Random Walks in Convex Sets

by

Benjamin James Morris

A.B. (University of California at Berkeley) 1995

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Statistics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:

Professor Alistair Sinclair, Chair
Professor David Aldous
Professor Jim Pitman
Professor Christos Papadimitriou

Fall 2000

The dissertation of Benjamin James Morris is approved:

Chair Date

Date

Date

Date

University of California, Berkeley

Fall 2000

Random Walks in Convex Sets

Abstract

Random Walks in Convex Sets

by

Benjamin James Morris
Doctor of Philosophy in Statistics

University of California, Berkeley

Professor Alistair Sinclair, Chair

We study mixing times for nearest-neighbor random walks on sets of the form $\mathbf{Z}^n \cap K$ for a convex set K, and the related problem of sampling (almost) uniformly at random from such sets.

In chapter 2, we solve an open problem concerning symmetric random walk on the n-dimensional cube truncated by a hyperplane, showing that the mixing time is polynomial in n. As a consequence, we obtain a fully-polynomial randomized approximation scheme for counting the feasible solutions of a 0-1 knapsack problem.

In chapter 3, we extend to the case where the cube is truncated by many hyperplanes, each of which corresponds to a knapsack constraint. We show that if the number of hyperplanes is fixed, then there is a polynomial upper bound for the mixing time. In chapters 2 and 3, the key ingredient in the analysis is a combinatorial construction we call a "balanced almost uniform permutation," which seems to be of independent interest.

In the final chapter we study the problem of sampling contingency tables (non-negative integer matrices with specified row and column sums) uniformly at random. We give an algorithm which runs in polynomial time provided that the row and column sums satisfy $r_i = \Omega(n^{3/2} m \log(m))$, and $c_j = \Omega(m^{3/2} n \log(n))$. This algorithm is based on a reduction to continuous sampling from a convex set. This is an approach which was taken by Dyer, Kannan and Mount in previous work. However, the algorithm we present is simpler and has

weaker requirements on the row and column sums.

Professor Alistair Sinclair
Dissertation Committee Chair

To my wife,

Maria Liu,

without whom I would not be married.

Contents

Acknowledgements

I would like to thank (i) my advisor Alistair Sinclair for being an excellent advisor, (ii) my committee, (iii) Persi Diaconis for telling me about [12], which greatly simplified chapter 4, (iv) Steve Evans, Yuval Peres and anybody else who talked with me about this thesis, (v) Mike Ostland, Rafael Irizarry and anybody else who helped me learn LaTeX.

Chapter 1

Introduction

Random walks on discrete grids inside convex sets tend to be rapidly mixing, i.e., converge quickly to their stationary distribution. For a convex set $K \in \mathbf{R}^n$ and positive real number δ, define *coordinate random walk* on the intersection $\Omega = \delta\mathbf{Z}^n \cap K$ by the following transition rule: If the current state is x, choose a neighbor $y \in \delta\mathbf{Z}^n$ uniformly at random. If $y \in K$, move there; otherwise, remain at x.

If the grid is very fine, we certainly should expect that coordinate random walk be rapidly mixing: the vertex set Ω will have roughly the same shape as K, so it won't contain any "bottlenecks" that would slow down convergence. This can be formalized using the combinatorial quantity called "conductance"; see [1, 28]. For any subset S of Ω, define

$$\Phi(S) = \frac{\text{number of edges in the } (S, \overline{S}) \text{ cut}}{n \min(|S|, |\overline{S}|)}, \tag{1.1}$$

where $\overline{S} = \Omega - S$. Define the conductance Φ of the Markov Chain by

$$\Phi = \min_{S \subset \Omega} \Phi(S).$$

The conductance, which gives a measure of the bottlenecks in the graph, governs the mixing time[1] of the Markov Chain. It's easy to see why a Markov chain with small conductance should not be rapidly mixing. Let $S \subset \Omega$ and suppose that $|S| \leq \frac{|\Omega|}{2}$. If $\Phi(S)$ is small, and the random walk is started in one of the vertices in S, then it is liable to remain in S for a long time. Therefore, the mixing time will be large.

[1] Informally, the mixing time is the number of steps needed for the Markov Chain to get close to stationarity, starting from any initial state. For a precise definition, see Section 2.2

More interestingly, the converse is also true [2, 28]: if Φ is large (and thus, there are no major bottlenecks in the graph) then the mixing time is small. More precisely, in the case of coordinate random walk the following holds.

- Coordinate random walk is *rapidly mixing*, i.e. has a mixing time which is a polynomial function of the dimension n, if and only if the conductance satisfies $\Phi > \frac{1}{p(n)}$ for some polynomial p.

In a famous paper, Dyer, Frieze and Kannan [9] proved that (under some "well-roundedness" conditions on the convex set K) coordinate random walk is indeed rapidly mixing if the width δ of the grid is sufficiently small.[2] This led to the first approximation algorithm for calculating the volume of a convex set. Their proof used ideas from geometry to give a lower bound on conductance. It is known (see [7]) that for any convex set K we have

$$\min_{S \subset K} \frac{\mathrm{vol}_{n-1}(\partial S \cap \mathrm{int}(K))}{\min(\mathrm{vol}_n(S), \mathrm{vol}_n(K - S))} \geq \frac{2}{\mathrm{diam}(K)}, \tag{1.2}$$

where vol_{n-1} and vol_n denote the surface area and (n-dimensional) volume, respectively, diam denotes the diameter and int denotes the interior. Inequalities like (1.2) are called *isoperimetric inequalities.* By comparing the numerator and denominator in (1.1) with those in (1.2), Dyer et al obtained a $\frac{1}{\mathrm{poly}(n)}$ lower bound on conductance, proving rapid mixing.

What happens, on the other hand, when δ is not small? In this case, there are examples where the random walk does not mix rapidly. The following example shows that if the grid is too coarse, then there can be bottlenecks in the graph.

Example: Two unit hypercubes joined by a single point. Let

$$S = \{0,1\}^n \cup \{0,-1\}^n \qquad\qquad K = \mathrm{conv}(S),$$

where conv denotes the convex hull. It's not hard to show that $Z^n \cap K = S$. Thus, if $\delta = 1$ then the state space Ω of coordinate random walk is just S. Clearly, the mixing time is exponential because of the bottleneck through the origin. Even though K has no bottlenecks because it is convex, there is a bottleneck in Ω; we can think of this as a "boundary effect" caused by the coarse discretization.

[2]We are using the term "sufficiently small" rather loosely here. Of course, if the width δ is too small, then the random walk won't be rapidly mixing because the diameter of the graph will be too large. What Dyer et al proved is that the walk is rapidly mixing for δ in a suitible range.

There are certain types of convex sets that give problems even with an arbitrarily fine grid. These sets have "sharp corners." Consider the following example.

EXAMPLE: CONVEX SET WITH A SHARP CORNER. Let

$$S = \bigcup_{l=1}^{\infty} \left\{1 - (\tfrac{1}{2})^{l-1}, 1 - (\tfrac{1}{2})^{l}\right\}^n \qquad\qquad K = \mathrm{conv}(S).$$

S is the union of an infinite sequence of hypercubes which decrease in size; the kth hypercube has side length $(\frac{1}{2})^k$ and is joined to the $(k+1)$th hypercube at the point $\left(1 - (\frac{1}{2})^k, \ldots, 1 - (\frac{1}{2})^k\right)$. Note that K has a sharp spike in the direction of $(1, \ldots, 1)$. Now, let $\Omega = \delta\mathbf{Z}^n \cap K$ for $\delta = (\frac{1}{2})^l$. In the resulting graph, the entire hypercube $\left\{1 - (\frac{1}{2})^{l-1}, 1 - (\frac{1}{2})^l\right\}^n$ is connected to the rest of Ω only through the single point $\left(1 - (\frac{1}{2})^{l-1}, \ldots, 1 - (\frac{1}{2})^{l-1}\right)$. This is the case for all l. Thus, even for arbitrarily small values of δ, the graph still has bottlenecks. In fact, if we let K contain the limit point $(1, \ldots, 1)$, then for $\delta = (\frac{1}{2})^l$ the graph wouldn't even be connected.

In the context of volume estimation, the restrictions on δ and K are not a problem. The goal there was to sample uniformly from the convex set K, not a discrete grid. Thus Dyer el al were free to choose a sufficiently small value of δ (provided it was large enough to keep the diameter of the graph reasonable; see the footnote on the previous page.) They were also free to alter K as necessary, removing any sharp corners, provided that this didn't change the volume too much.

However, there are many other interesting and important problems that involve sampling from a *fixed* grid inside a convex set. In these problems, the width δ of the grid may be large and the convex set may contain sharp corners, so new types of analysis are needed. This thesis focuses on problems of this type. We describe some of our results below.

1.1 Sampling knapsack solutions

For a positive real vector $\mathbf{a} = (a_i)_{i=1}^n$ and real number b, let Ω denote the set of 0-1 vectors $\mathbf{x} = (x_i)_{i=1}^n$ for which

$$\mathbf{a}'\mathbf{x} \equiv \sum_{i=1}^{n} a_i x_i \leq b.$$

Geometrically, we can view Ω as the set of vertices of the n-dimensional cube $\{0,1\}^n$ which lie on one side of the hyperplane $\mathbf{a}'\mathbf{x} = b$. Combinatorially, Ω is the set of feasible solutions to the 0-1 knapsack problem defined by $\mathbf{a}$ and b: if we think of the a_i as the weights of a set of n items, and b as the capacity (weight limit) of a knapsack, then there is a 1-1 correspondence between vectors $\mathbf{x} \in \Omega$ and subsets of items $X \subseteq \{1, \ldots, n\}$ whose aggregated weight does not exceed the knapsack capacity, given by $X = \{i : x_i = 1\}$.

It has been quoted as an open problem in several places [8, 14, 18, 26] to give a polynomial upper bound for the mixing time of coordinate random walk on Ω. Much of the interest in this problem stems from the connection to approximate counting: if coordinate random walk on Ω is rapidly mixing then there is a good approximation algorithm (specifically, a fully-polynomial randomized approximation scheme [21]) for the #P–hard problem of counting $|\Omega|$.

To put this into our earlier framework, we have $\Omega = \mathbf{Z}^n \cap K$ for $K = \{\mathbf{x} \in [0,1]^n : \mathbf{a}'\mathbf{x} \leq b\}$. This K is a very simple, yet very extreme example of a convex set; it has sharp corners, and it puts every vertex in Ω onto the boundary. Thus, it poses a challenge to existing methods. Nonetheless, there is strong geometric intuition that the walk should be rapidly mixing: random walk on the entire cube $\{0,1\}^n$ is well known to be rapidly mixing, and truncation by a hyperplane presumably cannot create bottlenecks that would severely slow down convergence. Nonetheless, the best known bound on the mixing time was $\exp(\mathrm{O}(\sqrt{n}(\log n)^{5/2}))$ [8], which beat the trivial bound of $\exp(\mathrm{O}(n))$ but was still exponential. In Chapter 2, we prove that the random walk is indeed rapidly mixing, with a mixing time of $\mathrm{O}(n^8)$ steps. This work appeared in preliminary form in [25].

1.1.1 Balanced almost uniform permutations

To prove rapid mixing we use a technique based on multicommodity flow (see [27]): if we can route unit flow between each pair of vertices X, Y in Ω simultaneously in such a way that no edge carries too much flow, then the random walk is rapidly mixing. Note that a shortest path between a pair of vertices X, Y in Ω can be viewed as a *permutation* of the symmetric difference $X \oplus Y$, the set of items that must be added to or removed from the knapsack in passing from X to Y. A natural approach to defining a good flow is to use a *random* permutation, so that the flow is spread evenly among all shortest paths and no edge is overloaded. However, a fundamental problem with this approach is that a random permutation may give a path along which the weight fluctuates too wildly; i.e., the permuation is "unbalanced." What we need, therefore, are permutations π that are both balanced (i.e., the weights stay in a bounded range), and also "sufficiently random" to ensure a well spread flow. More specifically, it turns out that we require the distribution of the initial segment $\{\pi(1), \ldots, \pi(k)\}$, viewed as an unordered set, to be "almost uniform." We call permutations with these properties *balanced almost uniform* permutations. We prove the following theorem.

Theorem. Let $\{w_i\}_{i=1}^n$ be a sequence of weights, with $\sum_i w_i = 0$, and let $M = \max_i |w_i|$. Then there is a random permutation π on $\{1, \ldots, n\}$, which satisfies

$$\max_{1 \le k \le m} \left| \sum_{i=1}^{k} w_{\pi(i)} \right| \le cM \qquad (\pi \text{ is } \textit{balanced});$$

and for all subsets $U \subset \{1, \ldots, n\}$

$$\Pr(\{\pi(1), \ldots, \pi(k)\} = U) \le \mathrm{poly}(n) \binom{n}{k}^{-1} \qquad (\pi \text{ is } \textit{almost uniform}),$$

where c is a universal constant and poly is a fixed polynomial function. Note that the balance condition is very strict. Nonetheless, there is always a balanced permutation that is also almost uniform. This theorem is the main technical result of Chapter 2 and we conjecture that it will have applications elsewhere.

1.1.2 Multiple constraints

In Chapter 3, we study random walk on the unit hypercube truncated by multiple hyperplanes. Let $\Omega = \{\mathbf{x} \in \{0,1\}^n : A\mathbf{x} \le \mathbf{b}\}$, where A is a non-negative $d \times n$ matrix, $\mathbf{b} \in \mathbf{R}^d$, and d is a constant. Then Ω is the set of feasible solutions to a knapsack problem with multiple contraints.

It turns out that these additional constraints cause some new complications. We can generalize the definition of balance for higher dimensions. Let $\{w_i\}_{i=1}^n$ be a set of weights in $\mathbf{R}^d$, with $w_i = (w_i^1, \ldots, w_i^d)$, and suppose that $\sum_i w_i = 0$. We say π is balanced if

$$\max_{k \le m} \left| \sum_{i=1}^{k} w_{\pi(i)}^j \right| \le c \times \max_i |w_i^j|,$$

for all $1 \le j \le d$, where c is a universal constant. Surprisingly, for $d \ge 2$ it is quite challenging to show that even a *single* balanced permutation exists. Nonetheless, we show, using a non-trivial extension of the analysis in Chapter 2, that for any set of weights $\{w_i\}$

there is a permutation which is not only balanced for these weights, but also almost uniform. Using this result and a similar strategy to the $d = 1$ case, we go on to prove that the random walk on Ω is rapidly mixing for any fixed d.

1.2 Sampling contingency tables

In Chapter 4, we study the problem of sampling contingency tables (non-negative integer matrices with given row and column sums) uniformly at random. In Statistics, contingency tables are used to hold the data from an experiment in which there are 2 variables. The sampling problem arises in the study of the dependence between the two variables, and was motivated by the influential paper by Diaconis and Efron [4]; it has been the subject of at least ten research papers.

The set of contingency tables Ω is of the form $\mathbf{Z}^n \cap K$ for a convex set K, and our method of sampling uses a random walk in a convex set. However, we will not actually use Ω as the state space. As we saw earlier, there exist polynomial-time algorithms for (continuous) sampling from a convex set. Thus, given *any* convex set K' which contains Ω, we can generate a random sample from Ω using the following algorithm:

1. Generate a random point in K', and "round" it to an integer lattice point z.

2. If $z \notin \Omega$, repeat.

Of course, we require that the distribution of the final sample be almost uniform. Thus, as x varies over Ω, we require that $\mathrm{vol}(x)$ is nearly constant, where $\mathrm{vol}(x)$ denotes the volume of points in K' which round to x. Thus we must make K' sufficiently large. (We cannot always naively set K' equal to K; the resulting distribution might place too little mass on the "boundary points" x where $\mathrm{vol}(x)$ is too small.)

However, there is a tradeoff between the the uniformity of the distribution and running time of the algorithm. As K becomes larger, it becomes more likely that each sample z will fall outside of Ω. Thus, if we make K too large, then the expected number of trials taken before we generate a sample $Z \in \Omega$ will be too high.

What we want therefore is a convex set K' which is a good continuous approximation to the discrete set Ω. Now, in the case of contingency tables, the convex set K is a polytope, and hence it can be defined in terms of bounding hyperplanes. We can therefore define K' in terms of another set of hyperplanes which are parallel to the original ones, only spread farther apart to ensure that $\mathrm{vol}(x) \approx 1$ for all $x \in \Omega$.

Using this technique, we show that a polynomial-time algorithm exists for the contingency tables sampling problem when the row and column sums satisfy $r_i = \Omega(n^{3/2} m \log(m))$, and $c_j = \Omega(m^{3/2} n \log(n))$. This improves on the previously best known bounds $r_i = \Omega(n^2 m)$ and $c_j = \Omega(m^2 n)$ obtained by Dyer, Kannan and Mount [11]. It also seems to be the best result obtainable using this technique; when the row and column sums are any smaller, boundary effects seem to make a reduction to continuous sampling impossible. This work appeared in preliminary form in [24]

Chapter 2

Random Walk on a Truncated Cube

2.1 Introduction

For a positive real vector $\mathbf{a} = (a_i)_{i=1}^n$ and real number b, let Ω denote the set of 0-1 vectors $\mathbf{x} = (x_i)_{i=1}^n$ for which

$$\mathbf{a}'\mathbf{x} \equiv \sum_{i=1}^{n} a_i x_i \leq b.$$

Geometrically, we can view Ω as the set of vertices of the n-dimensional cube $\{0,1\}^n$ which lie on one side of the hyperplane $\mathbf{a}'\mathbf{x} = b$; we will call this graph G_Ω. Combinatorially, Ω is the set of feasible solutions to the 0-1 knapsack problem defined by $\mathbf{a}$ and b: if we think of the a_i as the weights of a set of n items, and b as the capacity (weight limit) of a knapsack, then there is a 1-1 correspondence between vectors $\mathbf{x} \in \Omega$ and subsets of items X whose aggregated weight does not exceed the knapsack capacity, given by $X = \{i : x_i = 1\}$. We shall write $a(X)$ for the weight of X, i.e., $a(X) = \sum_{i \in X} a_i$.

In this setting, coordinate random is equivalent to the following Markov chain. If the current state is $X \subseteq \{1, \ldots, n\}$, then

1. pick an item $i \in \{1, \ldots, n\}$ uniformly at random (u.a.r.);
2. if $i \in X$, move to $X - \{i\}$; if $i \notin X$ and $a(X \cup \{i\}) \leq b$, move to $X \cup \{i\}$; otherwise, do nothing.

To avoid technical issues involving periodicity, we add to every state a holding probability of $\frac{1}{2}$: i.e., with probability $\frac{1}{2}$ do nothing, else make a move as described above.

Despite much recent activity in the analysis of mixing times of random walks, this deceptively simple example was still not known to be rapidly mixing. There is strong geometric intuition that it should be: random walk on the entire cube $\{0,1\}^n$ is rapidly mixing, and truncation by a hyperplane presumably cannot create "bottlenecks" that would severely slow down convergence. Nonetheless, the best known bound on the mixing time remained $\exp(\mathrm{O}(\sqrt{n}(\log n)^{5/2}))$ [8], which beat the trivial bound of $\exp(\mathrm{O}(n))$ but is still exponential.

Much of the interest in this random walk stems from the connection to approximate counting. The problem of computing $|\Omega|$ is #P-complete in its exact form. However, by a well-known relationship based on self-reducibility [17, 14], there is a good approximation algorithm, specifically a *fully-polynomial randomized approximation scheme* (fpras) for the counting problem if and only if there is a polynomial time algorithm for sampling elements of Ω (almost) uniformly at random. Therefore, if coordinate random walk on Ω is rapidly mixing then simulation of the walk for sufficiently many steps allows one to sample (almost) uniformly from Ω, and thus to approximate $|\Omega|$.

In recent years there has been a steady stream of results of this kind for #P-complete counting problems (see, for example, [18, 14, 14] for surveys); however, the 0-1 knapsack problem stood as one of a small handful of canonical problems that resisted attack. Indeed, it has been quoted as an open problem in several places [8, 14, 18, 26]. This interest stems in part from its combinatorial significance and its appealing geometric structure, and in part from the challenge it poses to existing methods. In this chapter, we resolve this issue by proving that the random walk is indeed rapidly mixing, with a mixing time of $\mathrm{O}(n^8)$ steps.[1] Along the way we introduce some new machinery that we believe will be useful for tackling other problems of a similar flavor, and possibly beyond.

To prove rapid mixing we use a technique based on multicommodity flow (see [27]): if we can route unit flow between each pair of vertices X, Y in G_Ω simultaneously in such a way that no edge carries too much flow, then the random walk is rapidly mixing. This technique is well known, but most previous applications (e.g., [16, 15]) have made use of "degenerate" flows in which all $X \to Y$ flow is routed along a single canonical path (though

[1] Our aim here is to provide a clear exposition of the arguments, and we have not attempted to optimize the exponent, or other constants appearing elsewhere.

see [3, 27] for exceptions). Our analysis seems to rely essentially on spreading out the flow along multiple paths.

The key ingredient in our analysis is the specification of these paths, which we achieve via an auxiliary combinatorial construction that we believe is of independent interest and will find further applications elsewhere. Note that a shortest path between a pair of vertices X, Y of G_Ω can be viewed as a *permutation* of the symmetric difference $X \oplus Y$, the set of items that must be added to or removed from the knapsack in passing from X to Y. A natural approach to defining a good flow is to use a *random* permutation, so that the flow is spread evenly among all shortest paths and no edge is overloaded. However, a fundamental problem with this approach is that a random permutation will tend to violate the knapsack constraint, as too many items will have been added at some intermediate point; i.e., the permuation is "unbalanced." What we need, therefore, are permutations π that are both balanced and also "sufficiently random" to ensure a well spread flow. More specifically, it turns out that we require the distribution of the initial segment $\{\pi(1), \dots, \pi(k)\}$, viewed as an unordered set, to be "almost uniform." We call permutations with these properties *balanced almost uniform* permutations. A main contribution of this chapter is to show the existence of such permutations.

The remainder of this chapter is structured as follows. We begin with some necessary background on flows and rapid mixing in section 2.2. We then establish rapid mixing of the knapsack random walk in the technically simpler case when the item weights a_i lie in the range $[1, B]$, for some constant B. This analysis is in two parts: in section 2.3 we show how to construct balanced almost uniform permutations, and in section 2.4 how to use these to define a good flow. We then extend everything to the general case in section 2.5.

2.2 The mixing time and multicommodity flow

As indicated earlier, we will view elements of Ω either as 0-1 vectors $\mathbf{x} = (x_i)_{i=1}^n$ or, more commonly, as subsets $X \subseteq \{1, \dots, n\}$, under the equivalence $X = \{i : x_i = 1\}$. Recall that $a(X) = \sum_{i \in X} a_i$ is the weight of X, so that $\Omega = \{X : a(X) \leq b\}$. Without loss of generality, we will assume that $a_i \leq b$ for all i.

We consider the symmetric random walk on the portion G_Ω of the hypercube $\{0,1\}^n$ defined in section 2.1. This walk is connected (all states communicate via the zero vector) and aperiodic (because of the holding probabilities), and since the transition probabilities are symmetric, the distribution at time t converges to the uniform distribution over Ω as $t \to \infty$, regardless of the initial state. Our goal is to bound the rate of convergence as measured by the *mixing time*, defined as

$$\tau_{\text{mix}} = \max_{X_0} \min\Big\{t : \|P_{t'} - \mathcal{U}\| \le \tfrac{1}{4} \quad \forall t' \ge t\Big\},$$

where X_0 is the initial state, P_t is the distribution of the walk at time t, $\mathcal{U}$ is the uniform distribution over Ω, and $\|\cdot\|$ denotes variation distance.[2] Thus τ_{mix} is the number of steps required, starting from any initial state, to get the variation distance from the uniform distribution down to $\frac{1}{4}$. By standard facts about geometric convergence, $\mathrm{O}(\tau_{\text{mix}} \log \epsilon^{-1})$ steps suffice to reduce the variation distance to any desired ϵ.

Fairly standard techniques (see [27]) allow us to estimate τ_{mix} by setting up a suitable multicommodity flow on the underlying graph G_Ω. Our task is to route one unit of flow from X to Y, for each ordered pair of vertices $X, Y \in \Omega$ simultaneously. For any such flow f, and any oriented edge e in G_Ω, let $f(e)$ denote the total flow along e. Define $\mathcal{C}(f) = \frac{1}{|\Omega|} \max_e f(e)$, the maximum flow along any edge normalized by $|\Omega|$, and $\mathcal{L}(f)$ to be the length of a longest flow-carrying path. The following theorem is a special case of results in [27] (see also [6, 5]):

Theorem 2.1 [27] *For any flow f, the mixing time is bounded by $\tau_{\text{mix}} \le 4n(n+1)\mathcal{C}(f)\mathcal{L}(f)$.*

We will bound τ_{mix} by constructing a flow with small values of $\mathcal{C}$ and $\mathcal{L}$. To bring out the main conceptual ideas, we will focus initially on what we term the *bounded ratio* case, where all weights a_i lie in the range $[1, B]$ for some constant B. We will derive a bound of the form $\tau_{\text{mix}} = n^{\mathrm{O}(B^2)}$ in this case. By introducing some additional technical complications, we will go on to get a uniform bound of $\tau_{\text{mix}} = \mathrm{O}(n^8)$ for the general case.

As explained in section 2.1, our flow will be based on the idea of a balanced almost uniform permutation. We devote the next section to this topic, and then return to the knapsack random walk in section 2.4.

2.3 Balanced almost uniform permutations

We begin by defining the notions of “balanced” and “almost uniform” permutations. We will write $\mathcal{S}_m$ to denote the set of all permutations of $\{1, \ldots, m\}$.

[2]For probability distributions μ, ν on Ω, the variation distance is defined as $\|\mu - \nu\| = \frac{1}{2}\sum_{x \in \Omega} |\mu(x) - \nu(x)| = \max_{S \subseteq \Omega} |\mu(S) - \nu(S)|$.

Definition 2.2 *Let $\{w_i\}_{i=1}^m$ be a set of real (not necessarily positive) weights, with $M = \max_i |w_i|$ and $W = \sum_i w_i$. A permutation $\pi \in \mathcal{S}_m$ is* balanced *if, for every k with $1 \leq k \leq m$,*

$$\min\{W, 0\} - M \leq \sum_{i=1}^{k} w_{\pi(i)} \leq \max\{W, 0\} + M. \tag{2.1}$$

Thus a balanced permutation is one whose partial sums do not fluctuate widely. In particular, if $\sum_i w_i = 0$ then condition (2.1) becomes $|\sum_{i=1}^k w_{\pi(i)}| \leq M$.

Definition 2.3 *Let π be a random variable taking values in $\mathcal{S}_m$, and let $\lambda \in \mathbf{R}$. We call π a* λ-uniform *permutation if*

$$\Pr[\pi\{1, \ldots, k\} = U] \leq \lambda \times \binom{m}{k}^{-1} \tag{2.2}$$

for every k with $1 \leq k \leq m$ and every $U \subseteq \{1, \ldots, m\}$ of cardinality k. (Here $\pi\{1, \ldots, k\}$ denotes the initial segment $\{\pi(1), \ldots, \pi(k)\}$.)

Note that, if π were a uniform random permutation, the probability in (2.2) would be exactly $\binom{m}{k}^{-1}$ for every U. In a λ-uniform permutation the probabilities are permitted to vary with U, but only by an amount specified by the parameter λ. In our applications, λ will be a fixed polynomial function of m; in this case we call π an *almost uniform* permutation.

The perhaps surprising result of this section is that, if the ratios of the weights are bounded, it is possible to construct an almost uniform permutation that is balanced (with probability 1). In section 2.5.1 we will show how to dispense with any restrictions on the weights.

Theorem 2.4 *Let $\{w_i\}_{i=1}^m$ be any set of weights with $|w_i| \in [1, B]$ for a constant B. Then there exists a balanced almost uniform permutation π on $\{w_i\}$.*

Proof: Let $M = \max_i |w_i|$ and $W = \sum_{i=1}^m w_i$. Assume first that $W = 0$; we will show how to discharge this assumption later. Let $I_1 = \{i : w_i > 0\}$, $I_2 = \{i : w_i < 0\}$, $m_1 = |I_1|$ and $m_2 = |I_2|$. Define the means $\mu_1 = \frac{1}{m_1} \sum_{i \in I_1} w_i$ and $\mu_2 = -\frac{1}{m_2} \sum_{i \in I_2} w_i$. Note that $m_1\mu_1 = m_2\mu_2$ since $W = 0$.

Consider an arbitrary permutation $\nu \in \mathcal{S}_m$. This induces permutations ν_1, ν_2 on I_1, I_2 respectively.[3] We call ν_1 *α-good* if, for every k_1 with $1 \leq k_1 \leq m_1$,

$$\left|\sum_{i=1}^{k_1} w_{\nu_1(i)} - k_1\mu_1\right| \leq \alpha(M-1)\sqrt{k_1^*}, \tag{2.3}$$

[3] Formally, we view ν_1 as a bijection from $\{1, \ldots, m_1\}$ to I_1, and similarly for ν_2. Throughout we shall adopt this convention where appropriate, without comment.

where $k_1^* = \min\{k_1, m_1 - k_1\}$, with an analogous definition for ν_2. We call ν α-*good* if both ν_1 and ν_2 are α-good. Thus in a good permutation, the partial sums of both positive and negative weights are close to their expected values.

Now suppose ν is chosen u.a.r. from $\mathcal{S}_m$. A routine application of Hoeffding's bound to the partial sums (see Lemma A.1 in Appendix A) yields

$$\Pr[\nu \text{ is not } \alpha\text{-good}] \le 2m \exp(-2\alpha^2). \tag{2.4}$$

If we set $\alpha = \sqrt{\ln m}$, this probability is at most $\frac{2}{m} \le \frac{1}{2}$ for $m \ge 4$.

Consider now a modified sample space in which ν is selected u.a.r. among all $\sqrt{\ln m}$-good permutations. We shall write $\Pr_{\text{unif}}$ for probabilities in the original uniform space to distinguish them from those in this modified space. By the above calculation, for any event $\mathcal{E} \subseteq \mathcal{S}_m$ we have

$$\Pr[\mathcal{E}] \le 2 \Pr_{\text{unif}}[\mathcal{E}]. \tag{2.5}$$

We are now in a position to construct our balanced almost uniform permutation. Let ν be chosen u.a.r. from all $\sqrt{\ln m}$-good permutations, and let ν_1, ν_2 be the induced permutations on I_1, I_2. To get a balanced permutation π, we *interleave* ν_1 and ν_2 as follows. We take the first element from ν_1, i.e., set $\pi(1) = \nu_1(1)$. Thereafter, for each $k > 1$ in turn we set $\pi(k)$ to be the next element in ν_2 if $\sum_{i=1}^{k-1} w_{\pi(i)} \ge 0$, and the next element in ν_1 otherwise. Since $\sum_i w_i = 0$ this process is well-defined and yields a permutation $\pi \in \mathcal{S}_m$. Moreover, since $|w_i| \le M$ for all i it is clear that π satisfies the balance condition (2.1).

We now need to verify the uniformity condition (2.2), for $\lambda = \text{poly}(m)$. Let $U \subseteq \{1, \ldots, m\}$ be arbitrary with $|U| = k$, and let $U_1 = U \cap I_1$, $U_2 = U \cap I_2$, $k_1 = |U_1|$, $k_2 = |U_2|$. Then we have

$$\begin{aligned}
&\Pr[\pi\{1, \ldots, k\} = U] \\
&\le \Pr[\nu_1\{1, \ldots, k_1\} = U_1 \text{ and } \nu_2\{1, \ldots, k_2\} = U_2] \\
&\le 2 \Pr_{\text{unif}}[\nu_1\{1, \ldots, k_1\} = U_1 \text{ and } \nu_2\{1, \ldots, k_2\} = U_2] \\
&= \frac{2}{\binom{m_1}{k_1}\binom{m_2}{k_2}},
\end{aligned} \tag{2.6}$$

where the second inequality follows from (2.5). Now some routine calculations involving Stirling's formula (see Lemma A.2 in Appendix A) allow us to relate $\binom{m_1}{k_1}\binom{m_2}{k_2}$ to $\binom{m_1+m_2}{k_1+k_2} = \binom{m}{k}$. Specifically, (2.6) becomes

$$\Pr[\pi\{1, \ldots, k\} = U] \le \frac{Cm^{1/2}}{\binom{m}{k}} \exp\left\{ \frac{l^2 + \frac{1}{2}|l|}{\alpha(1-\alpha)} \Big(\frac{1}{m_1} + \frac{1}{m_2} \Big) \right\}, \tag{2.7}$$

where $\alpha = \frac{k}{m}$, $l = \frac{m_1k_2 - m_2k_1}{m}$, and $C > 0$ is a universal constant. The quantity l measures the deviation of the numbers k_1, k_2 of positive and negative elements in U from the "expected" values $\alpha m_1, \alpha m_2$ respectively. But since π is balanced, ν is good, and the element sizes do not vary too much, $|l|$ cannot in fact be very large. To formalize this intuition, note first that

$$l = (k_2\mu_2 - k_1\mu_1)\frac{m_2}{\mu_1 m}, \tag{2.8}$$

since $\frac{m_2}{m_1} = \frac{\mu_1}{\mu_2}$. Now by the goodness condition (2.3) on ν_1, ν_2 we have

$$\begin{aligned}
&\Big|\sum_{i=1}^{k} w_{\pi(i)} - (k_1\mu_1 - k_2\mu_2)\Big| \\
&= \quad \Big|\Big(\sum_{i=1}^{k_1} w_{\nu_1(i)} + \sum_{i=1}^{k_2} w_{\nu_2(i)}\Big) - (k_1\mu_1 - k_2\mu_2)\Big| \\
&\le \quad 2(M-1)\sqrt{k^* \ln m},
\end{aligned}$$

where $k^* = \min\{k, m-k\}$. Since π is balanced we also know that $|\sum_{i=1}^{k} w_{\pi(i)}| \le M$, and therefore

$$|k_1\mu_1 - k_2\mu_2| \le 2(M-1)\sqrt{k^* \ln m} + M.$$

Together with (2.8) and our assumption that $M \le B$, this implies the following bound on $|l|$:

$$|l| \le \Big(2(B-1)\sqrt{k^* \ln m} + B\Big)\frac{m_2}{\mu_1 m}.$$

Plugging in this value for $|l|$, the exponent in (2.7) is bounded above, for sufficiently large m, by

$$\begin{aligned}
&5(B-1)^2 k^* \ln m \,\frac{m_2^2}{\mu_1^2 m^2}\,\frac{m^2}{k(m-k)}\,\frac{m}{m_1 m_2} \\
&= \quad 5(B-1)^2 \ln m \frac{k^* m}{k(m-k)}\frac{1}{\mu_1\mu_2} \\
&\le \quad 10(B-1)^2 \ln m,
\end{aligned} \tag{2.9}$$

since $k(m-k) \ge \frac{k^* m}{2}$ and $\mu_1, \mu_2 \ge 1$. Thus (2.7) becomes

$$\Pr[\pi\{1, \ldots, k\} = U] \le C' \tbinom{m}{k}^{-1} m^{10(B-1)^2 + 1/2}, \tag{2.10}$$

which verifies the uniformity condition (2.2) with $\lambda = C' m^{10(B-1)^2+1/2}$.

This concludes the proof of the theorem for the special case $W = \sum_i w_i = 0$. We can extend the argument to general values of W using a simple trick. We will assume $W > 0$;

the case $W < 0$ is entirely symmetrical. We begin by padding the sequence of weights with $d = \lceil W/M \rceil$ values $w_{m+1}, \ldots, w_{m+d}$ each of which (except possibly the last) is $-M$, so that $\sum_{i=1}^{m+d} w_i = 0$. Note that $d \leq m$. By the above argument for the $W = 0$ case, we can construct a balanced almost uniform permutation π' on this padded sequence (though see the remark immediately following this proof). Let π be the *induced* permutation on the weights $\{w_i\}_{i=1}^n$. We claim that π is also balanced and almost uniform.

To see that π is balanced, note that

$$\sum_{i=1}^{k} w_{\pi(i)} \geq \sum_{i=1}^{k'} w_{\pi'(i)} \geq -M; \quad \text{and}$$
$$\sum_{i=1}^{k} w_{\pi(i)} \leq \sum_{i=1}^{k'} w_{\pi'(i)} + W \leq M + W,$$

for some $k' \geq k$, using the balance property of π'.

To see that π is almost uniform, let us call the indices $\{1, \ldots, m\}$ *true* and the remainder *fake*. Let U be an arbitrary subset of true indices of cardinality k. We need to show that

$$\Pr[\pi\{1, \ldots, k\} = U] \leq \binom{m}{k}^{-1} \mathrm{poly}(m). \tag{2.11}$$

Since π is induced by π', this probability is bounded above by $\sum_S \Pr[\mathcal{E}_S]$, where for $S \subseteq \{1, \ldots, m+d\}$, $\mathcal{E}_S$ is the event that $\pi'\{1, \ldots, |S|\} = S$ and the sum is over all S of the form $U \cup U'$, where all elements of U' are fake. Now by the almost uniformity of π', this sum is at most

$$\mathrm{poly}(m+d) \sum_S \Pr_{\mathrm{unif}}[\mathcal{E}_S], \tag{2.12}$$

where $\Pr_{\mathrm{unif}}$ denotes probability under the uniform distribution on permutations in $\mathcal{S}_{m+d}$. But the sum in (2.12) is just the expectation, under the uniform distribution, of the random variable $X = \sum_S X_S$, where X_S is the indicator r.v. of $\mathcal{E}_S$. Thus X counts the number of events $\mathcal{E}_S$ that occur. We claim that

$$\mathrm{E}(X) = \binom{m}{k}^{-1} \left(1 + \tfrac{d}{m+1}\right). \tag{2.13}$$

This will complete the verification of condition (2.11); for replacing the sum in (2.12) by $\mathrm{E}(X)$ gives

$$\Pr[\pi\{1, \ldots, k\} = U] \leq \binom{m}{k}^{-1} \left(1 + \tfrac{d}{m+1}\right) \mathrm{poly}(m+d),$$

which is of the required form since $d \leq m$.

To see the claim in (2.13), let $\mathcal{E}$ be the event that $\pi\{1,\ldots,k\} = U$. Clearly $\Pr_{\text{unif}}[\mathcal{E}] = \binom{m}{k}^{-1}$, and $X = 0$ unless $\mathcal{E}$ occurs, so we have

$$\mathrm{E}(X) = \binom{m}{k}^{-1} \mathrm{E}(X|\mathcal{E}). \tag{2.14}$$

Conditioning now on $\mathcal{E}$, let r be the position in π' of the last element of U, so that $U \subseteq \pi'\{1,\ldots,r\}$ and $\pi'(r) \in U$. Also, let s be the position of the *next* true element, i.e., $\pi'(s)$ is true and $\pi'(t)$ is fake for $r < t < s$. (If no such element exists, let $s = m+d+1$.) Then $\mathcal{E}_S$ holds for precisely those sets $S = \pi'\{1,\ldots,t\}$, where $r \leq t < s$. The number of such sets is just the number of fake elements that fall between the true element at position r and the next true element (at position s), plus one. The expectation of this quantity under the uniform distribution is plainly $1 + \frac{d}{m+1}$. Plugging this into (2.14) we get the value claimed in (2.13), which concludes the proof that π is almost uniform. □

Remark: We should point out that the padded sequence we introduced in the second part of the above proof might contain one weight whose absolute value is less than one. Thus it is not, in a strict sense, a special case of the earlier $W = 0$ case, where we assumed that all the weights had absolute values in the range $[1, B]$. However, a more careful treatment of the analysis leading up to equation (2.10) shows that this equation still holds even when there is a single small weight (or even a constant number of small weights). Furthermore, we can make the constant C' that appears in (2.10) independent of B.

Now, following through the algebra in the second part of the proof, starting from equation (2.10), and noting that $d \leq m$, we see that the resulting permutation π is λ-uniform for $\lambda = 2C'(2m)^{10(B-1)^2+1/2} = C_B m^{10(B-1)^2+1/2}$, where the constant C_B increases with B. Moreover, it is also easy to verify that the permutation π actually satisfies a slightly stronger uniformity property, namely

$$\begin{aligned} &\Pr[\pi\{1,\ldots,i\} = U_1 \text{ and } \pi\{i+2,\ldots,m\} = U_2] \\ &\quad \leq\ C'_B m^{10(B-1)^2+1/2} \times \begin{pmatrix} m \\ i \ \ j \ \ 1 \end{pmatrix}^{-1}, \end{aligned} \tag{2.15}$$

where $j = m - i - 1$ and $C'_B = C_B(B+1)$. We will make use of these facts in section 2.5 when we discuss permutations of general weights. □

2.4 A good flow in the bounded ratios case

We now return to the random walk for the knapsack problem, and flesh out the sketch of a flow presented in section 2.1, making heavy use of the balanced almost uniform permutations from section 2.3. We continue to consider only the bounded ratio case, i.e., we assume that all weights a_i lie in the range $[1, B]$. To avoid trivialities, we also assume $B \leq b$.

Let X, Y be two arbitrary vertices of G_Ω, viewed as subsets of $\{1, \ldots, n\}$. We need to specify how to route one unit of flow from X to Y. First, write $X = X_0 \cup X_1$, where X_0, X_1 are disjoint, $a(X_1) \leq b - B$, and $|X_0| \leq B$; this can always be done since $a_i \in [1, B]$. Write $Y = Y_0 \cup Y_1$ similarly. All the flow leaving X will pass through X_1, and all the flow arriving at Y will pass through Y_1. Between X_1 and Y_1, we will route the flow using an almost uniform permutation. (Note that there is an obvious correspondence between unit flows from X_1 to Y_1 and probability distributions on paths between them.) Let $S = X_1 \oplus Y_1$ and $m = |S|$. Let $\{w_i\}_{i=1}^m$ be an arbitrary enumeration of the weights of the items in S, where weights in $S \cap Y_1$ appear with a positive sign and weights in $S \cap X_1$ with a negative sign. Thus $W = a(Y_1) - a(X_1)$.

We can now describe the flow from X to Y in three stages:

Stage 1: Send the entire unit flow along a single path from X to X_1 by removing the elements of X_0 in index order.

Stage 2: Distribute the unit flow along geodesic paths from X_1 to Y_1 according to a balanced almost uniform permutation π of the weights $\{w_i\}$ of the items in S.

Stage 3: Send the entire unit flow along a single path from Y_1 to Y by adding the elements of Y_0 in index order.

The role of stages 1 and 3 is simply to ensure that the endpoints of the random paths in stage 2 are at least a small distance below the bounding hyperplane, to accommodate the (small) fluctuations still present in balanced permutations.

Let us first observe that the above flow is valid. For this, we just need to check that all the flow-carrying paths remain within the set Ω. This is obvious for stages 1 and 3. For stage 2 it follows from the balance property of π: for if Z is the kth point along a flow-carrying path from X_1 to Y_1, then

$$a(Z) = a(X_1) + \sum_{i=1}^{k} w_{\pi(i)}$$

$$
\begin{aligned}
&\le a(X_1) + \max\{a(Y_1) - a(X_1), 0\} + B &(2.16)\\
&= \max\{a(Y_1), a(X_1)\} + B\\
&\le b,
\end{aligned}
$$

where in the last line we have used the fact that $a(X_1), a(Y_1)$ are both $\le b - B$. Hence $Z \in \Omega$.

Next we must bound the quantities $\mathcal{C}(f)$ and $\mathcal{L}(f)$ for this flow f, as defined in section 2.2. $\mathcal{L}(f)$, the length of a longest flow-carrying path, is plainly at most $n + 2B$. To estimate $\mathcal{C}(f)$, we must bound the flow along any edge of G_Ω. For convenience we will in fact bound the flow $f(Z)$ through any *vertex* Z; clearly this is also an upper bound on the flow along any edge.

So let Z be an arbitrary vertex of G_Ω. Define $\mathcal{P}(Z)$ to be the set of pairs (X, Y) such that some $X \to Y$ flow passes through Z. Note that $\mathcal{P}(Z) = \bigcup_{i=1}^{3} \mathcal{P}_i(Z)$, where $\mathcal{P}_i(Z)$ are the pairs whose paths pass through Z in stage i. We shall bound the contribution to $f(Z)$ from each $\mathcal{P}_i(Z)$ separately. For $i = 1, 3$ this is simple: since stage-1 paths have length at most B, the number of vertices X such that $(X, Y) \in \mathcal{P}_1(Z)$ is (crudely) at most $(B+1)n^B$, so the contribution to $f(Z)$ from such paths is no more than $(B+1)n^B|\Omega|$. The same bound holds symmetrically for $\mathcal{P}_3(Z)$. The main portion of the paths, $\mathcal{P}_2(Z)$, presents more of a challenge.

We shall actually work with $\widetilde{\mathcal{P}}_2(Z)$, the set of pairs (X_1, Y_1) such that Z lies on the stage-2 path with endpoints X_1, Y_1. By the observation in the previous paragraph, the flow contribution from $\mathcal{P}_2(Z)$ will be at most $(B+1)^2 n^{2B}$ times that from $\widetilde{\mathcal{P}}_2(Z)$. Recall that we are really interested in the ratio $\frac{f(Z)}{|\Omega|}$, rather than in $f(Z)$ itself. Accordingly, following earlier analyses of this general type (see, e.g., [16, 15]), we measure the set $\widetilde{\mathcal{P}}_2(Z)$ by associating with each of its elements (X_1, Y_1) an "encoding" Z', which belongs to Ω. This is defined as the complement of Z in the multiset $X_1 \cup Y_1$; more precisely,

$$Z' = (X_1 \oplus Y_1 - Z) \cup (X_1 \cap Y_1).$$

To see that $Z' \in \Omega$, we need to check that $a(Z') \le b$. But this follows because

$$
\begin{aligned}
a(Z') &= a(X_1) + a(Y_1) - a(Z)\\
&\le a(X_1) + a(Y_1) - (\min\{a(X_1), a(Y_1)\} - B)\\
&= \max\{a(X_1), a(Y_1)\} + B\\
&\le b,
\end{aligned}
$$

where in the second line we have used the balance property of π as in (2.16) to bound $a(Z)$, this time from below.

How many pairs (X_1, Y_1) could be mapped to a given Z'? First note that Z' uniquely determines both $S = X_1 \oplus Y_1$ and $I = X_1 \cap Y_1$ via the relations

$$S = Z' \oplus Z; \qquad I = Z' \cap Z.$$

Thus in particular such pairs share the same symmetric difference, S, of cardinality m, say. To determine X_1 and Y_1 uniquely, it suffices to specify the subset $U \subseteq S$ of elements that have already been affected (i.e., added or deleted) by the stage-2 path by the time it reaches Z; precisely, $U = Z \oplus X_1$. For then we know, from the fact that all stage-2 paths are geodesic, that X_1 agrees with Z on $S - U$ and with Z' on U, and vice versa for Y_1. More formally,

$$X_1 = Z \oplus U; \qquad Y_1 = Z' \oplus U.$$

The upshot of the foregoing discussion is that we can define a mapping from $\widetilde{\mathcal{P}}_2(Z)$ to pairs of the form (Z', U), where $Z' \in \Omega$ and U is a subset of $Z \oplus Z'$. Moreover, and crucially, this mapping is *injective*. It therefore effectively enumerates the set $\widetilde{\mathcal{P}}_2(Z)$.

Finally, we need to take account of the actual quantity of flow traveling along the paths. Consider a pair $(X_1, Y_1) \in \widetilde{\mathcal{P}}_2(Z)$, corresponding to the pair (Z', U). Recall that the flow distribution between X_1 and Y_1 is determined by a balanced almost uniform permutation π of the weights in $S = X_1 \oplus Y_1$. The proportion of this flow that passes through Z is precisely

$$\Pr[\pi\{1, \ldots, |U|\} = U] \leq \binom{m}{|U|}^{-1} \operatorname{poly}(m),$$

by the almost uniform property of π.

Putting all this together, we can bound the total contribution to $f(Z)$ from $\widetilde{\mathcal{P}}_2(Z)$ as follows:

$$\begin{aligned}
&\sum_{Z' \in \Omega} \sum_{U \subseteq Z \oplus Z'} \Pr[\pi\{1, \ldots, |U|\} = U] \\
&\quad \leq \sum_{Z' \in \Omega} \sum_{k} \sum_{\substack{U \subseteq Z \oplus Z' \\ |U| = k}} \binom{m}{k}^{-1} \operatorname{poly}(m) \\
&\quad \leq \operatorname{poly}(n) \sum_{Z' \in \Omega} \sum_{k} \binom{m}{k} \binom{m}{k}^{-1} \\
&\quad \leq n \operatorname{poly}(n) |\Omega|,
\end{aligned}$$

where in the summations $m = |Z \oplus Z'|$. The total contribution from all stage-2 paths is thus at most $(B+1)^2 n^{2B+1}\,\mathrm{poly}(n)|\Omega|$.

Combining this with our earlier bounds for stages 1 and 3, we obtain that $f(Z) \leq \mathrm{poly}(n)|\Omega|$ (for a different polynomial), and hence $\mathcal{C}(f) \leq \mathrm{poly}(n)$. Since both $\mathcal{L}(f)$ and $\mathcal{C}(f)$ are polynomially bounded in n, we now obtain immediately from Theorem 2.1 that the mixing time, τ_{mix}, is polynomial in n. By keeping track of the polynomial factors, we see that the exponent is dominated by the $\mathrm{poly}(n)$ term arising from the almost uniformity condition (2.2), which is of the form $n^{\mathrm{O}(B^2)}$ (see the Remark at the end of section 2.3).

We summarize our analysis in the following theorem.

Theorem 2.5 *Let Ω be the set of solutions to a knapsack problem whose weights a_i lie in the range $[1, B]$ for some constant B. The mixing time of the random walk on G_Ω is $\tau_{\mathrm{mix}} = n^{\mathrm{O}(B^2)}$.*

As mentioned in section 2.1, this immediately yields an fpras for computing $|\Omega|$ in this case, via a standard reduction to random sampling. The details are spelled out in [14].

2.5 The general case

We now generalize the results of the previous two sections to the case of arbitrary weights. The essential ideas are the same, but there are several non-trivial technical complications that need to be addressed.

2.5.1 Balanced almost uniform permutations

To handle arbitrary weights, we first need to extend our construction of balanced almost uniform permutations. The chief obstacle here is that it is no longer true (as in the bounded ratio case) that each item of positive weight can be balanced by a bounded number of items of negative weight. To overcome this difficulty, we will need to group items into "intervals" so that each interval has approximately the same (positive or negative) weight. We can then reduce to the bounded ratio case.

First we need a slightly more liberal balance condition:

Definition 2.6 *Let $\{w_i\}_{i=1}^m$ be a set of real weights, let $M = \max_{i \leq m} |w_i|$, let $W = \sum_i w_i$ and let ℓ be a nonnegative integer. A permutation $\pi \in \mathcal{S}_m$ is* ℓ-balanced *if, for all k with*

$1 \le k \le m$,

$$\min\{W, 0\} - \ell M \le \sum_{i=1}^{k} w_{\pi(i)} \le \max\{W, 0\} + \ell M. \tag{2.17}$$

Our earlier definition thus corresponds to $\ell = 1$. The following is a generalization of Theorem 2.4.

Theorem 2.7 *For any m and set of weights $\{w_i\}_{i=1}^m$, there exists a 10-balanced Cm^2-uniform permutation, where C is a universal constant.*

Proof: Let $M = \max_{i \le m} |w_i|$ and $W = sum_{i=1}^m w_i$. Let β be a uniform random element of $\mathcal{S}_m$. Let T_1 be the smallest t such that the partial sum $\sum_{i=1}^t w_{\beta(i)}$ has absolute value greater than $3M$ (or $T_1 = m$ if no such t exists). Similarly, let T_2 be the smallest $t > T_1$ such that $|\sum_{i=T_1+1}^t w_{\beta(i)}| > 3M$. Define $T_3, T_4, \ldots$ in the same way. Then let I_1 be the sequence $\{\beta(i)\}_{i=1}^{T_1}$, and I_2 the sequence $\{\beta(T_1 + i)\}_{i=1}^{T_2 - T_1}$. Continue this way, dividing β into *intervals* $I_1, \ldots, I_q$.

Now let α_i be the aggregated weight of interval I_i for $i = 1, 2, \ldots, q$. Note that for all $i < q$ we have $3M \le |\alpha_i| \le 4M$, so the ratio of the weights of any two of these intervals is at most $\frac{4}{3}$. Thus, by the results of section 2.3, there exists a 1-balanced λ-uniform permutation on $\{\alpha_i\}_{i=1}^{q-1}$ for $\lambda = Cq^{10(\frac{4}{3}-1)^2+1/2} \le Cq^2$, where C is a universal constant. By the Remark at the end of that section, we can in fact assume that this permutation has the slightly stronger uniformity property given in (2.15). Call this permutation π_I. We claim that the permutation

$$\pi = I_{\pi_I(1)} I_{\pi_I(2)} \cdots I_{\pi_I(q-1)} I_q$$

obtained by rearranging the first $q - 1$ intervals according to π_I is a 10-balanced Cm^2-uniform permutation on the original m weights.

Let $W' = \sum_{i=1}^{q-1} \alpha_i = W - \alpha_q$. Since π_I is 1-balanced, π satisfies

$$\min\{0, W'\} - 4M \le \sum_{i=1}^{T_j} w_{\pi(i)} \le \max\{0, W'\} + 4M$$

for all $1 \le j \le q$. Hence we have, for all j,

$$\min\{0, W'\} - 7M \le \sum_{i=1}^{j} w_{\pi(i)} \le \max\{0, W'\} + 7M,$$

since the intervals with positive weight never have partial sums less than $-3M$ and the intervals with negative weight never have partial sums greater than $3M$. Finally, note that

$|W - W'| = |\alpha_q| \leq 3M$. It follows that for all j,

$$\min\{0, W\} - 10M \leq \sum_{i=1}^{j} w_{\pi(i)} \leq \max\{0, W\} + 10M,$$

and hence π is 10-balanced.

To verify the uniformity property, consider first an alternative experiment in which the permutation π_I is chosen u.a.r. from $\mathcal{S}_{q-1}$ without regard to the balance property. Note that, conditional on the value of q, the distribution of $(I_1, \ldots, I_{q-1})$ is exchangeable. Thus, re-arranging the intervals according to a uniform π_I is a *measure-preserving* transformation, so π itself has the uniform distribution. Thus we need to show that for any U, the likelihood ratio

$$\frac{\Pr[\pi\{1, \ldots, k\} = U]}{\Pr_{\text{unif}}[\pi\{1, \ldots, k\} = U]} \leq Cm^2,$$

where we write $\Pr_{\text{unif}}$ for the probability when π_I is uniform and Pr for the probability when π_I is Cm^2-uniform. In fact, it suffices to show that the above bound on the likelihood ratio holds conditional on any β. So fix a permutation β. In order for the numerator to be non-zero, there can be at most one interval I_i containing elements from both U and U^c (the complement of U). Additionally, in this interval, all the U elements must come before all the U^c elements. Let A_1 be the collection of intervals in $\{I_i\}_{i=1}^{q-1}$ containing only U elements and A_2 the collection of intervals containing only U^c elements. Then $|A_1| + |A_2|$ is in $\{q-1, q-2\}$. Writing $\mathcal{E}_1$ for the event $\pi_I\{1, \ldots, |A_1|\} = A_1$ and $\mathcal{E}_2$ for the event $\pi_I\{q-1, \ldots, q-|A_2|\} = A_2$, the likelihood ratio is

$$\frac{\Pr[\mathcal{E}_1 \text{ and } \mathcal{E}_2]}{\Pr_{\text{unif}}[\mathcal{E}_1 \text{ and } \mathcal{E}_2]} \leq Cq^2 \leq Cm^2.$$

In the case where $|A_1|+|A_2| = q-1$ this is just the Cq^2-uniform property. When $|A_1|+|A_2| = q-2$ it is the slightly stronger property given in (2.15). Thus π is Cm^2-uniform, and the proof is complete. $\square$

Remark: We have just shown that for arbitrary weights there is a Cm^2-uniform permutation which is $10-$balanced. This is all we will need in our present application. However, a straightforward modification of the above proof (appealing to the properties of the constant C_B in the Remark at the end of section 2.3) yields:

Theorem 2.8 *For any $\epsilon > 0$, m and set of weights $\{w_i\}_{i=1}^m$, there exists a $Cm^{1/2+\epsilon}$-uniform permutation which is $O(\frac{1}{\sqrt{\epsilon}})$-balanced, where C is a universal constant.*

Thus, we can bound the uniformity parameter λ by a polynomial whose degree is arbitrarily close to $1/2$ (at the cost of a modest increase in the balance parameter ℓ). This is almost the best that one can hope for: it is easy to check that, if we have $m/2$ weights of $+1$ and $m/2$ of -1, then for any constants ℓ, C and $p < 1/2$, there is no ℓ-balanced Cm^p-uniform permutation if m is sufficiently large. $\square$

With a bit more work, we can further extend Theorem 2.7 to show that, when $|W|$ is sufficiently large, the weights $\{w_i\}_{i=1}^m$ actually admit a 0-*balanced* almost uniform permutation.

Theorem 2.9 *Let $\{w_i\}_{i=1}^m$ be a sequence of weights. Let $M = \max_i |w_i|$, let $W = \sum_i w_i$ and suppose $|W| > 21M$. Then there is a 0-balanced Cm^2-uniform permutation on $\{w_i\}$ for a universal constant C.*

Proof: W.l.o.g. assume that $W > 21M$. Let β be a uniform random permutation on $\{1, \ldots, m\}$. For $1 \le t \le m$, let $S_t = \sum_{i=1}^t w_{\beta(i)}$, and set

$$\begin{aligned} T_1 &= \min\{t : |S_t| > 10M\}; \\ T_2 &= \max\{t : |W - S_t| > 10M\}. \end{aligned}$$

Define the permutation π_1 by

$$\pi_1(i) = \begin{cases} \beta(i), & \text{if } i \le T_1 \text{ or } i > T_2; \\ \beta(T_1 + \alpha(i - T_1)), & \text{if } T_1 < i \le T_2, \end{cases}$$

where α is a 10-balanced Cm^2-uniform permutation on the weights $\{w_{\beta(T_1+j)}\}_{j=1}^{T_2-T_1}$. We claim that π_1 is a Cm^2-uniform permutation.

We can prove this in a way that is reminiscent of the proof of Theorem 2.7. Suppose that instead of choosing α as just described, we chose α to be a *uniform* random permuation of $\{1, \ldots, T_2 - T_1\}$. Then π_1 would in fact have the uniform distribution. Writing $\Pr_{\text{unif}}$ for probabilities under this scenario, we need to show that for any k and U,

$$\frac{\Pr[\pi\{1, \ldots, k\} = U]}{\Pr_{\text{unif}}[\pi\{1, \ldots, k\} = U]} \le Cm^2.$$

As before, this is easily seen by conditioning on β.

Next, define π_2 by

$$\pi_2(i) = \begin{cases} \pi_1(\alpha_1(i)), & \text{if } i \le T_1; \\ \pi_1(i), & \text{if } T_1 < i \le T_2; \\ \pi_1(T_2 + \alpha_2(i - T_2)), & \text{if } i > T_2, \end{cases}$$

where α_1 and α_2 are "rotations" (i.e., permutations of the form $x \to (x+r) \bmod s$), chosen so that the partial sums of π_2 stay above $\min\{0, S_{T_1}\}$ on the interval before T_1 and below $\max\{W, S_{T_2}\}$ on the interval after T_2. (Such rotations always exist; for example, for α_1 we could let $r = \mathrm{argmin}_{1 \le i \le T_1} S_i$.) We claim that π_2 is Cm^2-uniform. Let $\mathcal{E}$ be the event that $\pi_2\{1, \ldots, k\} = U$. Let $\mathcal{E}_1$ be the event that $\beta\{1, \ldots, k_1\} \cup \beta\{j, \ldots, j+k_2-1\} = U$, for some j and k_1, k_2 such that $k_1 + k_2 = k$. Let $\mathcal{E}_2$ be the event that $\pi_1\{1, \ldots, k\} = U$. Note that $\mathcal{E} \subseteq \mathcal{E}_1 \cup \mathcal{E}_2$. Therefore, we have $\Pr[\mathcal{E}] \le \Pr[\mathcal{E}_1] + \Pr[\mathcal{E}_2] \le (m^2 + Cm^2)\binom{m}{k}^{-1}$. Thus π_2 is Cm^2-uniform (for a slightly larger constant C).

Finally, note that π_2 is 0-balanced exactly when $S_{T_1} > 0$ and $S_{T_2} < W$. This is because the permutation is 10-balanced between T_1 and T_2, and the rotations have ensured that the partial sums remain between 0 and W on the intervals between 0 and T_1 and between T_2 and m. But $(S_{k \wedge T_1})_{k \ge 0}$ is a submartingale, and T_1 is a stopping time. Hence $E(S_{T_1}) \ge 0$. But $|S_{T_1}| \in [10M, 11M]$ with probability 1. An easy calculation shows that this implies that $\Pr[S_{T_1} > 0] \ge 10/21$. A similar argument shows that $\Pr[S_{T_2} < W] \ge 10/21$, even after conditioning on the value of S_{T_1}. Thus π_2 is 0-balanced with probability at least $(10/21)^2 \ge 1/5$, so if we modify our sample space so that we only accept π_2 if it is 0-balanced (and incorporate an extra factor of 5 into the constant C), then the result is still a Cm^2-uniform permutation. □

2.5.2 The flow

Now that we have balanced almost uniform permutations for general weights, we can follow a similar strategy to that in section 2.4 for constructing a good flow in the general case.

Lemma 2.10 *For arbitrary weights, it is possible to construct a multicommodity flow f in G_Ω with $\mathcal{C}(f) = O(n^5)$ and $\mathcal{L}(f) = O(n)$.*

Proof: Let X, Y be arbitrary vertices of G_Ω. Recall the scheme for constructing a flow from X to Y in the bounded ratio case in section 2.4: we essentially followed a balanced almost uniform permutation of $X \oplus Y$, except that we removed a constant number of elements from X and Y from consideration (processing them at the beginning and end of the path) to ensure that the path remained within G_Ω. The idea in the general case is the same, except that we will now remove a fixed number of elements from $X \cup Y$ and add/delete these repeatedly along the path to maintain fine balance. Note that the resulting flow-carrying paths will not in general be geodesics, though they will still have length $O(n)$.

Viewing X and Y as subsets of $\{1,\ldots,n\}$, let H be the 41 elements of $X \oplus Y$ having largest weight (or let $H = X \oplus Y$ if $|X \oplus Y| \leq 41$), with ties broken according to index order. Let $H_X = H \cap X$ and $H_Y = H \cap Y$. Define $S = X \oplus Y - H$, $m = |S|$, and let $\{w_i\}_{i=1}^m$ be an arbitrary enumeration of the weights of the items in S, with items in X, Y appearing with negative and positive signs respectively. We claim that there exists a Cm^2-uniform permutation π satisfying the following balance condition:

$$\begin{aligned} \min\{a(Y) - a(X), 0\} - a(H_Y) \leq \sum_{i=1}^{k} w_{\pi(i)} \\ \leq \max\{a(Y) - a(X), 0\} + a(H_X), \end{aligned} \tag{2.18}$$

for $1 \leq k \leq m$. As we shall see shortly, it is straightforward to construct such a permutation using the balanced almost uniform permutations from section 2.5.1. Our flow from X to Y will follow the random permutation π, except that along the way elements of H will be used to keep the knapsack as full as possible, but will be removed as necessary to make room for items in Y to be added. Thus each intermediate state will be of the form $H_0 \cup [X \oplus \{\pi(1),\ldots,\pi(k)\}]$, for some $H_0 \subseteq H$ and $k \leq m$. The transition rule will be as follows.

1. If $k < m$ and $w_{\pi(k+1)} > 0$ then add $\pi(k+1)$ if possible (i.e., if the result is an element of Ω); otherwise delete an element from H_0.

2. If $k < m$ and $w_{\pi(k+1)} < 0$ then add an element from $H - H_0$ if possible; otherwise delete $\pi(k+1)$.

3. If $k = m$ then add an element from H_Y if possible; otherwise delete an element from H_X.

The upper bound in (2.18) ensures that sufficient elements of H can always be removed so as to make room for adding the next element $\pi(k+1)$; hence the above rule defines a feasible random path from X to Y. Conversely, the lower bound in (2.18) ensures that, if all elements in H were added, the weight along the path would be at least $\min\{a(X), a(Y)\}$; since our rule keeps the weight as large as possible, this implies that, at any intermediate point Z along the path, there exist two elements $h_1, h_2 \in H$ such that $a(Z \cup \{h_1, h_2\}) \geq \min\{a(X), a(Y)\}$.

Now by analogy with the analysis in section 2.4, we define the "encoding" Z' by

$$Z' = (X \oplus Y - Z) \cup (X \cap Y) - \{h_1, h_2\}.$$

In similar fashion to the analysis in section 2.4 one can see that for a given Z, X and Y are completely specified by the 5-tuple (Z', U, h_1, h_2, H'), where $H' = H_X$ and $U = Z \oplus X$. We also have

$$\begin{aligned} a(Z') &= a(X) + a(Y) - a(Z \cup \{h_1, h_2\}) \\ &\leq a(X) + a(Y) - \min\{a(X), a(Y)\} \\ &= \max\{a(X), a(Y)\}, \end{aligned}$$

so $Z' \in \Omega$.

This encoding allows us to bound the flow $f(Z)$ through Z as in section 2.4. We have

$$\begin{aligned} f(Z) &\leq \sum_{Z' \in \Omega} \sum_{k, U, h_1, h_2, H'} Cm^2 \tbinom{m}{k}^{-1} \\ &\leq \sum_{Z' \in \Omega} 2^{41} n^3 \tbinom{m}{k} Cm^2 \tbinom{m}{k}^{-1} \\ &\leq C' n^5 |\Omega|, \end{aligned} \tag{2.19}$$

where the factors of 2^{41} and n^3 arise from summing over H', h_1, h_2 and k. This ensures that the above flow satisfies $\mathcal{C}(f) = \mathrm{O}(n^5)$, while plainly $\mathcal{L}(f) = \mathrm{O}(n)$ as claimed in the lemma.

To finish the proof, we must go back and validate our earlier claim that it is possible to construct a Cm^2-uniform permutation satisfying condition (2.18). Writing as usual $W = \sum_{i=1}^{m} w_i$, and using the fact that $W = a(Y) - a(X) + a(H_X) - a(H_Y)$, we may equivalently express (2.18) as

$$\begin{aligned} \min\{W, a(H_X) - a(H_Y)\} - a(H_X) \leq \sum_{i=1}^{k} w_{\pi(i)} \\ \leq \max\{W, a(H_X) - a(H_Y)\} + a(H_Y). \end{aligned} \tag{2.20}$$

We may assume w.l.o.g. that $W \geq a(H_X) - a(H_Y)$; the other case is symmetrical. Thus (2.20) becomes

$$-a(H_Y) \leq \sum_{i=1}^{k} w_{\pi(i)} \leq W + a(H_Y). \tag{2.21}$$

Now a little algebra shows that an ℓ-balanced permutation π satisfies (2.21) provided $\ell M \le \min\{a(H_Y), W + a(H_Y)\}$, where as usual $M = \max_i |w_i|$. Thus in particular a 0-balanced permutation always satisfies (2.21), so when $|W| > 21M$ we may appeal to Theorem 2.9. And when $|W| \le 21M$, the fact that $a(H_X) + a(H_Y) = a(H) \ge 41M$ implies that $a(H_Y) \ge \max\{10M, 10M - W\}$, so in this case a 10-balanced permutation satisfies (2.21) and we may appeal to Theorem 2.7. This completes the proof. $\square$

Given such a flow, we need only invoke Theorem 2.1 to derive our main result.

Theorem 2.11 *Let Ω be the set of solutions to an arbitrary instance of the 0-1 knapsack problem. The mixing time of the random walk on G_Ω is $\tau_{\text{mix}} = \mathrm{O}(n^8)$.*

This immediately implies the existence of an fpras for computing $|\Omega|$ in the general case.

Chapter 3

Multiple Hyperplanes

3.1 Introduction

In this Chapter, we will extend the results of Chapter 2 to handle multiple linear constraints. For a positive real $d \times n$ matrix A, and a positive real vector $\mathbf{b} = (b^1, \ldots, b^d)$, let Ω denote the set of 0-1 vectors $\mathbf{x} = (x_i)_{i=1}^n$ for which

$$A\mathbf{x} \leq \mathbf{b}.$$

The vertices in Ω constitute the set of feasible solutions to the multidimensional knapsack problem with the d constraints

$$\sum_{i=1}^{n} a_i^j x_i \leq b^j \text{ for } 1 \leq j \leq d, \tag{3.1}$$

where $a_i^j \equiv a_{ji}$. Following the notation of Chapter 2, we will write X for the set of indices $\{i : x_i = 1\}$, and $a(X) = (a^1(X), \ldots, a^d(X))$ for the (now d-dimensional) weight of X.

Geometrically, Ω is obtained by truncating the unit hypercube by d hyperplanes, each of which corresponds to a knapsack constraint. The essential geometric property of these "knapsack" hyperplanes is that their normal vectors all lie in the same quadrant. The results of this chapter will easily extend to any collection of hyperplanes with this property. However, we cannot allow the hyperplanes to be arbitrary. If arbitrary truncations were allowed, then just two hyperplanes could cause exponential bottlenecks or even disconnect the graph.

If coordinate random walk on Ω is rapidly mixing, then we immediately have a polynomial time algorithm for sampling elements from Ω (almost) uniformly at random. We will prove that when d is fixed, this is indeed the case.

Unfortunately, the degree of our polynomial upper bound for the mixing time will depend on d, but we can't avoid this: Consider the case where we have $\frac{n}{2d}$ of each of the weights $(n, 0, \ldots, 0), (0, n, \ldots, 0), \ldots$, and $(0, \ldots, 0, n)$, with the remaining $\frac{n}{2}$ weights equal to $(1, 1, \ldots, 1)$, and let $\mathbf{b} = (n, \ldots, n)$. Let S be the set of feasible solutions in Ω which don't contain any of the $(1, \ldots, 1)$ items. Then $|S| = \left(\frac{n}{2d} + 1\right)^d$, but S is connected to $\Omega - S$ only through the origin. It follows that the mixing time is at least $n^{\Omega(d)}$.

In fact, for d arbitrary there is no uniform polynomial upper bound for *any* sampling algorithm unless NP=RP. We can prove this using a reduction from the problem of sampling independent sets in a graph. By theorem (1.17) in [29], there is no algorithm for (almost) uniformly sampling independent sets in a graph unless NP=RP. So let $G = (V, E)$ be an arbitrary instance of the independent sets sampling problem. Then there is a 1-1 correspondence between the independent sets in G and the feasible solutions to the knapsack problem with $|V|$ variables and the E constraints $x_i + x_j \leq 1$ for all $(i, j) \in E$.

To prove rapid mixing of the random walk on Ω for fixed d we will use the multicommodity flow technique. As in Chapter 2, for every $X, Y \in \Omega$ we will spread the $X \to Y$ flow evenly using a balanced almost uniform permutation. However, since the weight function $a(\cdot)$ is now vector-valued, we first need to extend the definition of balance to higher dimensions. Let $\{w_i\}_{i=1}^m$ be a sequence of weights in $\mathbf{R}^d$, with $w_i = (w_i^1, \ldots, w_i^d)$, and suppose that $\sum_i w_i = 0$. For $1 \leq j \leq d$, let $M^j = \max_i |w_i^j|$. We say a permutation $\pi \in \mathcal{S}_m$ is *balanced* if

$$\max_k \left|\sum_{i=1}^{k} w_{\pi(i)}^j\right| \leq cM^j \text{ for all } 1 \leq j \leq d, \tag{3.2}$$

where c is a universal constant. Thus π is balanced with respect to vector weights $\{w_i\}$ if and only if it satisfies the d one-dimensional balance conditions given by (3.2). Constructing balanced almost uniform permutations is significantly more difficult in higher dimensions since one has to control fluctuations in all dimensions simultaneously. In fact, for $d \geq 2$, it is non-trivial to prove that even a *single* balanced permutation exists. Nonetheless, we will show that there is always a balanced permutation that is also almost uniform. To indicate the main ideas, we will now give a sketch of the proof in the special case where $d = 2$ and the weights satisfy $1 \leq |w_i^j| \leq 2$ for all i and j.

Proof Sketch: Suppose that $d = 2$ and the weights satisfy $|w_i^j| \in [1,2]$ for all i and j. Let $I_1 = \{i : w_i^2 \geq 0\}$, $I_2 = \{i : w_i^2 < 0\}$, and define $\nu = \sum_{i \in I_1} w_i$. For every $i \leq m$, let Y_i be the projection of w_i onto $\nu^\perp$. Let π_1 be an almost uniform permutation on I_1 which is balanced (in the one-dimensional sense) with respect to $\{Y_i\}_{i \in I_1}$, with a similar definition for π_2. Finally, interleave π_1 and π_2 to give a permutation on $\{1, \ldots, m\}$ which is balanced with respect to $\{w_i^2\}_{i=1}^m$. Since π_1 and π_2 are both almost uniform, so is π, by an argument similar to that in the proof of Theorem 2.4 in Chapter 2.

Furthermore, since π_1 and π_2 are each balanced with respect to projections onto $\nu^\perp$, so is π (note that the projections Y_i satisfy $\sum_{i \in I_1} Y_i = \sum_{i \in I_2} Y_i = 0$). Thus, for every k, the projections of $\sum_{i=1}^k w_{\pi(i)}$ onto both

1. the y-axis; and
2. $\nu^\perp$

are bounded, and since the w_i^j are all in $[1,2]$, the angle between the y-axis and $\nu^\perp$ can be bounded away from 0. Thus, the partial sums $\sum_{i=1}^k w_{\pi(i)}$ stay inside a parallelogram of bounded diameter. Hence π is balanced for the weights $\{w_i\}_{i=1}^m$. □

3.2 Balanced almost uniform permutations for high dimensions

We will now give a precise definition of balance for higher dimensions, and then prove the existence of balanced almost uniform permutations.

Definition 3.1 *Let d and l be positive integers, and let $\{w_i\}_{i=1}^m$ be a set of weights in $\mathbf{R}^d$ satisfying $\sum_{i=1}^m w_i = 0$. A permutation π is l–*balanced *if*

$$\left| \sum_{i=1}^{k} w_{\pi(i)}^j \right| \leq l M^j,$$

for all j and k, where $M^j = \max_{1 \leq i \leq m} |w_i^j|$.

Theorem 3.2 *For every positive integer d, there exists a constant c_d and a polynomial function p_d, such that for any set of weights $\{w_i\}_{i=1}^m$ in $\mathbf{R}^d$, there exists a c_d–balanced, $p_d(m)$–uniform permutation.*

Proof: The proof will be by induction on d. The base case $d = 1$ follows from Theorem 2.7 of Chapter 2, with $c_1 = 10$ and $p_1(m) = Cm^2$. Now, let d be arbitrary, and suppose that the result holds for $d-1$. Let $\{w_i\}_{i=1}^m$ be a set of weights in $\mathbf{R}^d$. Suppose first that the weights satisfy

$$\begin{aligned} M^j &= 2 \qquad \text{for all } j; &(3.3)\\ 1 \le \max_{1\le j\le d} |w_i^j| &\le 2 \qquad \text{for all } i. &(3.4)\end{aligned}$$

Thus each weight is at least half as large as the maximum (positive or negative) weight in some coordinate. Then

$$\max_{1\le j\le d} \sum_{i=1}^m |w_i^j| \ge \frac{m}{d}.$$

W.l.o.g., suppose that the sum in the LHS is maximized by $j = d$. Then we have

$$\sum_{i=1}^m (w_i^d)^+ = \sum_{i=1}^m (w_i^d)^- \ge \frac{m}{2d}.$$

Let $I_1 = \{i : w_i^d \ge 0\}$, $I_2 = \{i : w_i^d < 0\}$, $m_1 = |I_1|$, and $m_2 = |I_2|$. Define the means $\mu_1 = \frac{1}{m_1}\sum_{i\in I_1} w_i^d$, and $\mu_2 = -\frac{1}{m_2}\sum_{i\in I_2} w_i^d$. Note that $\mu_1, \mu_2 \ge \frac{1}{2d}$. For $1 \le j < d$, let

$$\gamma^j = \frac{\sum_{i\in I_1} w_i^j}{\sum_{i\in I_1} w_i^d} = \frac{\sum_{i\in I_2} w_i^j}{\sum_{i\in I_2} w_i^d}.$$

For all $i \le m$ and $j < d$, let $Y_i^j = w_i^j - \gamma^j w_i^d$, and let $Y_i = (Y_i^1, \ldots, Y_i^{d-1})$. Note that $|\gamma^j| \le 1$, $|Y_i^j| \le 4$, and $\sum_{i\in I_1} Y_i = \sum_{i\in I_2} Y_i = 0$.

Now, for $s = 1, 2$ let π_s be a $p_{d-1}(m)$–uniform permutation on I_s which is c_{d-1}–balanced with respect to $\{Y_i\}_{i\in I_s}$. Call π_1 α-good if for every k_1 with $1 \le k_1 \le m_1$ we have

$$\Big|\sum_{i=1}^{k_1} w_{\pi_1(i)}^d - k_1\mu_1\Big| \le 2\alpha\sqrt{k_1^*} \tag{3.5}$$

where $k_1^* = \min(k_1, m_1 - k_1)$. Now, for a particular value of k_1, Hoeffding's bounds [13] imply that we have

$$\Pr_{\text{unif}}(\pi_1 \text{ does not satisfy (3.5)}) \le 2\exp(-2\alpha^2),$$

and since the event depends only on the initial segment $\pi_1\{1, \ldots, k_1\}$, we also have

$$\begin{aligned}\Pr(\pi_1 \text{ does not satisfy (3.5)}) &\le p_{d-1}(m)\cdot \Pr_{\text{unif}}(\pi_1 \text{ does not satisfy (3.5)})\\ &\le p_{d-1}(m)\cdot 2\exp(-2\alpha^2). &(3.6)\end{aligned}$$

Suppose that for some constants C and r, the polynomial p_{d-1} satisfies $p_d(n) \leq Cn^r$ for all n. If we let $\alpha = \sqrt{(r+1)\ln(m)}$, then the RHS of (3.6) is at most $2Cm^{r-2(r+1)} \leq \frac{1}{2}$, for sufficiently large m. Thus, we can assume that π_1 is α-good with probability 1 and only increase C by a constant factor. Similar arguments apply to π_2.

Finally, note that it is always possible to interleave π_1 and π_2 to give a permutation on $\{1,\ldots,m\}$ which is 1-balanced with respect to $\{w_i^d\}_{i=1}^m$. Let π be such a permutation. Then we have $|\sum_{i=1}^k w_{\pi(i)}^d| \leq 2$, and

$$\begin{aligned}
\Big|\sum_{i=1}^k w_{\pi(i)}^j\Big| &= \Big|\sum_{\substack{i\in I_1:\\ i\leq k}} w_{\pi(i)}^j + \sum_{\substack{i\in I_2:\\ i\leq k}} w_{\pi(i)}^j\Big| \\
&= \Big|\sum_{\substack{i\in I_1:\\ i\leq k}} Y_{\pi(i)}^j + \sum_{\substack{i\in I_2:\\ i\leq k}} Y_{\pi(i)}^j + \gamma^j \sum_{i=1}^k w_{\pi(i)}^d\Big| \\
&\leq \Big|\sum_{\substack{i\in I_1:\\ i\leq k}} Y_{\pi(i)}^j\Big| + \Big|\sum_{\substack{i\in I_2:\\ i\leq k}} Y_{\pi(i)}^j\Big| + |\gamma^j|\cdot\Big|\sum_{i=1}^k w_{\pi(i)}^d\Big| \\
&\leq 4c_{d-1} + 4c_{d-1} + 2|\gamma^j| \\
&\leq 8c_{d-1} + 2,
\end{aligned}$$

for all $j < d$ and k. Hence π is c_d'–balanced for $c_d' = 4c_{d-1} + 1$ by assumption (3.3) .

To verify almost uniformity, we follow the proof of Theorem 2.4 in Chapter 2. Let $U \subseteq \{1,\ldots,m\}$ be arbitrary with $|U| = k$, and let $U_1 = U \cap I_1$, $U_2 = U \cap I_2$, $k_1 = |U_1|$, and $k_2 = |U_2|$. Then we have

$$\begin{aligned}
\Pr[\pi\{1,\ldots,k\} = U] &\leq \Pr[\pi_1\{1,\ldots,k_1\} = U_1 \text{ and } \pi_2\{1,\ldots,k_2\} = U_2] \\
&\leq (Cm^r)^2 \Pr_{\mathrm{unif}}[\pi_1\{1,\ldots,k_1\} = U_1 \text{ and } \pi_2\{1,\ldots,k_2\} = U_2] \\
&= \frac{C'm^{2r}}{\binom{m_1}{k_1}\binom{m_2}{k_2}}
\end{aligned}$$

Now we can bound the quantity $\binom{m_1}{k_1}\binom{m_2}{k_2}$ by mimicking the calculations from equation (2.7) to equation (2.10) in Theorem 2.4. In our current setting, we have $\mu_1, \mu_2 \geq \frac{1}{2d}$, and the $|w_i^d|$ are in $[0,2]$. We also have to make the substitution $\ln(m) \to (r+1)\ln(m)$, so equation (9) becomes

$$|l| \leq \frac{10\cdot 2^2}{(\frac{1}{2d})^2}(r+1)\ln(m) = 160d^2(r+1)\ln(m). \tag{3.7}$$

Hence π is p_d-uniform for $p_d(m) = C'm^{160d^2(r+1)+2r+1/2}$.

We have shown how to make balanced almost uniform permutations if the weights satisfy (3.4) and (3.3). To generalize to arbitrary weights $\{w_i\}_{i=1}^m$, we use the interval trick introduced in section (2.5.1). Let β be a uniform random permutation in $\mathcal{S}_m$, and let $T_1 = \min\{t : |\sum_{i=1}^t w_i^j| > M^j \text{ for some } j\}$. Define $T_2, T_3, \ldots$ similarly. Next, we use the T_i to divide β into intervals $I_1, \ldots, I_q$. Let $\{\alpha_i\}_{i=1}^{q-1}$ be the aggregated weights of the first $q-1$ intervals. Note that if we divide each α_i^j by $\frac{1}{2}\max_i |\alpha_i^j|$, then the resulting weights satisfy (3.4) and (3.3). Hence these weights admit a c_d'–balanced, $p_d(q)$–uniform permution (though see the remark immediately following this theorem.) Rearranging the intervals $\{I_i\}_{i=1}^{q-1}$ according to such a permutation gives a permutation on $\{1, \ldots, m\}$ which is p_d–uniform[1] and c_d–balanced for $c_d = 2c_d' + 2$. $\square$

Remark: We should point out that the weights $\{\alpha_i\}_{i=1}^{q-1}$ of the first $q-1$ intervals will not in general sum to 0. However, we can easily get around this by introducing a dummy weight α_q which is equal to the weight of interval I_q. The presence of this single small weight does not affect equation (3.7) for sufficiently large m. Hence there is a c_d'-balanced p_d-uniform permutation on this padded sequence $\{\alpha_i\}_{i=1}^q$. This induces a permutation π_I on $\{\alpha_i\}_{i=1}^{q-1}$ which is $(c_d'+1)$–balanced and cp_d–uniform for some constant c. Thus, if we incorporate an extra $+1$ into the constant c_d' and an extra factor of c into p_d then the argument in the above proof is still valid. $\square$

Before we define our flow, we need one more definition.

Definition 3.3 *Let $\{w_i\}_{i=1}^m$ be a sequence in $\mathbf{R}^d$, with $w_i = (w_i^1, \ldots, w_i^d)$, let $\mu = (\mu^1, \ldots, \mu^d) = \frac{1}{m}\sum_{i=1}^m w_i$, and let l be a non-negative integer. A permutation π is* strongly l–balanced *if, for all $k \le m$ and $j \le d$, there exists a set $S \subseteq \{1, \ldots, m\}$ with $|S \oplus \pi\{1, \ldots, k\}| < l$, such that $\left(\sum_{i=1}^k w_{\pi(i)}^j - k\mu^j\right)$ and $\left(\sum_{i\in S} w_{\pi(i)}^j - k\mu^j\right)$ have opposite signs (or either is 0).*

Thus, in a strongly balanced permutation, whenever the initial segment $\{\pi(i)\}_{i=1}^k$ is "above average" with respect to a particular coordinate j, it can be made "below average" by flipping at most a constant number of items, and vice-versa. As the name suggests, the strong balance condition is more strict than the usual balance condition. Nonetheless, the following lemma says that strongly balanced permutations always exist.

Lemma 3.4 *For any sequence $\{w_i\}_{i=1}^m$ in $\mathbf{R}^d$, there exists a strongly $16d^2$–balanced permutation π.*

[1]Here we are also using the strong uniformity property discussed in Remark 2.3

Note that this lemma only claims that a *single* strongly balanced permutation exists; it makes no claims regarding almost uniformity. The proof of Lemma 3.4 relies heavily on the following lemma which is due to Grinberg and Sevast'yanov [12].

Lemma 3.5 *(Grinberg and Sevast'yanov, 1980) Let $x_1, \ldots, x_n$ be vectors in $\mathbf{R}^d$ such that $\sum_i x_i = 0$. Then there exists a permutation $\nu \in \mathcal{S}_n$ such that*

$$\sum_{i=1}^{k} x_{\nu(i)} \in d \cdot \mathrm{conv}\{x_1, \ldots, x_n\} \qquad \textit{for all } k.$$

The proof of Lemma 3.4 is straightforward, but it is fairly long, so we will defer it to Appendix B.

3.3 A good flow

Now that we have multi-dimensional balanced almost uniform permutations and strongly balanced permutations, we are ready to contruct a good flow. As in Chapter 2, let G_Ω be the subgraph of $\{0,1\}^n$ induced by Ω.

Lemma 3.6 *For arbitrary weights, it is possible to construct a multicommodity flow f in G_Ω with polynomial $\mathcal{C}(f)$ and $\mathcal{L}(f)$.*

Proof: Recall that we identify each vertex $\mathbf{x} \in \Omega$ with the index set $X = \{i : x_i = 1\}$. Let $\hat{\Omega} = \{X \in \Omega : a^j(X) \leq b^j - 3c_d \max_{i \in X} a_i^j\}$. Our first step will be to construct a flow $\hat{f}$ which, simultaneously for every $X, Y \in \hat{\Omega}$, sends a unit of flow from X to Y. This flow will satisfy $\mathcal{C}(\hat{f}) \leq \mathrm{poly}(n)$ and $\mathcal{L}(\hat{f}) = \mathrm{O}(n)$.

Note that, from any vertex $X \in \Omega$, we can obtain a vertex $X' \in \hat{\Omega}$ by removing at most $3dc_d$ objects. Thus, we can use an approach similar to that in section 4 to extend $\hat{f}$ to a multicommodity flow f on the whole of Ω, and f will satisfy $\mathcal{C}(f) \leq n^{6dc_d}\mathrm{poly}(n) \leq \mathrm{poly}'(n)$ and $\mathcal{L}(f) \leq \mathcal{L}(\hat{f}) + 6dc_d = \mathrm{O}(n)$.

It remains to define the flow $\hat{f}$, and show that it has the properties claimed. Fix $X, Y \in \hat{\Omega}$. As usual, the path from X to Y will follow a permutation π on the symmetric difference $X \oplus Y$. However, as in the one-dimensional case, a simple balanced almost uniform permutation π will not do; such a permutation would not necessarily define a path that stayed in

Ω. The problem occurs when for some j, $\max_{i\in X} a_i^j$ is not comparable to $\max_{i\in Y} a_i^j$. (For example, if $\max_{i\in X} a_i^j \gg \max_{i\in Y} a_i^j$, then the path would have too much variation in the j–direction as it approached Y.) However, we can deal with this problem by considering the "large" and the "small" objects in $X \oplus Y$ separately.

Let $M = (M^1, \ldots, M^d)$, where $M^j = \min(\max_{i\in X} a_i^j, \max_{i\in Y} a_i^j)$. Let $L = \{i \in X \oplus Y : a_i^j > M^j \text{ for some } j\}$ and let $S = (X \oplus Y) - L$. Let $\{w_i\}_{i\in X\oplus Y}$ be an enumeration of the weights of the items in $X \oplus Y$, where weights from Y appear with a positive sign and weights from X appear with a negative sign. Let $\mu_1 = \frac{1}{|L|}\sum_{i\in L} w_i$, and let $\mu_2 = \frac{1}{|S|}\sum_{i\in S} w_i$. Let π_1 be a permutation on L which is strongly $16d^2$–balanced with respect to the weights $\{w_i\}_{i\in L}$, and let π_2 be a p_d–uniform permutation which is c_d–balanced with respect to the weights $\{w_i\}_{i\in S}$. To obtain π, we will interleave the strongly balanced permutation π_1 and the balanced permutation π_2. The rule for interleaving will be as follows. Suppose that $\pi(1), \ldots, \pi(k)$ have already been assigned, and that $\pi\{1, \ldots, k\} = \pi_1\{1, \ldots, k_1\} \cup \pi_2\{1, \ldots, k_2\}$. Now, either $\frac{k_1}{k} \le \frac{|L|}{|L|+|S|}$ or $\frac{k_2}{k} < \frac{|S|}{|L|+|S|}$, so we can define $\pi(k+1)$ by

$$\pi(k+1) = \begin{cases} \pi_1(k_1+1), & \text{if } \frac{k_1}{k} \le \frac{|L|}{|L|+|S|}; \\ \pi_2(k_2+1), & \text{if } \frac{k_2}{k} < \frac{|S|}{|L|+|S|}. \end{cases}$$

Now, let $\mu = \frac{1}{|X\oplus Y|}\sum_{i\in X\oplus Y} w_i = \frac{|L|\mu_1+|S|\mu_2}{|L|+|S|}$. We claim that π satisfies the following condition. Fix j and k. Then there exist sets of indices V_1 and V_2, with $|V_i \oplus \{1, \ldots, k\}| \le 17d^2$, such that

$$\sum_{i\in V_1} w^j_{\pi(i)} \le (k-1)\mu^j + 3c_d M^j \tag{3.8}$$

$$\sum_{i\in V_2} w^j_{\pi(i)} \ge (k-1)\mu^j - 3c_d M^j. \tag{3.9}$$

We will prove this in the case ${\mu_1}^j \ge {\mu_2}^j$. If ${\mu_1}^j < {\mu_2}^j$ the proof is similar. Let $k_1 = |L \cap \pi\{1, \ldots, k\}|$ and $k_2 = |S \cap \pi\{1, \ldots, k\}|$, so that $\pi\{1, \ldots, k\} = \pi_1\{1, \ldots, k_1\} \cup \pi_2\{1, \ldots, k_2\}$. The method of interleaving ensures that

$$\begin{aligned} \frac{k_1-1}{k-1} &\le \frac{|L|}{|L|+|S|} \\ \frac{k_2-1}{k-1} &\le \frac{|S|}{|L|+|S|}. \end{aligned}$$

Therefore,

$$(k_1-1)\mu_1^j + k_2\mu_2^j \le (k-1)\mu^j \tag{3.10}$$

$$k_1\mu_1^j + (k_2-1)\mu_2^j \ge (k-1)\mu^j. \tag{3.11}$$

Clearly, the strong balance condition on π_1 implies that there exist U_1, U_2, with $|U_i \oplus \{1,\ldots,k_1\}| \leq 16d^2+1$, such that

$$\begin{aligned} \sum_{i\in U_1} w^j_{\pi_1(i)} &\leq (k_1-1)\mu^j_1 & (3.12)\\ \sum_{i\in U_2} w^j_{\pi_1(i)} &\geq k_1\mu^j_1. & (3.13)\end{aligned}$$

By the balance condition on π_2 we have

$$\begin{aligned} \sum_{i=1}^{k_2} w^j_{\pi_2(i)} &\leq k_2\mu^j_2 + c_d \max_{i\in S}\{|w^j_i - \mu^j_2|\} \leq k_2\mu^j_2 + 3c_d M^j & (3.14)\\ \sum_{i=1}^{k_2} w^j_{\pi_2(i)} &\geq k_2\mu^j_2 - c_d \max_{i\in S}\{|w^j_i - \mu^j_2|\} \geq (k_2-1)\mu^j_2 - 3c_d M^j. & (3.15)\end{aligned}$$

Now, let $V_i = \pi^{-1}(\pi_1(U_i) \cup \pi_2\{1,\ldots,k_2\})$, for $i=1,2$. Then $|V_i \oplus \{1,\ldots,k\}| \leq 16d^2+1 \leq 17d^2$, and

$$\sum_{i\in V_i} w^j_{\pi(i)} = \sum_{i\in U_i} w^j_{\pi_1(i)} + \sum_{i=1}^{k_2} w^j_{\pi_2(i)}.$$

Combining this with equations (3.10)–(3.15) gives (3.8) and (3.9).

Now, π determines a path $\{Z_i\}_{i=0}^{|X\oplus Y|}$ from X to Y, where $Z_0 = X$ and $Z_i = X \oplus \{\pi(1),\ldots,\pi(i)\}$ for all $1 \leq i \leq |X\oplus Y|$. This path might not stay in Ω, but we can alter it slightly so that it does. Equations (3.8) and (3.9) imply that for every k and j, there exists a set of indices W^j_k, with $|W^j_k| \leq 34d^2$, such that

$$\begin{aligned} a^j(Z_k - W^j_k) &\leq \max(a(X), a(Y)) + 3c_d M^j \leq b^j & (3.16)\\ a^j(Z_k \cup W^j_k) &\geq \min(a(X), a(Y)) - 3c_d M^j. & (3.17)\end{aligned}$$

Now, let $W_0 = \emptyset$ and for $k = 1,\ldots,|X\oplus Y|$, let $W_k = \cup_{j=1}^d W^j_k$. Then, for all k, $|W_k| \leq 34d^3$, and $a(Z_k - W_k) \leq \mathbf{b}$. For $0 \leq k \leq |X\oplus Y|$, let

$$Z'_k = Z_k - W_k.$$

Then each $Z'_k \in \Omega$. Our flow from X to Y will pass through each of the Z'_k. To get from each Z'_k to Z'_{k+1}, we perform the following steps:

1. Remove each item in $Z'_k - (Z'_k \cap Z'_{k+1})$ in index order.

2. Add each item in $Z'_{k+1} - (Z'_k \cap Z'_{k+1})$ in index order.

Define $W_x = (W_k \cup W_{k+1}) \cap X$, and $W_y = (W_k \cup W_{k+1}) \cap Y$. By analogy with sections (2.4) and (2.5), for each intermediate point Z along the path, define the "encoding" $\overline{Z}$ by

$$\overline{Z} = (X \oplus Y - Z) \cup (X \cap Y) - (W_k \cup W_{k+1}),$$

and let $U = \pi\{1, \ldots, k\}$. In similar fashion to the earlier analysis one can see that for a given Z, X and Y are completely specified by the 4-tuple $(\overline{Z}, U, W_x, W_y)$. We also have

$$\begin{aligned} a^j(\overline{Z}) &= a^j(X) + a^j(Y) - a^j(Z \cup W_k \cup W_{k+1}) \\ &\leq a^j(X) + a^j(Y) - \min(a^j(Z_k \cup W_k), a^j(Z_{k+1} \cup W_{k+1})) \\ &\leq a^j(X) + a^j(Y) - (\min(a^j(X), a^j(Y)) - 3c_d M^j) \\ &= \max(a^j(X), a^j(Y)) + 3c_d M^j \\ &\leq b^j, \end{aligned}$$

where the second inequality follows from (3.17). Hence $\overline{Z} \in \Omega$. We can therefore bound the flow $\hat{f}(Z)$ through Z by

$$\hat{f}(Z) \leq \sum_{\overline{Z} \in \Omega} \sum_{W_x, W_y, U} \Pr[\{\pi(1), \ldots, \pi(|U|)\} = U]. \tag{3.18}$$

Now, for a given X and Y, let $L^j = \{i \in X \cup Y : a_i^j > M^j\}$, so that $L = \cup_{j=1}^d L^j$. Note that for every j, $L^j \cap Y$ is equal to either L^j or $\emptyset$. Thus, if we define $k_2 = |U \cap S|$, then for given values of M and k_2, there are at most $2^d \binom{|S|}{k_2}$ possible values for U in the inner sum of equation (3.18). Therefore, we have

$$\begin{aligned} \hat{f}(Z) &\leq \sum_{\overline{Z} \in \Omega} \sum_{M, W_x, W_y, k_2} \sum_{U : |U \cap S| = k_2} \Pr[\{\pi_2(1), \ldots, \pi_2(k_2)\} = U \cap S] \\ &\leq \sum_{\overline{Z} \in \Omega} \sum_{M, W_x, W_y, k_2} \sum_{U : |U \cap S| = k_2} p_d(|S|) \binom{|S|}{k_2}^{-1} \\ &\leq \sum_{\overline{Z} \in \Omega} \sum_{M, W_x, W_y, k_2} 2^d p_d(|S|) \\ &\leq \sum_{\overline{Z} \in \Omega} n^d \left[\binom{n}{68d^3} 2^{68d^3} \right] n \cdot 2^d p_d(n) \\ &= \mathrm{poly}(n) |\Omega| \end{aligned}$$

$\square$

Given such a flow, we need only invoke Theorem 2.1 in Chapter 2 to derive the main result of this chapter.

Theorem 3.7 *Let Ω be the set of solutions to an arbitrary instance of the d-dimensional 0-1 knapsack problem, for a fixed d. The mixing time of the random walk on G_Ω is* $\mathrm{poly}_d(n)$.

As in one dimension, this immediately implies the existence of an fpras for computing $|\Omega|$.

Chapter 4

Sampling Contingency Tables

4.1 Introduction

4.1.1 The problem

Given positive integer vectors $r = (r_i)_{i=1}^m$ and $c = (c_j)_{j=1}^n$ with $\sum_{i=1}^m r_i = \sum_{j=1}^n c_j$ and $m, n \geq 2$, let $I(r, c)$ denote the set of non-negative integer $m \times n$ matrices with row sums $r_1, \ldots, r_m$ and column sums $c_1, \ldots, c_n$. In this chapter, we consider the problem of generating an element of $I(r, c)$ uniformly at random.

4.1.2 Motivation

We will now give a brief sketch of how this problem arises in Statistics. The interested reader should consult [4] for a comprehensive account.

Suppose that we perform an experiment in which N independent samples are taken and classified according to two characteristics A and B, which take the values $1, \ldots, m$, and $1, \ldots, n$, respectively. For example, A might classify each subject's blood type and B might measure cholesterol level. We then assemble the results in an $m \times n$ matrix X such that X_{ij} is equal to the number of samples having $A = i$ and $B = j$. Such a matrix is called a *contingency table*. We will be interested in measuring the amount of dependence between the two variables in a contingency table. A traditional way to quantify this dependence is via the chi-squared statistic:

$$\chi^2(X) = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{(X_{ij} - \frac{r_i c_j}{N})^2}{\frac{r_i c_j}{N}}.$$

This is the sum of $(\text{Observed} - \text{Expected})^2/\text{Expected}$, where "Expected" refers to the quantity $\frac{r_i c_j}{N}$, which is the number of observations we would expect to see in a particular cell if A and B were independent with $\Pr(A = i) = \frac{r_i}{N}$ and $\Pr(B = j) = \frac{c_j}{N}$ for all i and j. The *p-value* is defined as the probability that N samples from a distribution with independent row and column variables (i.e., independent A and B) would give a chi-squared statistic which is at least as large as the observed value. Thus, when the p-value is small, there is evidence that the data in the contingency table is dependent.

Consider the following simple examples with $m = n = 2$. Let

$$X_1 = \begin{bmatrix} 10 & 20 \\ 40 & 30 \end{bmatrix}; \qquad X_2 = \begin{bmatrix} 20 & 40 \\ 80 & 60 \end{bmatrix}.$$

For X_1 we have $\chi^2(X_1) = 4.76$, which gives a p-value of about .03. For X_2, we have $\chi^2(X_2) = 9.52$, and a p-value of about .002. Thus, comparing p-values we would conclude that there is more dependence in the second data set than in the first. However, both data sets appear to have come from a similar underlying distribution, namely, something like

$$\begin{bmatrix} .1 & .2 \\ .4 & .3 \end{bmatrix}.$$

The second p-value is smaller only because the sample size is larger. This illustrates a problem with the traditional approach of using the p-value alone to measure dependence. While the p-value is useful for determining the *existence* of dependence, it should not be used to measure the *amount* of dependence. Thus it is unwise to compare a p-value or chi-squared statistic from one experiment with that of another. This and other considerations led Diaconis and Efron [4] to propose the following statistic:

$$T(X) = \frac{|\{X' \in I(r,c) : \chi^2(X') < \chi^2(X)\}|}{|I(r,c)|},$$

where r and c are the row and column sums respectively of X. $T(X)$ is the fraction of all contingency tables in $I(r,c)$ which have a smaller chi-squared statistic than the observed value. $T(X)$ is not highly sensitive to sample size, and is thus a better measure of dependence. In the examples above, we have $T(X_1) = \frac{9}{31} \approx \frac{19}{61} = T(X_2)$. Of course, when the contingency tables are large we cannot always calculate $T(X)$ exactly. This explains why we want an algorithm to sample uniformly from $I(r,c)$; given such an algorithm, we can estimate $T(X)$ in the following way:

1. Take a large number of independent samples from $I(r,c)$.

2. Compute the fraction of samples X' for which $\chi^2(X') < \chi^2(X)$.

4.1.3 Results

Our method of sampling will rely on the fact that $I(r,c)$, when viewed as a subset of $\mathbf{R}^{mn}$, is the intersection of a continuous convex set and the integer lattice. Specifically, we have $I(r,c) = \mathbf{Z}^{mn} \cap K$, where K is the set of non-negative *real* matrices with the given row and column sums. As stated in Chapter 1, a number of (random walk-based) polynomial-time algorithms have been developed in recent years for sampling (almost) uniformly from a convex set (see [9],[23], [19], and [22]). We will solve our discrete sampling problem by reducing it to a continuous one. Given em any convex set K' which contains $I(r,c)$, we can generate a random sample from $I(r,c)$ using the following algorithm:

1. Generate a random point Y in K', and "round" it to an integer point Z.

2. If $Z \notin I(r,c)$, repeat.

The two main ingredients in this sampling technique are the convex set K' and the rounding method. The choice of K' is a delicate matter. We require that the distribution of the final sample be almost uniform. Thus, as X varies over $I(r,c)$, we require that $\mathrm{vol}(X)$ is nearly constant, where $\mathrm{vol}(X)$ denotes the volume of points in K' which round to X. Thus we must make K' sufficiently large. (We could not, for example, naively set K' equal to K, since under any reasonable rounding procedure this would result in a distribution with too little mass on matrices with a large number of zero entries.)

However, there is a tradeoff between the uniformity of the distribution and the running time of the algorithm. As K' becomes larger, it becomes more likely that each sample Z will fall outside of $I(r,c)$. Thus, if we make K' too large, then the expected number of trials taken before $Z \in I(r,c)$ will be too high.

What we want therefore is a convex set K' which is a good continuous approximation to the discrete set $I(r,c)$. Now, since the convex set K is a polytope, it can be defined in terms of bounding hyperplanes. We can therefore define K' in terms of another set of hyperplanes which are parallel to the original ones, only spread farther apart to ensure that $\mathrm{vol}(X) \approx 1$ for all $X \in I(r,c)$.

The approach we have just described has formed the basis for previous work on sampling contingency tables. Dyer, Kannan and Mount [11] describe an algorithm similar to the one above and show that it samples (almost) uniformly from $I(r,c)$ in polynomial time, provided that the row and column sums satisfy $r_i = \Omega(n^2 m)$ for all $1 \leq i \leq m$, and $c_j = \Omega(m^2 n)$ for all $1 \leq j \leq n$. These were essentially the best known bounds for general m and n (although when $m = 2$ there is a polynomial-time algorithm for any values of r and c; see [10]). In this chapter, we will show that the requirements can be loosened to $r_i = \Omega(n^{3/2} m \log(m))$ and $c_i = \Omega(m^{3/2} n \log(n))$. We accomplish this using the following method for rounding. For $Y \in K'$, we will round Y to the integer matrix Z which has $|Z_{ij} - Y_{ij}| \leq 1/2$ for all $i < m$, $j < n$, and has the appropriate row and column sums. This rounding method is quite simple, and it allows us to prove easy bounds on $\mathrm{vol}(X)$ for $X \in I(r,c)$. In turn, this allows us to determine the best choice for the convex set K', and leads to the improved requirements on the row and column sums.

The Contingency Tables problem is a special case of the problem of sampling from the set of integer points in a polytope. This is a class of problems that was studied by Kannan and Vempala in [20], where they give conditions on the polytope which guarantee a polynomial-time algorithm. (When they apply their results to the Contingency Tables problem, they improve on the row and column sum requirements given in [11], but only by logarithmic factors.) We believe that the techniques in this chapter may extend to other problems of this general type.

4.2 The Algorithm

We will now describe the algorithm for sampling from $I(r,c)$ in detail. Two things are needed to specify the algorithm. First, we need to define the convex set K' from which we perform continuous sampling. Second, we need to describe the rounding method which maps elements of K' to integer lattice points. Since the choice of K' will depend on the rounding method, we will discuss the rounding method first.

We will assume, without loss of generality, that $m \leq n$. Note that any $m \times n$ matrix X whose row and column sums are fixed can be completely specified by $(X_{ij})_{i<m, j<n}$. It will be helpful to think of $I(r,c)$ as a subset of $\mathbf{R}^{(m-1)(n-1)}$, indexed by $\{(i,j) : i < m, j < n\}$. Thus, we will define $I(r,c)$ as the set of $(m-1) \times (n-1)$ non-negative integer matrices satisfying the constraints:

$$\sum_{j=1}^{n-1} X_{ij} \leq r_i \quad \text{for all } i < m; \qquad \sum_{i=1}^{m-1} X_{ij} \leq c_j \quad \text{for all } j < n;$$

$$\sum_{i=1}^{m-1}\sum_{j=1}^{n-1} X_{ij} \geq \sum_{i=1}^{m-1} r_i - c_n \quad .$$

For $X \in I(r,c)$, we may still at times refer to the quantities X_{in} and X_{mj}, but it will be with the understanding that they are defined in terms of the other entries, via

$$X_{in} \equiv r_i - \sum_{j=1}^{n-1} X_{ij} \quad \text{for all } i < m, \qquad X_{mj} \equiv c_j - \sum_{i=1}^{m-1} X_{ij} \quad \text{for all } j \leq n \quad .$$

Using this convention, the rounding method (which we described in section 4.1) simply consists of rounding each of the $(m-1)(n-1)$ coordinates to the nearest integer. Furthermore, the convex set K' from which we perform continuous sampling will be a full-dimensional subset of $\mathbf{R}^{(m-1)(n-1)}$.

Recall that for integer matrices X and a convex set K', vol(X) is defined as the volume of points in K' which round to X. Thus, vol(X) is equal to the volume of the intersection of K' and the $(m-1)(n-1)$-dimensional unit hypercube centered at X. This leads to the following appealing characterization of vol(X). Let $\mathcal{E}$ be a random $(m-1) \times (n-1)$ matrix, whose entries are mutually independent and have the uniform distribution over $[-1/2, 1/2]$. Then for all integer matrices X, we have vol$(X) = \Pr(X+\mathcal{E} \in K')$. Thus, vol$(X)$ is equal to the probability that, if we perturb X by adding a small random variable to each entry, then the result is in K'.

Now, in order for the output of our algorithm to have an almost uniform distribution, K' must be large enough so that vol(X) is nearly 1 for all $X \in I(r,c)$. In light of the above, this means that if we take any X in $I(r,c)$ and add $\mathcal{E}$, then the result must be in K' with high probability. We are now ready to define K'. Let $0 < \epsilon < 1/2$ be an error parameter, and let

$$C_1 = \frac{\log(4/\epsilon)}{2\log(m)} + 1/2, \quad C_2 = \frac{\log(4/\epsilon)}{2\log(n)} + 1/2, \quad C_3 = \frac{1}{2}\log(2/\epsilon) \quad .$$

Let K' be the set of real, $(m-1) \times (n-1)$ matrices Y satisfying

$$Y_{ij} > -1/2 \tag{4.1}$$

$$Y_{in} \geq -\sqrt{C_1 n \log(m)} \qquad Y_{mj} \geq -\sqrt{C_2 m \log(n)},$$

$$Y_{mn} \geq -\sqrt{C_3 mn}$$

for all $i < m$ and $j < n$. The reasons behind our choices for the above parameters should become clear after we prove the following lemma.

Lemma 4.1 *For any $X \in I(r,c)$, we have $1 \geq \mathrm{vol}(X) \geq 1 - \epsilon$.*

Proof: Let $X \in I(r,c)$ and let $X' = X + \mathcal{E}$. We want to show that $\Pr(X' \notin K')$ is less than ϵ. Now, $X'_{ij} \geq -1/2$ for all $i < m$ and $j < n$, since $|\mathcal{E}_{ij}| < 1/2$. We also have $X'_{in} = X_{in} - \sum_{j=i}^{n-1} \mathcal{E}_{ij} \geq -\sum_{j=i}^{n-1} \mathcal{E}_{ij}$ for all $i < m$. Thus, Hoeffding's bounds [13] give

$$\Pr\left(X'_{in} < -\sqrt{C_1 n \log(m)}\right) \leq e^{-2C_1 \log(m)} = m^{-2C_1}, \text{ for all } i < m.$$

Hence the probability that some X'_{in} is too small is at most $m^{1-2C_1} = \epsilon/4$. Applying Hoeffding's bounds again, this time to the column sums, gives

$$\Pr\left(X'_{mj} < -\sqrt{C_2 m \log(n)}\right) \leq e^{-2C_2 \log(n)} = n^{-2C_2}, \text{ for all } j < n.$$

Hence the probability that some X'_{mj} is too small is also at most $\epsilon/4$. Finally, note that $X'_{mn} = X_{mn} + \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} \mathcal{E}_{ij}$. Hence, Hoeffding's bounds imply that $\Pr(X'_{mn} < -\sqrt{C_3 mn}) \leq e^{-2C_3} = \epsilon/2$. Putting this all together, we conclude that $\Pr(X' \notin K') \leq \epsilon/4 + \epsilon/4 + \epsilon/2 = \epsilon$, so $\mathrm{vol}(X) > 1 - \epsilon$. $\square$

We will call the algorithm that repeatedly samples from K' and rounds each sample to an integer point algorithm A. Let Z be the random integer matrix produced by a single trial of algorithm A. The preceding lemma implies that the conditional distribution of Z, given that it is in $I(r,c)$, is almost uniform. Thus, the final output of algorithm A will indeed be an almost uniform sample from $I(r,c)$.

4.3 Main theorem

In order to bound the running time of the algorithm, we must bound the expected number of trials taken before the random sample Z is in $I(r,c)$. We now state the main result of this chapter.

Theorem 4.2 *Suppose that the row and column sums satisfy $r_i = \Omega(n^{3/2} m \log(m))$ and $c_j = \Omega(m^{3/2} n \log(n))$ for all i and j. Then the expected number of trials before algorithm A generates a sample point $Z \in I(r,c)$ is $O(1/\epsilon^2)$.*

Proof: First, assume for simplicity that $r_i \geq 2n^{3/2}m\log(m)$ and $c_j \geq 2m^{3/2}n\log(n)$ for all i and j. Algorithm A repeatedly generates random integer matrices Z until one of them is in $I(r,c)$. Thus, the number of trials is a geometric random variable with parameter $\Pr(Z \in I(r,c))$. Let $\bar{I}(r,c)$ denote the set of integer matrices X having $\mathrm{vol}(X) > 0$, i.e., the points which have some $Y \in K'$ rounding to them. Then the expected number of trials $\frac{1}{\Pr(Z\in I(r,c))}$ is equal to

$$\frac{\mathrm{vol}(K')}{\mathrm{vol}(I(r,c))} = \frac{\mathrm{vol}(\bar{I}(r,c))}{\mathrm{vol}(I(r,c))}.$$

Suppose that $X \in \bar{I}(r,c)$. Then by (4.1), X must satisfy $X_{ij} \geq 0$ for all $i < m$ and $j < n$. Thus, X is in $I(r,c)$ if and only if it satisfies $X_{in} \geq 0$ for all $1 \leq i \leq m$ and $X_{mj} \geq 0$ for all $1 \leq j \leq n$. Thus X, when thought of as an $m \times n$ matrix, is in $I(r,c)$ when it has no negative entry anywhere in its last row or column. We must show that such points form a non-negligible fraction of $\bar{I}(r,c)$.

A sketch of our argument is as follows. First, consider the random variable Z_{1n}. Since the row and column sums are large, the probabilities $\Pr(Z_{1n} = k)$ will remain roughly constant over a long interval in k. Thus, since the number of possible negative values for Z_{1n} is limited to about n, the probability that Z_{1n} will take a non-negative value is quite large. Of course, a similar argument will hold for the other Z_{in} and Z_{mj}, so Z will stand a good chance of being in $I(r,c)$.

Instead of working directly with the vol function, we will find it easier to work with an upper bound on vol which is based on Hoeffding's bounds. Let $X \in \bar{I}(r,c)$. Then Hoeffding's bounds give the following three upper bounds on $\mathrm{vol}(X)$:

$$\begin{aligned}
\mathrm{vol}(X) &\leq \exp\left(\frac{-2\sum_{i<m}((X_{in}+C_1')^-)^2}{n}\right)\\
\mathrm{vol}(X) &\leq \exp\left(\frac{-2\sum_{j<n}((X_{mj}+C_2')^-)^2}{m}\right)\\
\mathrm{vol}(X) &\leq \exp\left(\frac{-2((X_{mn}+C_3')^-)^2}{mn}\right)
\end{aligned}$$

where $C_1' = \sqrt{C_1 n\log(m)}$, $C_2' = \sqrt{C_2 m\log(n)}$, $C_3' = \sqrt{C_3 mn}$, and $x^- = \max(0,-x)$. Putting these together (taking the geometric mean of these three bounds), we get $\mathrm{vol}(X) \leq w(X)$, where

$$w(X) = \exp\left\{-\frac{2}{3}\left[\frac{\sum_{i<m}((X_{in}+C_1')^-)^2}{n} + \frac{\sum_{j<n}((X_{mj}+C_2')^-)^2}{m} + \frac{((X_{mn}+C_3')^-)^2}{mn}\right]\right\} \tag{4.2}$$

We will call w a *weight* function. Note that for all integer points X we have $0 \le \mathrm{vol}(X) \le w(X) \le 1$, and since $\epsilon < \frac{1}{2}$, we have $w(I(r,c)) \le 2 \cdot \mathrm{vol}(I(r,c))$ by Lemma 4.1. This implies that the expected number of trials $\frac{\mathrm{vol}(\bar{I}(r,c))}{\mathrm{vol}(I(r,c))} \le 2 \cdot \frac{w(\bar{I}(r,c))}{w(I(r,c))}$. Thus, our task reduces to giving an upper bound on the quantity $\frac{w(\bar{I}(r,c))}{w(I(r,c))}$.

Recall that $\bar{I}(r,c) - I(r,c)$ consists of the points in $\bar{I}(r,c)$ which have a negative entry somewhere in their last row or column. Let W_0 denote the set of matrices X in $\bar{I}(r,c)$ which have $X_{mn} \ge 0$. Next, for all i in $\{1, \dots, m-1\}$, define $W_i = \{X \in W_{i-1} : X_{in} \ge 0\}$. Finally, for all i in $\{m, m+n-2\}$, define $W_i = \{X \in W_{i-1} : X_{m,i-m+1} \ge 0\}$. We have $\bar{I}(r,c) \supset W_0 \supset W_1, \dots, \supset W_{m+n-2} = I(r,c)$. Our strategy will be to write $\frac{w(\bar{I}(r,c))}{w(I(r,c))}$ as the product

$$\left(\frac{w(\bar{I}(r,c))}{w(W_0)}\right)\left(\frac{w(W_0)}{w(W_1)}\right)\cdots\left(\frac{w(W_{m+n-3})}{w(W_{m+n-2})}\right) \tag{4.3}$$

and then show that each factor is not too large.

First, we will bound $w(\bar{I}(r,c))/w(W_0)$, which amounts to showing that there is not too much weight on the points X which have $X_{mn} < 0$. For integers s, let $V_s = \{X \in \bar{I}(r,c) : X_{mn} = s\}$. Then we want to bound

$$\frac{\sum_s w(V_s)}{\sum_{s \ge 0} w(V_s)} \tag{4.4}$$

We will do this by giving an upper bound on $\frac{w(V_{s-1})}{w(V_s)}$ for each s. What we want to show is that $w(V_{s-1}) \le \alpha w(V_s)$, for some α only slightly larger than 1. We will do this in a way that is reminiscent of the standard technique in Combinatorics in which one shows that two sets S_1 and S_2 satisfy $|S_1| \le k|S_2|$ by giving a k-to-1 function from S_1 to S_2. What we will do here is give a *random* function (i.e., a function which is itself a random variable) from V_{s-1} to V_s with the property that for every $X \in V_s$, the *expected* value of the total weight of points mapping to X is at most $\alpha \cdot w(X)$.

For positive integers $a < m$ and $b < n$, let $T_{ab}(\cdot)$ be the transformation, acting on a matrix X, that increases X_{ab} and X_{mn} by 1 and reduces X_{an} and X_{mb} by 1. Let $M = m^{3/2} n \log(n)$, and for $0 \le s \le M$, let f_s be a random function from V_{s-1} to V_s such that

$$f_s(X) = T_{ab}(X) \ \text{ with probability } \frac{X_{an}^+ X_{mb}^+}{\left(\sum_{i<m} X_{in}^+\right)\left(\sum_{j<n} X_{mj}^+\right)}$$

for all $a < m$ and $b < n$, where $x^+ = \max(0, x)$. Note that if $X \in V_s$ for $s \le -C_3'$, and

$X' = T_{ab}^{-1}(X)$ for some a and b, then

$$\frac{w(X')}{w(X)} = \exp\left\{-\frac{2}{3} \cdot \frac{(s+C_3'-1)^2-(s+C_3')^2}{mn}\right\} = \exp\left\{\frac{2}{3} \cdot \frac{2(s+C_3')-1}{mn}\right\}.$$

(Note that the random a and b in the definition of X' will be concentrated on points having X_{an} and X_{nb} greater than 0. Thus their contribution to the first two terms in the exponent in (4.2) is 0.) Let

$$p(s) = \begin{cases} \exp\left\{\frac{2}{3} \cdot \frac{2(s+C_3')-1}{mn}\right\} & \text{if } s \le -C_3'; \\ 1 & \text{otherwise.} \end{cases}$$

Then for all $s \le M$ and $X \in V_s$ we have

$$\begin{aligned} \mathrm{E}(w(f_s^{-1}(X))) &\le \sum_{a=1}^{m-1}\sum_{b=1}^{n-1} \frac{(X_{an}^+ + 1)(X_{mb}^+ + 1)}{\left(\sum_i X_{in}^+ + 1\right)\left(\sum_j X_{mj}^+ + 1\right)} \cdot p(s) \cdot w(X) \\ &= \frac{[\sum_a X_{an}^+ + m] \cdot \left[\sum_b X_{mb}^+ + n\right]}{\left[\sum_a X_{an}^+ + 1\right] \cdot \left[\sum_b X_{mb}^+ + 1\right]} \cdot p(s) \cdot w(X) \\ &\le (1 + \frac{1}{n\sqrt{m}\log(n)})(1 + \frac{1}{m\sqrt{n}\log(m)}) \cdot p(s) \cdot w(X) \\ &\le (1 + \frac{3}{m\sqrt{n}\log(m)}) \cdot p(s) \cdot w(X) \end{aligned} \tag{4.5}$$

where the second inequality follows from the facts that $[\sum_a X_{an}^+] \ge m^{3/2} n \log(n)$ and $\left[\sum_b X_{mb}^+\right] \ge n^{3/2} m \log(m)$ (recall that $X_{mn} \le nm^{3/2} \le mn^{3/2}$).

It follows that for all $s \le M$ we have $\frac{w(V_{s-1})}{w(V_s)} \le (1 + \frac{3}{m\sqrt{n}\log(m)}) \cdot p(s)$. A calculation shows that this bound is less than $(1 - \frac{1}{\sqrt{mn}})$ when $s \le -C_3' - 3\sqrt{mn}$. Thus

$$\frac{w(V_{s-1})}{w(V_s)} \le \begin{cases} (1 - 1/\sqrt{mn}) & \text{if } s \le -C_3' - 3\sqrt{mn}; \\ 1 + \frac{3}{m\sqrt{n}\log(m)} & \text{otherwise.} \end{cases}$$

Next, by assuming that the preceding inequality is actually an equality (which can only increase the quantity (4.4) we are trying to bound), we obtain

$$\frac{\sum_s w(V_s)}{\sum_{s \ge 0} w(V_s)} \le \frac{\sum_{j\ge 0}(1 - 1/\sqrt{mn})^j + \sum_{j=1}^{M+C_3'+3\sqrt{mn}}(1 + \frac{3}{m\sqrt{n}\log(m)})^{-j}}{\sum_{j=C_3'+3\sqrt{mn}}^{M+C_3'+3\sqrt{mn}}(1 + \frac{3}{m\sqrt{n}\log(m)})^{-j}}.$$

A straightforward calculation (using the fact that $C_3' = \sqrt{C_3 mn} = \sqrt{\frac{1}{2}\log(2/\epsilon)mn}$) shows that this upper bound is $O(1/\epsilon)$, where the implied constant does not depend on m and n.

Next, we will bound the second ratio in (4.3), namely $\frac{w(W_0)}{w(W_1)}$. For $s \leq M$, let $T_s = \{X \in W_0 : X_{1n} = s\}$. Using an argument very similar to that above, we obtain

$$\frac{w(T_{s-1})}{w(T_s)} \leq \begin{cases} (1 - n^{-1/2}) & \text{if } s \leq -C_1' - 3\sqrt{n}; \\ 1 + \frac{3}{m\sqrt{n}\log(m)} & \text{otherwise.} \end{cases}$$

Hence

$$\frac{w(W_0)}{w(W_1)} = \frac{\sum_s w(T_s)}{\sum_{s\geq 0} w(T_s)} \quad \leq \quad \frac{\sum_{j\geq 0}(1-n^{-1/2})^j + \sum_{j=1}^{M+C_1'+3\sqrt{n}}(1+\frac{3}{m\sqrt{n}\log(m)})^{-j}}{\sum_{j=C_1'+3\sqrt{n}}^{M+C_1'+3\sqrt{n}}(1+\frac{3}{m\sqrt{n}\log(m)})^{-j}},$$

and a calculation shows that this upper bound is $O\left(\exp\{\frac{\log(\frac{4}{\epsilon})}{2m}\}\right)$. An identical argument gives the same bound for $\frac{w(W_i)}{w(W_{i+1})}$ for all $1 \leq i < m-1$.

Next, we must bound the ratios $\frac{w(W_i)}{w(W_{i+1})}$ for $i = m-1, \ldots, m+n-2$. It turns out that we run into a minor technical problem here. In order for our calculations to proceed in a similar fashion to those above, we need the last row sum r_m to satisfy $r_m \geq 2n^2\sqrt{m}\log(n)$. Unfortunately, this is not directly implied by any of our assumptions, since m could be much less than n. However, the conditions on the column sums imply that the average entry in any column is at least $2n\sqrt{m}\log(n)$. Thus, the same lower bound holds for the average entry of the entire matrix. Hence there must be at least one row sum larger than $n \cdot (2n\sqrt{m}\log(n)) = 2n^2\sqrt{m}\log(n)$. It is no loss of generality to assume that one such row is the last row.

With this additional assumption on r_m, we can obtain a bound which is analogous to (4.5), only with m and n interchanged. For integers $s < M$, let $U_s = \{X \in W_{m-1} : X_{m1} = s\}$. Proceeding as before, only with the roles of m and n interchanged, we obtain the bound

$$\frac{w(U_{s-1})}{w(U_s)} \leq \begin{cases} (1 - m^{-1/2}) & \text{if } s \leq -C_2' - 3\sqrt{m}; \\ 1 + \frac{3}{n\sqrt{m}\log(n)} & \text{otherwise.} \end{cases}$$

Hence

$$\frac{w(W_{m-1})}{w(W_m)} \quad \leq \quad \frac{\sum_{j\geq 0}(1-m^{-1/2})^j + \sum_{j=1}^{M+C_2'+3\sqrt{m}}(1+\frac{3}{n\sqrt{m}\log(n)})^{-j}}{\sum_{j=C_2'+3\sqrt{m}}^{M+C_2'+3\sqrt{m}}(1+\frac{3}{n\sqrt{m}\log(n)})^{-j}} = O\left(\exp\{\frac{\log(\frac{4}{\epsilon})}{2n}\}\right).$$

An identical argument gives the same bound for $\frac{w(W_i)}{w(W_{i+1})}$ for all $m-1 < i < m+n-2$. Putting this all together, we get

$$\frac{w(\bar{I}(r,c))}{w(I(r,c))} = O\left(\frac{1}{\epsilon} \cdot \left(\exp\left\{\frac{\log(\frac{4}{\epsilon})}{2m}\right\}\right)^{m-1} \left(\exp\left\{\frac{\log(\frac{4}{\epsilon})}{2n}\right\}\right)^{n-1}\right) = O(1/\epsilon^2).$$

We have shown that the expected number of trials is $O(1/\epsilon^2)$ when we assume that $r_i \geq 2n^{3/2} m \log(m)$, and $c_j \geq 2m^{3/2} n \log(n)$ for all i and j. An easy modification of our proof allows us to handle the case when we only have $r_i = \Omega(n^{3/2} m \log(m))$ and $c_j = \Omega(m^{3/2} n \log(n))$. $\square$

We conclude that the (expected) running time of algorithm A is of the form

$$q(m,n) \cdot O(1/\epsilon^2),$$

where $q(m,n)$ is the (polynomial) running time of a single trial. Hence the algorithm runs in time which is a polynomial function of m, n and $1/\epsilon$.

Note that we do not have the logarithmic dependence on $1/\epsilon$ that can be achieved in many sampling algorithms that use direct applications of Markov Chains. However, in most applications of random sampling (e.g. [29, 17]), logarithmic dependence on $1/\epsilon$ is not necessary.

If the continuous samples from the convex set K' were *perfectly* uniform, then by Lemma 4.1 the parameter ϵ would be an upper bound on the total variation distance between the sample produced by algorithm A and the uniform distribution over $I(r,c)$. However, in all algorithms for sampling continously from convex set, the samples are only *almost* uniform. Algorithm A will therefore only guarantee that the distance from uniform is at most $\epsilon + \delta$ for some small δ.

Bibliography

[1] N. Alon. Eigenvalues and expanders. *Combinatorica*, 6:83–96, 1986.

[2] N. Alon and V. D. Milman. λ_1, isoperimetric inequalities for graphs, and superconcentrators. *J. Combin. Theory Ser. B*, 38(1):73–88, 1985.

[3] P. Dagum, M. Luby, M. Mihail, and U. Vazirani. Polytopes, permanents and graphs with large factors. In *Proceedings of the 29th IEEE Symposium on Foundations of Computer Science*, pages 412–421, 1988.

[4] P. Diaconis and B. Efron. Testing for independence in a two-way table: new interpretations of the chi-squared statististic. *Annals of Statistics*, 13:845–913, 1985.

[5] P. Diaconis and L. Saloff-Coste. Comparison techniques for reversible Markov chains. *Annals of Applied Probability*, 3:696–730, 1993.

[6] P. Diaconis and D. Stroock. Geometric bounds for eigenvalues of Markov chains. *Annals of Applied Probability*, 1:36–61, 1991.

[7] M. Dyer and A. Frieze. Computing the volume of convex bodies: a case where randomness provably helps. In *Probabilistic Combinatorics and its Applications*, volume 44 of *Proceedings of AMS Symposia in Applied Mathematics*.

[8] M. Dyer, A. Frieze, R. Kannan, A. Kapoor, L. Perkovic, and U. Vazirani. A mildly exponential time algorithm for approximating the number of solutions to a multidimensional knapsack problem. *Combinatorics, Probability and Computing*, 2:271–284, 1993.

[9] M. E. Dyer, A. M. Frieze, and R. Kannan. A random polynomial time algorithm for approximating the volume of convex bodies. *Journal of the ACM*, 38:1–17, 1991.

[10] M. E. Dyer and C. Greenhill. A genuinely polynomial-time algorithm for sampling two-rowed contingency tables.

[11] M. E. Dyer, R. Kannan, and J. Mount. Sampling contingency tables. *Random Structures & Algorithms*, 10:487–506, 1997.

[12] V. S. Grinberg and S. V. Sevast'yanov. Value of the Steinitz constant. *Funktsional. Anal. i Prilozhen.*, 14(2):56–57, 1980.

[13] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.

[14] M. R. Jerrum. Mathematical foundations of the Markov chain Monte Carlo method. In J. Ramirez-Alfonsin M. Habib, C. McDiarmid and B. Reed, editors, *Probabilistic Methods for Algorithmic Discrete Mathematics*, volume 16, pages 116–165. Springer-Verlag, 1998.

[15] M. R. Jerrum and A. J. Sinclair. Polynomial-time approximation algorithms for the Ising model. *SIAM Journal on Computing*, 22:1087–1116, 1993.

[16] M. R. Jerrum and A.J. Sinclair. Approximating the Permanent. *SIAM Journal on Computing*, 22:1149–1178, 1989.

[17] M. R. Jerrum, L. G. Valiant, and V. V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Science*, 43:169–188, 1986.

[18] R. Kannan. Markov chains and polynomial time algorithms. In *Proceedings of the 35th IEEE Conference on Foundations of Computer Science*, pages 656–671, 1994.

[19] R. Kannan, L. Lovász, and M. Simonovits. Random walks and an $O^*(n^5)$ volume algorithm for convex bodies. *Random Structures & Algorithms*, 11:1–50, 1997.

[20] R. Kannan and S. Vempala. Sampling lattice points. In *Proceedings of the 29th Annual Symposium on the Theory of Computing*, pages 696–700, 1997.

[21] R. M. Karp and M. Luby. Monte-carlo algorithms for enumeration and reliability problems. In *Proceedings of the 24th Annual IEEE Symposium on Foundations of Computer Science*, pages 56–64, 1983.

[22] L. Lovász. Hit-and-run mixes fast. *Mathematical Programming*, 86(3, Ser. A):443–461, 1999.

[23] L. Lovász and R. Kannan. Faster mixing via average conductance. In *Proceedings of ACM Symposium on Theory of Computing*, pages 282–287, 1999.

[24] B. Morris. Improved bounds for sampling contingency tables. In *3rd International Workshop on Randomization and Approximation*, volume 1671 of *Lecture Notes in Computer Science*, pages 121–129, 1999.

[25] B. J. Morris and A. J. Sinclair. Random walks on truncated cubes and sampling 0-1 knapsack solutions. In *40th IEEE Symposium on Foundations of Computer Science*, pages 230–240, 1999.

[26] L. Saloff-Coste. *Lectures on finite Markov chains.* Lecture notes, Ecole d'été de probabilités de St. Flour. 1996.

[27] A. J. Sinclair. Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combinatorics, Probability and Computing*, 1:351–370, 1992.

[28] A. J. Sinclair and M. R. Jerrum. Approximate counting, uniform generation and rapidly mixing Markov chains. *Information and Computation, 82:93–133*, 82:93–133, 1989.

[29] A.J. Sinclair. *Randomised algorithms for counting and generating combinatorial structures.* Advances in Theoretical Computer Science. Birkhäuser, Boston, 1993.

Appendix A

Combinatorial Lemmas for Chapter 2

We give here two technical lemmas that were used in the proof of Theorem 2.4.

Lemma A.1 *Let ν be a uniform random permutation in $\mathcal{S}_m$. Then*

$$\Pr[\nu \text{ is not } \alpha\text{-good}] \leq 2m \exp(-2\alpha^2).$$

Proof: We adopt the notation of the proof of Theorem 2.4. Let $1 \leq k_1 \leq m_1$. It suffices to show that

$$\Pr\Big[\Big|\sum_{i=1}^{k_1} w_{\nu_1(i)} - k_1\mu_1\Big| > \alpha(M-1)\sqrt{k_1}\Big] \leq 2\exp(-2\alpha^2),$$

for then the lemma follows from the union bound and from symmetry (which allows us to replace k_1 by $k_1^* = \min\{k_1, m_1 - k_1\}$). But this inequality is a direct consequence of Hoeffding's bound on deviations in sampling without replacement [13]. $\square$

Lemma A.2 *Let m_1, m_2, k_1, k_2 be non-negative integers and $m = m_1 + m_2$, $k = k_1 + k_2$. Then*

$$\frac{\binom{m_1}{k_1}\binom{m_2}{k_2}}{\binom{m}{k}} \geq Cm^{-1/2} \exp\left\{-\frac{l^2 + \frac{1}{2}|l|}{\alpha(1-\alpha)}\Big(\frac{1}{m_1} + \frac{1}{m_2}\Big)\right\},$$

where $\alpha = \frac{k}{m}$, $l = \frac{m_1 k_2 - m_2 k_1}{m}$, and $C > 0$ is a universal constant.

Proof: Note that $k = \alpha m$, $k_1 = \alpha m_1 - l$, and $k_2 = \alpha m_2 + l$. By the symmetry of binomial coefficients, we may assume that $l \geq 0$. We shall prove the lemma by showing the two inequalities

$$\frac{\binom{m_1}{\alpha m_1}\binom{m_2}{\alpha m_2}}{\binom{m}{\alpha m}} \geq C_1 m^{-1/2} \tag{A.1}$$

and

$$\frac{\binom{m_1}{\alpha m_1 - l}\binom{m_2}{\alpha m_2 + l}}{\binom{m_1}{\alpha m_1}\binom{m_2}{\alpha m_2}} \geq C_2 \exp\left\{-\frac{l^2+\frac{1}{2}|l|}{\alpha(1-\alpha)}\left(\frac{1}{m_1}+\frac{1}{m_2}\right)\right\} \tag{A.2}$$

for positive constants C_1, C_2.

The first inequality is an immediate consequence of Stirling's approximation, $\sqrt{2\pi n}(\frac{n}{\mathrm{e}})^n \leq n! \leq C_3\sqrt{2\pi n}(\frac{n}{\mathrm{e}})^n$, where $C_3 = 1 + \mathrm{e}^{1/12}$ is a constant. To prove the second inequality, we apply Stirling's approximation to all four binomial coefficients to get the following lower bound on the left-hand side of (A.2):

$$\left[\frac{\mathrm{P}(\alpha m_1)\mathrm{P}((1-\alpha)m_1)\mathrm{P}(\alpha m_2)\mathrm{P}((1-\alpha)m_2)}{\mathrm{P}(\alpha m_1 - l)\mathrm{P}((1-\alpha)m_1+l)\mathrm{P}(\alpha m_2+l)\mathrm{P}((1-\alpha)m_2-l)}\right] \times \left[\frac{\alpha m_1(1-\alpha)m_1\alpha m_2(1-\alpha)m_2}{(\alpha m_1-l)((1-\alpha)m_1+l)(\alpha m_2+l)((1-\alpha)m_2-l)}\right]^{1/2}, \tag{A.3}$$

where $\mathrm{P}(x)$ denotes x^x. Now we have

$$\begin{aligned}\frac{\mathrm{P}(\alpha m_1)}{\mathrm{P}(\alpha m_1 - l)} &= (\alpha m_1)^l\left(1+\frac{l}{\alpha m_1 - l}\right)^{\alpha m_1 - l}\\ &\geq (\alpha m_1)^l \exp\left\{\frac{l(\alpha m_1 - l)}{\alpha m_1}\right\},\end{aligned}$$

where we have used the inequality $(1+\frac{x}{y})^y \geq \exp(\frac{xy}{x+y})$, valid for $x, y > 0$. Handling the three other pairs of factors in the numerator and denominator in similar fashion (using in addition the inequality $(1-\frac{x}{y})^y \geq \exp(\frac{-xy}{y-x})$, valid for $y > x > 0$) we see that the first parenthesis in (A.3) is bounded below by

$$\exp\left\{-\frac{l^2}{\alpha(1-\alpha)}\left(\frac{1}{m_1}+\frac{1}{m_2}\right)\right\}. \tag{A.4}$$

A similar calculation bounds the second parenthesis in (A.3) by

$$\exp\left\{-\frac{|l|}{\alpha(1-\alpha)}\left(\frac{1}{m_1}+\frac{1}{m_2}\right)\right\}. \tag{A.5}$$

Combining (A.4) and (A.5) completes the verification of inequality (A.2) above, and hence the proof of the lemma. $\square$

Appendix B

Strongly Balanced Permutations

Here we prove the existence of strongly balanced permutations, which are used in the proof of Lemma 3.6 in Chapter 3.

Theorem B.1 *For any sequence $\{w_i\}_{i=1}^m$ in $\mathbf{R}^d$, there exists a strongly $16d^2$–balanced permutation π.*

Proof: First, suppose that $\sum_{i=1}^m w_i = 0$. We will show that, in this case, there exists a strongly $8d^2$–balanced permutation π. Let L be the set containing the $4d$ indices i which maximize w_i^j and the $4d$ indices i which maximize $-w_i^j$, for each $j \leq d$. Then $|L| \leq 8d^2$.

The permutation π we construct will satisfy $\{\pi(m), \ldots, \pi(m - |L| + 1)\} = L$. It will be enough to check that the strong balance condition holds for $1 \leq k \leq m - |L|$. It suffices to show that, for all $j \leq d$ and $k \leq m$ we have,

$$-s^{j+} \leq \sum_{i=1}^{k} w_{\pi(i)}^j \leq s^{j-}, \tag{B.1}$$

where

$$s^{j+} \equiv \sum_{i \in L} (w_i^j)^+$$
$$s^{j-} \equiv \sum_{i \in L} (w_i^j)^-.$$

We will need the following Lemma which is due to Grinberg and Sevast'yanov [12].

Lemma B.2 *(Grinberg and Sevast'yanov, 1980) Let $x_1, \ldots, x_n$ be vectors in $\mathbf{R}^d$ such that $\sum_i x_i = 0$. Then there exists a permutation $\nu \in \mathcal{S}_n$ such that*

$$\sum_{i=1}^{k} x_{\nu(i)} \in d \cdot \mathrm{conv}\{x_1, \ldots, x_n\} \qquad \textit{for all } k.$$

Note that the permutation ν' with $\nu'(i) = \nu(n+1-i)$ for all i satisfies

$$\sum_{i=1}^{k} x_{\nu'(i)} \in -d \cdot \mathrm{conv}\{x_1, \ldots, x_n\} \qquad \text{for all } k.$$

Let $S_1 = \{1, \ldots, m\} - L$, let $m' = m - |L|$, and let π_1 be a permutation on S_1 such that

$$\sum_{i=1}^{k} (w_{\pi_1(i)} - \mu_1) \in -d \cdot \mathrm{conv}\{w_i - \mu_1 : i \in S_1\},$$

for all k, where $\mu_1 = \frac{1}{m'} \sum_{i \in S_1} w_i$. Suppose that m' is even and $m' = 2r$. If m' is odd the proof is similar. Now, let $S_2 = \{\pi_1(r+1), \ldots, \pi_1(m')\}$, and let π_2 be a permutation on S_2 such that

$$\sum_{i=1}^{k} (w_{\pi_2(i)} - \mu_2) \in d \cdot \mathrm{conv}\{w_i - \mu_2 : i \in S_2\},$$

where $\mu_2 = \frac{1}{r} \sum_{i \in S_2} w_i$. Define the permutation π by

$$\pi(i) = \begin{cases} \pi_1(i), & \text{if } i \le r; \\ \pi_2(i-r), & \text{if } r < i \le m'. \end{cases}$$

We must check that π satisfies (B.1). Fix j. W.l.o.g. $s^{j+} \le s^{j-}$, so that $\mu_1^j \ge 0$. For $k \le r$ we have

$$\begin{aligned} \sum_{i=1}^{k} w_{\pi(i)}^j &= \sum_{i=1}^{k} w_{\pi_1(i)}^j \\ &\ge \sum_{i=1}^{k} (w_{\pi_1(i)}^j - \mu_1^j) \\ &\ge -d \max_{1 \le i \le m'} \{w_i^j - \mu_1^j\} \\ &\ge -d \max_{1 \le i \le m'} \{w_i^j\} \\ &\ge -s^{j+}/4 \\ &\ge -s^{j+}, \end{aligned}$$

and

$$
\begin{aligned}
\sum_{i=1}^{k} w^j_{\pi(i)} &= \sum_{i=1}^{k} w^j_{\pi_1(i)} \\
&= k\mu^j_1 + \sum_{i=1}^{k}(w_{\pi_1(i)} - \mu^j_1) \\
&\leq r\mu^j_1 + d \max_{1\leq i\leq m'}\{-(w^j_i - \mu^j_1)\} \\
&= \tfrac{1}{2}(s^{j-} - s^{j+}) + d \max_{1\leq i\leq m'}\{-w^j_i\} + d\mu^j_1 \\
&\leq \tfrac{1}{2}(s^{j-} - s^{j+}) + (\tfrac{1}{4}s^{j-}) + (\tfrac{1}{4}s^{j+}) \\
&= \tfrac{3}{4}s^{j-} - \tfrac{1}{4}s^{j+} \\
&\leq s^{j-} - \tfrac{1}{4}s^{j+}. \qquad \text{(B.2)}
\end{aligned}
$$

(We will need the extra $-\frac{1}{4}s^{j+}$ in the second part of this proof.) If $r < k \leq 2r$ we have

$$
\begin{aligned}
\sum_{i=1}^{k} w^j_{\pi(i)} &= \sum_{i=1}^{r} w^j_{\pi_1(i)} + \sum_{i=1}^{k-r} w^j_{\pi_2(i)} \\
&= \left[\sum_{i=1}^{r} w^j_{\pi_1(i)} + (k-r)\mu^j_2\right] + \sum_{i=1}^{k-r}(w^j_{\pi_2(i)} - \mu^j_2) \\
&\geq \left[\tfrac{1}{2}(s^{j-} - s^{j+}) - d \max_{1\leq i\leq m'}\{w^j_i - \mu^j_1\}\right] - d \max_{i\in S_2}\{-(w^j_i - \mu^j_2)\} \\
&\geq \left[\tfrac{1}{2}(s^{j-} - s^{j+}) - d \max_{1\leq i\leq m'}\{w^j_i\}\right] - d \max_{i\in S_2}\{-w^j_i\} - d\mu^j_2 \\
&\geq \left[\tfrac{1}{2}(s^{j-} - s^{j+}) - s^{j+}/4\right] - s^{j-}/4 - s^{j-}/4 \\
&= -\frac{3}{4}s^{j+} \\
&\geq -s^{j+},
\end{aligned}
$$

and

$$
\sum_{i=1}^{k} w^j_{\pi(i)} \leq \sum_{i=1}^{r} w^j_{\pi_1(i)} + (k-r)\mu^j_2 + \min(d, k-r)\cdot \max_{i\in S_2}\{w^j_i - \mu^j_2\}.
$$

If $\mu^j_2 \geq 0$ we have

$$
\begin{aligned}
\text{RHS} &\leq (s^{j-} - s^{j+}) + d \cdot \max_{i\in S_2}\{w^j_i\} \\
&\leq (s^{j-} - s^{j+}) + s^{j+}/4 \\
&\leq s^{j-}.
\end{aligned}
$$

If $\mu_2^j < 0$ we have

$$\begin{aligned}
\text{RHS} &\leq \sum_{i=1}^{r} w^j_{\pi_1(i)} + \min(d, k-r) \cdot \max_{i \in S_2}\{w^j_i\} \\
&\leq (s^{j-} - s^{j+}/4) + s^{j+}/4 \qquad\qquad \text{by equation (B.2)} \\
&\leq s^{j-}.
\end{aligned}$$

Thus π is strongly $8d^2$–balanced. Furthermore, for all k and j there exists a set of indices $S \supseteq \{1, \ldots, k\}$, with $|S| \leq k + 8d^2$, such that $\sum_{i=1}^{k} w^j_{\pi(i)}$ and $\sum_{i \in S} w^j_{\pi(i)}$ have opposite signs.

Now, let $\{w_i\}_{i=1}^m$ be arbitrary. By above, there exists a permutation π which is strongly $8d^2$–balanced with respect to the sequence $\{w_i - \mu\}_{i=1}^m$, where $\mu = \frac{1}{m}\sum_{i=1}^m w_i$. We claim that π is also strongly $16d^2$–balanced with respect to the original sequence $\{w_i\}_{i=1}^m$. Fix k and j. W.l.o.g. $\mu^j > 0$. Then there exists $S \supseteq \{1, \ldots, k\}$, with $|S| \leq k + 8d^2$ such that

$$\sum_{i \in S} w^j_{\pi(i)} - k\mu^j \geq \sum_{i \in S} w^j_{\pi(i)} - |S|\mu^j \geq 0.$$

In addition, there is an $S \supseteq \{1, \ldots, k\}$, with $|S| \leq k + 8d^2$, such that $\sum_{i \in S} w^j_{\pi(i)} \leq |S|\mu^j$. It follows that for some $S' \subseteq S$ with $|S'| = k$ we have $\sum_{i \in S'} w^j_{\pi(i)} \leq k\mu^j$. Since $|S' \oplus \{1, \ldots, k\}| \leq 16d^2$ this completes the proof. $\square$