On sampling with Markov chains

F. R. K. Chung
University of Pennsylvania
Philadelphia, PA 19104

R. L. Graham
AT&T Bell Laboratories
Murray Hill, NJ 07974

S.-T. Yau
Harvard University
Cambridge, MA 02138

1. Introduction

There are many situations in which one would like to select a random element (nearly) uniformly from some large set. One method for doing this which has received much attention recently is the following. Suppose we can define a Markov chain M on X which has the uniform distribution as its stationary distribution. Then starting at some (arbitrary) initial point x_0, and applying M for sufficiently many (say t) steps, the resulting point $M^{(t)}(x_0)$ will be (almost) uniformly distributed over X, provided t is large enough. In order for this approach to be effective, however, M should be "rapidly mixing", i.e., $M^{(t)}$ should be very close to the uniform distribution U (in some appropriate sense) in polynomially many steps t, measured by the size of X.

There are in use a variety of methods for estimating the "mixing rate" of M, i.e., the rate of convergence of M to its stationary distribution, e.g., coupling, path-counting, strong stationary time, eigenvalue estimates and even Stein's method (cf. [A87], [SJ89], [S93], [DFK89], [FKP94]). However, for many problems of current interest such as volume estimation, approximate enumeration of linear extensions of a partial order and sampling contingency tables, eigenvalue methods have not been effective because of the difficulty in obtaining good (or any!) bounds on the dominant eigenvalue of the associated process M.

In this paper we remedy this problem to some extent by applying new eigenvalue bounds of two of the authors [CY1], [CY2] for random walks on what we call "convex subgraphs" of homogeneous graphs (to be defined in later sections).

A principal feature of these techniques is the ability to obtain eigenvalue estimates in situations in which there is a nonempty boundary obstructing the walk, typically a more difficult situation then the boundaryless case. In particular, we will give the first proof of rapid convergence for the "natural" walk on contingency tables, as well as generalizations to restricted contingency tables, symmetric tables, compositions of an integer, and so-called knapsack solutions. We point out that these methods can also be applied to other problems, such as selecting random points in a given polytope (which is a crucial component of many recent volume approximation algorithms for polytopes). Also, the same ideas can be carried out in the context of weighted graphs and biased random walks (see [CY1] for the basic statements). However, we have deliberately restricted ourselves to some of the simpler applications for ease of exposition.

Before proceeding to our main results, it will be necessary to introduce a certain amount of background material. This we do in the next section.

2. Background

We begin by recalling a variety of concepts from graph theory and differential geometry. Any undefined terms can be found in [SY94] or [BM76].

For a graph $G = (V, E)$ with vertex set V and edge set E, we let d_v denote the degree of the vertex $u \in V$, i.e., the number of $w \in V$ such that $\{u, w\} \in E$ is an edge (which we also indicate by writing $v \sim w$). For the most part, our graphs will be undirected, and without loops or multiple edges. Define the square matrix $L = L_G$, with rows and columns indexed by V by:

$$L(u,v) = \begin{cases} d_u & \text{if } u = v, \\ -1 & \text{if } u \sim v, \\ 0 & \text{otherwise} . \end{cases} \tag{1}$$

Let T denote the diagonal matrix with (v, v) entry $T(v, v) = d_v$, $v \in V$ (and, of course, 0 otherwise).

A key concept for us will be the following operator: The Laplacian $\mathcal{L} = \mathcal{L}_G$ for G is defined by

$$\mathcal{L} := T^{-1/2} L T^{-1/2} . \tag{2}$$

Thus, $\mathcal{L}$ can be represented as a $|V|$ by $|V|$ matrix given by

$$\mathcal{L}(u,v) = \begin{cases} 1 & \text{if } u = v, \\ -\frac{1}{\sqrt{d_u d_v}} & \text{if } u \sim v, \\ 0 & \text{otherwise} . \end{cases} \tag{3}$$

Considering $\mathcal{L}$ as an operator acting on $\{f : V \to \mathbb{C}\}$ we have:

$$\mathcal{L}f(v) = \frac{1}{\sqrt{d_v}} \sum_{\substack{u \in V \\ u \sim v}} \left(\frac{f(v)}{\sqrt{d_v}} - \frac{f(u)}{\sqrt{d_u}} \right) . \tag{4}$$

In the case that G is d-regular, i.e., $d_v = d$ for all $v \in V$, (3) and (4) take the more familiar forms

$$\mathcal{L} = I - \frac{1}{d} A \tag{3'}$$

where I is the identity matrix and $A = A_G$ is the adjacency matrix for G, and

$$\mathcal{L}f(v) = \frac{1}{d} \sum_{\substack{u \\ u \sim v}} (f(v) - f(u)) . \tag{4'}$$

Denote the eigenvalues of $\mathcal{L}$ by

$$0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{n-1}$$

where we let n denote $|V|$ for now. It is known that $\lambda_1 \leq 1$ unless $G = K_n$, the complete graph on n vertices, and $\lambda_{n-1} \leq 2$ with equality if and only if G is bipartite.

The most important of these eigenvalues from our perspective will be λ_1 which we call the Laplacian eigenvalue of G, and denote by $\lambda = \lambda_G$. Note that with $\mathbf{1} : V \to \mathbb{C}$ denoting the constant $\mathbf{1}$ function, $T^{1/2}\mathbf{1}$ is an eigenfunction of $\mathcal{L}$ with eigenvalue 0. It follows (see [CY1]) that

$$\begin{aligned} \lambda := \lambda_1 &= \inf_{f \perp T\mathbf{1}} \sum_{u \sim v} \frac{(f(v) - f(u))^2}{\sum_v d_v f(v)^2} \\ &= \inf_f \sup_c \frac{\sum_{u \sim v} (f(v) - f(u))^2}{\sum_v d_v (f(v) - c)^2} . \end{aligned} \tag{5}$$

For a (nonempty) set $S \subset V$, consider the subgraph of G induced by S. This subgraph, which we also denote by S, has edge set $E(S) = \{\{u, v\} \in E(G) : u, v \in S\}$. The edge boundary ∂S is defined by

$$\partial S := \{\{u, v\} \in E(G) : u \in S, u \in V \setminus S\} .$$

The vertex boundary δS is defined by

$$\delta S := \{v \in V \setminus S : \{u, v\} \in E(G) \ \text{ for some } \ v \in V\} .$$

δS

$V \setminus S$

∂S v

S

u

Figure 1: Various boundaries of S

Let $E' := E(S) \cup \partial S$.

We illustrate these definitions in Figure 1.

The next key concept we will need is that of the Neumann eigenvalues for S on G. First, define (parallel to (5))

$$\begin{aligned} \lambda_S &= \inf_{\substack{f \\ \sum\limits_{x\in S} d_x f(x)=0}} \frac{\sum\limits_{\{x,y\}\in E'} (f(x)-f(y))^2}{\sum\limits_{x\in S} d_x f^2(x)} \\ &= \inf_f \sup_c \frac{\sum\limits_{\{x,y\}\in E'} (f(x)-f(y))^2}{\sum\limits_{x\in S} d_x (f(x)-c)^2} . \end{aligned} \tag{6}$$

In general, define

$$\lambda_{S,i} = \inf_f \sup_{f' \in C_{i-1}} \frac{\sum\limits_{\{x,y\}\in E'} (f(x)-f(y))^2}{\sum\limits_{x\in S} d_x (f(x)-f'(x))^2} \tag{7}$$

where C_k is the subspace spanned by the eigenfunctions ϕ_i achieving $\lambda_{S,i}$, $0 \le i \le k$. In particular,

$$\lambda_S = \lambda_{S,1} = \inf_{g \perp T^{1/2}\mathbf{1}} \frac{\langle g, \mathcal{L} g\rangle_S}{\langle g, g\rangle_S} \tag{8}$$

where $\mathcal{L}$ is the Laplacian of G and $\langle f_1, f_2\rangle_S$ denotes the inner product $\sum\limits_{x\in S} f_1(x) f_2(x)$. Note that when $S = V$, then λ_S is just the usual Laplacian eigenvalue of G. The $\lambda_{S,i}$ are eigenvalues of a matrix $\mathcal{L}_S$ defined by the following steps.

For $X \subset V$, let L_X denote the submatrix of L restricted to rows and columns induced by elements of X. Define the matrix N with rows indexed by $S \cup \delta S$ and columns indexed by S

given by

$$N(x,y) = \begin{cases} 1 & \text{if } x = y, \\ 0 & \text{if } x \in S,\ x \neq y, \\ \frac{1}{d'_x} & \text{if } x \in \delta S,\ x \sim y \in S, \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

where d'_x denotes the degree of x to points in S. Then the matrix

$$\mathcal{L}_S := T^{-1/2} N^{tr} L_S N T^{-1/2} \tag{10}$$

has the $\lambda_{S,i}$ as its eigenvalues (together with 0, of course), with $\lambda_S = \lambda_{S,1}$ being the least eigenvalue exceeding 0. The matrix $\mathcal{L}_S$ is associated with the following random walk P on $S \subset G$, which we call the *Neumann walk* on S. A particle at vertex u moves to each of the neighbors v of u with equal probability $1/d_u$ except if $v \notin S$ (so that $v \in \delta S$). In this case, the particle then moves to each of the d'_v neighbors of v which are in S, with equal probability $\frac{1}{d'_v}$. Thus, $P(u,x)$, the probability of moving from u to x is

$$P(u,x) = \left\{ \begin{array}{cc} 1/d_u & \text{if } u \sim x \\ 0 & \text{otherwise} \end{array} \right\} + \frac{1}{d_u} \sum_{\substack{v \in \delta S \\ v \sim u}} \frac{1}{d'_v} \ .$$

It is easy to see that when G is regular, the stationary distribution of P is uniform for any choice of S connected. The rate of convergence of P to its stationary distribution is controlled by the Neumann Laplacian eigenvalue λ_S, which we eventually will bound from below.

The final key concept we introduce here is the *heat kernel* H_t of S. To begin with, let us decompose $\mathcal{L}_S$ in the usual way as a linear combination of its projections P_i onto the corresponding eigenfunctions ϕ_i of $\mathcal{L}_S$:

$$\mathcal{L}_S = \sum_i \lambda_{S,i} P_i \ . \tag{11}$$

For $t \geq 0$, the heat kernel H_i is defined to be the operator of $\{f : S \cup \delta S \to \mathbb{C}\}$ given by

$$\begin{aligned} H_t : \quad &= \quad \sum_i e^{-\lambda_{S,i}} P_i \\ &= \quad e^{-t\mathcal{L}_S} \\ &= \quad I - t\mathcal{L}_S + \frac{t^2}{2}\mathcal{L}_S^2 - \frac{t^3}{6}\mathcal{L}_S^3 + \cdots \end{aligned} \tag{12}$$

Thus, $H_0 = I$.

For $f : S \cup \delta S \to \mathbb{C}$, set

$$F = H_t f = e^{-t\mathcal{L}_S} f \ .$$

That is,

$$F(t,x) = (H_t f)(x) = \sum_{y \in S \cup \delta S} H_t(x,y) f(y) \ .$$

Some useful facts concerning H_t and F are the following (see [CY1] for proofs):

(i) $F(0,x) = f(x)$

(ii) For $x \in S \cup \delta S$,

$$\sum_{y \in S \cup \delta S} H_t(x,y)\sqrt{d_y} = \sqrt{d_x}$$

(iii) F satisfies the heat equation $\frac{\partial F}{\partial t} = -\mathcal{L}_S F$

(iv) For any $x \in \delta S$, $\mathcal{L}_S F(t,x) = 0$

(v) For all $x, y \in S \cup \delta(S)$, $H_t(x,y) \geq 0$

The connection of the heat kernel to the Neumann eigenvalue λ_S of S is given by the following result:

Theorem [Chung, Yau [CY1]]. *For all $t > 0$,*

$$\lambda_S \geq \frac{1}{2t} \sum_{x \in S} \inf_{y \in S} \frac{H_t(x,y)\sqrt{d_x}}{\sqrt{d_y}} \ . \tag{13}$$

This inequality will be an essential ingredient in eventually obtaining lower bounds for λ_S (and upper bounds on the mixing rates) for the various Markov chains we consider.

3. A direct application to heat kernel inequality

The main use of the heat kernel inequality (13) in lower bounding λ_S will depend on controlling the behavior of H_t by connecting it to an associated continuous heat kernel h_t for an appropriate Riemannian manifold $M = M_S$ containing the points of $S \cup \delta S$. This we do in Section 4.

However, there are situations in which we can get bounds on λ_S directly from (13) without going through this process. We describe several of these in this section.

Let us consider the special case for which $S = V$ (so that $\mathcal{L}_S = \mathcal{L}$) and, further, suppose the graph G has a "covering" vertex x_0 with the property that x_0 is adjacent to every $g \in V \setminus \{x_0\}$ (so that $d_x = n - 1$ where $n := |V|$. We will apply (13) with $t \to 0$. Thus, by (13)

$$H_t = I - t\mathcal{L} + O(t^2), \quad t \to 0$$

$$
= \begin{array}{ccccc}
 & & y' & & y \\
 & 1-t & & & \\
 & & 1-t & & \\
 & & & & \\
x & \cdots & 0 & \cdots & \frac{t}{\sqrt{d_x d_y}} \\
 & & x \not\sim y' & & x \sim y
\end{array} + O(t^2) \tag{14}
$$

Thus,

$$
\left(H_t(x,y) \frac{\sqrt{d_x}}{\sqrt{d_y}} \right) = \begin{array}{ccccc}
 & & y' & & y \\
 & 1-t & & & \\
 & & 1-t & & \\
 & & & & \\
x & \cdots & 0 & \cdots & \frac{t}{d_y} \\
 & & x \not\sim y' & & x \sim y
\end{array} + O(t^2)
$$

$$
= \begin{array}{ccccc}
 & & y' & & y \\
 & 1-t & & & \\
 & & 1-t & & \\
 & & & & \\
x_0 & \cdots & \frac{t}{d_{y'}} & \cdots & \frac{t}{d_y}
\end{array} + O(t^2) \tag{15}
$$

Thus, by (13)

$$
\begin{aligned}
\lambda_S = \lambda &\geq \frac{1}{2t} \sum_x \inf_y \frac{H_t(x,y)\sqrt{d_x}}{\sqrt{d_y}} \\
&= \frac{1}{2t} \left(\inf_{y \neq x_0} \frac{t}{d_y} + O(t^2) \right) \\
&= \frac{1}{2t} \left(\frac{t}{\delta_2} \right) + O(t), \ t \to 0
\end{aligned} \tag{16}
$$

where δ_2 denotes the second largest degree in G. Thus, we have

$$\lambda \geq \frac{1}{2\delta_2} \tag{17}$$

for this situation. For example, for $G = P_3$, the path with 3 vertices, it is true that $\lambda_0 = 0$, $\lambda = \lambda_1 = 1$ and $\lambda_2 = 2$, while our estimate in (17) gives $\lambda \geq 1/2$.

A similar analysis shows that of G has $k > 1$ covering vertices then (13) implies

$$\lambda \geq \frac{k}{2(n-1)} \ . \tag{18}$$

Applying this to $G = K_n$, the complete graph on n vertices yields

$$\lambda \geq \frac{n}{2(n-1)}$$

while the truth is $\lambda = \frac{n}{n-1}$ (again off by a factor of 2).

4. The basic set-up

As hinted at in the previous section, the real use of (13) in boundary λ_S will proceed along the following lines:

(a) Embed G and S as "convex" sets into some Riemannian manifold $M = M_S$ (usually $\cong \mathbb{E}^N$ for our applications).

(b) Relate the heat kernel H_t on $S \cup \delta S$ to the continuous heat kernel h_t on M.

(c) Relate h_t to various properties of M (and S), such as the "density" of points of S in M, the diameter of M, the dimension of M, etc.

We next describe a specific situation to which we can apply this procedure.

We begin with an infinite "lattice" graph $\Gamma = (V, E)$ with $V \subset \mathcal{M} \cong \mathbb{E}^N$, with the property that the automorphism group $\mathcal{H}$ of G is transitive. Thus, $\mathcal{H} : V \to V$ so that $u \sim v \Leftrightarrow gu \sim gv$, $g \in \mathcal{H}$, and for all $u, v \in V$,

$$v = gu \ \text{ for some } \ g \in \mathcal{H} \ .$$

Suppose Γ has an edge-generating set $\mathcal{K} \subset \mathcal{H}$, so that every edge of Γ is of the form $\{v, gv\}$ for some $v \in V$, $g \in \mathcal{K}$, and any such pair is an edge (hence the term lattice graph). We will further assume (since we think of Γ as undirected) that $g \in \mathcal{K} \Rightarrow g^{-1} \in \mathcal{K}$. Finally, we assume that for any x, any lattice point of $y \in \mathcal{M}$ in the convex hull of $\{gx : g \in \mathcal{K}\}$ is also adjacent to x in Γ. Let S be some finite subset of V. Assume S is "convex", which means that for some submanifold $M \subset \mathcal{M}$ with nonempty convex boundary, S consists of all lattice points of V which are in M. Let ℓ denote the minimum length of any of the edges $\{x, gx\}$, $g \in \mathcal{K}$, and assume that for each $x \in S$, the ball $B_x(\ell/3)$ of radius $\ell/3$ centered at x is contained in M.

For $x \in V$, let $U(x)$ denote the Voronoi origin for x, i.e., the set of all points of $\mathcal{M}$ closer to x than to any other point $y \in S$. Since $\mathcal{H}$ is transitive, all $U(x)$ have the same volume, denoted by vol U. Finally, set

$$\begin{aligned} \operatorname{diam} M &:= \text{diameter of } M \\ \dim M &:= \text{dimension of } M \end{aligned}$$

Under the preceding assumptions and notation, we have the following estimate.

Theorem [Chung, Yau [CY1]]. *For convex S,*

$$\lambda_S \geq c_0 \left(\frac{\ell}{\dim M \ \operatorname{diam} M} \right)^2 \frac{|S| \operatorname{vol} U}{\operatorname{vol} \ M} \tag{19}$$

for constant $c_0 > 0$ depending only on Γ and not on S.

The proof of (19) depends upon a variety of ideas and techniques from differential geometry, linear algebra and combinatorics (as well as (13)), and can be found in [CY1]. The bulk of our results presented here will depend on applying (19) and its generalizations to specific cases of interest.

We remark that the results in [CY1] are similar in spirit to those in [CY2] which also gives lower bounds for λ_S for Neumann walks. There, S is required to satisfy more restrictive conditions (e.g., being "strongly convex") but the bounds on λ_S are sharper. In particular, it is shown that under the appropriate hypotheses on S (and Γ),

$$\lambda_S \geq \frac{1}{8kD^2} \tag{20}$$

where $k = |\mathcal{K}|$ and $D :=$ graph diameter of S (see Section 8 for a comparison of (19) and (20) in a specific case).

5. Contingency tables

Given integer vectors $\overline{r} = (r_1, \ldots, r_m)$, $\overline{c} = (c_1, \ldots, c_n)$ with $r_i, c_j \geq 0$ and $\sum_i r_i = \sum_j c_j$, we can consider the space of all $m \times n$ arrays T with the property that

$$\begin{aligned} \sum_j T(i,j) &= r_i, \qquad 1 \leq i \leq m \\ \sum_i T(i,j) &= c_j, \qquad 1 \leq j \leq n \ . \end{aligned}$$

Let us denote by $\mathcal{T} = \mathcal{T}(\overline{r}, \overline{c})$ the set of all such arrays. The arrays $T \in \mathcal{T}$ are often called contingency tables (with given row and column sums). These tables arise in a variety of applications, such as goodness of fit tests in statistics, enumeration of permutations by descents, describing tensor product decompositions, counting double cosets, etc., and have a long history. (An excellent survey can be found in [DG].) It seems to be a particularly difficult problem to obtain good estimates of the size of $\mathcal{T}(\overline{r}, \overline{c})$ for large $\overline{r}$ and $\overline{c}$. In order to attack this problem, a standard (by now — see [A87], [SJ89], [S93], [G91], [DSt]) technique depends on rapidly generating random tables from $\mathcal{T} = \mathcal{T}(\overline{r}, \overline{c})$ with nearly equal probability.

To do this, we first consider the following natural random walk P on $\mathcal{T}$. ¿From any given table $T \in \mathcal{T}$, select uniformly at random a pair of rows $\{i, i'\}$ and a pair of columns $\{j, j'\}$, and move to the table T', obtained from T by changing four entries of T as follows:

$$\begin{aligned} T'(i,j) &= T(i,j) + 1, \qquad & T'(i',j') &= T(i,j) + 1 \\ T'(i',j) &= T(i,j') - 1, \qquad & T'(i,j') &= T(i,j') - 1 \ . \end{aligned}$$

Such a move we call a basic move. The table T' clearly has the same line sums (i.e., row and column sums) as T. The only problem is that T' may have negative entries (because we might have $T(i,j') = 0$, for example) and so is not in $\mathcal{T}$. To deal with this "boundary" problem, we instead execute the corresponding Neumann walk in $\mathcal{T}$, as described in Section 2.

We first need to place our contingency table problem into the framework of the preceding section. The manifold $\mathcal{M}$ will consist of all real mn-triples $\overline{x} = (x_{11}, x_{12}, \ldots, x_{mn})$ satisfying

$$\sum_j x_{ij} = r_i, \quad \sum_i x_{ij} = c_j \ .$$

Since $\sum\limits_i r_i = \sum\limits_j c_j$ then

$$\dim \mathcal{M} = N := (m-1)(n-1) \ .$$

The graph Γ has as vertices all the integer points in $\mathcal{M}$, i.e., all $\overline{x}$ with all $x_{ij} \in \mathbb{Z}$. The edge generating set $\mathcal{K}$ consists of all the basic moves described above. Thus, $|\mathcal{K}| = \binom{m}{2}\binom{n}{2}$. The set S will just be $\mathcal{T} = \mathcal{T}(\overline{r}, \overline{c})$, the set of all $T \in \Gamma$ with all entries nonnegative. Thus,

$$S = \bigcap_{i,j} \{T \in \Gamma : x_{ij} \geq 0\} \ .$$

Similarly, the manifold $M \subset \mathcal{M}$ is defined by

$$M = \bigcap_{i,j} \{\overline{x} \in \mathcal{M} : x_{ij} \geq -2/3\} \ .$$

It is clear that M is an N-dimensional convex polytope and $S = \mathcal{T}$ is the set of all lattice points in M, and consequently convex in the sense needed for (19). It is easy to see that $\mathcal{T}$ is connected by the basic moves generated by $\mathcal{K}$, and that each edge of Γ has length 2. Our next problem is to deal with the term $\frac{|S| \text{ vol } U}{\text{vol } M}$ in (19). In particular, we would like to show this is close to 1, provided that r_i and c_j are not too small. To do this we need the following two results.

Claim 1. Suppose $L \subset \mathbb{E}^N$ is a lattice generated by vectors $v_1, \dots, v_N$. Then the covering radius of L is at most $R := \frac{1}{2}\left(\sum_{i=1}^{N} \|v_i\|^2\right)^{1/2}$.

Proof. The assertion clearly holds for $N = 1$. Assume it holds for all dimensions less than N. It is enough to prove that any point $\overline{x} = (x_1, \dots, x_N)$ in the fundamental domain generated by the v_i is at most a distance of R from some v_i. Let $\overline{x}_0$ be the projection of $\overline{x}$ on either the hyperplane generated by $v_1, \dots, v_{N-1}$, or a translate of the hyperplane by v_N, whichever is closer (these are two bounding hyperplanes of the fundamental domain). Thus, $d(\overline{x}_1, \overline{x}_0) \le \frac{1}{2}\|v_N\|$. By the induction hypothesis,

$$d(\overline{x}_0, v_j) \le \frac{1}{2}\left(\sum_{i=1}^{N-1} \|v_i\|^2\right)^{1/2} \quad \text{for some} \quad j < N .$$

Hence

$$d(\overline{x}, v_j) \le \left(\frac{1}{4}\|v_N\|^2 + \frac{1}{4}\sum_{i=1}^{N-1} \|v_i\|^2\right)^{1/2} = R$$

as claimed. ∎

Claim 2. If M is convex and contains a ball $B(cRN)$ of radius cRN, $c > 0$, then

(21) $$e^{-1/c} < \frac{|S| \operatorname{vol} U}{\operatorname{vol} M} < e^{1/c}$$

where $v_1, \dots v_N$ generate Γ, and $R = \frac{1}{2}\left(\sum_{v=1}^{N} \|v_i\|^2\right)^{1/2}$.

Sketch of proof: Consider an enlarged copy $(1+\delta)M$ of M expanded about the center of the ball $B(cRN)$, $\delta > 0$. Let L be some bounding hyperplane of M, and let $(1+\delta)L$ be the corresponding expanded copy of L (see Figure 2)

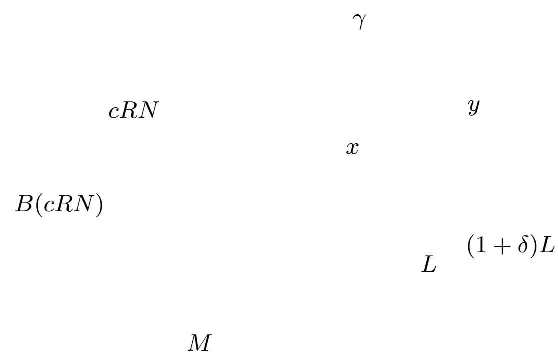

Figure 2: A large ball in M

Let $x \in S \subset M$ and suppose there exists $y \in U(x)$, the Voronoi region for x, with $y \notin (1+\delta)M$. Thus

$$d(x,y) \geq \gamma > c\delta RN \ .$$

However, by Claim 1, each point in $\mathcal{M}$ has distance at most R from lattice point in $\mathcal{M}$. This is a contradiction if we take $\delta = \frac{1}{cN}$.

Thus, for all $x \in S$, $U(x) \subset (1+\delta)M$ and so

$$|S| \text{ vol } U \leq \text{vol}(1+\delta)M = (1+\delta)^N \text{ vol } M$$

i.e.,

$$\frac{|S| \text{ vol } U}{\text{vol } M} \leq (1+\delta)^N = \left(1 + \frac{1}{cN}\right)^N < e^{1/c} \ .$$

A similar argument shows that

$$\frac{|S| \text{ vol } U}{\text{vol } M} > e^{-1/c}$$

and Claim 2 is proved. ■

In order to apply the result in Claim 2, we must find a large ball in M. Let s_0 denote the smallest *line sum average*, i.e.,

$$s_0 = \min\left(\min_i \frac{r_i}{n},\ \min_j \frac{c_j}{m}\right) \ .$$

We begin constructing an element $T_0 \in M$ recursively as follows. Suppose without loss of generality that

$$s_0 = \frac{r_1}{m} \ .$$

Then, in T_0, set all elements of the first row equal to s_0, and subtract s_0 from each value c_j to form $c'_j = c_j - s_0$, $1 \le j \le n$. Now, to complete T_0, we are reduced to forming an $(m-1)$ by n table T'_0 with row sums $r_2, \ldots r_m$ and column sums $c'_1, \ldots, c'_n$. The key point here is that all the line sum averages for T'_0 are still at least as large as s_0. Hence, continuing this process we can eventually construct a table T_0 (with rational entries) having least entry equal to s_0. Consequently there is a ball $B(s_0)$ of radius s_0 centered at $T_0 \in M$ which is contained in M (since to leave M, some entry must become negative). Therefore, if we assume $s_0 > cN^{3/2}$ then by (19) and (21),

$$\lambda_S \ge \frac{c_0 e^{-1/c}}{N^2 (\mathrm{diam}\, M)^2} \tag{22}$$

for some absolute constant $c_0 > 0$ (since for tables, all the generators have length 2, so that $R \le N^{1/2}$).

In the Appendix, we illustrate how a specific value can be derived here for c_0 (as well as in several other cases of interest, as well). In particular, for contingency tables, we can take $c_0 = 1/800$.

Since

$$\mathrm{diam}\, M < 2 \min \left\{ \Big(\sum_i r_i^2\Big)^{1/2},\ \Big(\sum_j c_j^2\Big)^{1/2} \right\}$$

then (22) can be written as follows:

For the natural Neumann walk P on the space of tables $\mathcal{T}(\overline{r}, \overline{c})$ where

$$\min \left\{ \min_i \frac{r_i}{n},\ \min_j \frac{c_j}{m} \right\} > c(m-1)^{3/2}(n-1)^{3/2}$$

we have

$$\lambda_S > \left[3200 e^{1/c} (m-1)^2 (n-1)^2 \min \left\{ \sum_i r_i^2, \sum_j c_j^2 \right\} \right]^{-1} . \tag{23}$$

To convert the estimate in (23) to an estimate for the rate of convergence of P to its stationary (uniform) distribution π, we use the following (standard) techniques (e.g., see [S93]).

Define the *relative pointwise distance* of $P^{(t)}$ to π to be

$$\Delta(t) := \max_{x,y} \frac{|P^{(t)}_{y,x} - \pi(x)|}{\pi(x)} \tag{24}$$

where $P^{(t)}_{y,x}$ denotes the probability of being at x after t steps starting at y. It is not hard to show (see [S93]) that

$$\begin{aligned}\Delta(t) &< (1-\lambda)^t \frac{\text{vol } S}{\min\limits_x \deg_\Gamma x} \\ &\le e^{-\lambda t}\frac{\text{vol } S}{\deg\ \Gamma} = e^{-\lambda t}|S|\end{aligned} \tag{25}$$

where

$$\begin{aligned}\text{vol } S: &= \sum_{x\in S}\deg_\Gamma x, \\ \deg G &= \deg_\Gamma x \ \text{ for any } x\end{aligned}$$

and λ is the eigenvalue of $\mathcal{L}_S$ which maximizes $|1-\lambda|$ for $\lambda \ne 0$.

In order to guarantee that λ is in fact λ_S, we can modify P to be "lazy", i.e., so that the modified walk $\overline{P}$ stays put with probability $1/2$, and moves with probability $1/2$ times what P did. The eigenvalues for $\overline{P}$ are just $1/2$ times those for P, and so, are contained in $[0,1]$. Thus, if

$$t > \frac{2}{\lambda_S}\ln\frac{|\mathcal{T}|}{\epsilon} \tag{26}$$

then $\Delta(t) < \epsilon$. Note that

$$|\mathcal{T}| \le \min\left\{\prod_i r_i^n,\ \prod_j c_j^m\right\}\ .$$

Thus, by (19) and (26), if

$$t > 6400e^{1/c}m^2n^2\min\left\{\sum_i r_i^2, \sum_j c_j^2\right\}\left(\ln\frac{1}{\epsilon} + \min\left\{n\sum_i \ln r_i, m\sum_j \ln c_j\right\}\right) \tag{27}$$

then $\Delta(t) < \epsilon$, provided

$$\min\left\{\min_i \frac{r_i}{n}, \min_j \frac{c_j}{n}\right\} > c(m-1)^{3/2}(n-1)^{3/2}\ .$$

As remarked earlier, this shows that the natural Neumann walk on $\mathcal{T}(\overline{r},\overline{c})$ converges to uniform in time polynomial in the dimensions of the table and the sizes of the line sums (and as the square of the diameter of the space). This strengthens a recent result of Diaconis and Saloff-Coste [DS1] who showed that for *fixed* dimension the natural walk on $\mathcal{T}$ (where any step which might create a negative entry is simply not taken) converges to uniform in time polynomial in the sizes of the line sums and as the square of diameter of the graph $\mathcal{T}$ (they use total variation distance instead of relative pointwise distance) but with constants that grow exponentially in mn.

By taking successive relaxations of the line sum constraints (as described in [DG] or [DKM]), it is possible to approximately enumerate $\mathcal{T}$ in polynomial time as well.

We also note that Dyer, Kannan and Mount [DKM] have developed a rather different (continuous) random walk which is rapidly mixing on $\mathcal{T}$. They show that the dominant eigenvalue λ for their walk satisfies

$$\lambda > \frac{c}{(m-1)^4(n-1)^4(m+n-2)}$$

for $m, n > 1$.

6. Restricted contingency tables

A natural extension of a contingency table is one in which certain entries are restricted, e.g., required to be 0. In this section we indicate how such restricted tables can be dealt with.

Given $m, n > 0$, let $A \subseteq [1, m] \times [1, n]$ be some nonempty index set. By a A-table T we mean an $m \times n$ array in which $T(i, j) \equiv 0$ if $(i, j) \notin A$. For given row sum and column sum vectors $\overline{r} = (r_1, \dots, r_m)$ and $\overline{c} = (c_1, \dots, c_n)$, respectively, with $\sum_i r_i = \sum_j c_j$, we let

$$\mathcal{T}_A = \mathcal{T}_A(\overline{r}, \overline{c}) = \left\{ A\text{-tables } T : \sum_j T(i,j) = r_i, \sum_i T(i,j) = c_j, 1 \le i \le m, 1 \le j \le n \right\} .$$

As before, we want to execute a "natural" Neumann walk P on $\mathcal{T}_A$ and show that it is rapidly mixing. However, several new complications arise from what had to be considered in the preceding section.

To begin with, what will we use for steps in our walk? Let us associate to A a bipartite graph $B = B_A$ with vertex sets $[1, m]$ and $[1, n]$, and with (i, j) an edge of B if and only if $(i, j) \in A$. In the case of unrestricted tables, B is the complete bipartite graph on these vertex sets, and the basic steps just occur on the 4-cycles of B. For our more general bipartite graph B, we first normalize it as follows. Let C be a connected component of B. For each *cut-edge* $e = (i, j)$ of C, the corresponding value $T(i, j)$ is easily seen to be determined by $\overline{r}$ and $\overline{c}$, say, it is $w(e)$. Then replace r_i by $r_i - w(e)$ and c_j by $c_j - w(e)$. Now continue this process recursively until we finally arrive at the 2-connected components C' of C, with correspondingly reduced row and column sums $\overline{r}'$ and $\overline{c}'$. It is not hard to show that there are always feasible assignments satisfying the modified line sum constraints.

Next, we have to describe the basic moves of our Neumann walk on $\mathcal{T}_A$. For each 2-connected component C of B, let T_C denote a fixed spanning tree on the vertex set $V(C) = V_1 \cup V_2$ of C, where $V_1 \subset [1, m]$, $V_2 \subset [1, n]$. For each edge $i = \{i, j\}$ *not* in T_C, with $i \in V_1$, $j \in V_2$, the addition of e to T_C creates some simple even cycle $Z(e)$. We then assign alternating ± 1's on the consecutive edges of $Z(e)$ to generate a possible move $\mathbb{Z}^{\pm}(e)$ on the table. It is not hard to show that the set of cycles $\bigcup_e Z(e)$ form a cycle basis over $\mathbb{Z}$ for the even cycles on $V(C)$ (in fact, only coefficients of 0 or ± 1 are ever needed). It is also not difficult to see that the set of moves $\mathcal{K} = \bigcup_C \bigcup_e Z^{\pm}(e)$, C a 2-connected component of B, connects the set $\mathcal{T}_A$ of all A-tables. Note that $\mathcal{K}$ has size $O(mn)$, and so is smaller than what was used in the unrestricted (complete) case.

Of course, if A' denotes the set of non-cut-edges of B then our tables $T \in \mathcal{T}_A$ only have possibly varying entries $T(i, j)$ where $(i, j) \in A'$. Thus, our A'-tables can now be represented as integer points in $\mathbb{E}^{|A'|}$,. In fact, because of the line sum constraints, the set $\mathcal{T}_{A'}$ of all A'-tables actually lies in a subspace $\mathcal{M}$ of dimension $N = \sum_C (m_C - 1)(n_C - 1)$ where C ranges over the 2-connected components of B, and m_C and n_C are the sizes of the vertex sets of C.

The same "minimum average line sum' technique from the preceding section now applies to each (2-connected) component C of B to produce a "central point" in $M_{A'}$, the expanded submanifold of $\mathcal{M}$ which allows real entries $\geq -2/3$ in A'-tables. Of course, $S_{A'} = \mathcal{T}_{A'}$ is now the set of all lattice points in $A_{A'}$ (and so, consists of A'-tables with all entries ≥ 0). This shows that there is a ball of radius s_0 in $M_{A'}$, where s_0 is the minimum average line sum occurring in all the components C. Of course, as before, we must restrict s_0 to be reasonably large. The final result is an estimate for λ_S of the form (22) where now the "constant" c_0 depends on the geometry of the specific generators in $\mathcal{K}$. At worst, c_0 decreases by a factor of at most $O((m^2 + n^2)m^2n^2)$ from the unrestricted case. We sketch how this comes about in the Appendix.

Another interesting special case we mention here is that of *symmetric* contingency tables, i.e., with $m = n$, $\overline{r} = \overline{c}$ and $T(i, j) = T(j, i)$ for all i and j. We can use as basic moves in this case the $\binom{n}{2}$ symmetric transformations

$$\begin{aligned} T(i, i) &\rightarrow T(i, i) + 1 \\ T(j, j) &\rightarrow T(j, j) + 1 \\ T(i, j) &\rightarrow T(i, j) - 1 \end{aligned}$$

$$T(j,i) \quad\rightarrow\quad T(j,i)-1$$

and their inverses, for any $i \neq j$. Any symmetric table can be transformed to the diagonal table with $T(i,i) = r_i$ for all i (and 0 otherwise), so that these moves connect the space $\mathcal{S} = \mathcal{S}(\overline{r})$ of symmetric contingency tables for $\overline{r}$. Easy calculations show that $\operatorname{diam} M < 2(\sum_i r_i)^{1/2}$ and that M contains a ball of radius $\frac{1}{n}\min_i r_i$. Thus, by the same arguments that led to (23), we have

$$\lambda_S > \frac{c_0 e^{-1/c}}{n^4 \sum_i r_i^2} \tag{28}$$

for an absolute constant c_0, provided $\min_r r_i > c\binom{n}{2}^{3/2}$. As usual, the translation of this bound to one for the rate of convergence of the walk to uniform is straightforward.

We remark here that a similar analysis can be applied to the more general problem in which our space of objects on which to walk is now a general graph G with nonnegative integer weights assigned to its edges so that for each vertex v of G, the sum of the weights on the edges incident to v is exactly some preassigned value $S(v)$. Again, the steps of the walk will consist of modifying edge weights by alternating ± 1's on certain simple even cycles (forming a cycle basis over $\mathbb{Z}$ of all even cycles in G, analogous to what was done for general bipartite graphs). Not surprisingly, the bound on λ_S for the corresponding Neumann walk has the same general form as for the bipartite case.

7. Compositions of an integer

An easy application of the preceding ideas concerns compositions of an integer T with a fixed number of parts. These are just ordered partitions $(r_1, r_2, \ldots, r_n)$ with integers $r_i > 0$ so that $\sum_i r_i = T$. The basic moves for the walk will be the $n(n-1)$ transformations of the type: $(x_1, \ldots, x_i, \ldots, x_j, \ldots, x_n) \rightarrow (x_1, \ldots, x_i + 1, \ldots, x_j - 1, \ldots, x_n)$. Of course, Γ consists of all points $\overline{x} = (x_1, \ldots, x_n) \in \mathbb{E}^n$, and $\mathcal{M} = \mathcal{M}_T = \{\overline{x} \in \mathbb{E}^n : x_1 + \cdots + x_n = T\}$. We let $S = \{\overline{x} \in \mathcal{M} : x_i \geq -2/3\}$. Then $\operatorname{diam} M \leq \sqrt{2}(T + \frac{2n}{3})$, $\dim M = n - 1$ and M contains a ball of radius T/n. Thus, by (19) if $T > cn^{5/2}$ then

$$\lambda_S > \frac{c_0 e^{-1/c}}{n^2 T^2} \tag{29}$$

for an absolute constant c_0. In the Appendix, we show that we can take $c_0 = 1/200$. Hence, for $T > cn^{5/2}$, if $t > 400 e^{1/c} n^2 T^2 (n \ln T + \ln 1/\epsilon)$ then $\Delta(t) < \epsilon$. We point out that compositions can also be treated somewhat more directly by the methods in [CY2]. This is essentially because S in this case is what is called there "strongly convex", and, in fact, satisfies the stronger condition that if $x, y \in S$, $z \in \partial S$ with $x \sim z \sim y$ then $x \sim y$. As was pointed out in [CY2], we therefore can conclude that

$$\lambda_S \geq \frac{1}{8kD^2}$$

where $k = |\mathcal{K}|$ and $D :=$ graph diameter of S. Thus,

$$\lambda_S \geq \frac{1}{8n(n-1)(T-n)^2} > \frac{1}{8n^2T^2}$$

for this Neumann walk on compositions (with no restrictions on T). We note that Diaconis/Saloff-Coste [DS1] also treat compositions (by quite different methods) for n and T in restricted ranges. In particular, they conjecture that $t = \Omega((nT + T^2)(n \ln T + \ln 1/\epsilon))$ steps suffice to guarantee that $\Delta(t) < \epsilon$.

8. Knapsacks

In our final example, we will consider the following natural generalization of compositions, which we call "knapsack" solutions (e.g., see [DFKKPV], [DGS]). We are given an integer vector $\overline{a} = (a_1, \ldots, a_n)$ with $a_i > 0$ and $gcd(a_1, \ldots, a_n) = 1$. For an integer T, we consider the set $S = S_T$ of all integer vectors $\overline{r} = (r_1, \ldots, r_n)$, $r_i > 0$, such that

$$\overline{r} \cdot \overline{a} = \sum_i r_i a_i = T \ . \tag{30}$$

As usual, our goal will be to (approximately) select a random element from (and enumerate) the set of "knapsack" vectors in S. We do this by constructing a rapidly mixing Markov chain on S, and estimating the corresponding Neumann eigenvalue λ_S. The manifold $\mathcal{M}$ is just $\{\overline{x} \in \mathbb{E}^n : \sum_i a_i x_i = T\}$, and $M := \{\overline{x} \in \mathcal{M} : x_i > -2/3,\ 1 \leq i \leq n\}$. Let $e_i = (0, \ldots, 1, \ldots, 0)$ with 1 in the i^{th} component, and define $g_{ij} = a_j e_i - a_i e_j$. Our generator set will be $\mathcal{K} = \{\pm g_{ij} : 1 \leq i \neq j \leq n\}$.

The first problem is to show that S is connected with the generators from $\mathcal{K}$, provided T is sufficiently large. For ease of exposition we are going to make the assumption that $gcd(a_1, a_2) = 1$. The general case in which we only assume $gcd(a_1, a_2, \ldots, a_n) = 1$ is somewhat more technical but offers no real new difficulty. Define $A = a_1 + a_2 + \cdots a_n$, and suppose $T > 2A \max_i \{a_i\}$. Then it is easy to see that by repeated application of generators of the form

g_{3i}, $i \neq 3$, we can reach a vector $(r_1, r_2, \dots, r_n) \in S$ so that $r_3 a_3 > 2a_1 a_2$. Now it is well known (see [EG80]) that any $w > 2a_1 a_2$ can be represented as

$$w = w_1 a_1 + w_2 a_2, \quad w_1, w_2 \text{ integers } \geq 0\ .$$

Hence, we can apply g_{13} w_1 times, and then g_{23} w_2 times to reach the vector

$$(r_1 + w_1 a_3, r_2 + w_2 a_3, 0, \dots, r_n)$$

which has 0 for its 3rd component. Now, can repeat this argument for each of the other components r_i, $i \geq 4$, and eventually arrive at a vector $(r_1', r_2', 0, 0, \dots, 0) \in S$. Finally, by applying g_{12} appropriately we can finally reach $(r_1^*, r_2^*, 0, 0, \dots, 0) \in S$ where $0 \leq r_2^* < a_1$. Such a vector is unique, and in reaching it we always remained in S. Thus, S is connected by the generators in $\mathcal{K}$. Also, $\dim M = n-1$ and $\operatorname{diam} M \leq T$. Since the point $(T/A, T/A, \dots, T/A) \in S$, then M contains a ball of radius T/A. Hence, by the usual arguments (by now) where $R < (2(a_1^2 + \cdots + a_n^2))^{1/2}$ in Claim 2 by using generators $g_{i,i+1}$, if $T > cn^{3/2}\left(\sum_i a_i^2\right)^{1/2}$ then

$$\lambda_S > \frac{c_0 e^{-1/c}}{n^2 T^2} \tag{31}$$

where c_0 depends only on the geometry of the generators in $\mathcal{K}$, and not on T. (We estimate c_0 in the Appendix.) Thus, by (26), if $T > cn^{3/2}\left(\sum_i a_i^2\right)^{1/2}$ then for

$$t > c_1 e^{1/c} n^2 T^2 (n \ln T + \ln 1/\epsilon) \tag{32}$$

we have $\Delta(t) < \epsilon$ (since $|S| < T^n$) where c_1 depends only on the geometry of the vectors in $\mathcal{K}$.

Appendix

In this section, we provide additional details needed for bounding the constant c_0 occurring in (19). The arguments will extend (and depend on) those in [CY1]. Briefly, in [CY1] we have

$$\begin{aligned} -\frac{2d}{\ell^2}\mathcal{L}_s &= \frac{2d}{|\mathcal{K}|}\sum_{g\in\mathcal{K}^*}\left(\frac{\mu(x,gx)}{\ell}\right)^2\frac{\partial^2}{\partial g^2} \\ &= \sum_{i,j} a_{ij}\frac{\partial^2}{\partial x_i \partial x_j} \end{aligned}$$

where $d := \dim M$, $\ell := \min_{g\in\mathcal{K}} \mu(x, gx)$, μ denotes (Euclidean) length, and $\mathcal{K}^* \subset \mathcal{K}$ consists of exactly one element from each pair $\{g, g^{-1}\}$, $g \in \mathcal{K}$.

Suppose C_1 and C_2 are constants so that

$$C_1 I \leq (a_{ij}) \leq C_2 I$$

where I is the identity operator on M, and $X \leq Y$ means that the operator $Y - X$ is positive definite. In particular, we can take for C_1 and C_2 the least and greatest eigenvalues, respectively, of (a_{ij}) restricted to M. Now, it follows from the arguments in [CY1] that when $\mathcal{M}$ is Euclidean then the constant c_0 in (19) can be taken to be

$$c_0 = \frac{1}{100}\min(C_1, C_2^{-1}) \ .$$

Thus, to determine c_0 in various applications, our job becomes that of bounding the eigenvalues of the corresponding matrix (a_{ij}).

First, we consider $m \times n$ contingency tables. With each edge generator $g = x_{ij} - x_{i'j} - x_{ij'} - x_{i'j'}$ we consider $\frac{\partial^2}{\partial g^2}$ in terms of the x's.

Expanding, we have

$$\begin{aligned} \frac{\partial^2}{\partial g^2} &= \frac{\partial^2}{\partial x_{ij^2}} + \frac{\partial^2}{\partial x_{i'j}^2} + \frac{\partial^2}{\partial x_{ij'}^2} + \frac{\partial^2}{\partial x_{i'j'}^2} - 2\frac{\partial^2}{\partial x_{ij}\partial x_{i'j}} \\ &\quad - 2\frac{\partial^2}{\partial x_{ij} dx_{ij'}} - 2\frac{\partial^2}{\partial x_{i'j'} dx_{i'j}} - 2\frac{\partial^2}{\partial x_{i'j'} dx_{ij'}} \\ &\quad + 2\frac{\partial^2}{\partial x_{ij}\partial x_{i'j'}} + 2\frac{\partial^2}{\partial x_{i'j} dx_{ij'}} \ . \end{aligned}$$

We can abbreviate this in matrix form as

	x_{ij}	$x_{i'j}$	$x_{ij'}$	$x_{i'j'}$
x_{ij}	1	−1	−1	1
$x_{i'j}$	−1	1	1	−1
$x_{ij'}$	−1	1	1	−1
$x_{i'j'}$	1	−1	−1	1

We need to consider the operator

$$\sum_{g\in\mathcal{K}^*} \frac{\partial^2}{\partial g^2} \ .$$

The corresponding matrix Q has the following coefficient values for its various entries:

Entry	Coefficient
(x_{ij}, x_{ij})	$(m-1)(n-1)$
$(x_{ij}, x_{i'j})$	$-(m-1)$
$(x_{ij}, x_{ij'})$	$-(m-1)$
$(x_{ij}, x_{i'j'})$	1

Thus, Q has two distinct eigenvalues: one is mn with multiplicity $(m-1)(n-1)$, and the other is 0 with multiplicity $m+n-1$. Now, $\dim M = (m-1)(n-1)$ and the operator corresponding to Q when restricted to M has all eigenvalues equal to mn. So the matrix (a_{ij}) has all eigenvalues equal to $2\frac{mn(m-1)(n-1)}{\binom{m}{2}\binom{n}{2}} = 8$, and consequently we can take $C_1 = C_2 = 8$, and $c_0 = 1/800$.

Next, in the case of restricted tables, we associate our restrictions with a bipartite graph B (which we assume for now is 2-connected; the general case involves taking the union of such graphs). As usual, let T denote a fixed spanning tree of B, and let C denote the associated (even) cycle basis for B. For each cycle Z in C with edges $e_1, e_2, \ldots, e_{2r}$ we have the edge generator $g = x_{e_1} - x_{e_2} + \cdots - x_{e_{2r}}$. We consider the matrix (a_{ij}) associated with the operator

$$\frac{2d}{|\mathcal{K}|} \sum_{g\in\mathcal{K}^*} \left(\frac{\mu(x, gx)}{\ell}\right)^2 \frac{\partial^2}{\partial g^2} = \sum_{i,j} a_{ij} \frac{\partial^2}{\partial x_i \partial x_j} \ .$$

Clearly, for $f : E(B) \to \mathbb{R}$, we can express (a_{ij}) as a quadratic form:

$$\langle f, (a_{ij}) f\rangle = \sum_{Z\in C} (f(e_1) - f(e_2) + \cdots - f(e_{2r}))^2 \cdot r$$

since $\frac{2d\|Z\|^2}{|\mathcal{K}|\ell^2} = r$.

To upper-bound the eigenvalues of (a_{ij}) we have

$$C_2 \le \frac{2d}{|\mathcal{K}|}(m+n)^2 mn \le 2(m+n)^2 mn$$

since

$$\begin{aligned}
\frac{\langle f, (a_{ij}) f\rangle}{\langle f, f\rangle} &\le \sum_{Z\in C} \frac{(f(e_1) - f(e_2) + \cdots - f(e_{2r}))^2 r}{\sum_i f^2(e_i)} \\
&\le \frac{\sum_{Z\in C} (f^2(e_1) + \cdots + f^2(e_{2r})) r^2}{\sum_i f^2(e_i)} \\
&\le r^2 |C| \le (m+n)^2 mn \ .
\end{aligned}$$

To establish a bound for C_1, more work is needed. We will use the following modified discrete version of Cheeger's theorem (see [C]):

For a graph G, suppose f is the eigenfunction associated with some nonzero eigenvalue λ of the Laplacian of G. Then λ satisfies

$$\lambda = \sum_{u \sim v} \frac{(f(u) - f(v))^2}{\sum_v f^2(v) d_v} \geq \frac{h}{2} \lambda^2$$

where $h_\lambda = \min_v h(v)$ and

$$h(v) := \frac{|\{\{u, w\} \in E(G) : f(u) \leq f(v) < f(w)\}|}{\min \left\{ \sum\limits_{\substack{u \\ f(u) \leq f(v)}} d_u, \quad \sum\limits_{\substack{w \\ f(v) < f(w)}} d_w \right\}} .$$

In particular, for $\lambda \neq 0$,

$$h_\lambda \geq \frac{1}{E(G)} .$$

Before applying the above result, we will modify $A = (a_{ij})$. First, choose a root ρ in T. For each tree edge $e = \{u, v\}$ with $d_T(\rho, u) < d_T(\rho, v)$ and $d_T(\rho, u)$ odd (where d_T denotes the graph distance in T), we define

$$e' = e - e_1 + e_2 - \cdots$$

where the unique path from u to ρ consists of the edges $e_1, e_2, \ldots$. Let X denote the matrix corresponding to this change of coordinates. Thus,

$$\frac{\langle f, X^{tr}(a_{ij})Xf \rangle}{\langle fX^{tr}, Xf \rangle} \cdot \frac{\langle fX^{tr}, Xf \rangle}{\langle f, f \rangle} = \frac{\langle f, A'f \rangle}{\langle f, f \rangle} \tag{33}$$

where $A' = X^{tr}(a_{ij})X$ corresponds to the quadratic form

$$\langle f, (a'_{ij})f \rangle = \sum_{Z \in C} (f(e'_1) - f(e'_2) + f(e) - f(e_1))^2 r .$$

Note that there are just four terms in each of the squared terms in the sum. By (33), the eigenvalues of (a'_{ij}) are products of eigenvalues of (a_{ij}) and eigenvalues of $X^{tr}X$. So, to lower-bound the nonzero eigenvalues of A', we apply the above result twice, since the 4-term sum can be interpreted as two 2-term sums. Thus, we have a lower bound in this case of

$$\frac{1}{(m+n)^2} \cdot \frac{1}{(2mn)^2} .$$

This implies that we can take

$$c_0 = \frac{1}{400(m+n)^2 m^2 n^2}$$

for restricted tables.

For compositions, our generators are of the form

$$g_{ij} = (0, \ldots, 1, \ldots, -1, \ldots, 0) = x_i - x_j \ .$$

Then

$$\frac{2d}{|\mathcal{K}|} \sum_{\{i,j\}} \frac{\partial^2}{\partial g_{ij}^2} = \sum_{ij} a_{ij} \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j}$$

where

$$a_{ij} = \begin{cases} \frac{2(n-1)}{n} & \text{if} \quad i = j \\ -\frac{2}{n} & \text{if} \quad i \neq j \end{cases} .$$

Thus, (a_{ij}) has eigenvalues 0 of multiplicity one, and 2 of multiplicity $n-1$. This implies that C_1 and C_2 are both equal to 2 so that we can take $c_0 = 1/200$.

Finally, for the knapsack problem, we have edge generators $(\ldots, a_j, \ldots, -a_i, \ldots)$. These correspond to $g_{ij} = a_j x_i - a_i x_j$. Therefore

$$\begin{aligned} \sum_{\{i,j\}} \left(\frac{a_i^2 + a_j^2}{2} \right) \frac{\partial^2}{\partial g_{ij}^2} &= \sum_{\{i,j\}} \left(a_i^2 \frac{\partial^2}{\partial x_i^2} - 2 a_i a_j \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} + a_j^2 \frac{\partial}{\partial x_j^2} \right) \\ &= \frac{|\mathcal{K}|}{2d} \sum_{i,j} a_{ij} \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} \end{aligned}$$

where the sum is taken over all unordered pairs i, j. The matrix (a_{ij}) corresponds to the quadratic form

$$\frac{2d}{|\mathcal{K}|\ell^2} \sum_{i,j} \left(\frac{a_i^2 + a_j^2}{2} \right) (a_j x_i - a_i x_j)^2 = \langle \overline{x}, A\overline{x} \rangle$$

where $d = n-1$, $|\mathcal{K}| = n(n-1)$, and $\ell = \min\limits_{i,j} (a_i^2 + a_j^2)^{1/2}$ is the minimum edge length. Set $\beta_i = \sum\limits_j \frac{1}{2}(a_i^2 + a_j^2) a_i^2 a_j^2$. Then

$$\begin{aligned} \min_{i,j}(a_i^2 + a_j^2) \cdot \frac{n}{2} \frac{\langle \overline{x}, A\overline{x} \rangle}{\langle \overline{x}, \overline{x} \rangle} &= \frac{\sum\limits_{\{i,j\}} \frac{1}{2}(a_i^2 + a_j^2)(a_j x_i - a_i x_j)^2}{\sum\limits_i x_i^2} \\ &= \frac{\sum\limits_{\{i,j\}} \frac{1}{2}(a_i^2 + a_j^2) a_i^2 a_j^2 (y_i - y_j)^2}{\sum\limits_i a_i^2 y_i^2} \quad \text{where} \quad y_i = \frac{x_i}{a_i} \\ &\geq \frac{\min\limits_i a_i^2 \sum\limits_{\{i,j\}} a_i^2 a_j^2 (y_i - y_j)^2}{\sum\limits_i a_i^2 y_i^2} \\ &\geq \min_i a_i^2 \cdot \min_i \sum_{\substack{j \\ j \neq i}} a_j^2 \frac{\sum\limits_{\{i,j\}} a_i^2 a_j^2 (y_i - y_j)^2}{\sum\limits_i a_i^2 y_i^2 \sum\limits_{\substack{k \\ k \neq i}} a_k^2} . \end{aligned}$$

Now, by the modified Cheeger theorem referred to previously, we have for any eigenvalue $\lambda \neq 0$ of A

$$\min_{i,j}(a_i^2 + a_j^2)\frac{n}{2}\lambda \geq \frac{1}{2}h^2 \left(\sum_j a_j^2 - \max_i a_i^2\right) \min_i a_i^2$$

where h is the Cheeger constant for the complete graph K_n with edge weights $a_i^2 a_j^2$, which is defined by

$$h = \inf_{I \subset V} \frac{\sum\limits_{i \in I} a_i^2 \sum\limits_{j \notin I} a_j^2}{\sum\limits_{i \in I} a_i^2 \sum\limits_{j \neq i} a_j^2}$$

taken over all $I \subset V = V(K_n)$ satisfying

$$\sum_{i \in I} a_i^2 \left(\sum_{k \neq i} a_k^2\right) \leq \frac{1}{2} \sum_i \sum_{\substack{k \\ k \neq i}} a_i^2 a_k^2$$

i.e.,

$$\sum_{j \notin I} a_j^2 \sum_{\substack{k \\ k \neq j}} a_k^2 \geq \sum_{i \in I} a_i^2 \sum_{\substack{k \\ k \neq i}} a_k^2 \ .$$

First, suppose $\sum\limits_{i \in I} a_i^2 \leq \sum\limits_{j \notin I} a_j^2$. Then

$$h \geq \frac{\sum\limits_{i \in I} a_i^2 \sum\limits_{j \notin I} a_j^2}{\sum\limits_{i \in I} a_i^2 \sum\limits_{\substack{k \\ k \neq i}} a_k^2} \geq 1/2 \ .$$

On the other hand, suppose $\sum\limits_{j \notin I} a_j^2 < \sum\limits_{i \in I} a_i^2$. Then

$$\begin{aligned} h &\geq \frac{\sum\limits_{i \in I} a_i^2 \sum\limits_{j \notin I} a_j^2}{\sum\limits_{j \notin I} a_j^2 \sum\limits_{\substack{k \\ k \neq j}} a_i^2} \\ &\geq \frac{\sum\limits_{i \in I} a_i^2 \sum\limits_{j \notin I} a_j^2}{\sum\limits_{j \notin I} a_j^2 \sum\limits_{k} a_k^2} \geq \frac{1}{2} \ . \end{aligned}$$

Thus, in both cases we get $h \geq 1/2$ so that

$$\frac{\langle \overline{x}, A\overline{x} \rangle}{\langle \overline{x}, \overline{x} \rangle} \geq \left(\frac{2}{n}\right) \cdot \left(\frac{1}{2}\right)^2 \frac{\min\limits_i a_i^2 \min\limits_i \sum\limits_{j \neq i} a_j^2}{\min\limits_{i,j}(a_i^2 + a_j^2)} = C_1$$

For C_2, we compute

$$\frac{\langle \overline{x}, A\overline{x} \rangle}{\langle \overline{x}, \overline{x} \rangle} \quad \leq \quad \frac{2}{n \min\limits_{i,j}(a_i^2 + a_j^2)} \frac{\sum\limits_{\{i,j\}} \frac{1}{2}(a_i^2 + a_j^2)(a_j x_i - a_i x_j)^2}{\sum\limits_i x_i^2}$$

$$\begin{aligned} &\leq \frac{2}{n\min\limits_{i,j}(a_i^2+a_j^2)} \frac{\sum\limits_{\{i,j\}}(a_i^2+a_j^2)(a_j^2x_i^2+a_i^2x_j^2)}{\sum\limits_i x_i^2} \\ &\leq \frac{2}{n\min\limits_{i,j}(a_i^2+a_j^2)} \max_i \sum_{\substack{j \\ j\neq i}} a_j^2(a_i^2+a_j^2) = C_2 \end{aligned}$$

and, as usual, we take $c_0 = \frac{1}{100}\min(C_1, C_2^{-1})$.

References

[A87] D. Aldous, On the Markov-chain simulation method for uniform combinatorial simulation and simulated annealing, Prob. Eng. Info. Sci. 1 (1987), 33-46.

[BM76] J. A. Bondy and U. S. R. Murty, Graph Theory with Applications, American Elsevier, New York, 1976.

[C] F. R. K. Chung, Laplacians of graphs and Cheeger's inequalities (to appear).

[CY1] F. R. K. Chung and S.-T. Yau, A Harnack inequality for homogeneous graphs and subgraphs, Comm. Analysis and Geometry, 2 (1994) 628–639.

[CY2] F. R. K. Chung and S.-T. Yau, Heat kernel estimates and eigenvalue inequalities for convex subgraphs (preprint).

[DG] P. Diaconis and A. Gangolli, Rectangular arrays with fixed margins (preprint).

[DGS] P. Diaconis, R. L. Graham and B. Sturmfels, Primitive partition identities (to appear).

[DS1] P. Diaconis and L. Saloff-Coste, Random walk on contingency tables with fixed row and column sums (preprint).

[DSt] P. Diaconis and B. Sturmfels, Algebraic algorithms for sampling from conditional distributions (to appear in Ann. of Statistics).

[DFK89] M. Dyer, A. Freeze and R. Kannan, A random polynomial time algorithm for approximating the volume of convex bodies, in Proc. 21st ACM Symp. on Theory of Computing, 1989, pp. 375–381.

[DFKKPV] M. Dyer, A. Freeze, R. Kannan, A. Kapour, L. Perkovic, U. Vazirani, A mildly exponential time algorithm for approximating the number of solutions to a multi-dimensional knapsack problem, (to appear in Combinatorics, Probability and Computing).

[DKM] M. Dyer, R. Kannan and J. Mount, Sampling contingency tables (preprint).

[EG80] P. Erdős and R. L. Graham, Old and new problems and results in combinatorial number theory, L'Enseignement Mathematique, Monographe N°28, Genève, 1980, p. 86.

[FKP94] A. Frieze, R. Kannan and N. Polson, Sampling from log-concave distributions, Annals Appl. Prob. 4 (1994), 812–837.

[G91] A. Gangolli, Convergence bounds for Markov chains and applications to sampling, Ph.D. Dissertation, Dept. of Comp. Sci., Stanford, 1991.

[JS] M. Jerrum and S. Sinclair, Fast uniform generation of regular graphs, Th. Comp. Sci. 73 (1990), 91–100.

[LS90] L. Lovàsz and M. Simonovits, The mixing rate of Markov chains, an isoperimetric inequality and computing the volume, in Proc. 31st IEEE Symp. on Foundations of Computer Science, 1990, pp. 346–354.

[MW] B. McKay and N. Wormald, Uniform generation of random regular graphs of moderate degree (to appear in J. of Algorithms).

[SY94] R. Schoen and S-T. Yau, Differential Geometry, International Press, Cambridge, 1994.

[S93] A. Sinclair, Algorithms for random generation and counting: a Markov chain approach, Birkhauser, Boston, 1993.

[SJ89] A. Sinclair and M. Jerrum, Approximate counting, uniform generation and rapidly mixing Markov chains, Infor. and Comput. 82 (1989), 93–133.