{z \in K-K} \max_{|a|=1} az = \max_{|a|=1} \max_{z \in K-K} az = \max_a W(K, a).\end{aligned}$$

□

We will also need the following technical result.

Lemma 4.5.9 *Let* $a_1, a_2, \ldots, a_{n-1}$ *be mutually orthogonal unit vectors and suppose that* $a \in LIN(a_1, a_2, \ldots, a_{n-1})$. *Then diameter* $K(s) < n^{1/2} \max_i W(K, a_i)$ *for all* s.

Proof If $a, |a| = 1$ is in the subspace generated by the a_i then $W(K(s), a) \leq W(K, a)$. But $W(K, a) \leq \sqrt{n-1} \max_i W(K, a_i)$, since K can clearly be contained in an (infinite) cubical cylinder of side $\max_i W(K, a_i)$. Applying Lemma 4.5.8 now gives the conclusion. □

Let $\alpha(s) = \text{vol}_{n-1}(K(s))$ and $V(s) = \text{vol}_n(K^+(s))$, and assume, without loss, that $\beta_1 = 0$ and $\beta_2 = w$. Note then $V(w) = \text{vol}_n(K)$. It is a consequence of the Brunn-Minkowski theorem, that $\alpha(s)^{1/(n-1)}$ is a concave function of s in $[0, w]$. Then we have

Lemma 4.5.10 $V(s)/V(w) \leq ns/w$.

Proof First we show that if $0 < u < s$, $\alpha(u)/\alpha(s) \geq (u/s)^{n-1}$. This follows since if $u = \lambda 0 + (1 - \lambda)s$, then Brunn-Minkowski implies

$$\begin{aligned}\alpha(u)^{1/(n-1)} &\geq \lambda\alpha(0)^{1/(n-1)} + (1 - \lambda)\alpha(s)^{1/(n-1)} \\ &\geq (1 - \lambda)\alpha(s)^{1/(n-1)} = (u/s)\alpha(s)^{1/(n-1)}.\end{aligned}$$

Thus

$$V(s) \geq \int_0^s (u/s)^{n-1}\alpha(s)\, du = (s/n)\alpha(s), \tag{4.21}$$

$$V(w) - V(s) \leq \int_s^w (u/s)^{n-1}\alpha(s)\, du = (w^n - s^n)/(ns^{n-1})\alpha(s). \tag{4.22}$$

Dividing (4.22) by (4.21) gives $V(s)/V(w) \geq (s/w)^n$. By symmetry, this inequality in turn implies

$$(V(w) - V(s))/V(w) \geq ((w-s)/w)^n = (1 - s/w)^n \geq 1 - ns/w,$$

since $(1-x)^n \geq 1 - nx$ for $x \in [0,1]$. This gives the result. □

We will need the following simple lemma asserting the existence of a hyperplane simultaneously "bisecting the measure" of two arbitrary sets.

Lemma 4.5.11 *Let $S_1, S_2 \subseteq \mathbb{R}^n$, be measurable and L a two-dimensional linear subspace of $\mathbb{R}^n$. Let f be continuous on K. Then there exists a hyperplane H, with normal $a \in L$, such that the half-spaces H^+, H^- determined by H satisfy $f(S_i \cap H^+) = f(S_i \cap H^-)$ for $i = 1, 2$.*

Proof Let α_1, α_2 be a basis for L. For each $\theta \in [-1, +1]$, let $b_i(\theta)$ be such that the hyperplane $(\theta\alpha_1 + (1 - |\theta|)\alpha_2)x = b_i(\theta)$ bisects the f-measure of S_i for $i = 1, 2$. (If S_i is disconnected in such a way that the possible b_i form an interval, $b_i(\theta)$ will be its midpoint.) It clearly suffices to show that $b_1(\theta_0) = b_2(\theta_0)$ for some θ_0. If $b_1(-1) = b_2(-1)$ we are done, so suppose w.l.o.g. that $b_1(-1) > b_2(-1)$. We clearly have $b_i(1) = -b_i(-1)$ for $i = 1, 2$, so $b_1(1) < b_2(1)$. But since f is a continuous measure, it follows easily that $b_i(\theta)$ is a continuous function of θ. The existence of $\theta_0 \in (-1, 1)$ now follows. □

Three lemmas on logconcavity

Lemma 4.5.12 *Let $f : \mathbb{R}_+ \to \mathbb{R}_+$ be log-concave and let $F(x) = \int_0^x f(t)dt$. Then F is also log-concave.*

Proof It suffices to show that for $0 < a < b$ and $c = (a+b)/2$ that $F(c)^2 \geq F(a)F(b)$. Now

$$\Delta = F(c)^2 - F(a)F(b) = B(A + B) - AC$$

where $A = \int_0^a f(t)dt$, $B = \int_a^c f(t)dt$ and $C = \int_c^b f(t)dt$.

Let $g(x) = Ge^{\theta x}$ where $\theta = (c-a)^{-1}\log(f(c)/f(a))$ and $G = f(a)e^{-\theta a}$ so that $g(x) = f(x)$ for $x = a, c$. If $f(c) \leq f(a)$ then $C \leq B$ and then clearly $\Delta \geq 0$. We can therefore assume that $f(c) > f(a)$ and hence that $\theta > 0$.

The log-concavity of f implies that

$$f(x) \geq g(x) \text{ for } x \in [a, c] \text{ and } f(x) \leq g(x) \text{ for } x \notin [a, c].$$

Thus

$$\Delta \geq \hat{B}(A + \hat{B}) - A\hat{C}$$

where $\hat{B} = \int_a^c g(t)dt = G\theta^{-1}(e^{\theta c} - e^{\theta a})$, $\hat{C} = \int_c^b g(t)dt = G\theta^{-1}(e^{\theta b} - e^{\theta c})$.
We can therefore prove the lemma by showing that

$$\hat{B} + \frac{\hat{B}^2}{\hat{A}} \geq \hat{C} \tag{4.23}$$

where $\hat{A} = \int_0^a g(t)dt = G\theta^{-1}(e^{\theta a} - 1) \geq A$. But (4.23) is equivalent to

$$(e^{\theta a} - 1)(e^{\theta b} - e^{\theta c}) \leq (e^{\theta a} - 1)(e^{\theta c} - e^{\theta a}) + (e^{\theta c} - e^{\theta a})^2$$

or after simplification,

$$e^{\theta a}e^{\theta b} + 2e^{\theta c} \leq e^{2\theta c} + e^{\theta a} + e^{\theta b}$$

which follows from $e^{\theta a}e^{\theta b} = e^{2\theta c}$ and the convexity of $e^{\theta x}$. □

Corollary 4.5.3 *Let f, F be as in Lemma 4.5.12. Let $0 \leq x \leq d$ and $0 \leq t \leq d - x$. Then*

$$F(x + t) - F(x) \geq \frac{t}{d} F(x) \log\left(\frac{F(d)}{F(x)}\right).$$

Proof Let $\tilde{F}(x) = F(x)/F(d)$. Lemma 4.5.12 implies that $\tilde{F}$ is log-concave. Write $x + t = \lambda x + (1 - \lambda)d$ where $\lambda = \frac{d-x-t}{d-x}$. Then the log-concavity of $\tilde{F}$ implies

$$\tilde{F}(x + t) \geq \tilde{F}(x)^{1-t/(d-x)}$$

and so

$$\begin{aligned}
\tilde{F}(x + t) - \tilde{F}(x) &\geq \tilde{F}(x)(\tilde{F}(x)^{-t/(d-x)} - 1) \\
&= \tilde{F}(x)\left(\exp\left\{\frac{t}{d-x}\log\left(\frac{1}{\tilde{F}(x)}\right)\right\} - 1\right) \\
&\geq \tilde{F}(x) \cdot \frac{t}{d-x}\log\left(\frac{1}{\tilde{F}(x)}\right)
\end{aligned}$$

and the lemma follows. □

Lemma 4.5.13 *The local conductance ℓ is a log-concave function on K.*

Proof This follows from Corollary 4.5.2 and (4.19). □

The main argument

Let S_1, S_2 be a partition of K into two measurable sets. Let

$$h(x) = \begin{cases} \frac{\text{vol}((x+\delta B)\cap S_2)}{\text{vol}(\delta B)} & x \in S_1 \\ \\ \frac{\text{vol}((x+\delta B)\cap S_1)}{\text{vol}(\delta B)} & x \in S_2 \end{cases}$$

Then

$$\frac{h(S_1)}{\ell(K)} = \int_{x\in S_1} \frac{\ell(x)}{\ell(K)} \cdot \frac{\text{vol}((x+\delta B)\cap S_2)}{\text{vol}((x+\delta B)\cap K)} dx$$

is the probability $Q(S_1, S_2)$ that $x \in S_1$ and $y \in S_2$ where x is a point chosen from K with distribution $\ell(x)/\ell(K)$ and y is obtained from x by one *continuous* speedy step. It follows as in (2.3) that $h(S_1) = h(S_2)$. The conclusion of Lemma 4.5.7 is that

$$h(x) + h(y) \geq \frac{1}{5} \min\{\ell(x), \ell(y)\} \text{ for } x \in S_1, y \in S_2, |x-y| \leq \delta/\sqrt{n}. \tag{4.24}$$

Theorem 4.2.2 will follow from

Theorem 4.5.2 *Let $B \subseteq K \subseteq dB$ be a closed convex set in $\mathbb{R}^n$. Let K be partitioned into two measurable subsets S_1, S_2. Let ℓ be a log-concave function which is strictly positive on K. Let h be a non-negative integrable function which satisfies (4.24). Then*

$$\frac{h(K)\ell(K)}{2\ell(S_1)\ell(S_2)} \geq \frac{1}{5000\sqrt{n}} \min\left\{ \frac{\delta}{d} \log\left(\frac{\ell(K)^2}{\ell(S_1)\ell(S_2)} \right), 1 \right\}. \tag{4.25}$$

Proof We first prove

$$h(K) \geq \frac{1}{2500\sqrt{n}} \min\left\{ \frac{\delta}{d} \max_{i=1,2} \left\{ \ell(S_i) \log\left(\frac{\ell(K)}{\ell(S_i)} \right) \right\}, \ell(S_1), \ell(S_2) \right\}. \tag{4.26}$$

Let $\ell_0 = \min\{\ell(x) : x \in K\} > 0$. We let $\epsilon = \min\left\{ \frac{\ell_0}{100}, \frac{\delta}{\sqrt{n}} \right\}$ and then let $\delta_1 \leq \delta/(10\sqrt{n})$ be such that

$$|\ell(x) - \ell(x')| \leq \epsilon \text{ whenever } x, x' \in K, |x - x'| \leq \delta_1.$$

We fix a line $L = \{x_s = a + su : s \in \mathbb{R}\}$ in direction u, where $a, u \in \mathbb{R}^n, |u| = 1$.

Let $\alpha(s), K(s)$ be as in the previous section and let $I = \{s : K(s) \neq \emptyset\} = [\beta_1, \beta_2]$. Let $\bar{h}(s) = h(K(s))/\alpha(s)$ be the average of h over $K(s)$.

We consider first the case where K is *needle-like* i.e. each $K(s)$ has diameter at most $\delta_1/2$.

It follows that $\ell(x) \in [.99\ell(x_s), 1.01\ell(x_s)]$ for $x \in K(s)$.

Let $H = \left\{ s \in I : \bar{h}(s) \geq \frac{1}{30}\ell(x_s) \right\}$. Let $J_i = \{s \in I \setminus H : \text{vol}_{n-1}(K(s) \cap S_i) \geq 2\alpha(s)/5\}$.

Claim 4.5.1

(a) *If* $s \in I \setminus H$ *then* $\min_{i=1,2}\{vol_{n-1}(K(s) \cap S_i)\} \le \frac{1}{3}\alpha(s)$.

(b) $s \in J_i$ *and* $y \in S_{3-i}$ *implies* $h(y) \ge \frac{1}{12}\ell(x_s)$.

Proof (a) Suppose $s \in I \setminus H$ and that $\mathrm{vol}_{n-1}(K(s) \cap S_i) \ge 2\alpha(s)/5$ for $i = 1, 2$. Choose $x \in K(s) \cap S_1$, $y \in K(s) \cap S_2$. Then we have

$$h(x) + h(y) \ge \frac{1}{5}\min\{\ell(x), \ell(y)\} \ge \frac{1}{6}\ell(x_s). \tag{4.27}$$

If $h(x) \ge \frac{1}{12}\ell(x_s)$ for all $x \in K(s) \cap S_1$ then $\bar{h}(s) \ge \frac{1}{30}\ell(x_s)$, contradiction. Otherwise it follows that $h(y) \ge \frac{1}{12}\ell(x_s)$ for all $y \in K(s) \cap S_2$ and we get a similar contradiction.
(b) If $h(y) < \frac{1}{12}\ell(x_s)$ then (4.27) implies $h(x) \ge \frac{1}{12}\ell(x_s)$ for all $x \in S_i \cap K(s)$, which implies $\bar{h}(s) \ge \frac{1}{30}\ell(x_s)$, contradiction.
End of proof of Claim 4.5.1

It is clear that $J_1 \cup J_2 = I \setminus H$ and it follows from Claim 4.5.1 that J_1 and J_2 are disjoint.

Now let $\mu(s) = \ell(x_s)\alpha(s)$ for $s \in I$. Then

$$\int_{s\in H} \bar{h}(s)\alpha(s)ds \ge \frac{1}{30}\int_{s\in H} \ell(x_s)\alpha(s)ds = \frac{1}{30}\mu(H). \tag{4.28}$$

On the other hand

$$\int_{s\in H} \bar{h}(s)\alpha(s)ds \le h(K). \tag{4.29}$$

Now for $i = 1, 2$,

$$\ell(S_i) = \int_{s\in J_i}\int_{x\in K(s)\cap S_i} \ell(x)dx + \int_{s\in J_{3-i}}\int_{x\in K(s)\cap S_i} \ell(x)dx + \int_{s\in H}\int_{x\in K(s)\cap S_i} \ell(x)dx$$
$$\le 1.01\mu(J_i) + 12h(K) + 1.01\mu(H). \tag{4.30}$$

The term $12h(K)$ is a consequence of Claim 4.5.1.
Similarly,

$$\ell(S_i) \ge \frac{99}{100}\mu(J_i) \text{ and } \ell(K) \le \frac{101}{100}\mu(I). \tag{4.31}$$

I don't know where I got $\frac{99}{100}$ from. $\frac{1}{2}$ is clear. It affects constants, maybe

If (4.26) fails then (4.30) gives

$$\ell(S_i) \le 1.02(\mu(J_i) + \mu(H))$$

and then together with (4.28), (4.29) we obtain

$$\mu(H) < \frac{31}{2500\sqrt{n}}\min\left\{\frac{\delta}{n}\max_{i=1,2}\left\{(\mu(J_i) + \mu(H))\log\left(\frac{101\mu(I)}{99\mu(J_i)}\right)\right\},\right.$$
$$\left.\mu(J_1) + \mu(H), \mu(J_2) + \mu(H).\right\}$$

We see immediately that $\mu(H) = o(\mu(J_i))$ so that $\mu(J_i) + \mu(H) = (1 + o(1))\mu(J_i)$ for $i = 1, 2$ and then dividing through by $1 - o(1)$ we can write

$$\mu(H) < (1 + o(1))\frac{31}{2500\sqrt{n}} \min\left\{\frac{\delta}{d}\max_{i=1,2}\left\{\mu(J_i)\log\left(\frac{101\mu(I)}{99\mu(J_i)}\right)\right\}, \mu(J_1), \mu(J_2)\right\} \quad (4.32)$$

Assuming $\mu(J_1) \le \mu(J_2)$ we have

$$\mu(J_1)\log\left(\frac{101\mu(I)}{99\mu(J_1)}\right) \le \mu(J_1)\log\left(\frac{\mu(I)}{\mu(J_1)}\right) \times \frac{\log(202/99)}{\log 2} \le 1.03\mu(J_1)\log\left(\frac{\mu(I)}{\mu(J_1)}\right) \quad (4.33)$$

and

$$\mu(J_2)\log\left(\frac{101\mu(I)}{99\mu(J_2)}\right) \le \mu(J_2)\log(202/99) < \mu(J_2). \quad (4.34)$$

Using (4.33), (4.34) in (4.32), we obtain

$$\mu(H) < \frac{1}{80\sqrt{n}} \min\left\{\frac{\delta}{d}\max_{i=1,2}\left\{\mu(J_i)\log\left(\frac{\mu(I)}{\mu(J_i)}\right)\right\}, \mu(J_1), \mu(J_2)\right\}. \quad (4.35)$$

We first dismiss the *degenerate* case where, say, $\mu(J_1) = 0$. It follows from (4.28–4.30) that $h(K) \ge \frac{1}{43}\ell(S_1)$ and the theorem is clearly true.

Claim 4.5.2 *If $s \in J_1$ and $t \in J_2$ then $|s - t| > \delta_1$.*

Proof $\bar{h}(s) < \frac{1}{30}\ell(x_s)$ implies that there exists $x \in S_1 \cap K(s)$ such that $h(x) \le \frac{1}{18}\ell(x_s)$. Similarly there exists $y \in S_2 \cap K(t)$ such that $h(y) \le \frac{1}{18}\ell(x_t)$. If $|s - t| \le \delta_1$ then $|x - y| \le \frac{\delta}{\sqrt{n}}$ and so $h(x) + h(y) \ge \frac{1}{5}\ell(x) \ge \frac{1}{6}\ell(x_s)$, assuming that $\ell(x) \le \ell(y)$. It follows that $\ell(x_t) \ge 2\ell(x_s)$ and so $|\ell(x_s) - \ell(x_t)| \ge \ell_0$, contradicting $|x_s - x_t| \le \delta_1$.
End of proof of Claim 4.5.2

We now show that we can assume w.l.o.g. the existence of an interval $(\sigma, \tau) \subseteq H$ such that if $A_1 = [\beta_1, \sigma]$ and $A_2 = [\tau, \beta_2]$ then

$$\mu(A_i \cap J_i) \ge \frac{1}{2}\mu(J_i),\ i = 1, 2. \quad (4.36)$$

Let

$$b_1 = \inf\left\{s :\ \mu(J_i \cap [\beta_1, s]) \ge \frac{1}{2}\mu(J_i),\ i = 1, 2\right\}.$$

Assume w.l.o.g. that

$$\mu(J_1 \cap [\beta_1, s]) \ge \frac{1}{2}\mu(J_1) \text{ and } \mu(J_2 \cap [s, \beta_2]) \ge \frac{1}{2}\mu(J_2).$$

It follows from Claim 4.5.2 that $b_1 \in J_1 \cup H$. Let

$$b_2 = \inf([b_1, \beta_2] \cap J_2).$$

If $b_2 \geq \beta_2$ we are in the degenerate case dealt with following (4.35). Let

$$b_3 = \sup([b_1, b_2] \cap J_1)$$

and let $\sigma = b_3$ and $\tau = b_2$. Equation (4.36) is satisfied and Claim 4.5.2 implies that $\tau - \sigma \geq \delta_1$.

Suppose now that $\tau - \sigma \geq \frac{2\delta}{3\sqrt{n}}$. The Brunn-Minkowski Theorem implies that $\alpha^{1/(n-1)}$ is concave and so α is log-concave. It follows that μ is also log-concave. Applying Corollary 4.5.3 (twice) we have

$$\mu(H) \geq \max_{i=1,2} \left\{ \frac{2\delta}{3d\sqrt{n}} \mu(A_i) \log \left(\frac{\mu(I)}{\mu(A_i)} \right) \right\}. \tag{4.37}$$

Suppose next that $\xi = \tau - \sigma < \frac{2\delta}{3\sqrt{n}}$. Let $\lambda(s) = \ell(x_\sigma) e^{-\theta(s-\sigma)/\xi}$ where $e^\theta = \ell(x_\sigma)/\ell(x_\tau)$, so that $\lambda(s) = \ell(x_s)$ for $s = \sigma, \tau$. The log-concavity of ℓ implies that $\lambda(s) \leq \ell(x_s)$ for $s \in H$ and $\lambda(s) \geq \ell(x_s)$ for $s \notin H$. Our aim is to find a contradiction to (4.35) and so we can assume in fact that $\ell(x_s) = \lambda(s)$ for $s \in I$.

Suppose that $\mu(\tau) \geq (1+\zeta)\mu(\sigma)$ for some $\zeta > 0$. The log-concavity of μ implies that $\mu(s) \geq \mu(\sigma)$ for $s \in H$ and that $\mu(\sigma - t) \leq (1+\zeta)^{-t/\xi}\mu(\sigma)$ for $t > 0$. But then

$$\mu(A_1) \leq \mu(\sigma)(\log(1+\zeta))^{-1}\xi \text{ and } \mu(H) \geq \mu(\sigma)\xi.$$

This implies

$$\mu(H) \geq \frac{1}{10\sqrt{n}} \mu(A_1) \tag{4.38}$$

if $\zeta \geq \frac{1}{5\sqrt{n}}$. Using the same argument when $\mu(\sigma) \geq (1+\zeta)\mu(\tau)$ we can now assume that

$$\left| \frac{\mu(\sigma)}{\mu(\tau)} - 1 \right| \leq \frac{1}{10\sqrt{n}}. \tag{4.39}$$

It follows immediately that

$$\mu(H) \geq \frac{\mu(\sigma)\xi}{2}. \tag{4.40}$$

We choose $u \in J_1 \cap [\sigma - \delta_1, \sigma]$ and $v \in J_2 \cap [\tau, \tau + \delta_1]$ and argue as in Claim 4.5.2 to prove that

$$\frac{\ell(x_\sigma)}{\ell(x_\tau)} > 3/2, \tag{4.41}$$

which implies

$$\theta \geq \log 3/2.$$

Now let η be a super-gradient of $\alpha^{1/(n-1)}$ at the point σ i.e. $\alpha(\sigma - t)^{1/(n-1)} \leq \alpha(\sigma)^{1/(n-1)} - \eta t$ for $t \in \mathbb{R}$. Then

$$\mu(\sigma - t) \leq (\alpha(\sigma)^{1/(n-1)} - \eta t)^{n-1} \ell(x_\sigma) e^{\theta t/\xi}$$

for $t \in \mathbb{R}$.

(4.39) and (4.41) imply that $\alpha(\tau) > \alpha(\sigma)$ and so $\eta > 0$. Now, putting $\alpha_0 = \alpha(\sigma)^{1/(n-1)}$, using $1 - x \le e^{-x-x^2}$ for $0 \le x \le 1$ and noting that necessarily $\alpha_0 \ge (\sigma - \beta_1)\eta$, we obtain

$$\begin{aligned}\mu(A_1) &\le \ell(x_\sigma)\int_0^{\sigma-\beta_1}(\alpha_0-\eta t)^{n-1}e^{\theta t/\xi}dt\\ &\le \mu(\sigma)\int_0^{\sigma-\beta_1}\exp\left\{\frac{\theta t}{\xi}-(n-1)\left(\frac{\eta t}{\alpha_0}+\frac{\eta^2t^2}{\alpha_0^2}\right)\right\}dt. \end{aligned} \tag{4.42}$$

Now

$$\begin{aligned}\eta\xi &\ge \alpha(\tau)^{1/(n-1)}-\alpha(\sigma)^{1/(n-1)}\\ &= \alpha_0\left(\left(\frac{\alpha(\tau)}{\alpha(\sigma)}\right)^{1/(n-1)}-1\right) \ge \alpha_0\left(\left(\frac{\ell(x_\sigma)}{\ell(x_\tau)}\left(1-\frac{1}{4\sqrt{n}}\right)\right)^{1/(n-1)}-1\right)\\ &\ge \alpha_0\left(\exp\left\{\frac{\theta}{n-1}-\frac{1}{3n^{3/2}}\right\}-1\right) \ge \alpha_0\frac{\theta}{n-1}\left(1-\frac{1}{2\sqrt{n}}\right).\end{aligned}$$

So

$$\frac{\eta}{\alpha_0} \ge \frac{\theta}{(n-1)\xi}\left(1-\frac{5\delta}{\sqrt{n}}\right).$$

Going back to (4.42) we obtain

$$\mu(A_1) \le \mu(\sigma)\int_{-\infty}^{\infty}\exp\left\{-\frac{\theta^2t^2}{2\xi^2(n-1)}+\frac{\theta t}{2\xi\sqrt{n}}\right\}dt = (1+o(1))\xi\mu(\sigma)\sqrt{\pi n/(2\theta^2)},$$

Comparing with (4.40) we see that

$$\mu(H) \ge (1-o(1))\frac{\theta}{\sqrt{\pi n}}\mu(A_1). \tag{4.43}$$

It follows from (4.36), (4.37), (4.38) and (4.43) that

$$\begin{aligned}\mu(H) &\ge \frac{1}{10\sqrt{n}}\min\left\{\frac{\delta}{d}\max_{i=1,2}\left\{\mu(A_i)\log\left(\frac{\mu(I)}{\mu(A_i)}\right)\right\},\mu(A_1),\mu(A_2)\right\}\\ &\ge \frac{1}{20\sqrt{n}}\min\left\{\frac{\delta}{d}\max_{i=1,2}\left\{\mu(A_i)\log\left(\frac{\mu(I)}{\mu(A_i)}\right)\right\},\mu(J_1),\mu(J_2)\right\}\end{aligned} \tag{4.44}$$

Now for $i = 1, 2$,

$$\mu(A_i) \le \mu(J_i) \text{ implies } \mu(A_i)\log\left(\frac{\mu(I)}{\mu(A_i)}\right) \ge \frac{1}{2}\mu(J_i)\log\left(\frac{\mu(I)}{\mu(J_i)}\right). \tag{4.45}$$

$x \log x^{-1}$ has a unique maximum over $x \in [0,1]$ at e^{-1} and so

$$\mu(J_i) \le \mu(A_i) \le \frac{9}{10}\mu(I) \text{ implies } \mu(A_i)\log\left(\frac{\mu(I)}{\mu(A_i)}\right) \ge \frac{1}{4}\mu(J_i)\log\left(\frac{\mu(I)}{\mu(J_i)}\right). \quad (4.46)$$

Finally, if $\mu(J_i) \le \mu(A_i) = (1-\alpha)\mu(I)$, $\alpha \le \frac{1}{10}$ then $\mu(J_{3-i}) = \beta\mu(I)$ where $\alpha \le \beta \le 2\alpha$ and $\mu(J_i) = (1-\beta)\mu(I) - \mu(H)$. Now (4.35) implies $\mu(H) = o(\alpha)$ and then we have

$$\mu(J_i)\log\left(\frac{\mu(I)}{\mu(J_i)}\right) \le \mu(I)(\beta+\beta^2) \text{ and } \mu(A_i)\log\left(\frac{\mu(I)}{\mu(A_i)}\right) \ge \frac{9}{10}\alpha\mu(I)$$

and so

$$\mu(A_i)\log\left(\frac{\mu(I)}{\mu(A_i)}\right) \ge \frac{1}{3}\mu(J_i)\log\left(\frac{\mu(I)}{\mu(J_i)}\right). \quad (4.47)$$

It follows from (4.44) – (4.47) that

$$\mu(H) \ge \frac{1}{80\sqrt{n}} \min\left\{ \frac{\delta}{d} \max_{i=1,2}\left\{ \mu(J_i)\log\left(\frac{\mu(I)}{\mu(A_i)}\right)\right\}, \mu(J_1), \mu(J_2)\right\}$$

contradicting (4.35) and completing the proof of the needle-like case.

We now turn to the general case where K is not necessarily needle-like. Let $\ell_1 = \max\{\ell(x) : x \in K\}$ and $M = \max\{\ell_0^{-1}, \ell_1\}$. Suppose there is a convex body K with sets S_1, S_2 such that (4.26) fails. Suppose that there exist mutually orthogonal directions $a_1, \ldots, a_j$ such that $\max_{1\le i\le j} W(K, a_i) < \delta_1/(2\sqrt{n})$. If $j \ge n-1$, by Lemma 4.5.9 the needle-like case applies and we have a contradiction. Thus suppose $j \le n-2$ is maximal such that a counter-example can be found. Let L be a two-dimensional linear subspace orthogonal to $a_1, \ldots, a_j$. By Lemma 4.5.11 there is a hyperplane P with normal $a \in L$, $|a| = 1$, which bisects the ℓ-measure of both S_1, S_2. We choose P^+ to be the half-space such that $h(K \cap P^+)$ is smaller. Let us write K' for $K \cap P^+$ etc. If the theorem fails for K, S_1, S_2, then it follows that it must also fail for K', S_1', S_2'. (The diameter can only decrease, and the value of ℓ_0 increase, so the same d, δ_1, ϵ will apply.) Also, if $K^* = K \setminus K'$,

$$\mathrm{vol}(K^*) \ge \frac{\ell(K^*)}{M} = \frac{\ell(K)}{2M} \ge \frac{\mathrm{vol}(K)}{2M^2}.$$

Thus, by Lemma 4.5.10, $W(K', a) \le \rho W(K, a)$ where $\rho = 1 - \frac{1}{2nM^2}$.

Suppose we iterate this bisection, obtaining a sequence of bodies

$$K = K^{(1)} \supset K^{(2)} \supset \cdots K^{(m)} \supset \cdots,$$

where $K^{(m)} = P^{(m)} \cap K^{(m-1)}$, containing sets for which the theorem fails. Now $K^{(m)}$ clearly converges to a compact convex set K^*. If $a^{(m)}$ is the normal to $P^{(m)}$, by compactness $a^{(m)}$ has a cluster point $a^* \in L$. By continuity, taking the limit in $0 \le W(K^{(m+1)}, a^{(m)}) \le \rho W(K^{(m)}, a^{(m)})$ gives $0 \le W(K^*, a^*) \le \rho W(K^*, a^*)$. Thus $W(K^*, a^*) = 0$, and hence for some m, $W(K^{(m)}, a^{(m)}) < \delta_1/(2\sqrt{n})$, contradiction.

Assuming that $\ell(S_1) \le \ell(S_2)$ and using (4.26) we obtain

$$\begin{aligned}\frac{h(K)\ell(K)}{2\ell(S_1)\ell(S_2)} &\ge \frac{1}{2500\sqrt{n}}\min\left\{\frac{\delta}{d}\log\left(\frac{\ell(K)}{\ell(S_1)}\right),1\right\}\\ &\ge \frac{1}{2500\sqrt{n}}\min\left\{\frac{\delta}{d}\log\left(\frac{\ell(K)^2}{2\ell(S_1)\ell(S_2)}\right),1\right\}\\ &\ge \frac{1}{2500\sqrt{n}}\min\left\{\frac{\delta}{2d}\log\left(\frac{\ell(K)^2}{\ell(S_1)\ell(S_2)}\right),1\right\}\end{aligned}$$

and the theorem follows. □

We now complete the proof of Theorem 4.2.2. Suppose that we have a partition $S_{\mathcal{L}}^{(1)}, S_{\mathcal{L}}^{(2)}$ of $K_{\mathcal{L}}$ with $S_{\mathcal{L}}^{(1)} \le S_{\mathcal{L}}^{(2)}$. We need to bound the following quantity from below:

$$\Phi_{\mathcal{L}}(S_{\mathcal{L}}^{(1)}) = \frac{\ell_{\mathcal{L}}(K_{\mathcal{L}})}{\ell_{\mathcal{L}}(S_{\mathcal{L}}^{(1)})\ell_{\mathcal{L}}(S_{\mathcal{L}}^{(2)})}\sum_{x\in S_{\mathcal{L}}^{(1)}}\frac{|B_{\mathcal{L}}(x,\delta)\cap S_{\mathcal{L}}^{(2)}|}{|B_{\mathcal{L}}(x,\delta)|}.$$

We define

$$\delta' = \delta - \eta\sqrt{n},\ K^* = (1-\eta\sqrt{n})K,\ S_1^* = K^* \cap \bigcup_{x\in S_{\mathcal{L}}^{(1)}} C(x) \text{ and } S_2^* = K^* \setminus S_1^*.$$

Then denoting the local conductance of K^* by ℓ^* we see that Theorem 4.26 implies

$$\frac{\ell^*(K^*)}{\ell^*(S_1^*)\ell^*(S_2^*)}\int_{x\in S_1^*}\frac{\mathrm{vol}(B(x,\delta')\cap S_2^*)}{\mathrm{vol}(B(x,\delta'))}\,dx \ge \frac{1}{5000\sqrt{n}}\min\left\{\frac{\delta'}{d}\log\left(\frac{\ell(K^*)}{\ell^*(S_1^*)}\right),1\right\} \quad (4.48)$$

Arguing as in Lemma 4.5.2 we get

$$\ell^*(K^*) \le \left(1+\frac{\epsilon}{n}\right)\eta^n\ell_{\mathcal{L}}(K_{\mathcal{L}}), \ell^*(S_i^*) \ge \left(1+\frac{\epsilon}{n}\right)^{-1}\eta^n\ell_{\mathcal{L}}(S_{\mathcal{L}}^{(i)*}),\ i=1,2 \quad (4.49)$$

where $S_{\mathcal{L}}^{(i)*} = \left\{x \in S_{\mathcal{L}}^{(i)} :\ C(x)\cap K^* \ne \emptyset\right\}$.

Furthermore,

$$\begin{aligned}\int_{x\in S_1^*}\frac{\mathrm{vol}(B(x,\delta')\cap S_2^*)}{\mathrm{vol}(B(x,\delta'))}\,dx &= \int_{x\in S_1^*}\int_{x\in S_2^*} 1_{|x-y|\le\delta'}\,dxdy \le\\ &\sum_{x\in S_{\mathcal{L}}^{(1)*}}\sum_{x\in S_{\mathcal{L}}^{(2)*}}\eta^{2n}1_{|x-y|\le\delta} \le \sum_{x\in S_{\mathcal{L}}^{(1)*}}|B_{\mathcal{L}}(x,\delta)\cap S_{\mathcal{L}}^{(2)}|. \quad (4.50)\end{aligned}$$

It follows from (4.49), (4.50) and $\eta^n|B_{\mathcal{L}}(x,\delta)| \le \left(1+\frac{\epsilon}{n}\right)\mathrm{vol}(B(x,\delta'))$ – Lemma 4.3.1 – that

$$\Phi_{\mathcal{L}}(S_{\mathcal{L}}^{(1)}) \ge \frac{\left(1+\frac{\epsilon}{n}\right)^{-4}}{5000\sqrt{n}}\frac{\ell_{\mathcal{L}}^*(S_{\mathcal{L}}^{(1)*})}{\ell_{\mathcal{L}}(S_{\mathcal{L}}^{(1)})}\cdot\frac{\ell_{\mathcal{L}}^*(S_{\mathcal{L}}^{(2)*})}{\ell_{\mathcal{L}}S_{\mathcal{L}}^{(2)})}\min\left\{\frac{\delta'}{d}\log\left(\frac{\left(1+\frac{\epsilon}{n}\right)^{-1}\ell_{\mathcal{L}}(K_{\mathcal{L}})}{\left(1+\frac{\epsilon}{n}\right)\ell_{\mathcal{L}}(S_{\mathcal{L}}^{(1)*})}\right),1\right\} \quad (4.51)$$

Arguing as in Lemma 4.5.2 we get

$$\ell_{\mathcal{L}}(S_{\mathcal{L}}^{(2)*}) \geq \ell_{\mathcal{L}}(S_{\mathcal{L}}^{(2)}) - 2\frac{\epsilon}{n}\ell_{\mathcal{L}}(K_{\mathcal{L}}) \geq \ell_{\mathcal{L}}(S_{\mathcal{L}}^{(2)})\left(1 - 4\frac{\epsilon}{n}\right)$$

since $\ell_{\mathcal{L}}(S_{\mathcal{L}}^{(2)}) \geq \frac{1}{2}\ell_{\mathcal{L}}(K_{\mathcal{L}})$.

Case 1: $\ell_{\mathcal{L}}(S_{\mathcal{L}}^{(1)*}) \geq \frac{1}{2}\ell_{\mathcal{L}}(S_{\mathcal{L}}^{(1)})$.
It follows from (4.51) that

$$\Phi_{\mathcal{L}}(S_{\mathcal{L}}^{(1)}) \geq \frac{1}{10001\sqrt{n}} \min\left\{\frac{\delta}{d}\log\left(\frac{\ell_{\mathcal{L}}(K_{\mathcal{L}})}{\ell_{\mathcal{L}}(S_{\mathcal{L}}^{(1)})}\right), 1\right\}. \tag{4.52}$$

Case 1: $\ell_{\mathcal{L}}(S_{\mathcal{L}}^{(1)*}) < \frac{1}{2}\ell_{\mathcal{L}}(S_{\mathcal{L}}^{(1)})$.
We show that

$$x \in S_{\mathcal{L}}^{(1)B} = S_{\mathcal{L}}^{(1)} \setminus S_{\mathcal{L}}^{(1)*} \text{ implies } |B_{\mathcal{L}}(x,\delta) \cap S_{\mathcal{L}}^{(2)}| \geq \frac{1}{2}|B_{\mathcal{L}}(x,\delta) \cap K_{\mathcal{L}}|. \tag{4.53}$$

As a consequence

$$\Phi_{\mathcal{L}}(S_{\mathcal{L}}^{(1)}) \geq \frac{\ell_{\mathcal{L}}(K_{\mathcal{L}})}{\ell_{\mathcal{L}}(S_{\mathcal{L}}^{(1)})\ell_{\mathcal{L}}(S_{\mathcal{L}}^{(2)})} \sum_{x \in S_{\mathcal{L}}^{(1)B}} \frac{|B_{\mathcal{L}}(x,\delta) \cap K_{\mathcal{L}}|}{2|B_{\mathcal{L}}(x,\delta)|} = \frac{\ell_{\mathcal{L}}(K_{\mathcal{L}})\ell_{\mathcal{L}}(S_{\mathcal{L}}^{(1)B})}{2S_{\mathcal{L}}^{(1)})\ell_{\mathcal{L}}(S_{\mathcal{L}}^{(2)})} \geq \frac{1}{4}.$$

Suppose that in contradiction to (4.53) that $\exists x \in S_{\mathcal{L}}^{(1)B}$ such that

$$|B_{\mathcal{L}}(x,\delta) \cap S_{\mathcal{L}}^{(1)}| > \frac{1}{2}|B_{\mathcal{L}}(x,\delta) \cap K_{\mathcal{L}}| \geq \frac{1}{3}\eta^{-n}(10d/\delta)^{-n} \tag{4.54}$$

where the last inequality is from Theorem 4.3.1 and Lemma 4.5.2.

Now

$$\ell_{\mathcal{L}}(S_{\mathcal{L}}^{(1)B}) \leq |S_{\mathcal{L}}^{(1)B}| \leq 3\eta^{-(n-1)}d^n n^{1/2} \tag{4.55}$$

and (4.53) implies

$$\ell_{\mathcal{L}}(S_{\mathcal{L}}^{(1)*}) \geq (\eta^{-n}(2d)^{-n} - 3\eta^{-(n-1)}d^n\sqrt{n})(2d)^{-n}. \tag{4.56}$$

(4.55) and (4.56) contradict $\ell_{\mathcal{L}}(S_{\mathcal{L}}^{(1)*}) < \ell_{\mathcal{L}}(S_{\mathcal{L}}^{(1)B})$.

Thus (4.52) holds in general and so

$$\begin{aligned}\Phi_{\mathcal{L}}(S_{\mathcal{L}}^{(1)}) &\geq \frac{1}{10001\sqrt{n}} \min\left\{\frac{\delta}{d}\log\left(\frac{\ell_{\mathcal{L}}(K_{\mathcal{L}})^2}{2\ell_{\mathcal{L}}(S_{\mathcal{L}}^{(1)})\ell_{\mathcal{L}}(S_{\mathcal{L}}^{(2)})}\right), 1\right\} \\ &\geq \frac{1}{10001\sqrt{n}} \min\left\{\frac{\delta}{2d}\log\left(\frac{\ell_{\mathcal{L}}(K_{\mathcal{L}})^2}{\ell_{\mathcal{L}}(S_{\mathcal{L}}^{(1)})\ell_{\mathcal{L}}(S_{\mathcal{L}}^{(2)})}\right), 1\right\}\end{aligned}$$

Thus the conductance function $\Phi_{\mathcal{L}}(x)$ for the speedy chain satisfies the conditions of Theorem 2.6.4 with $A = \frac{\delta}{20002d\sqrt{n}}$ and $B = \frac{1}{20002\sqrt{n}}$ and Theorem 4.2.2 follows. □

4.6 Deferred proofs of Section 4.4

4.6.1 Proof of Theorem 4.4.3

(a) The claim is invariant under affine transformations and so we can assume that K is in isotropic position. We therefore have to show that $g \in -\phi K$ which is implied by $|g| \leq \frac{\phi}{n}$. Now

$$\mathbf{E}(|g|^2) = \frac{1}{m_2^2}\left(\sum_i \mathbf{E}(|z^{(i)}|^2) + \sum_{i \neq j} \mathbf{E}({z^{(i)}}^T z^{(j)})\right). \tag{4.57}$$

Let $\xi_i = W^{(i)} - U_{\mathcal{L}}$ where $W^{(i)}$ will be the conditional distribution of $z^{(i)}$ given $z^{(j)}, j \neq i$. Then

$$\mathbf{E}(|z^{(i)}|^2) = \sum_{x \in K_{\mathcal{L}}} |x|^2 U_{\mathcal{L}}(x) + \sum_{x \in K_{\mathcal{L}}} |x|^2 \xi(x) \leq n + 2n^2\epsilon_2.$$

Also, for $i < j$,

$$\begin{aligned}\mathbf{E}({z^{(i)}}^T z^{(j)}) &= \sum_{x \in K_{\mathcal{L}}} \mathbf{E}(x^T z^{(j)} \mid z^{(i)} = x) W^{(i)}(x) \\ &= \sum_{x \in K_{\mathcal{L}}} \sum_{y \in K_{\mathcal{L}}} x^T y \mathbf{Pr}(z^{(j)} = y \mid z^{(i)} = x) W^{(i)}(x) \\ &= \left(\sum_{x \in K_{\mathcal{L}}} x\xi_i(x)\right)^T \left(\sum_{y \in K_{\mathcal{L}}} y\xi_j(y)\right) \leq 4n^2\epsilon_2^2.\end{aligned}$$

Hence

$$\mathbf{E}(|g|^2) \leq \frac{1}{m_2^2}(n^2 + 2n^2\epsilon_2 + 4n^4\epsilon_2^2) \leq \frac{\gamma\phi^2}{4}$$

and so we can use the Markov inequality to complete the proof of (a).

The proof of (b) is similar. □

4.6.2 Proof of Theorem 4.4.2

Replacing K by TK for some non-singular affine transformation T yields the same value for K' and so we can assume that K is in isotropic position.

We start by proving the second condition of θ-isotropy. We want to prove that with probability at least $1 - \eta$ every $w \in \mathbb{R}^n$ satisfies

$$(1-\theta)|w|^2 \leq \frac{1}{\mathrm{vol}(K')} \int_{K' - b(K')} (w^T y)^2 dy \leq (1+\theta)|w|^2. \tag{4.58}$$

By change of variables, $y \to Y^{1/2}y + \bar{y}$, (4.58) can be written as

$$(1-\theta)|v|^2 \leq \frac{1}{\mathrm{vol}(K)} \int_K (v^T y)^2 dy \leq (1+\theta)|v|^2.$$

We can assume that $|v| = 1$ in (4.59) and so the middle term is 1. So we have to prove

$$\frac{1}{1+\theta} \leq v^T Y v \leq \frac{1}{1-\theta}. \tag{4.59}$$

Putting $Y = Z - \bar{y}\,\bar{y}^T$ where $Z = \frac{1}{m}\sum_{i=1}^m y^{(i)} y^{(i)^T}$ we see that we have to show that with probability at least $1-\eta$, for every $v \in \mathbb{R}^n, |v| = 1$,

$$\frac{1}{1+\theta} + (v^T\bar{y})^2 \leq v^T Z v \leq \frac{1}{1-\theta} + (v^T\bar{y})^2.$$

Now by Theorem 4.4.3 we have that with probability at least $1-\gamma$, we have

$$|\bar{y}| \leq \theta/4 \tag{4.60}$$

and so it suffices to prove that for all $|v| = 1$,

$$|v^T \mathbf{E}(Z) v - 1| \leq \frac{\theta}{4}. \tag{4.61}$$

Indeed,

$$v^T\mathbf{E}(Z)v = \frac{1}{m_1}\sum_{i=1}^{m_1} v^T\mathbf{E}(y^{(i)}y^{(i)^T})v = \frac{1}{m_1}\sum_{i=1}^{m_1}\sum_{x\in K_{\mathcal{L}}} (v^Tx)^2 W^{(i)}(x)$$

$$= \frac{1}{m_1}\sum_{i=1}^{m_1}(v^Tx)^2(U_{\mathcal{L}}(x) + \xi^{(i)}(x)) = 1 + \frac{1}{m_1}\sum_{i=1}^{m_1}(v^Tx)^2\xi^{(i)}(x)$$

and (4.61) follows from

$$\left|\frac{1}{m_1}\sum_{i=1}^{m_1}(v^Tx)^2\xi^{(i)}(x)\right| \leq n^4\epsilon_1.$$

Next we prove that with probability at least $1 - \frac{\gamma}{2}$

$$||Z - \mathbf{E}(Z)|| \leq \frac{\theta}{4}. \tag{4.62}$$

To prove this we use

$$||Z - \mathbf{E}(Z)||^2 \leq \mathrm{Tr}((Z - \mathbf{E}(Z))^2).$$

We compute the expectation of this trace.

$$m_1^2(Z - \mathbf{E}(Z))^2 = \sum_{i=1}^{m_1}(y^{(i)}y^{(i)^T} - \mathbf{E}(y^{(i)}y^{(i)^T})^2 +$$

$$\sum_{i\neq j}(y^{(i)}y^{(i)^T} - \mathbf{E}(y^{(i)}y^{(i)^T}))(y^{(j)}y^{(j)^T} - \mathbf{E}(y^{(j)}y^{(j)^T})). \tag{4.63}$$

The first term is handled as follows: Fix any i; then

$$\mathbf{E}((y^{(i)}y^{(i)^T} - \mathbf{E}(y^{(i)}y^{(i)^T})^2) = \mathbf{E}((y^{(i)}y^{(i)^T})^2) - \mathbf{E}(y^{(i)}y^{(i)^T})^2$$

and hence

$$\begin{aligned}\mathbf{E}(\mathrm{Tr}((y^{(i)}y^{(i)^T} - \mathbf{E}(y^{(i)}y^{(i)^T})^2))) &\leq \mathbf{E}(\mathrm{Tr}((y^{(i)}y^{(i)^T})^2)) = \mathbf{E}(|y^{(i)}|^4) \\ &= \sum_{x\in K_{\mathcal{L}}} |x|^4 U_{\mathcal{L}}(x) + \sum_{x\in K_{\mathcal{L}}} |x|^4 \xi^{(i)}(x) \leq 8n^2 + \epsilon_1 n^4,\end{aligned}$$

where we have used the inequality $\mathbf{E}_K(|x|^4) \leq 8\mathbf{E}_K(|x|^2)$ by a theorem of Gromov and Milman.

which theorem?

The second term in (4.63)is handled as follows: Fix any $i \neq j$; then

$$\begin{aligned}\mathrm{Tr}((y^{(i)}y^{(i)^T} - \mathbf{E}(y^{(i)}y^{(i)^T}))(y^{(j)}y^{(j)^T} - \mathbf{E}(y^{(j)}y^{(j)^T}))) = \\ \sum_{k=1}^{n}\sum_{l=1}^{n}(y_k^{(i)}y_l^{(i)} - \mathbf{E}(y_k^{(i)}y_l^{(i)}))(y_k^{(j)}y_l^{(j)} - \mathbf{E}(y_k^{(j)}y_l^{(j)})).\end{aligned}$$

The expectation of each term here can be bounded by $\epsilon_1(n+1)^4$ and the expectation of the second term in (4.63)c an be bounded by $\epsilon_1 m_1^2(n+1)^4$. So,

$$\mathbf{E}(||Z - \mathbf{E}(Z)||^2) \leq m_1^{-1}(8n^2 + \epsilon_1 n^4) + \epsilon_1(n+1)^4 \leq \frac{\gamma^2\phi^2}{8}$$

and (4.61) follows.

To complete the proof it suffices to remark that $b(K') = -Y^{1/2}\bar{y}$ and hence if (4.59) and (4.60) hold, then

$$|b(K')| = \sqrt{\bar{y}^T Y \bar{y}} \leq \sqrt{\frac{\theta^2}{16(1-\theta)}} < \theta.$$

□

4.6.3 Proof of Theorem 4.4.5

We first prove some preliminary lemmas.

Lemma 4.6.1 *If $b(K) \in -\alpha K$ for $\alpha > 0$ and $h^T x \leq c$ for $h \in \mathbb{R}^n, |h| = 1$ and $x \in K$ then $\mathbf{E}_K((h^T x)^2) \leq (1 + 2\alpha + 2\alpha^2)c^2$.*

Proof We first observe that if $f = \rho^T x + \sigma$ is an arbitrary linear function on K then

$$\mathbf{E}_K(f^2) \leq \mathbf{E}_K(f)^2 + \left(\max_K f - \mathbf{E}_K(f)\right)^2. \tag{4.64}$$

Indeed, (4.64) is clear if K is in isotropic position for then $\mathbf{E}_K(f) = \sigma$, $\mathbf{E}_K(f^2) = |\rho|^2$ and $\max_K f \geq |\rho|$ since $K \supseteq B$. Now note thet if (4.64) is true for some K and all f then it remains true for AK and all f, where A is an affine transformation.

Now let $f = h^T x$. Then $\mathbf{E}_K(h^T x) = h^T b \geq -\alpha c$ since $b \in -\alpha K$. So, applying (4.64), we get

$$\mathbf{E}_K((h^T x)^2) \leq c^2 + (c + \alpha c)^2.$$

□

Lemma 4.6.2 *Let $h \in \mathbb{R}^n, |h| = 1$ for which $h^T x < \frac{1}{2}$ for all $x \in K$. Let U_h be as in (4.11). Then*

(a) *If K satifies (4.7) and (4.9) then so does $U_h K$.*

(b) *$vol(U_h K) \geq \frac{9}{8} vol(K)$.*

Proof For (4.7) we use

$$b(U_h K) = U_h b(K) \in U_h \left(-\frac{1}{10} K\right) = -\frac{1}{10} U_h K.$$

For (4.9) we see that Lemma 4.6.1 implies $\mathbf{E}_K((h^T x)^2) \leq \frac{1}{3}$. Hence

$$\mathbf{E}_{U_h K}(|x|^2) = \mathbf{E}_K(|U_h x|^2) = \left(1 - \frac{1}{2n}\right)^2 E_K(|x + (x^T h)h|^2) =$$
$$\left(1 - \frac{1}{2n}\right)^2 E_K(|x|^2 + 3(h^T x)^2) < \left(1 - \frac{1}{2n}\right)^2 (100n + 1) < 100n.$$

This completes the proof of (a).

Since $\det(U_h) = 2\left(1 - \frac{1}{2n}\right)^n \geq \frac{9}{8}$ we have (b). □

Lemma 4.6.3 *Let K be a convex body containing the origin and let v be chosen uniformly from K. Make one step of a lazy random walk starting from v. Then the probability that this step is a nonflat improper step is at most $4\delta_0\sqrt{n}$.*

Proof Put $K_1 = conv(K \cup \frac{1}{2}B)$. Assume that the attempted step $v \to u$ is nonflat improper. Then trivially $u \notin K_1$. We prove that the $(2n)$-dimensional measure of the set S of pairs v, u with $v \in K$, $u \in \mathbb{R}^n \setminus K_1$ and $|v - u| \leq \delta_0$ is at most $4\delta_0\sqrt{n}\mathrm{vol}(K)\mathrm{vol}(\delta_0 B)$ and this will prove the lemma.

Let q' be the point of intersection of the segment $[v, u]$ and ∂K_1. Then clearly $q' \in F' = \partial K_1 \cap (2^{1/n} K)$. Applying Lemma 4.5.1 to K_1 we get that

$$\mathrm{vol}_{2n}(S) \leq \delta_0 \mathrm{vol}_{n-1}(F') \frac{c_{n-1}}{(n+1)c_n} \mathrm{vol}(\delta_0 B) < \frac{\delta_0}{\sqrt{n}} \mathrm{vol}_{n-1}(F') \mathrm{vol}(\delta_0 B).$$

The hyperplane supporting K_1 at any point of F' has distance at least $1/2$ from the origin. Hence the union U of segments connecting 0 to F' has volume at least $\text{vol}_{n-1}(F')/(2n)$. On the other hand, clearly $U \subseteq 2^{1/n}K$. This implies that

$$\text{vol}_{n-1}(F') \leq 4n\text{vol}(K)$$

and so

$$\text{vol}_{2n}(S) \leq 4\delta_0\sqrt{n}\text{vol}(K)\text{vol}(\delta_0 B).$$

□

Lemma 4.6.4 *Let $K \subseteq d_0 B$, $d_0 \geq 1$ be a convex body with average local conductance λ with respect to δ_0 moves where $0 < \delta_0 < \frac{1}{32}$. Let $u \in K$. Starting from u, do a lazy random walk in K until at least*

$$T = \left\lceil \kappa n^2 d_0^2 \delta_0^{-2} \log(\delta_0\sqrt{n}) \right\rceil$$

proper steps are made. Then the probability that no flat steps were attempted is at most $\lambda + 6\delta_0\sqrt{n}$.

Proof We may assume that $\delta_0 < 1/(6\sqrt{n})$. Consider a random walk in the body $K_1 = conv(K \cup \frac{1}{2}B)$ starting at u. Until this walk hits $K_1 \setminus K$ it can be considered a random walk in K. Conversely, a random walk in K can be considered a random walk in K_1 until the first flat step is attempted, because until then, any time we attempt to step out of K, we are actually stepping out of K_1. Hence the probability that a random walk of length T in K attempts a flat step is at least as large as the probability that a random walk in K_1 of length T hits $K_1 \setminus K$.

Now $(1/2)B \subseteq K_1 \subseteq d_0 B$ and so (4.16) implies that the average local conductance of K_1 is at least $1 - \delta_0\sqrt{n}$. Theorem 4.2.2 and Lemma 4.2.1 imply that the distribution of the point w at the $(T-1)$th proper step is within variation distance

$2\delta_0\sqrt{n}$ of uniform. Lemma 4.6.3 then implies that the probability the step from w is proper or nonflat is at most $\lambda + 2\delta_0\sqrt{n} + 4\delta_0\sqrt{n}$. □

We can now prove Theorem 4.4.5. The first assertion is clear by Lemma 4.6.2. This
plausible implies that the volume of every K_i is at most that of $B(0, 10n)$ which is $(10n)^{n/2}c_n$. On the other hand (4.9) implies that $\text{vol}(K_0) \geq n^{-n/4}c_n$. Since each flat step increase the volume by at least $9/8$ we see that at most $8n \log n$ flat steps can occur.

Consider the algorithm going on for M rather than N iterations. Let L_i, L_i' be the average local conductances of K_i, K_i' respectively. It follows from (4.9), $K_i' \subseteq dB$ and the Markov inequality that

$$\frac{\text{vol}(K_i')}{\text{vol}(K_i)} \geq 1 - 50\theta.$$

So,

$$L_i \geq L'_i \frac{\mathrm{vol}(K'_i)}{\mathrm{vol}(K_i)} \geq L'_i(1-50\theta).$$

Let $\lambda_i = \mathbf{E}(L_i)$ (L_i is a random variable). Let X_i be the indicator variable of the event that the ith random walk ended with a flat step. Then $\sum_i X_i$ is the number of such walks and hence

$$\sum_i X_i \leq 8n \log n. \tag{4.65}$$

On the other hand, from Lemma 4.6.4, we get that

$$\mathbf{Pr}(X_{i+1} = 1 \mid \textit{previous history}) \geq 1 - 6\delta_0\sqrt{n} - L'_i \geq 1 - \frac{\theta}{4} - \frac{1}{1-50\theta} L_i$$

and so

$$\sum_{i=0}^{M-1} \mathbf{E}(X_i) \geq \sum_{i=0}^{M-1} \left(1 - \frac{\theta}{4} - \frac{1}{1-50\theta} L_i\right)$$

and so by (4.65)

$$\frac{1}{M} \sum_{i=0}^{M-1} \lambda_i \geq (1-50\theta)\left(1 - \frac{8n\log n}{M} - \frac{\theta}{4}\right) \geq 1 - 100\theta.$$

Since N is chosen randomly from $\{0, 1, \ldots, M-1\}$ we see that

$$\mathbf{E}(L_N) = \lambda_N \geq 1 - 100\theta.$$

For simplicity imagine that the last walk goes on if necessary until a total of at least $3MT$ steps are made. If the number of nonflat improper steps during the algorithm is larger than MT then their number among the first $3MT$ steps is larger than MT. Since u and therefore evey given point in the sequence has a distribution that is closer to uniform than $\gamma/6$ (in total variation distance), the probability that a given step is nonflat improper is at most $\gamma/6 + 4\delta_0\sqrt{n} < \gamma/3$ by Lemma isolem4. Thus the expected number of nonflat improper steps is at most γMT. The probability bound on the number of steps then follows from the Markov inequality. □

Chapter 5

Matroids

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 $S = \{u_0, \ldots, u_{m-1}\}$ 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. A matroid that is representable over every field is called *regular*. 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 [**?**] or Welsh [**?**].

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 unlabelled 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*. 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*. All graphic matroids are regular, but the converse does not hold: the co-graphic matroid of a non-planar graph is regular but not graphic. A rather trivial

class of matroids we shall encounter are the "uniform matroids." 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 or size r.

Two absolutely central operations on matroids are contraction and deletion. If $e \in E$ is an element of the ground set of M then the matroid $M \setminus e$ obtained by *deleting* e has ground set $E_{-e} = E \setminus \{e\}$ and bases $\mathcal{B}(M \setminus e) = \{X \subseteq E_{-e} : X \in \mathcal{B}(M)\}$; the matroid M/e obtained by *contracting* e has ground set E_{-e} and bases $\mathcal{B}(M/e) = \{X \subseteq E_{-e} : X \cup \{e\} \in \mathcal{B}(M)\}$. 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 base (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 [**?**], as is the proof of rapid mixing. We follow their treatment quite closely.

5.1 Balanced matroids

For 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. Suppose a base $X \in \mathcal{B}$ 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 $\mathbf{Pr}(ef) \leq \mathbf{Pr}(e)\mathbf{Pr}(f)$ holds for all pairs of distinct elements $e, f \in E$. Another way of expressing negative correlation is by writing $\mathbf{Pr}(e \mid f) \leq \mathbf{Pr}(e)$; 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 §5.1.2 that regular matroids are always balanced. But there are balanced matroids that are not regular: it is easy to check that all uniform matroids satisfy the negative correlation property and that the

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

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 [**?**, Theorem 13.1.1].)

5.1.1 Efficiently sampling bases of balanced matroids

It is convenient in this section to work with a combinatorial version of conductance rather than conductance itself. The *cutset expansion* of a graph G is the minimum, over all subsets $\mathcal{S} \subset V(G)$ with $0 < |\mathcal{S}| \leq |V(G)|/2$ of the ratio $|\operatorname{cut}(\mathcal{S})|/|\mathcal{S}|$, where $\operatorname{cut}(\mathcal{S}) \subseteq E(G)$ denotes the set of edges with one endpoint in $\mathcal{S}$ and one in the complement of S. The main result of this section is a lower bound on cutset expansion of the bases-exchange graph.

Theorem 5.1.1 *The cutset expansion of the bases-exchange graph $G(M)$ of any balanced matroid M is at least* 1.

Suppose we implement the random walk on the bases-exchange graph $G(M)$ in the following natural way. The current state (base) is X.

Step 1 With probability $\frac{1}{2}$ set $Y = X$.

Step 2 Otherwise, choose e u.a.r. from $E \setminus X$.

Step 3 Choose $f \in E$ u.a.r. from the elements of the ground set satisfying $Y = X \cup \{e\} \setminus \{f\} \in \mathcal{B}$.

The new state is Y. Call this the *bases-exchange walk*. Note that the transition matrix implicitly described by the above implementation is symmetric. Since we have already observed that $G(M)$ is connected, we see that the bases-exchange walk converges to a stationary distribution that is uniform over states. Furthermore, the non-zero transition probabilities (corresponding to edges of $G(M)$) are all at least $1/2mr$. Thus, according to Theorem 5.1.1, the conductance of the random walk is bounded below by $1/2mr$, and by Theorem 2.2.1 we obtain:

Corollary 5.1.1 *The mixing time of the bases-exchange walk on any balanced matroid of rank r on a ground set of size m is at most $4m^2r^2(r \ln m + \ln \epsilon^{-1})$.*

We approach the proof of Theorem 5.1.1 via a couple of lemmas. 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.

Lemma 5.1.1 *Let M be a balanced matroid and let $e \in E$.*

(a) *If μ is an increasing property over E_{-e}, then $\mathbf{Pr}(\mu \mid e) \leq \mathbf{Pr}(\mu \mid \bar{e})$.*

(b) *If μ is a decreasing property over E_{-e}, then $\mathbf{Pr}(\mu \mid e) \geq \mathbf{Pr}(\mu \mid \bar{e})$.*

Proof We prove (a), (b) follows by consideration of $\bar{\mu}$. The proof is by induction on the size of the ground set. We may assume that $\mathbf{Pr}(\mu e) > 0$, 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 $\mathbf{Pr}(\mu \mid e) \leq \mathbf{Pr}(\mu)$. If the rank of M is 1 then either (i) $\emptyset \in \mu$ and $\mathbf{Pr}(\mu) = 1$ or (ii) $\emptyset \notin \mu$ and $\mathbf{Pr}(\mu e) = 0$. Thus we may assume that the rank r of M is at least 2.

From the identity

$$\mathbf{E}(|X \setminus e| \mid \mu e) = \sum_{f \neq e} \mathbf{Pr}(f \mid \mu e) = r - 1 = \mathbf{E}(|X \setminus e| \mid e) = \sum_{f \neq e} \mathbf{Pr}(f \mid e),$$

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

$$\mathbf{Pr}(\mu \mid ef) \geq \mathbf{Pr}(\mu \mid e); \tag{5.1}$$

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

$$\mathbf{Pr}(f \mid e) \leq \mathbf{Pr}(f) \tag{5.2}$$

and

$$\mathbf{Pr}(\mu \mid ef) \leq \mathbf{Pr}(\mu \mid f); \tag{5.3}$$

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 $\mathbf{Pr}(\bar{f} \mid e) = 0$. It follows from (5.2) that $\mathbf{Pr}(f) = 1$, and then from (5.3) that $\mathbf{Pr}(\mu \mid e) \leq \mathbf{Pr}(\mu)$, as desired. So we may now assume $\mathbf{Pr}(\bar{f} \mid e) > 0$ and hence that probabilities conditional on the event $e\bar{f}$ are well defined. In particular,

$$\mathbf{Pr}(\mu \mid e\bar{f}) \leq \mathbf{Pr}(\mu \mid \bar{f}), \tag{5.4}$$

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 (5.1) may be re-expressed as

$$\mathbf{Pr}(\mu \mid ef) \geq \mathbf{Pr}(\mu \mid e\bar{f}). \tag{5.5}$$

The inductive step is now achieved through a chain of inequalities based on (5.2)–(5.5):

$$\begin{aligned}
\mathbf{Pr}(\mu \mid e) &= \mathbf{Pr}(\mu \mid ef)\mathbf{Pr}(f \mid e) + \mathbf{Pr}(\mu \mid e\bar{f})\mathbf{Pr}(\bar{f} \mid e) \\
&= \mathbf{Pr}(\mu \mid ef)\mathbf{Pr}(f \mid e) + \mathbf{Pr}(\mu \mid e\bar{f})(1 - \mathbf{Pr}(f \mid e)) \\
&= \big[\mathbf{Pr}(\mu \mid ef) - \mathbf{Pr}(\mu \mid e\bar{f})\big]\mathbf{Pr}(f \mid e) + \mathbf{Pr}(\mu \mid e\bar{f}) \\
&\leq \big[\mathbf{Pr}(\mu \mid ef) - \mathbf{Pr}(\mu \mid e\bar{f})\big]\mathbf{Pr}(f) + \mathbf{Pr}(\mu \mid e\bar{f}) \qquad (5.6) \\
&= \mathbf{Pr}(\mu \mid ef)\mathbf{Pr}(f) + \mathbf{Pr}(\mu \mid e\bar{f})\mathbf{Pr}(\bar{f}) \\
&\leq \mathbf{Pr}(\mu \mid f)\mathbf{Pr}(f) + \mathbf{Pr}(\mu \mid \bar{f})\mathbf{Pr}(\bar{f}) \qquad (5.7) \\
&= \mathbf{Pr}(\mu),
\end{aligned}$$

where inequality (5.6) uses (5.2) and (5.5), and inequality (5.7) uses (5.3) and (5.4). □

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. 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. For $\mathcal{S} \subseteq \mathcal{B}$ we let $\mathcal{S}_e = \{X \in \mathcal{S} : e \in X\}$ and $\mathcal{S}_{\bar{e}} = \mathcal{S} \setminus \mathcal{S}_e$.

Lemma 5.1.2 *Suppose M is a balanced matroid, $e \in E$, and that the partition $\mathcal{B} = \mathcal{B}_e \cup \mathcal{B}_{\bar{e}}$ is non-trivial. Then for all $\mathcal{S} \subseteq \mathcal{B}$,*

$$\frac{|\Gamma_e(\mathcal{S}_e)|}{|\mathcal{B}_{\bar{e}}|} \geq \frac{|\mathcal{S}_e|}{|\mathcal{B}_e|}, \text{ and}$$
$$\frac{|\Gamma_e(\mathcal{S}_{\bar{e}})|}{|\mathcal{B}_e|} \geq \frac{|\mathcal{S}_{\bar{e}}|}{|\mathcal{B}_{\bar{e}}|}.$$

Proof $\mu_1 = \{Y \subseteq E_{-e} : \exists X \in \mathcal{S}_e \text{ s.t. } Y \supseteq X \setminus \{e\}\}$ is an increasing property. The collection of all bases in $\mathcal{B}_e$ satisfying μ_1 is precisely $\mathcal{S}_e$, while the collection of all bases in $\mathcal{B}_{\bar{e}}$ satisfying μ_1 is precisely $\Gamma_e(\mathcal{S}_e)$. Hence the first part of the lemma is equivalent to the inequality $\mathbf{Pr}(\mu_1 \mid \bar{e}) \geq \mathbf{Pr}(\mu_1 \mid e)$, which follows from Lemma 5.1.1. Similarly, $\mu_2 = \{Y \subseteq E_{-e} : \exists X \in \mathcal{S}_{\bar{e}} \text{ s.t. } Y \subseteq X \cup \{e\}\}$ is a decreasing property. The set of all bases in $\mathcal{B}_{\bar{e}}$ satisfying μ_2 is precisely $\mathcal{S}_{\bar{e}}$, while the set of all bases in $\mathcal{B}_e$ satisfying μ_2 is precisely $\Gamma_e(\mathcal{S}_{\bar{e}})$. Hence the second part of the lemma is equivalent to the inequality $\mathbf{Pr}(\mu_2 \mid e) \geq \mathbf{Pr}(\mu_2 \mid \bar{e})$, which again follows from Lemma 5.1.1. □

We now have the tools needed to bound the cutset expansion of the bases-exchange graph.

Proof of Theorem 5.1.1 We proceed by induction on $|E|$. Let $\mathcal{S} \subset \mathcal{B}$ be a collection of bases, with $|\mathcal{S}| \leq |\mathcal{B}|/2$, defining a cut in the bases-exchange graph of M. Let $\mathcal{S}_e = \mathcal{S} \cap \mathcal{B}_e$ and $\mathcal{S}_{\bar{e}} = \mathcal{S} \cap \mathcal{B}_{\bar{e}}$, and define α and β by $|\mathcal{S}_e| = \alpha|\mathcal{B}_e|$ and $|\mathcal{S}_{\bar{e}}| = \beta|\mathcal{B}_{\bar{e}}|$.

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 $\mathcal{B}_{\bar{e}}$. By the induction hypothesis, the numbers of edges of kinds (i) and (ii) are at least $\min\{\alpha, 1-\alpha\}|\mathcal{B}_e|$ and $\min\{\beta, 1-\beta\}|\mathcal{B}_{\bar{e}}|$, respectively. To lower bound the number of edges of kind (iii), assume first that $\alpha \geq \beta$. By Lemma 5.1.2, there are at least $\alpha|\mathcal{B}_{\bar{e}}|$ bases in $\mathcal{B}_{\bar{e}}$ adjacent to some base in $\mathcal{S}_e$; of these, at least $(\alpha-\beta)|\mathcal{B}_{\bar{e}}|$ must lie outside $\mathcal{S}_{\bar{e}}$. Thus there are at least $(\alpha-\beta)|\mathcal{B}_{\bar{e}}|$ edges of type (iii). This argument can equally well be applied in the opposite direction, starting at the set $\mathcal{B}_{\bar{e}} \setminus \mathcal{S}_{\bar{e}}$, yielding a second lower bound of $((1-\beta)-(1-\alpha))|\mathcal{B}_e| = (\alpha-\beta)|\mathcal{B}_e|$. Thus the number of edges of kind (iii) is at least $(\alpha-\beta)\max\{|\mathcal{B}_e|, |\mathcal{B}_{\bar{e}}|\}$. Since the case $\alpha < \beta$ is entirely symmetric, we obtain, summing the contributions from edges of kinds (i)–(iii):

$$|\operatorname{cut}(\mathcal{S})| \geq \min\{\alpha, 1-\alpha\}|\mathcal{B}_e| + \min\{\beta, 1-\beta\}|\mathcal{B}_{\bar{e}}| + |\alpha-\beta|\max\{|\mathcal{B}_e|, |\mathcal{B}_{\bar{e}}|\}. \tag{5.8}$$

To complete the proof we must show that $|\operatorname{cut}(\mathcal{S})|$ is always at least $\alpha|\mathcal{B}_e| + \beta|\mathcal{B}_{\bar{e}}| = |\mathcal{S}|$, whenever $|\mathcal{S}| \leq |\mathcal{B}|/2$. Note that this last condition may be expressed as

$$(\tfrac{1}{2}-\alpha)|\mathcal{B}_e| + (\tfrac{1}{2}-\beta)|\mathcal{B}_{\bar{e}}| \geq 0. \tag{5.9}$$

If $\alpha, \beta \leq \frac{1}{2}$, the required lower bound on $|\operatorname{cut}(\mathcal{S})|$ follows immediately from (5.8). We therefore just need to treat the cases when one of α or β is greater than $\frac{1}{2}$. To simplify the working, we'll exploit the symmetry of (5.8) and assume, without loss of generality, that

$$|\mathcal{B}_e| \geq |\mathcal{B}_{\bar{e}}|. \tag{5.10}$$

Suppose first that $\alpha > \frac{1}{2}$. Then inequalities (5.9) and (5.10) entail $\beta < 1-\alpha < \frac{1}{2}$, and inequality (5.8) simplifies to

$$|\operatorname{cut}(\mathcal{S})| \geq (1-\alpha)|\mathcal{B}_e| + \beta|\mathcal{B}_{\bar{e}}| + (\alpha-\beta)|\mathcal{B}_e|.$$

Hence,

$$|\operatorname{cut}(\mathcal{S})| \geq (1-\beta)|\mathcal{B}_e| + \beta|\mathcal{B}_{\bar{e}}| \geq \alpha|\mathcal{B}_e| + \beta|\mathcal{B}_{\bar{e}}| = |\mathcal{S}|,$$

as required.

Finally, suppose that $\beta > \frac{1}{2}$. Then necessarily $\alpha < \frac{1}{2}$ and inequality (5.8) simplifies to

$$\begin{aligned}|\operatorname{cut}(\mathcal{S})| &\geq \alpha|\mathcal{B}_e| + (1-\beta)|\mathcal{B}_{\bar{e}}| + (\beta-\alpha)|\mathcal{B}_e| \geq \beta|\mathcal{B}_e| + (1-\beta)|\mathcal{B}_{\bar{e}}| \\ &\geq (\alpha+\beta-\tfrac{1}{2})|\mathcal{B}_e| + \tfrac{1}{2}|\mathcal{B}_{\bar{e}}| \geq \alpha|\mathcal{B}_e| + \beta|\mathcal{B}_{\bar{e}}| = |\mathcal{S}|.\end{aligned}$$

This completes the inductive step. □

5.1.2 Regular matroids are balanced

A natural question now presents itself: how big is the class of balanced matroids? Recall that a regular matroid is one that is representable over every field. In this section we

prove 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 [**?**, 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 base. A *cut* is a minimal set of elements whose complement does not contain a base. 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 $D \in \mathcal{D}$:

$$\begin{aligned} \gamma(C, g) &\neq 0 \text{ iff } g \in C, \\ \delta(D, g) &\neq 0 \text{ iff } g \in D, \text{ and} \\ \sum_{g \in E} \gamma(C, g)\delta(D, g) &= 0. \end{aligned} \tag{5.11}$$

We work in this section towards the following result.

Theorem 5.1.2 *Orientable (and hence regular) matroids are balanced.*

A *near base* of M is a set $N \subseteq E$ that can be augmented to a base 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 base by the removal of a single element. A near base N defines a unique cut D_N consisting of all elements of the ground set whose addition to N results in a base. A unicycle U defines a unique cycle C_U consisting of all elements which whose removal from U results in a base. Let e, f be distinct elements of the ground set E. We claim that

$$\gamma(C_U, e)\gamma(C_U, f) + \delta(D_N, e)\delta(D_N, f) = 0, \tag{5.12}$$

for all near-bases N and unicycles U that are related by $U = N \cup \{e, f\}$. To see this, note that the equation (5.11) simplifies in this situation to

$$\gamma(C_U, e)\delta(D_N, e) + \gamma(C_U, f)\delta(D_N, f) = 0, \tag{5.13}$$

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 (5.13) are zero, in which case we are done. Otherwise, some quantity, say $\delta(D_N, e)$, is non-zero, in which case $D_N \cup \{e\} = C_U \setminus \{f\}$

[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.

is a base and $\gamma(C_U, f)$ is non-zero also. Multiplying (5.13) 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 (5.12) 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 (5.12), 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 (5.11) for orientability, nor changing the value of Δ_{ef}. With this convention we have

$$\sum_{g \neq e} \gamma(C, g)\delta(D, g) = 1, \quad \text{provided } C \ni e \text{ and } D \ni e; \tag{5.14}$$

$$\gamma(C_U, f) = \delta(D_N, f), \quad \text{provided } U = N \cup \{e, f\}; \tag{5.15}$$

and

$$\Delta_{ef} = \sum_{U : e \in C_U} \gamma(C_U, f) = \sum_{N : e \in D_N} \delta(D_N, f), \tag{5.16}$$

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 tends 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 5.1.3 *The bases $\mathcal{B} = \mathcal{B}(M)$ of an oriented 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 $g \neq e$:

$$X' = X \cup \{e\} \setminus \{g\} \tag{5.17}$$

$$Y' = Y \cup \{g\} \setminus \{e\}. \tag{5.18}$$

With each adjacent pair we associate a weight

$$\gamma(C_{X \cup \{e\}}, g)\delta(D_{Y \setminus \{e\}}, g). \tag{5.19}$$

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 (5.17) and (5.18) is precisely that the weight (5.19) is non-zero. Note that whenever this occurs, $(X', Y') \in \mathcal{B}_e \times \mathcal{B}_{\bar{e}f}$. Thus

$$\begin{aligned} |\mathcal{B}_{\bar{e}}| \cdot |\mathcal{B}_{ef}| &= \sum_{(X,Y) \in \mathcal{B}_{\bar{e}} \times \mathcal{B}_{ef}} \left[\sum_{g \neq e} \gamma(C_{X \cup \{e\}}, g) \delta(D_{Y \setminus \{e\}}, g) \right] \\ &= W, \end{aligned} \tag{5.20}$$

where W is the total weight of adjacent pairs. Here we have used equation (5.14).

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

$$\delta(D_{X' \setminus \{e\}}, g) \gamma(C_{Y' \cup \{e\}}, g) \tag{5.21}$$

to each adjacent pair, which is consistent, by (5.15), with the weight (5.19) applied earlier. Again, starting at (X', Y'), the condition that the pair (X, Y) obtained by inverting the exchange given in (5.17) and (5.18) is that the weight (5.21) in non-zero. But now, even if the weight is non-zero, there is a possibility that the new pair of bases (X, Y) will not be a member of $\mathcal{B}_{\bar{e}} \times \mathcal{B}_{ef}$; this will happen precisely when $g = f$. Thus

$$|\mathcal{B}_e| \cdot |\mathcal{B}_{\bar{e}f}| = \sum_{(X',Y') \in \mathcal{B}_e \times \mathcal{B}_{\bar{e}f}} \left[\sum_{g \neq e} \delta(D_{X' \setminus \{e\}}, g) \gamma(C_{Y' \cup \{e\}}, g) \right] \tag{5.22}$$

$$\begin{aligned} &= \sum_{(X',Y') \in \mathcal{B}_e \times \mathcal{B}_{\bar{e}f}} \left[\sum_{g \neq e,f} \delta(D_{X' \setminus \{e\}}, g) \gamma(C_{Y' \cup \{e\}}, g) \right] \\ &\qquad + \sum_{(X',Y') \in \mathcal{B}_e \times \mathcal{B}_{\bar{e}f}} \delta(D_{X' \setminus \{e\}}, f) \gamma(C_{Y' \cup \{e\}}, f) \\ &= W + \sum_{(X',Y') \in \mathcal{B}_e \times \mathcal{B}_{\bar{e}}} \delta(D_{X' \setminus \{e\}}, f) \gamma(C_{Y' \cup \{e\}}, f) \end{aligned} \tag{5.23}$$

$$\begin{aligned} &= W + \sum_{X' \in \mathcal{B}_e} \delta(D_{X' \setminus \{e\}}, f) \sum_{Y' \in \mathcal{B}_{\bar{e}}} \gamma(C_{Y' \cup \{e\}}, f) \\ &= W + \Delta_{ef}^2. \end{aligned} \tag{5.24}$$

Here, step (5.22) is by (5.14); step (5.23) uses the observation that terms are non-zero only when $f \in Y'$; and (5.24) is from the definition (5.16) of Δ_{ef}.

Comparing 5.20 and 5.24 we have

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

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

Proof of Theorem 5.1.2 According to Theorem 5.1.3, 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. □

Remark 1: (Flesh this out.) Number of bases of a regular matroid may be computed exactly in time ? by matrix-tree theorem + Gaussian elimination. This gives alternative polynomial-time sampling procedure. However, as we have seen, the class of balanced matroids is strictly larger than the class of regular matroids.

Remark 2: (Flesh this out.) There exist non-balanced matroids. Let M be a matroid of rank r on ground set E. For any $0 < r' < r$,

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

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

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

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

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

contradicting negative correlation.

5.2 Graphic matroids in particular

Since graphic matroids are balanced, the bases-exchange walk may be used to sample, efficiently and almost u.a.r., spanning trees in an undirected graph. However there are a number of other procedures for sampling bases in this special case, some of them providing *exactly* uniform samples. Perhaps the most efficient proposal is the "cycle-popping" technique of Wilson. We describe this now in the somewhat more general setting of sampling a directed tree in a directed graph.

5.2.1 Cycle popping: the general setting

Let $G = (V, A, r)$ be a directed graph with vertices V, arcs A and a distinguished root r. A *directed tree with root* r in G is a subgraph (V, T) in which there is a unique path from each vertex $v \in V$ to the root r. Note that a tree (we drop the qualifier "directed" at this point) has $n - 1$ arcs, and every vertex other than r has outdegree 1 (the root has

outdegree 0). Thus another way of viewing a tree is as a function $f : V \setminus \{r\} \to V$ that is cycle free: that is, $f^i(v) = v$ entails $i = 0$, for all v, i such that $f^i(v)$ is defined. One way to sample a tree is to select u.a.r. a function $f : V \setminus \{r\} \to V$ and accept if it is cycle-free. However the rejection probability will in general be high, as can be seen by considering the $n \times n$ grid: there are $O(n^2)$ disjoint 4-cycles in the grid, and each of them will, independently with probability 1/256, lead to a cycle in f. Thus the probability that f is cycle-free is exponentially small in the number of vertices. The idea behind the cycle-popping strategy is to remove cycles and re-randomise f on the affected vertices.

We first describe cycle-popping in a setting that is convenient for proof, but not for implementation. For each vertex $u \in V \setminus \{r\}$, we postulate a sequence $(S_u^0, S_u^1, S_u^2, \ldots) \in \Gamma(u)^\omega$ of r.v's, where $\Gamma(u) = \{v : (u, v) \in A\}$ is the set of neighbours of u. Each r.v. S_u^i is distributed uniformly over $\Gamma(u)$, and is independent of all the other r.v's. We call the indices i "colours." At any instant there is a visible colour $c(u)$ at vertex u; initially, $c(u) = 0$ for all $u \in V \setminus \{r\}$. As time progresses, higher colours become visible, corresponding to r.v's further along the lists being revealed.

Consider the following procedure, guided by the r.v's (S_u^i). Let the currently visible colours be $c : V \setminus \{r\} \to \mathbf{N}$, and consider the function $f : V \setminus \{r\} \to V$ given by $f(u) = S_u^{c(u)}$ for all $u \in V \setminus \{r\}$. The digraph $D_f = (V, \{(v, f(v)) : v \in V \setminus \{r\}\})$ has the following structure. The weak component containing r is a directed tree with root r. Every other weak component consists of a single cycle plus disjoint directed trees rooted at a vertex of the cycle. If f is cycle-free we are done. Otherwise, select an ℓ-cycle $C = (u, f(u), f^2(u), \ldots, f^{\ell-1}(u))$ and "pop" it; that is, increment $c(v)$ for all vertices v on the cycle C, revealing a fresh set of colours/r.v's. This process, if iterated, might continue indefinitely, but if it terminates, f will define a tree in G with root r. We shall argue that the process terminates with probability 1, and that the tree produced is exactly uniform.

The cycle-popping process is nondeterministic, since a number of cycles may be available for popping at any instant. The key observation is that the order of popping does not matter: if the process terminates then it always terminates with the colour labelling c (and hence the same cycle-free function f). Consider any configuration of the process, uniquely determined by the colour assignment c. A number of cycles $C_0, \ldots, C_{s-1}$ may be available for popping. Assume $s \geq 2$ and that C_j and C_k are distinct cycles. Necessarily, C_j and C_k are disjoint, so that if we decide to pop C_j first we can then pop C_k and be in exactly the same configuration as if we had popped C_k first and then C_j. Thus the process has the "diamond property" and hence is Church-Rosser: either the process continues indefinitely, or terminates at a well defined configuration independent of the order in which cycles are popped. This is a result of Newman [**?**], see also Sperschneider and Antoniou [**?**].

In fact, more is true. Label each transition of the cycle-popping process by the *coloured* cycle—i.e., the sequence of vertices u on the cycle together with their corresponding

colours $c(u)$—whose popping generates that transition. In each diamond, the same two coloured cycles are involved in the two paths through the diamond. Thus in any sequence of transitions leading to the unique terminating configuration (assuming it exists) exactly the same set of coloured cycles are popped, only the order of popping varies. Thus we can think of the cycle-popping process (assuming it terminates) as defining an underlying tree T, rooted at r, on which are superimposed a partially ordered set $\mathcal{C}$ of coloured cycles. Conditioned on the set $\mathcal{C}$ of cycles, the tree T is uniform. Thus, conditioned on termination, the cycle-popping process generates a rooted tree u.a.r. We collect these discoveries in the following theorem.

Theorem 5.2.1 *The order in which cycles are popped in the cycle-popping process is of no consequence: for a given collection of r.v's (S_u^i) the process either always continues indefinitely, or always terminates at the same configuration (colouring) c. Conditioned on termination, the tree defined by the colouring c is distributed uniformly.*

In fact, termination occurs with probability 1, but this is easier to appreciate once we move to an alternative, more implementation-friendly version of the process.

5.2.2 Cycle popping: the implementation

The process of the previous section is straightforwardly implementable, provided we view the r.v's (S_u^i) as being revealed to us on demand. We know from Theorem 5.2.1 that the order in which the cycles are popped is of no consequence. A particularly elegant way of performing the computations is to perform a random walk on G, popping cycles as soon as they are discovered. (Refer to Figure 5.1.) We shall refer to this particular implementation of cycle-popping as the *cycle-erased random walk*, even though it is more commonly called the *loop-erased random walk* in the literature. Note that by storing the current function f as an array $Tree[\,]$, the effect of popping a cycle is achieved automatically through overwriting an array element.

Perhaps we should relax this.

Assume that G is strongly connected and contains at least one odd cycle, so that the random walk on G has a well defined stationary distribution π.

Theorem 5.2.2 *The procedure call* TreeSample(G) *halts with probability 1, returning a uniform random tree in G, rooted at r. The expected running time of* TreeSample(G) *is proportional to $\sum_u \pi(u)C_{u,r}$, i.e., the expected commute time between r and a π-random vertex.*

Proof The running time of TreeSample is proportional to the number of steps in the simulated random walk. For each $u \neq r$ we estimate the number of steps that are taken *from* u; the total number of steps will be the sum of these. The key observation

```
TreeSample(G, r)
begin
    InTree[u] ← false, for all u ∈ V \ {r};
    InTree[r] ← true;
    for all s ∈ V:
        u ← s;
        while not InTree[u]:
            Select v ∈ Γ(u), u.a.r.;
            Tree[u] ← v;
            u ← v;
        u ← s;
        while not InTree[u]:
            InTree[u] ← true;
            u ← Tree[u];
    return Tree
end
```

Figure 5.1: An implementation of the cycle-popping strategy.

is that the number of steps from u is one greater that the number of coloured cycles containing u that are popped. So the number of steps from u is dependent on u and the r.v's (S_u^i), but *not* on the order in which the starting points are considered by TreeSample, i.e., the order in which vertices s are taken in the outer loop. Thus we may assume without loss of generality that $s = u$ is the first vertex to be selected. The expected number of steps from u is the expected number of visits to u (including the visit at time 0) made by a random walk started at u before hitting r. The latter quantity is $\pi(u)C_{u,r}$, see Lemma 2.5.2. The result now follows from Theorem 5.2.1. □

The commute time between any pair of vertices is bounded by the twice the cover time. Thus the expected running time of TreeSample(G) when applied to an undirected graph G (i.e., all directed edges occur in antiparallel pairs) is $O(nm)$, where m is the number of edges, see Lemma 2.5.4.

5.3 Independent sets in matroids: forests

The *independent sets* $\mathcal{I}(M)$ of a matroid $M = (E, \mathcal{B})$ are the subsets of the ground set E that may be extended to a base; thus, $\mathcal{I}(M) = \{I \subseteq E : \exists X \in \mathcal{B}(M).\, X \supseteq I\}$. In the case of the cycle matroid of a graph G, the independent sets are the forests in G. There is a natural random walk on the independent sets of a matroid. Suppose the current independent set is I. Select an element $e \in E$ of the ground set u.a.r., and let $I' = I \oplus \{e\}$; if $I' \in \mathcal{I}(M)$ then move to I', otherwise remain at I. (If desired,

exchange moves akin to those employed in the bases-exchange walk may be added.) At first sight, performing a random walk on all independent sets rather than just the maximal independent sets (i.e., bases) appears to allow more freedom, only increasing the potential for rapid mixing. However, this initial impression is misleading, and it is not known whether the natural random walk on independent sets is rapidly mixing, even in the special case of graphic matroids. Exponential mixing time is consistent with our present knowledge.

There is another way to connect forests with matroids. The set of all k-edge forests in a graph G can be viewed as the set of bases of a matroid: a truncation of the cycle matroid of G. Unfortunately, we saw at the end of §5.1.2 that the truncation of a graphic matroid is not necessarily balanced, so we cannot employ the machinery so far established.

However, there is a special situation where we do know how to sample forests in a graph G, and that is when G is sufficiently dense. For $\alpha > 0$ we say that a graph G is *α-dense* if every vertex in G has degree at least αn. The main result of this section is that there is a polynomial-time uniform sampler for forests in α-dense graphs. (The degree of the polynomial governing the runtime of the sampler grows unboundedly as $\alpha \to 0$.) The idea, due to Annan [**?**], is to reduce the current problem to the already solved problem of sampling spanning trees.

Theorem 5.3.1 *Suppose we have a procedure* TreeSample(H) *for sampling, u.a.r., spanning trees in a graph H. There is a polynomial-time algorithm* ForestSample *that, given access to* TreeSample*, takes an α-dense graph G as input and satisfies the following specification:*

- ForestSample *either produces a forest in G or no output; the output distribution, conditioned on there being an output, is uniform over all forests in G.*
- ForestSample *produces an output with probability at least $\frac{1}{2}$.*
- *The number of calls to* TreeSample *is bounded by $n^{4/\alpha}$.*

Let $G = (V, E)$ be an n-vertex graph with vertex set V and edge set E. Denote by G^+ the derived graph with vertex set $V^+ = V \cup \{t\}$ and edge set $E^+ = E \cup \{\{v, t\} : v \in V\}$. Each spanning tree (V^+, T) in G^+ projects to a forest (V, F) in G, where the edge set of the forest is simply $F = T \cap E$. Moreover, every forest in G may be derived from at least one spanning tree in G^+ by projecting in this way. This observation in itself does not provide a reduction from forest sampling to spanning tree sampling, as the number of distinct spanning trees in G^+ projecting to given forest F varies widely as a function of F. At one extreme, the forest consisting of n trivial components arises in just one way; while, at the other, the forest consisting of $n/2$ components of size 2 (i.e., a perfect matching, assuming n is even) arises in $2^{n/2}$ ways. In general, the number of spanning

```
ForestSample(G)
begin
    Construct G+ as described in the text;
    repeat 2n^{4/α} times, or until successful:
        (V+, T) ← TreeSample(G+);
        F ← T ∩ V
        if (V+, T) is the canonical tree for forest (V, F) return (V, F)
end
```

Figure 5.2: A procedure for sampling forests in an α-dense graph.

trees corresponding to a specified forest is the product of the sizes of the connected components forming that forest.

To overcome this problem we nominate a canonical spanning tree in G^+ for each forest F.[3] For example, assume a linear ordering on the vertices of V and deem a spanning tree T canonical if it contains edges from t to the least vertex in every connected component of G. Clearly, this rule results in one canonical tree for each forest. Certainly, then, the output distribution of the procedure presented in Figure 5.2 is uniform over forests in G. What is not immediately clear is that the procedure will ouput *some* forest with probability at least $\frac{1}{2}$. The key fact we need to prove this is the following.

Lemma 5.3.1 *Suppose the n-vertex graph G is α-dense, and let G^+ be the derived graph as defined above. Let $\{v,t\}$ be any of the n edges in G^+ with an endpoint at t. Select T u.a.r. from the set of all spanning trees in G^+. Then the probability that T contains the edge $\{v,t\}$ is at most $2/(\alpha n+2)$.*

Proof In Annan's proof, this lemma is established using connections between spanning trees and resistances in electrical networks. For us, it is more convenient to appeal to the property of balance, since we have already set up the machinery. Let $e \in E$ be any edge in G. Since graphic matroids are balanced, the probability that edge $\{v,t\}$ is contained in a random spanning tree of G^+ is not decreased by the removal of edge e. Removing in turn all edges $e \in E$ that are not incident at v yields a subgraph of G^+ whose only remaining edges are $(\{t\} \times V) \cup (\{v\} \times \Gamma(v))$, where $\Gamma(v)$ denotes the set of neighbours of v in G. By direct calculation, the number of spanning trees in this vestigial graph which contain (respectively, do not contain) the edge $\{t,v\}$ is 2^k (respectively, $k2^{k-1}$), where $k = |\Gamma(v)| \geq \alpha n$. Thus the probability that $\{t,v\}$ is contained in a random spanning tree of G^+ is at most $2^k/(k2^{k-1}+2^k) \leq 2/(\alpha n+2)$, as claimed. □

Lemma 5.3.1 assures us that a typical random spanning tree in G^+ projects to a forest with few components.

[3]For convenience, in the remainder of the section, we blur the distinction between a forest or tree and the edges that compose it. Since all forests and trees are spanning, this will cause no confusion.

Proof of Theorem 5.3.1 It is only the second of the three claims in the statement of the theorem that remains to be proved. Let T be a spanning tree in G^+ selected u.a.r., and let F be the derived forest in G. According to Lemma 5.3.1, the expected degree of vertex t in T is at most $2/\alpha$. By Markov's inequality, with probability at least $\frac{1}{2}$, the degree of t is no greater than $4/\alpha$. Conditioned on the event that the degree of t is at most $4/\alpha$, the probability that T is canonical for F is at least $n^{-4/\alpha}$. So the probability that *some* forest is output is at least $\frac{1}{2}n^{-4/\alpha}$. The probability that *none* of the $n^{4/\alpha}$ trials produces an output is therefore bounded above by $(1-\frac{1}{2}n^{-4/\alpha})^{2n^{4/\alpha}} \leq 1/e \leq \frac{1}{2}$. □

As a simple corollary of the above, we can show how to generate and count *trees* (not necessarily spanning) in a dense graph.

Corollary 5.3.1 *Let G be an α-dense graph. There is a good sampler and an FPRAS for the set of* all *trees of G.*

Proof Consider the following algorithm:

(i) Choose forest F at random.

(ii) Accept if F contains one non-trivial tree plus a collection of isolated vertices.

We claim that

$$\mathbf{Pr}(\text{accept in (ii)}) \geq n^{-\lfloor 4/\alpha \rfloor}. \tag{5.25}$$

Let f_k denote the number of forests with k non-trivial trees and $f = \sum_{k=1}^{n} f_k$ and let $b = \lfloor 4/\alpha \rfloor$. Then

(a) $\mathbf{Pr}(\text{accept in (ii)}) \geq \frac{f_1}{f}$.

(b) $f_1 + f_2 + \cdots + f_b \geq f/2$.

(c) $f_{k+1} \leq n f_k$.

Here, (a) is clear, (b) follows from Lemma 5.3.1 and (c) is a consequence of the fact that we can obtain all $k+1$ tree forests by deleting an edge of a k tree forest. So

$$\begin{aligned} f_1 + f_2 + \cdots + f_b &< (1 + n + n^2 + \cdots + n^{b-1}) f_1 \\ &= \frac{n^b - 1}{n-1} f_1 \end{aligned}$$

and the result follows. □

Chapter 6

Some other approaches

In this chapter we describe some approximate counting problems that can be solved without the use of Markov chains.

6.1 Satisfiability

Here we are given a Boolean function F in *Disjunctive Normal Form* (DNF) e.g.

$$F = x_1x_2x_3 + \bar{x}_1x_2x_4 + x_3\bar{x}_5x_6x_7$$

and our task is to estimate the number of satisfying assignments of the variables.

Thus assume we have n Boolean variabes, $x_1, x_2, \ldots, x_n$ and

$$F = m_1 + m_2 + \cdots + m_r \text{ where } m_i = \prod_{j=1}^{n} x_j^{\alpha_{i,j}}$$

and $\alpha_{i,j} \in \{0, \pm 1\}$ and $x_j^0 = 1, x_j^1 = x_j, x_j^{-1} = \bar{x}_j$.

Let $A = \{0,1\}^n$ be the set of possible assignments of 0/1 values to the variables and let $A^* = \{a \in A : \ F(a) = 1\}$. The task is then to estimate $|A^*|$.

Let $A_i = \{a \in A : \ m_i(a) = 1\}$, $i = 1, 2, \ldots, r$. Then we are faced with the following problem:
Cardinality of the Union Problem
Given sets $A_1, A_2, \ldots, A_r \subseteq A$, estimate $|A^*|$ where $A^* = \bigcup_{i=1}^{r} A_i$.

The approach given here is valid if

(i) r is small, i.e. polynomial in the description of the problem.

(ii) $|A_i|$ is known for each i.

(iii) It is possible to efficiently choose a random element of A_i for each i.

(iv) It is possible to efficiently decide whether a given $a \in A$ lies in A_i for each i.

For the problem to be interesting, the A_i need to be large, i.e. exponential in the description of the problem.
It is clear that the DNF problem satisfies the conditions (i)–(iv).

Another way of looking at this problem is that we have a $r \times |A|$ 0/1 matrix M where $M(i,a) = 1$ iff $a \in A_i$. Let ρ_i be the number of 1's in row i and let $\rho = \rho_1 + \rho_2 + \cdots + \rho_r$ be the total number of 1's in M. Let c_a denote the number of 1's in column a and let $A^* = \{a : \ c_a > 0\}$.. Now we want to estimate $\nu = |A^*|$. Consider the following algorithm: N is a parameter to be determined later.

MATRIX COLUMN WEIGHT ALGORITHM

begin
Compute $p_i = \rho_i/\rho$ for $i = 1, 2, \ldots, r$.
For $t = 1$ **to** N **do**
begin
(1a) Choose i_t randomly from $[r]$ according to distribution $p_1, p_2, \ldots, p_r$.
(1b) Choose a_t randomly from $\{a : \ M(i_t, a) = 1\}$.
(1c) Compute $Z_t = \rho/c_{a_t}$.
Output $\bar{Z} = \frac{Z_1 + Z_2 + \cdots + Z_N}{N}$.
end
end

The next lemma evaluates the accuracy of this procedure.

Lemma 6.1.1

$$\mathbf{Pr}(|\bar{Z} - \nu| \geq \epsilon\nu) \leq \frac{r}{\epsilon^2 N}.$$

Proof Fix $t = 1$. We claim that (i_t, a_t) is chosen uniformly at random from the set $\{(i,a) : \ M(i,a) = 1\}$. Indeed (1a) determines row i_t with probability proportional to the number of 1's in a row and then (1b) chooses a random member o f the row. Thus we see that for $a \in A$

$$\mathbf{Pr}(a_1 = a) = \frac{c_a}{\rho}.$$

Thus

$$\mathbf{E}(Z_1) = \sum_{a \in A} \mathbf{E}(Z_1 \mid a_t = a)\mathbf{Pr}(a_t = a) = \sum_{a \in A^*} \frac{\rho}{c_a} \cdot \frac{c_a}{\rho} = |A^*|.$$

Now we estimate the variance.

$$\mathbf{E}(Z_1^2) = \sum_{a \in A^*} \frac{\rho^2}{c_a^2} \cdot \frac{c_a}{\rho} = \rho \sum_{a \in A^*} \frac{1}{c_a} \leq \rho\nu \leq r\nu^2.$$

Thus

$$\mathbf{Var}(Z_1) \leq r\mathbf{E}(Z_1)^2$$

which implies

$$\mathbf{Var}(\bar{Z}) \leq \frac{r}{N}\mathbf{E}(\bar{Z})^2$$

and the result follows from the Chebychef inequality. □

Putting $N = 4r\epsilon^{-2}$ we see that

$$\mathbf{Pr}(|\bar{Z} - \nu| \geq \epsilon\nu) \leq \frac{1}{4}.$$

This probability can be reduced to δ as we did in (1.2) by repeating the algorithm $\lceil 12\ln(2/\delta) \rceil$ times and taking the median result.

6.1.1 Random assignments

We consider a generalisation of the DNF problem which will be useful in Section 6.2. We consider the problem of estimating the probability that a randomly generated assignment a satisfies a Boolean formula F, given in DNF. Thus we are given $0 < p < 1$ and suppose that assignment $a \in A$ is chosen by independently putting $x_j = 1$ with probability p. We wish to estimate

$$\Delta_p = \mathbf{Pr}(F(a) = 1).$$

If $p = 1/2$ then $\Delta_p = 2^{-n}|A^*|$ and so this problem generalises the problem of the previous section.

For $a \in A$ let $s(a) = |\{j : a_j = 1\}$. Let $A_k = \{a \in A : s(a) = k\}$ and $A_k^* = A^* \cap A_k$ for $k = 0, 1, 2, \ldots n$. Then

$$\Delta_p = \sum_{k=0}^{n} |A_k^*| p^k (1-p)^{n-k}$$

and so we can estimate Δ_p efficiently if we can estimate the $|A_k^*|$ efficiently. So let $A_{i,k}^* = \{a \in A_k^* : m_i(a) = 1\}$ so that $A_k^* = \bigcup_{i=1}^r A_{i,k}^*$. It remains to check that the sets $A_{i,k}$ satisfy requirements (i)–(iv) above. (i) holds. Let $n_i = |\{j : \alpha_{i,j} = 1\}|$ and $\bar{n}_i = |\{j : \alpha_{i,j} = -1\}|$ then $|A_{i,k}| = \binom{n-n_i-\bar{n}_i}{k-n_i}$ and so (ii) holds. This calculation shows that (iii) holds and (iv) still holds. It follows that the $|A_k^*|$ can be estimated, along with Δ_p.

It is more efficient to estimate Δ_p directly by modifying the MATRIX COLUMN WEIGHT ALGORITHM. We briefly indicate the steps. We write

$$\Delta_p = (1-p)^n \sum_{a \in A^*} x^{s(a)} \qquad \text{where } x = \frac{p}{1-p}.$$

Then define the $r \times |A|$ matrix M with $M(i,a) = x^{s(a)}$ if $m_i(a) = 1$ and $M(i,a) = 0$ otherwise. Now define $\rho_i = \sum_{a \in A} M(i,a)$ and p_i and ρ as above. The only change that we need to make to the algorithm is to replace (1b) by
(1b') Choose a_t from $\{a : \ M(i_t, a) = 1\}$ with probability proportional to $x^{s(a)}$.
With these changes, (i_t, a_t) is chosen with probability proportional to $x^{s(a)}$ and the output of the algorithm has expectation equal to $\sum_{a \in A^*} x^{s(a)}$. The proof of Lemma 6.1.1 goes through minor changes.

6.2 Reliability

Here we are given a graph $G = (V, E)$, $n = |V|, m = |E|$ and a probability $0 < p < 1$. Let $G_p = (V, E_p)$ be the random subgraph of G obtained by independently including each $e \in E$ with probability p. This models a network where each link (edge) *fails* independently with probability $q = 1 - p$. The task is to estimate

$$FAIL(p) = \mathbf{Pr}(G_p \text{ is not connected}).$$

If $FAIL(p)$ is large then this can be estimated easily. We simply generate random copies $G_1, G_2, \ldots, G_N$ of G_p and let

$$\delta_i = \begin{cases} 1 & G_i \text{ is not connected} \\ 0 & G_i \text{ is connected} \end{cases}$$

We can then estimate $FAIL(p)$ by $\bar{\delta} = \frac{\delta_1 + \cdots + \delta_N}{N}$ where N is given in Lemma 6.2.1 below.

Let κ denote the minimum size of a cut in G i.e. $\min_{S \subseteq V} |S : \bar{S}|$ where $S : \bar{S}$ is the set of edges with one end in S and the other in $\bar{S} = V \setminus S$. We see immediately that

$$FAIL(p) \geq q^\kappa$$

since this is the probability that any given minimum cut *fails* in G_p i.e. contains no G_p edges.

Lemma 6.2.1 *Assume $q^\kappa \geq n^{-4}$ and let $N = 4n^4\epsilon^{-2}$. Then*

$$\mathbf{Pr}(|FAIL(p) - \bar{\delta}| \geq \epsilon FAIL(p)) \leq \frac{1}{4}.$$

Proof $\bar{\delta}$ has mean $\phi = FAIL(p)$ and variance $\phi(1-\phi)N^{-1}$ and so by the Chebychef inequality

$$\mathbf{Pr}(|FAIL(p) - \bar{\delta}| \geq \epsilon FAIL(p)) \leq \frac{\phi(1-\phi)}{\epsilon^2\phi^2 N}$$

and the result follows from $\phi \geq n^{-4}$. □

Note that each δ_i can be computed in $O(m)$ time and that κ can be computed in $O(n^3)$ time so that we can decide when to use Lemma 6.2.1.

The interesting case is of course when $FAIL(p)$ is small. A cut $S : \bar{S}$ with $\alpha\kappa$ edges is called an α-minimum cut. The algorithm we describe rests on the following two theorems:

Theorem 6.2.1 *G has at most $12n^{2\alpha}$ cuts of size at most $\alpha\kappa$.*

Proof (Deferred to Section 6.3).

Theorem 6.2.2 *Suppose $q^\kappa = n^{-(2+\delta)}$ for some $\delta > 0$. Then*

$$\mathbf{Pr}(\exists \text{ an } (\geq \alpha)\text{-minimum cut which fails}) \leq n^{-\alpha\delta}\gamma 12^\gamma$$

where $\gamma = 1 + 2/\delta$.

Proof (Deferred to Section 6.3).

Now consider the following algorithm:

$$q^\kappa = n^{-(2+\delta)} \text{ and } \alpha_0 = 2 - \frac{\log(\epsilon/1000)}{\log n}.$$

Reliability algorithm

1. Enumerate the cuts $S_i : \bar{S}_i$, $i = 1, 2, \ldots, \nu$ of size at most $\alpha_0\kappa$.
2. Compute an $\epsilon/2$-approximation Φ to
$$\Phi_0 = \mathbf{Pr}(\exists 1 \leq i \leq \nu : \ S_i : \bar{S}_i \text{ fails}).$$
3. Output Φ.

The cuts $S_i : \bar{S}_i$, $i = 1, 2, \ldots, \nu$ can be found in polynomial time (see Section 6.3).

Step 2 is executed as follows: Suppose we assign a set of Boolean variables $x_e.e \in E$. Consider the Boolean formula

$$F = F_1 + F_2 + \cdots + F_\nu$$

where

$$F_i = \prod_{e \in S_i : \bar{S}_i} x_e.$$

The edges of G_p define a (random) assignment of values to the x_e i.e. $x_e = 1$ iff edge e does *not* occur in G_p. Then $F_i = 1$ iff cut $S_i : \bar{S}_i$ fails and so $F = 1$ iff $\exists i \;\; S_i : \bar{S}_i$ fails. Thus we can use the algorithm of Section 6.1.1 to compute an $\epsilon/2$-approximation to $\Phi_0 = \mathbf{Pr}(F = 1)$ and so carry out Step 2. We execute this algorithm so the probability of failure is at most $\frac{1}{4}$. Thus the RELIABILITY ALGORITHM can be executed in polynomial time. It remains to prove

Theorem 6.2.3 *If $q^\kappa = n^{-(2+\delta)}$, $\delta \geq 2$, then*

$$\mathbf{Pr}(|\Phi - FAIL(p)| \geq \epsilon FAIL(p)) \leq \frac{1}{4}.$$

Proof Let

$$\Phi_1 = \mathbf{Pr}(\exists \text{ an } (\geq \alpha)\text{-minimum cut which fails}).$$

It follows from Theorem 6.2.2 that

$$\frac{\Phi_1}{FAIL(p)} \leq \frac{288 n^{-\alpha_0 \delta}}{n^{-(2+\delta)}} \leq \frac{\epsilon}{3}.$$

Now

$$\Phi_0 \leq FAIL(p) \leq \Phi_0 + \Phi_1$$

and so the $\frac{\epsilon}{2}$-approximation Φ to Φ_0 is an ϵ-approximation to $FAIL(p)$ and the theorem follows. □

6.3 Deferred Proofs

6.3.1 Proof of Theorem 6.2.1

We use the following CONTRACTION ALGORITHM to produce a cut. Each cut of size at most $\alpha\kappa$ will have probability at least $\frac{1}{12} n^{-2\alpha}$ of being chosen and the theorem follows.

CONTRACTION ALGORITHM

begin
  $k = \lceil 2\alpha \rceil$, $H \leftarrow G$.
  while $|V(H)| > k$ **do**
  begin
A Choose e randomly from $E(H)$.

$H \leftarrow H \setminus e$ – contract e.
end
B **begin**
Let $K : \bar{K}$ be a random partition of $V(H)$ into 2 non-empty subsets.
"Expand" K into $S \subseteq V(G)$.
Output S
end
end

We need to explain "expand" K. When we contract edge $\{v, w\}$, the two vertices v, w are replaced by a single new vertex. Thus, in general, the vertices of H at Step B correspond to (disjoint) subsets of V. Thes $S = \bigcup_{v \in K} v$.

We note next that the minimum cut size of H is at least κ throughout. (H contains parallel edges.) This because the cutsets of H are a subset of the cutsets of G. In particular, H has minimum degree at leat κ.

We now consider a fixed α-minimum cut C of G. We will output S if (i) no edge of C is chosen at Step A and (ii) the contracted version of C is chosen at Step B.

After t executions of Step A, H will have $n - t$ vertices. Assume that no edge of C has been contracted. H has at least $\frac{1}{2}(n-t)\kappa$ edges and so the probability we do not choose $e \in C$ at the next iteration is at least $1 - \frac{2\alpha}{n-t}$. Thus the probability we choose C is at least

$$2^{1-k} \prod_{r=k+1}^{n} \left(1 - \frac{2\alpha}{r}\right) = \prod_{i=0}^{k-1} \frac{2k - 2\alpha - i}{n-i} \prod_{i=k}^{n-k-1} \frac{n - i + k - 2\alpha}{n-i}$$
$$\geq \frac{2^{1-k}k!}{n^k} \exp\left\{ \sum_{i=k}^{n-k-1} \frac{f}{n-i} - \frac{1}{2} \sum_{i=k}^{n-k-1} \frac{f^2}{(n-i)^2} \right\}$$

where $f = k - 2\alpha$. Now

$$\exp\left\{ \sum_{i=k}^{n-k-1} \frac{f}{n-i} - \frac{1}{2} \sum_{i=k}^{n-k-1} \frac{f^2}{(n-i)^2} \right\} \geq n^f (ek)^{-1} e^{-\pi^2/12}$$

and so the probability we choose C is at least

$$\frac{2^{1-k}k!}{ekn^{2\alpha}e^{\pi^2/12}} \geq \frac{1}{12n^{2\alpha}}.$$

□

We see immediately that if we run the CONTRACTION ALGORITHM $O(n^{2\alpha} \log n)$ times then **whp** we will produce all cuts of size $\alpha\kappa$ or less.

6.3.2 Proof of Theorem 6.2.2

Let $\kappa = \kappa_1 \leq \kappa_2 \leq \cdots \leq \kappa_r$ be an enumeration of the cut sizes in G. We bound

$$\Delta = \sum_{i=i_0}^{r} q^{\kappa_i} \qquad \text{where } i_0 = \min\{i : \ \kappa_i \geq \alpha\kappa\}$$

which bounds the probability that a large cut fails.

Theorem 6.2.1 implies that

$$\kappa_i \geq \max\left\{\alpha, \frac{\log(i/12)}{2\log n}\right\}\kappa \qquad \text{for } i \geq i_0.$$

Thus

$$\begin{aligned}\Delta &\leq \sum_{i\leq 12n^{2\alpha}} q^{\alpha\kappa} + \sum_{i>12n^{2\alpha}} n^{-(2+\delta)\log(i/12)/(2\log n)} \\ &\leq 12n^{-\alpha\delta} + \int_{12n^{2\alpha}}^{\infty} \left(\frac{x}{12}\right)^{-(1+\delta/2)} dx \\ &\leq 12n^{-\alpha\delta} + 12^{1+\delta/2}(12n^{2\alpha})^{-\delta/2}\end{aligned}$$

and the result follows.

6.4 Tutte Polynomial in Dense Graphs

For a graph G, the *Tutte Polynomial* $T_G(x, y)$ is a bivariate polynomial which for many values of x, y evaluates to interesting graph invariants e.g. $T_G(1, 1)$ equals the number of spanning trees of G. We define it here by

$$T_G(x, y) = \sum_{A\subseteq E} (x-1)^{\kappa(A)-1}(y-1)^{|A|+\kappa(A)-n} \tag{6.1}$$

where $\kappa(A)$ is the number of components of $G_A = (V, A)$.

Some more interesting evaluations

- $T_G(2, 1)$ is the number of forests of G.
- $T_G(1, 2)$ is the number of forests of edge sets which contain a spanning tree of G.
- $T_G(2, 0)$ is the number of orientations of the edges of G which do not contain a directed cycle.

- $(-1)^{n-\kappa(E)}\lambda^{\kappa(E)}T_G(1-\lambda)$ is the chromatic polynomial of G i.e. the coefficient of λ^k in this polynomial is the number of proper k-colourings of the vertices of G.
- $1 - FAIL(p) = q^{|E|-n+1}p^{n-1}T_G(1, 1/q)$ when G is connected.

It turns out that the hyperbolae H_α defined by

$$H_\alpha = \{(x, y) : (x-1)(y-1) = \alpha\}$$

play a special role in the theory.

- Along H_1, $T_G(x, y) = x^{|E|}(x-1)^{n-\kappa(E)-|E|}$.
- Along H_Q, for general positive integer Q, T_G specialises to the partition function of the Potts model of statistical physics.

There are several other important evaluations. Given the expressive power of this polynomial, it is not surprising that apart from a few special points and 2 special hyperbolae, the exact evaluation of T_G is $\#P$-hard even for the very restricted class of planar bipartite graphs. Here we consider *dense* graphs and prove the existence of an FPRAS for $T_G(x, y)$ whenever $x, y > 1$.

For $0 < \alpha < 1$, let $\mathcal{G}_\alpha$ denote the set of graphs $G = (V, E)$ with $|V| = n$ and minimum degree $\delta(G) \geq \alpha n$. A graph is α-dense if it is a member of $\mathcal{G}_\alpha$ or, somewhat loosely, dense if we omit the α.

A first easy, but essential, observation is the following. Let G_p denote the random graph obtained by selecting edges of G independently with probability p.

Lemma 6.4.1 *Assume G is connected with n vertices and m edges. Assume $x, y > 1$ and let $p = (y-1)/y$ and $Q = (x-1)(y-1)$. Let $\kappa = \kappa(G_p)$ be the number of components of G_p. Then*

$$T_G(x, y) = \frac{y^m}{(x-1)(y-1)^n}\mathbf{E}(Q^\kappa).$$

Proof It follows from (6.1) that

$$\begin{aligned} T_G(x, y) &= \frac{y^m}{(x-1)(y-1)^n}\sum_{A\subseteq E}\left(\frac{y-1}{y}\right)^{|A|}\left(\frac{1}{y}\right)^{m-|A|}((x-1)(y-1))^{\kappa(A)} \\ &= \frac{y^m}{(x-1)(y-1)^n}\sum_{A\subseteq E} Q^{\kappa(A)}\mathbf{Pr}\{G_p = G_A\}. \end{aligned}$$

□

We now describe a property of dense graphs which is the key to much of the ensuing analysis. Let $N(v), v \in V$ denote the set of neighbours of v. Define $G^* = (V, E^*)$ by $(u, v) \in E^*$ if and only if $|N(u) \cap N(v)| \geq \alpha^2 n/2$. Let

$$s = \lceil 2/\alpha \rceil - 1.$$

Lemma 6.4.2 *Among any s+1 vertices of G, there are two which are adjacent in G^*.*

Proof Suppose there exist $v_1, v_2, \ldots v_{s+1}$ such that $|N(v_i) \cap N(v_j)| < \alpha^2 n/2$ if $i \neq j$. But then

$$\begin{aligned}\left|\bigcup_{i=1}^{s+1} N(v_i)\right| &\geq \sum_{i=1}^{s+1} |N(v_i)| - \sum_{i \neq j} |N(v_i) \cap N(v_j)| \\ &> (s+1)\alpha n - \binom{s+1}{2}\frac{\alpha^2 n}{2} \\ &= (s+1)\alpha n\left(1 - \frac{s\alpha}{4}\right) \\ &\geq n.\end{aligned}$$

□

Let $\hat{Q} = \max\{Q, Q^{-1}\}$ and $\zeta = y^m/((x-1)(y-1)^n)$.

We claim that the following algorithm estimates $T_G(x, y)$ for $G \in \mathcal{G}_\alpha$.

Algorithm EVAL
begin
  $p := \frac{y-1}{y}$; $Q := (x-1)(y-1)$; $t := \lceil 16\hat{Q}^{2s}\epsilon^{-2} \rceil$;
  for $i = 1$ **to** t **do**
  begin
    Generate G_p; $Z_i := Q^{\kappa(G_p)}$
  end
  $\tilde{Z} := \frac{Z_1 + Z_2 + \cdots + Z_t}{t}$;
  Output $Z = \zeta\tilde{Z}$
end

We first prove

Lemma 6.4.3 *In the notation of Lemma 6.4.1, let*

$$n_0 = \min\left\{n : n \geq \max\left\{\frac{24\ln(n\hat{Q})}{\alpha^2 p^2}, Q^{20/\alpha^2}\right\}\right\}.$$

If $n \geq n_0$ then

$$Q \geq 1 \text{ implies } \mathbf{E}(Q^{2\kappa}) \leq 2Q^{2s}.$$
$$Q < 1 \text{ implies } \mathbf{E}(Q^{\kappa}) \geq Q^s/2.$$

Proof Let $\mathcal{E}_u$ denote the event $\{\kappa(G_p) \geq u+s+1\}$ for $1 \leq u \leq u_0 = \lfloor \alpha^2 n/8 \rfloor$. If $\mathcal{E}_u$ occurs we choose $X = \{x_1, x_2, \ldots, x_{u+s+1}\}$ with each x_i from a different component of G_p. Lemma 6.4.2 implies that we can choose $y_1, y_2 \in X$ such that y_1, y_2 are adjacent in G^*. Repeating the argument yields a matching $\{y_1, y_2\}, \ldots, \{y_{2t-1,2t}\}$ in G^* where $t = \lceil (u+1)/2 \rceil$ and $y_1, y_2, \ldots, y_{2t}$ each lie in different components of G_p. The probability that G_p contains no path of length 2 connecting y_{2i-1} to y_{2i} for each i, $1 \leq i \leq t$ is at most $(1-p^2)^K$, where $K = (\alpha^2 n/2 - 2u)t$. Hence for $u \leq u_0, n \geq n_0$

$$\mathbf{Pr}(\mathcal{E}_u) \leq n^{2t}(1-p^2)^K \leq (n^2 e^{-\alpha^2 p^2 n/8})^u.$$

Thus for $u \leq u_0, n \geq n_0$

$$\mathbf{Pr}(\mathcal{E}_u) \leq (n^2 \exp\{-3\ln(n\hat{Q})\})^u = n^{-u}\hat{Q}^{-3u}.$$

Suppose first that $Q \geq 1$. Then

$$\begin{aligned}
\mathbf{E}(Q^{2\kappa}) &\leq Q^{2s}\left(1 + Q^2 \sum_{u=1}^{u_0} Q^{2u}\mathbf{Pr}(\mathcal{E}_u)\right) + Q^{2n}\mathbf{Pr}(\mathcal{E}_{u_0}) \\
&\leq Q^{2s}\left(1 + Q^2 \sum_{u=1}^{u_0} (n^{-1}Q^{-1})^u\right) + Q^{2n}n^{-\alpha^2 n/8} \\
&\leq 2Q^{2s}.
\end{aligned}$$

Suppose now that $Q < 1$. Then

$$\begin{aligned}
\mathbf{E}(Q^{\kappa}) &\geq Q^s(1 - \mathbf{Pr}(\mathcal{E}_1)) \\
&\geq Q^s/2
\end{aligned}$$

for $n \geq n_0$. □

Theorem 6.4.1 *For fixed rational x, y, and $\epsilon > 0$, if $T = T_G(x,y)$ and Z is the output of Algorithm EVAL, then*

$$\mathbf{Pr}(|Z - T| \geq \epsilon T) \leq \frac{1}{4}.$$

Proof Since $Z = \zeta\left(\frac{Z_1 + \ldots + Z_t}{t}\right)$, from Lemma 6.4.1 we see that $T = \mathbf{E}(Z)$. From Chebychev's inequality

$$\mathbf{Pr}\{|Z - T| \geq \epsilon T\} \leq \frac{\mathbf{Var}(Z)}{\epsilon^2 T^2} \leq \frac{\zeta^2}{\epsilon^2 t}\frac{\mathbf{Var}(Z_i)}{T^2} \leq \frac{\zeta^2}{\epsilon^2 t}\frac{\mathbf{E}(Z_i^2)}{T^2}.$$

Case $Q < 1$

Lemma 6.4.3 gives

$$\mathbf{E}(Z_i^2) = \mathbf{E}(Q^{2\kappa(G_p)}) \leq 1.$$

$$T^2 = \zeta^2(\mathbf{E}(Z_i))^2 = \zeta^2(\mathbf{E}(Q^{\kappa(G_p)}))^2 \geq \zeta^2 Q^{2s}/4.$$

giving

$$\mathbf{Pr}\{|Z - T| \geq \epsilon T\} \leq \frac{4}{\epsilon^2 t Q^{2s}}.$$

Case $Q \geq 1$

$$\mathbf{Pr}\{|Z - T| \geq \epsilon T\} \leq \frac{\zeta^2}{\epsilon^2 t}\frac{\mathbf{E}(Q^{2\kappa})}{T^2} \leq \frac{2Q^{2s}}{\epsilon^2 t}$$

using Lemma 6.4.3, and noticing that for $Q \geq 1$, $T \geq \zeta$.

The result follows provided

$$t \geq \frac{16}{\epsilon^2 Q^{2s}} \quad (Q < 1) \text{ and } t \geq \frac{8Q^{2s}}{\epsilon^2} \quad (Q \geq 1),$$

which it is by choice of t in EVAL. □

Note: although polynomially bounded the running time grows when $(x-1)(y-1)$ or its inverse grow.

6.5 Permanent via Determinant

We consider here an algorithm for estimating the permanent of a 0-1 matrix based on evaluating a determinant. Let A be an $n \times n$ 0-1 matrix:

The KKLLL estimator
The estimator is defined to be the random variable Z that results from the simple experiment described below. The idea is due to Karmarker, Karp, Lipton, Lovász and Luby [**?**] and it is an improvement on a method due to Godsil and Gutman [**?**].

(1) Form a matrix B from A as follows. Let $\{1, \omega, \omega^2\}$ be the cube roots of unity. For each pair i, j in the range $1 \leq i, j \leq n$:

- **(a)** If $A_{i,j} = 0$ then set $B_{i,j}$ equal to 0;
- **(b)** If $A_{i,j} = 1$ then choose $B_{i,j}$ independently and randomly from the set $\{1, \omega, \omega^2\}$.

(2) Set Z equal to $|\det B|^2$, where $|z|$ denotes the modulus of complex number z.

Let G be the bipartite graph on vertex set $U + V$, where $U = V = [n]$ and (i, j) is an edge of G iff $A_{i,j} = 1$. Let $\mathcal{M}$ denote the set of all perfect matchings in G. Clearly, $\operatorname{per} A = |\mathcal{M}|$. For $M \in \mathcal{M}$ let $\operatorname{sgn}(M)$ be the sign of the associated permutation σ_M where $\sigma_M(i) = j$ iff $(i, j) \in \mathcal{M}$. Let $\beta(M) = \prod_{(i,j)\in M} B_{i,j}$. With these notational definitions we can prove that Z is an unbiassed estimator of $\operatorname{per} A$.

Theorem 6.5.1

$$\mathbf{E}(Z) = \operatorname{per} A.$$

Proof

$$\begin{aligned} Z &= \sum_{M,M'\in\mathcal{M}} \operatorname{sgn}(M)\operatorname{sgn}(M')\beta(M)\overline{\beta(M')} \\ &= \sum_{M\in\mathcal{M}} 1 + \sum_{M\neq M'\in\mathcal{M}} \operatorname{sgn}(M)\operatorname{sgn}(M')\beta(M)\overline{\beta(M')}. \end{aligned}$$

The result now follows from

$$\mathbf{E}(\beta(M)\overline{\beta(M')}) = 0 \qquad \text{for } M \neq M' \in \mathcal{M}. \tag{6.2}$$

If $M \neq M'$ then $M \oplus M'$ (the symmetric difference of M and M') contains at least one cycle C, say. Let $(i_1, j_1) \in M$ be an edge of C. Then we can write $\beta(M)\overline{\beta(M')} = B_{i_1,j_1}\Delta$ where Δ depends only on the values of $B_{i,j}$, $(i, j) \neq (i_1, j_1)$. Then, by the independence of the $B_{i,j}$'s,

$$\mathbf{E}(\beta(M)\overline{\beta(M')}) = \mathbf{E}(B_{i,j})\mathbf{E}(\Delta) = \frac{1}{3}(1 + \omega + \omega^2)\mathbf{E}(\Delta) = 0$$

which confirmes (6.2). □

The efficiency of the KKLLL estimator will depend on its variance.

Let M and M' be perfect matchings in G. Denote by $c(M, M')$ the number of connected components (cycles) in $M \oplus M'$. Define $\gamma(G) = \mathbf{E}\left(2^{c(M,M')}\right)$ to be the expected value of $2^{c(M,M')}$ when M and M' are selected randomly from $\mathcal{M}$. (If G has no perfect matchings then define $\gamma(G) = 1$.)

Theorem 6.5.2

$$\frac{\mathbf{E}(Z^2)}{\mathbf{E}(Z)^2} = \gamma(G).$$

Proof Let

$$\xi(M_1, M_2, M_3, M_4) = \prod_{i\in\{1,3\}} \operatorname{sgn}(M_i)\beta(M_i) \prod_{i\in\{2,4\}} \operatorname{sgn}(M_i)\overline{\beta(M_i)}.$$

Then

$$Z^2 = \sum_{\mathcal{M}^4} \xi(M_1, M_2, M_3, M_4).$$

If there exists (i_1, j_1) which appears an odd number of times in the product ξ then $\mathbf{E}(\xi) = 0$. Indeed, if it appears 3 times then it occurs at least once as $B_{i,j}$ and at least once as $\overline{B_{i,j}}$ and as $B_{i,j}\overline{B_{i,j}} = 1$ we can reduce to the case where (i_1, j_1) appears exactly once and then $\mathbf{E}(\xi) = 0$ as in the proof of (6.2).

So now assume that every (i, j) appears an even number of times in the product ξ. If there exists (i_1, j_1) which appears twice as $B_{i,j}^2$ or as $\overline{B_{i,j}}^2$ then $\mathbf{E}(\xi) = 0$ as $\mathbf{E}(B_{i,j}^2) = \mathbf{E}(\overline{B_{i,j}}^2) = 0$. (Here we see the advantage of taking ω as a cube root of unity, as opposed to -1 as in [**?**]).

i?

We are left with the case where each $B_{i,j}$ occurs with an accompanying $\overline{B_{i,j}}$. But now we have $M_1 \oplus M_2 = M_3 \oplus M_4$, for if say $(i_1, j_1) \in M_1 \oplus M_2 \setminus M_3 \oplus M_4$ then (i_1, j_1) appears an odd number of times in ξ. Observe that in this case $\prod_{i=1}^4 \mathrm{sgn}(M_i) = 1$ since $\pi_{M_1}\pi_{M_2}^{-1}$ and $\pi_{M_3}\pi_{M_4}^{-1}$ have the same cycle structure (defined by $M_1 \oplus M_2$). Also $B_{i,j}\overline{B_{i,j}} = 1$ and so $\xi = 1$ here.

Now given M_1, M_2 there are $2^{c(M_1,M_2)}$ choices of M_3, M_4 which satisfy $M_1 \oplus M_2 = M_3 \oplus M_4$. Thus

There is something wrong here

$$\mathbf{E}(Z^2) = \sum_{\mathcal{M}^2} 2^{c(M_1,M_2)}$$

and the result follows from $\mathbf{E}(Z)^2 = |\mathcal{M}|^2$. □

We will now restrict our attention to *dense* matrices. We assume that each row and column of A has at least $(\frac{1}{2} + \alpha)n$ non-zeros for some constant $\alpha > 0$.

Theorem 6.5.3 *Suppose $\alpha > 0$ is a constant, and that bipartite graph G has minimum vertex degree $\delta(G) \geq (\frac{1}{2} + \alpha)n$; then $\gamma(G) \leq O(n^{1+(2\ln 2)/\alpha})$.*

Proof Fix a perfect matching $M_0 \in \mathcal{M}$, and for $M \in \mathcal{M}$ let

$$\iota(M) = |M \cap M_0|, \text{ and}$$
$$c(M) = \text{number of cycles in } M \oplus M_0.$$

Let $\mathcal{M}_{k,\ell} = \{M \in \mathcal{M} : \iota(M) = k,\ c(M) = \ell\}$, and $N_{k,\ell} = |\mathcal{M}_{k,\ell}|$. We show that perfect matchings of G are concentrated in sets $\mathcal{M}_{k,\ell}$ with k and ℓ small.

Lemma 6.5.1 *Let $N_{k,\ell}$ be as defined above. Then*

(a) $k\alpha N_{k,\ell} \leq N_{k-2,\ell+1} + 2N_{k-1,\ell}$, *and*

(b) $(2\alpha\ell - 1 - k\ell/n)\, N_{k,\ell} \le 2(\ln n) N_{k,\ell-1}$.

Proof We use a quantitative version of Dirac's [3] argument for demonstrating the existence of a Hamilton cycle in a dense graph; the same basic technique was used in Section 3.3.3 to verify an fpras for counting Hamilton cycles in a dense graph.

We first show part (a) of the lemma. Fix k, ℓ and consider pairs (M, M') with $M \in \mathcal{M}_{k,\ell}$ and $M' \in \mathcal{M}_{k-2,\ell+1} \cup \mathcal{M}_{k-1,\ell}$ such that for some $a_1, a_2 \in U$ and $b_1, b_2 \in V$,

$$M \setminus M' = \{(a_1, b_1), (a_2, b_2)\},$$
$$M' \setminus M = \{(a_1, b_2), (a_2, b_1)\},$$

and

$$(a_1, b_1) \in M \cap M_0.$$

There are two types of pair satisfying these conditions:

(i) If $(a_2, b_2) \in M_0$, then $M' \in \mathcal{M}_{k-2,\ell+1}$; moreover, $M' \cap M_0$ is obtained from $M \cap M_0$ by deleting the two edges (a_1, b_1) and (a_2, b_2), and $M' \oplus M_0$ is obtained from $M \oplus M_0$ by adding the 4-cycle $(a_1, b_1, a_2, b_2, a_1)$.

(ii) If $(a_2, b_2) \notin M_0$, then $M' \in \mathcal{M}_{k-1,\ell}$; moreover, $M' \cap M_0$ is obtained from $M \cap M_0$ by deleting the single edge (a_1, b_1), and $M' \oplus M_0$ is obtained from $M \oplus M_0$ by replacing the edge (a_2, b_2) of some cycle by the path (a_2, b_1, a_1, b_2) of length three.

Let $E_{k,\ell}$ denote the set of all such pairs (M, M'). For $M \in \mathcal{M}_{k,\ell}$, let $\zeta(M)$ denote the number of perfect matchings $M' \in \mathcal{M}_{k-2,\ell+1} \cup \mathcal{M}_{k-1,\ell}$ such that $(M, M') \in E_{k,\ell}$. For $M' \in \mathcal{M}_{k-2,\ell+1} \cup \mathcal{M}_{k-1,\ell}$, let $\eta(M')$ denote the number of perfect matchings $M \in \mathcal{M}_{k,\ell}$ such that $(M, M') \in E_{k,\ell}$.

Fix $M \in \mathcal{M}_{k,\ell}$ and $(a, b) \in M \cap M_0$. There are $s \ge 2\alpha n - 1$ edges (a', b') of M, other than (a, b) itself, such that both (a, b') and (a', b) are edges of G. Suppose s_1 are such that $(a', b') \in M \cap M_0$, and let $s_2 = s - s_1$. Then (a, b) contributes to s_1 type (i) pairs and s_2 type (ii) pairs involving M. Hence,

$$\zeta(M) \ge \sum_{(a,b)} (\tfrac{1}{2} s_1 + s_2) \tag{6.3}$$
$$\ge \tfrac{1}{2} k\alpha n, \tag{6.4}$$

provided $n \ge \alpha^{-1}$. The $\frac{1}{2}$ in inequality (6.3) comes from the fact that two edges of $M \cap M_0$ contribute to the same type (i) pair.

On the other hand, if $M' \in \mathcal{M}_{k-2,\ell+1}$ then $\eta(M')$ is at most the number of 4-cycles in $M' \oplus M_0$, and so $\eta(M) \le \frac{1}{2} n$. If $M' \in \mathcal{M}_{k-1,\ell}$ then $\eta(M')$ is at most the number of

paths of length three in $M' \oplus M_0$ with middle edge in M_0, and so $\eta(M') \le n$. Hence,

$$\tfrac{1}{2}k\alpha n N_{k,\ell} \le |E_{k,\ell}| \le \tfrac{1}{2}N_{k-2,\ell+1} + nN_{k-1,\ell},$$

and (a) follows.

We now turn to part (b) of the lemma. Let $E'_{k,\ell}$ denote the set of pairs $(M, M') \in \mathcal{M}_{k,\ell} \times \mathcal{M}_{k,\ell-1}$ such that, for some $a_1, a_2 \in U$ and $b_1, b_2 \in V$,

$$M \setminus M' = \{(a_1, b_1), (a_2, b_2)\},$$
$$M' \setminus M = \{(a_1, b_2), (a_2, b_1)\},$$

and

$$(a_1, b_1), (a_2, b_2), (a_2, b_1), (a_1, b_2) \notin M_0.$$

Here $M' \cap M_0 = M \cap M_0$ and $M' \oplus M_0$ is obtained from $M \oplus M_0$ as follows: take two disjoint cycles, C_1 containing (a_1, b_1) and C_2 containing (a_2, b_2). Replace the edges $(a_1, b_1), (a_2, b_2)$ by $(a_1, b_2), (a_2, b_1)$ creating one large cycle out of the vertices of C_1 and C_2. If C_i has $2m_i$ vertices, for $i = 1, 2$, we define $w(M, M') = m_1^{-1} + m_2^{-1}$.

For $M \in \mathcal{M}_{k,\ell}$, let

$$\mu(M) = \sum_{M':(M,M')\in E'_{k,\ell}} w(M, M'),$$

and for $M' \in \mathcal{M}_{k,\ell-1}$, let

$$\nu(M') = \sum_{M:(M,M')\in E'_{k,\ell}} w(M, M').$$

Fix $M \in \mathcal{M}_{k,\ell}$ and $(a, b) \in M \setminus M_0$, and suppose the cycles of $M \oplus M_0$ have size $2m_i$, for $1 \le i \le \ell$. If (a, b) is in a cycle of size $2m$ then there are $s \ge 2\alpha n - m - k$ edges (a', b') of $M \setminus M_0$ such that (a, b') and (a', b) are edges of G, and (a', b') and (a, b) are in different cycles. Putting $(a_1, b_1) = (a, b)$ and $(a_2, b_2) = (a', b')$ yields a member of $E'_{k,\ell}$. Apportioning weight m^{-1} to (a, b):

$$\mu(M) \ge \sum_{i=1}^{\ell} m_i(2\alpha n - m_i - k)m_i^{-1} \ge (2\alpha\ell - 1)n - k\ell.$$

Now fix $M' \in \mathcal{M}_{k,\ell-1}$ and suppose the cycles of $M' \oplus M_0$ have size $2m_i$, for $1 \le i \le \ell-1$. Fix a cycle C of size $2m$ in $M' \oplus M_0$. At worst, each pair of edges of $C \setminus M_0$ could contribute a pair (M, M') to $E'_{k,\ell}$. This observation gives

$$\nu(M) \le \sum_{i=1}^{\ell-1} m_i \left(\sum_{j=2}^{m_i-2} \frac{1}{j} + \frac{1}{m_i - j} \right) \le \sum_{i=1}^{\ell-1} 2m_i \ln m_i \le 2n \ln n.$$

Finally,

$$((2\alpha\ell - 1)n - k\ell)N_{k,\ell} \leq \sum_{(M,M') \in E'_{k,\ell}} w(M, M') \leq 2n(\ln n)N_{k,\ell-1},$$

and (b) follows. □

Proof of Theorem 6.5.3
Let $N = |\mathcal{M}|$, and

$$\Delta = \sum_{k=0}^{n} \sum_{\ell=0}^{n} N_{k,\ell} 2^{\ell}.$$

Our aim is to find a uniform bound on Δ/N, which will also be a bound on $\gamma(G)$. Let $s_{k,\ell} = N_{k,\ell}2^{\ell}$. It follows from Lemma 9(a) that

$$k\alpha s_{k,\ell} \leq \tfrac{1}{2} s_{k-2,\ell+1} + 2s_{k-1,\ell}. \tag{6.5}$$

Let $S_k = \sum_{\ell=0}^{n} s_{k,\ell}$. Then inequality (6.5) implies $k\alpha S_k \leq \frac{1}{2}S_{k-2} + 2S_{k-1}$. It follows by an easy induction on k that for $k > k_0 = \lceil 4/\alpha \rceil$,

$$S_k \leq \left(\frac{1+\sqrt{3}}{4} \right)^{k-k_0} (S_{k_0} + S_{k_0-1}),$$

and hence

$$\sum_{k=k_0}^{n} S_k = O(S_{k_0} + S_{k_0-1}). \tag{6.6}$$

Now assume $k \leq k_0$. From Lemma 9(b),

$$\frac{N_{k,\ell}}{N_{k,\ell-1}} \leq \frac{2\ln n}{(2\alpha - k/n)\ell - 1} \leq \frac{1}{2},$$

provided

$$\ell \geq \ell_0 = \left\lceil \frac{4\ln n + 1}{2\alpha - k_0/n} \right\rceil.$$

Thus, for $k \leq k_0$,

$$S_k \leq n s_{k,\ell_0} + \sum_{\ell=0}^{\ell_0} s_{k,\ell} \leq (n + \ell_0) 2^{\ell_0} N. \tag{6.7}$$

Hence, from (6.6) and (6.7), $\Delta/N = \sum_{k=0}^{n} S_k/N = O(n^{1+(2\ln 2)/\alpha})$. □

It follows from Theorem 6.5.3 that $\mathbf{Var}(Z) = O(n^{1+(2\ln 2)/\alpha})$ and so by taking the average of $t = O(\epsilon^{-2} n^{1+(2\ln 2)/\alpha})$ independently generated values of Z we can estimate $\operatorname{per} A$ to within $1+\epsilon$, with probability at least $3/4$ as required.

Bibliography

[1] D. Aldous also J. Fill Reversible Markov chains and random walks on graphs Monograph in preparation, available from `http://www.stat.berkeley.edu/users/aldous/book.html`, Department of Statistics, University of California, Berkeley.

[2] T. Lindvall Lectures on the Coupling Method Wiley-InterscienceNew York 1992

[3] C. H. Papadimitriou Computational Complexity Addison–WesleyReading, MA 1994

[Bro86] A.Z. Broder, How hard is it to marry at random? (On the approximation of the permanent), *Proceedings of the 18th Annual ACM Symposium on Theory of Computing*, ACM Press, 1986, 50–58. Erratum in *Proceedings of the 20th Annual ACM Symposium on Theory of Computing*, 1988, p. 551.

[Friez89] A.M. Frieze, *A note on computing random permanents* (unpublished manuscript), 1989.

[FS92] A. Frieze and S. Suen, Counting the number of Hamiltonian cycles in random digraphs, *Random Structures and algorithms*, 3:235–241, 1992.

[HL72] O.J. Heilmann and E.H. Lieb, Theory of monomer-dimer systems, *Communications in Mathematical Physics*, 25:190–232, 1972.

[JMS92] M. Jerrum, B. McKay and A. Sinclair, When is a graphical sequence stable?, *Random Graphs* 2 (A. Frieze and T. Łuczak, eds), Wiley, 1992, 101–115.

[JS89] M.R. Jerrum and A.J. Sinclair, Approximating the permanent, *SIAM Journal on Computing*, 18:1149–1178, 1989.

[JS90a] M.R. Jerrum and A.J. Sinclair, Fast Uniform Generation of Regular Graphs, *Theoretical Computer Science*, 73:91–100, 1990.

[4] Mark Jerrum and Alistair Sinclair, The Markov chain Monte Carlo method: an approach to approximate counting and integration. In *Approximation Algorithms for NP-hard Problems* (Dorit Hochbaum, ed.), PWS, 1996, 482–520.

[KRS93] C. Kenyon, D. Randall and A. Sinclair, Matchings in lattice graphs, *Proceedings of the 25th Annual ACM Symposium on Theory of Computing*, ACM Press, 1993, 738–746. Full version to appear in *Journal of Statistical Physics*, 1996.

[Met53] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, Equation of state calculation by fast computing machines, *Journal of Chemical Physics*, 21:1087–1092, 1953.

[Mih89a] M. Mihail, On coupling and the approximation of the permanent, *Information Processing Letters*, 30:91–95, 1989.

[MW91] M. Mihail and P. Winkler, On the number of Eulerian orientations of a graph, *Proceedings of the 3rd Annual ACM-SIAM Symposium on Discrete Algorithms*, ACM Press, 1992, 138–145.

[Mot89] R. Motwani, Expanding graphs and the average-case analysis of algorithms for matchings and related problems, *Proceedings of the 21st Annual ACM Symposium on Theory of Computing*, ACM Press, 1989, 550–561.

[Val79a] L.G. Valiant, The complexity of computing the permanent, *Theoretical Computer Science*, 8:189–201, 1979.