Random Walks on Truncated Cubes and Sampling 0-1 Knapsack Solutions *

BEN MORRIS[†] ALISTAIR SINCLAIR[‡]

July 12, 2002

Abstract

We solve an open problem concerning the mixing time of symmetric random walk on the n-dimensional cube truncated by a hyperplane, showing that it 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. The results extend to the case of any fixed number of hyperplanes. The key ingredient in our analysis is a combinatorial construction we call a "balanced almost uniform permutation," which seems to be of independent interest.

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

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

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 $\mathbf{a} \cdot \mathbf{x} = b$. 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 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., $a(X) = \sum_{i \in X} a_i$.

This paper is concerned with the problem of computing $|\Omega|$, i.e., counting the number of feasible solutions to the knapsack problem. The problem is #P-complete in exact form, so we aim for a good approximation algorithm, specifically a *fully-polynomial randomized approximation scheme* (fpras). By a well-known relationship based on self-reducibility [12, 11], this is equivalent to constructing a polynomial time algorithm for sampling elements of Ω (almost) uniformly at random.

In recent years there has been a steady stream of results of this kind for #P-complete counting problems (see, for example, [13, 11, 8] for surveys); however, the 0-1 knapsack problem still stands

*A preliminary version of this paper appeared in *Proceedings of the 40th IEEE Symposium on Foundations of Computer Science*, October 1999, pp. 230–240.

[†]Department of Statistics, Evans Hall, University of California, Berkeley CA 94720. Email: morris@stat.berkeley.edu. Supported by NSF graduate and post-doctoral fellowships and by NSF grant ECS-9873086.

[‡]Computer Science Division, Soda Hall, University of California, Berkeley CA 94720-1776. Email: sinclair@cs.berkeley.edu. Supported in part by NSF grants CCR-9505448 and CCR-9820951.

as one of a small handful of canonical problems that have so far resisted attack. Indeed, it has been quoted as an open problem in several places [4, 11, 13, 15]. 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 paper, we resolve this issue by constructing an fpras for the 0-1 knapsack problem. 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.

Almost all known approximate counting algorithms proceed by simulating a suitable random walk on the set of interest Ω. The walk is constructed so that it converges to the uniform distribution over Ω; simulation of the walk for sufficiently many steps therefore allows one to sample (almost) uniformly from Ω, and thus to approximate $|\Omega|$. In any application of this method, the key step is to establish that the random walk is *rapidly mixing*, i.e., gets close to the uniform distribution after a polynomial number of steps.

In the case of 0-1 knapsack solutions, a particularly simple and natural random walk on Ω has been proposed. 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.

This process may equivalently be viewed as a nearest neighbor random walk on the portion of the cube $\{0,1\}^n$ truncated by the hyperplane $\mathbf{a} \cdot \mathbf{x} = b$, in which the probability of moving to any neighbor is $\frac{1}{n}$; we will call this graph G_Ω. 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.

It is easy to check that this random walk converges to the uniform distribution over Ω. However, despite much recent activity in the analysis of mixing times of random walks, this deceptively simple example is 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 remains $\exp(\mathrm{O}(\sqrt{n}(\log n)^{5/2}))$ [4], which beats the trivial bound of $\exp(\mathrm{O}(n))$ but is still exponential.

In this paper we prove that the above random walk is indeed rapidly mixing, with a mixing time of $\mathrm{O}(n^{9/2+\epsilon})$ steps for any $\epsilon > 0$. This immediately implies the existence of an fpras for counting 0-1 knapsack solutions.

We also present a non-trivial extension of these results to the case of multiple hyperplanes (more precisely, multiple constraints of the form $\mathbf{a}_j \cdot \mathbf{x} \leq b_j$ for non-negative vectors $\mathbf{a}_j$).[1] Here we are also able to prove a mixing time of $\mathrm{O}(n^c)$ (where c is a constant) for any fixed number of hyperplanes. (The exponent c depends on the number d of hyperplanes, but this is inevitable as it is not hard to prove a lower bound of $n^{\Omega(d)}$ on the mixing time. Moreover, it is possible to encode NP-hard problems if the number of hyperplanes is permitted to depend on n, so we would not expect any polynomial time sampling algorithm for this case.)

To prove rapid mixing we use a technique based on multicommodity flow (see [16]): 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., [9, 10]) have made use of "degenerate" flows in which all $X \to Y$

[1] We mention in passing that all our results extend from the 0-1 case to more general cubes of the form $[0, \ldots, L]^n$. This extension is purely technical and does not require any substantial new ideas, so we omit the details.

flow is routed along a single canonical path (though see [1, 16] 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. Slightly less obviously, a symmetric problem arises because a random permutation will tend to remove too many items at some intermediate point, causing congestion among edges of the hypercube near the origin. To avoid these problems, we want our permutations to remain "balanced," in the sense that items are added and removed at approximately the correct rates throughout the path; but we also want them to be "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. A main contribution of this paper is to show the existence of such permutations.

The remainder of this paper is structured as follows. We begin with some necessary background on flows and rapid mixing in section 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 3 we show how to construct balanced almost uniform permutations, and in section 4 how to use these to define a good flow. We then extend everything to the general case in section 6. The extension to multiple constraints is handled in section 6; this involves extending our construction of balanced almost uniform permutations from scalar weights to vectors in arbitrary dimension. This again may be of independent interest.

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 $X \subseteq \{1, \ldots, n\}$, 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 $\Omega = \{X : a(X) \le b\}$. Without loss of generality, we will assume that $a_i \le b$ for all i.

We consider the symmetric random walk on the portion G_Ω of the hypercube $\{0,1\}^n$ defined in the Introduction. 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

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

standard facts about geometric convergence, $\mathrm{O}(\tau_{\mathrm{mix}} \log \epsilon^{-1})$ steps suffice to reduce the variation distance to any desired ϵ.

Fairly standard techniques (see [16]) allow us to estimate τ_{mix} by setting up a suitable multi-commodity 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; i.e., $f(e)$ is the sum over all ordered pairs X, Y of the $X \to Y$ flow carried by 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[3] is a special case of results in [16] (see also [3, 2]):

Theorem 2.1 [16] *For any flow f, the mixing time is bounded by $\tau_{\mathrm{mix}} \leq 4n(n+1)\mathcal{C}(f)\mathcal{L}(f)$.*

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 $\tau_{\mathrm{mix}} = n^{\mathrm{O}(B^2)}$ in this case. By introducing some additional technical complications, we will go on to get a uniform bound of $\tau_{\mathrm{mix}} = \mathrm{O}(n^{9/2+\epsilon})$ for the general case, for any $\epsilon > 0$.

Remark: We note that our bound on the mixing time is only slightly larger than the upper bound of $\mathrm{O}(n^3)$ which one obtains by applying Theorem 2.1 to the hypercube itself (without the hyperplane constraint): see, e.g., [17]. This is in turn somewhat off from the true mixing time of $\mathrm{O}(n \log n)$. On the other hand, it is fairly easy to obtain a lower bound of $\Omega(n^2/\log n)$ for the mixing time of the truncated cube: consider, for example, an instance in which $\log n$ items have weight 1, the other $n - \log n$ items have weight n, and the knapsack capacity is $b = n$. $\square$

As explained in the Introduction, 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 4.

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 $\{1, \ldots, m\}$.

Definition 3.1 *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 $1 \leq k \leq m$,*

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

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

Definition 3.2 *Let π be a random permutation in $\mathcal{S}_m$, and let $\lambda \in \mathbf{R}$. We call π a* λ-uniform *permutation if*

$$\Pr[\pi\{1, \ldots, k\} = U] \leq \lambda \times \binom{m}{k}^{-1} \tag{2}$$

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 $\{\pi(1), \ldots, \pi(k)\}$.)

[3] We note that this theorem applies to symmetric random walk on *any* connected subgraph of the hypercube $\{0,1\}^n$, in which transitions are made to each neighbor with probability $\frac{1}{2n}$.

Note that, if π were a uniform random permutation, the probability in (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 guaranteed to be balanced. In section 5.1 we will show how to dispense with any restrictions on the weights.

Theorem 3.3 *Let $\{w_i\}_{i=1}^m$ be any set of weights with $|w_i| \in [1, B]$ for a constant $B > 1$. 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.[4] We call ν_1 *α-good* if, for every k_1 with $1 \leq k_1 \leq m_1$,

$$\Big|\sum_{i=1}^{k_1} w_{\nu_1(i)} - k_1\mu_1\Big| \leq \alpha(M-1)\sqrt{k_1^*}, \tag{3}$$

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.1.1 in the Appendix) yields

$$\Pr[\nu \text{ is not } \alpha\text{-good}] \leq 2m\exp(-2\alpha^2). \tag{4}$$

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

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

$$\Pr[\mathcal{E}] \leq 2\Pr_{\text{unif}}[\mathcal{E}]. \tag{5}$$

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 $\sum_{i=1}^{k-1} w_{\pi(i)} \geq 0$, 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 $|w_i| \leq M$ for all i it is clear that π satisfies the balance condition (1).

We now need to verify the uniformity condition (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

$$\begin{aligned}
\Pr[\pi\{1,\ldots,k\} = U] &\leq \Pr[\nu_1\{1,\ldots,k_1\} = U_1 \text{ and } \nu_2\{1,\ldots,k_2\} = U_2] \\
&\leq 2\Pr_{\text{unif}}[\nu_1\{1,\ldots,k_1\} = U_1 \text{ and } \nu_2\{1,\ldots,k_2\} = U_2] \\
&= \frac{2}{\binom{m_1}{k_1}\binom{m_2}{k_2}}, \qquad (6)
\end{aligned}$$

[4] 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 the second inequality follows from (5). Now some routine calculations involving Stirling's formula (see Lemma A.1.2 in the Appendix) 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, (6) becomes

$$\Pr[\pi\{1,\ldots,k\} = U] \le \frac{Cm^{1/2}}{\binom{m}{k}} \exp\left\{\frac{l^2 + \frac{1}{2}|l|}{\alpha(1-\alpha)}\Big(\frac{1}{m_1} + \frac{1}{m_2}\Big)\right\}, \tag{7}$$

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

$$l = (k_2\mu_2 - k_1\mu_1)\frac{m_2}{\mu_1 m}, \tag{8}$$

since $\frac{m_2}{m_1} = \frac{\mu_1}{\mu_2}$. Now by the goodness condition (3) on ν_1, ν_2 we have

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

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 (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 (7) is bounded above, for sufficiently large m, by

$$\begin{aligned} & 5(B-1)^2 k^* \ln m \, \frac{m_2^2}{\mu_1^2 m^2} \frac{m^2}{k(m-k)} \frac{m}{m_1 m_2} \\ = \;& 5(B-1)^2 \ln m \frac{k^* m}{k(m-k)} \frac{1}{\mu_1\mu_2} \\ \le \;& 10(B-1)^2 \ln m, \end{aligned} \tag{9}$$

since $k(m-k) \ge \frac{k^* m}{2}$ and $\mu_1, \mu_2 \ge 1$. Thus (7) becomes

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

which verifies the uniformity condition (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 $d \le m$. 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

$$\Pr[\pi\{1, \ldots, k\} = U] \leq \binom{m}{k}^{-1} \mathrm{poly}(m). \tag{11}$$

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

$$\mathrm{poly}(m+d) \sum_S \Pr_{\mathrm{unif}}[\mathcal{E}_S], \tag{12}$$

where $\Pr_{\mathrm{unif}}$ denotes probability under the uniform distribution on permutations in $\mathcal{S}_{m+d}$. But the sum in (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

$$\mathrm{E}(X) = \binom{m}{k}^{-1} \left(1 + \tfrac{d}{m+1}\right). \tag{13}$$

This will complete the verification of condition (11); for replacing the sum in (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 (13), let $\mathcal{E}$ be the event that $\pi\{1, \ldots, k\} = U$. Clearly $\Pr_{\mathrm{unif}}[\mathcal{E}] = \binom{m}{k}^{-1}$, and $X = 0$ unless $\mathcal{E}$ occurs, so we have

$$\mathrm{E}(X) = \binom{m}{k}^{-1} \mathrm{E}(X|\mathcal{E}). \tag{14}$$

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 (14) we get the value claimed in (13), which concludes the proof that π is almost uniform. $\square$

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 (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 (10) independent of B.

Now, following through the algebra in the second part of the proof, starting from equation (10), and noting that $d \leq m$, it is not hard to check 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

$$\Pr[\pi\{1,\ldots,k\} = U \text{ and } \pi(k+1) = l] \;\leq\; C'_B m^{10(B-1)^2+1/2} \times \binom{m}{k \;\; m-k-1 \;\; 1}^{-1}, \qquad (15)$$

for any U with $|U| = k$ and any $l \notin U$, where $C'_B = C_B(B+1)$. (To get this value for C'_B, note that π must first choose U and then l; this second choice introduces the factor $B+1$.) We will make use of these facts in section 5 when we discuss permutations of general weights. □

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 the Introduction, making heavy use of the balanced almost uniform permutations from section 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 $S = X_1 \oplus Y_1$ (where $\oplus$ denotes symmetric difference) 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

$$\begin{aligned} a(Z) &= a(X_1) + \sum_{i=1}^{k} w_{\pi(i)} \\ &\le a(X_1) + \max\{a(Y_1) - a(X_1), 0\} + B \qquad (16) \\ &= \max\{a(Y_1), a(X_1)\} + B \\ &\le b, \end{aligned}$$

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. $\mathcal{L}(f)$, 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 Bn^B, so the contribution to $f(Z)$ from such paths is no more than $Bn^B|\Omega|$. 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 $B^2 n^{2B}$ 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., [9, 10]), 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,

$$Z' = X_1 \oplus Y_1 \oplus Z.$$

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 (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 processed (i.e., added or deleted) by the stage-2 path by the time it reaches Z. 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

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

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:

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

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

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 bounded polynomially 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), which is of the form $n^{\mathrm{O}(B^2)}$ (see the Remark at the end of section 3).

We summarize our analysis in the following theorem.

Theorem 4.1 *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 the Introduction, this immediately yields an fpras for computing $|\Omega|$ in this case, via a standard reduction to random sampling (whose details are spelled out in [11]).

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.

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

$$\min\{W,0\} - \Delta M \le \sum_{i=1}^{k} w_{\pi(i)} \le \max\{W,0\} + \Delta M. \tag{17}$$

Our earlier definition (Definition 3.1) thus corresponds to $\Delta = 1$.

Relaxing our earlier terminology slightly, we shall call $\pi \in \mathcal{S}_m$ a "balanced almost uniform permutation" if π is Δ-balanced for a fixed constant Δ, and λ-balanced for λ a fixed polynomial function of m. The following theorem is a generalization of Theorem 3.3; it says that we can construct a balanced almost uniform permutation for an arbitrary set of weights. Moreover, 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 Δ. 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 Δ, C and $p < 1/2$, there can be no Δ-balanced Cm^p-uniform permutation if m is sufficiently large.

For technical reasons, we shall actually prove a slightly stronger uniformity property. Call π *strongly λ-uniform* if

$$\Pr[\pi\{1,\ldots,k\} = U \text{ and } \pi(k+1) = l] \le \lambda \times \binom{m}{k \quad m-k-1 \quad 1}^{-1} \tag{18}$$

for every k with $1 \le k \le m$, every $U \subseteq \{1,\ldots,m\}$ of cardinality k, and every $l \notin U$. Note that the expression on the right-hand side of (18) is just λ times the probability of the given event if π were chosen uniformly at random. Plainly (18) is a strengthening of equation (2) in Definition 3.2; recall from equation (15) that our permutations in the previous section also had this stronger property.

Theorem 5.2 *Fix $0 < \epsilon < 1$ and let $\Delta = 1 + \sqrt{90/\epsilon}$. For any m and set of weights $\{w_i\}_{i=1}^m$, there exists a Δ–balanced strongly $Cm^{1/2+\epsilon}$–uniform permutation, where C is a universal constant.*

Proof: Let $M = \max_i |w_i|$ and set $\widehat{\Delta} = \frac{\Delta-1}{3}$. Let β be a uniform random permutation in $\mathcal{S}_m$. Let T_1 be the smallest t such that the partial sum $\sum_{i=1}^t w_{\beta(i)}$ has absolute value greater than $\widehat{\Delta}M$ (or $T_1 = m$ if no such t exists). Similarly, let T_2 be the smallest $t > T_1$ such that $|\sum_{i=T_1+1}^t w_{\beta(i)}| > \widehat{\Delta}M$. 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 in this way, dividing β into *intervals* $I_1, \ldots, I_q$ (so that $T_q = m$).

Now let α_i be the aggregated weight of interval I_i for $i = 1, 2, \ldots, q-1$. Note that $|\alpha_i| \in [\widehat{\Delta}M, (\widehat{\Delta}+1)M]$ for all $i < q$, so the ratio of the weights of any two of these intervals is at most $(\widehat{\Delta}+1)/\widehat{\Delta}$. Thus, by the results of section 3, there exists a 1-balanced λ-uniform permutation on $\{\alpha_i\}_{i=1}^{q-1}$ for $\lambda = Cq^{10\left((\widehat{\Delta}+1)/\widehat{\Delta}-1\right)^2+1/2} = Cq^{1/2+\epsilon}$. By the Remark at the end of that section, we can in fact assume that this permutation is *strongly* λ-uniform and (since $(\widehat{\Delta}+1)/\widehat{\Delta}$ is bounded above by a constant, namely $1+\sqrt{1/10}$) that the constant C does not depend on ϵ. 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 Δ-balanced $Cm^{1/2+\epsilon}$–uniform permutation on the original m weights.

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

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

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

$$\min\{0, W'\} - (2\widehat{\Delta}+1)M \leq \sum_{i=1}^{j} w_{\pi(i)} \leq \max\{0, W'\} + (2\widehat{\Delta}+1)M,$$

since the partial sums within any interval lie in the range $[-\widehat{\Delta}M, \widehat{\Delta}M]$. Finally, note that $|W - W'| = |\alpha_q| \leq \widehat{\Delta}M$. It follows that for all j,

$$\min\{0, W\} - (3\widehat{\Delta}+1)M \leq \sum_{i=1}^{j} w_{\pi(i)} \leq \max\{0, W\} + (3\widehat{\Delta}+1)M,$$

and hence π is Δ-balanced.

To verify the strong 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 and any index $l \notin U$, the likelihood ratio

$$\frac{\Pr[\pi\{1, \ldots, k\} = U \text{ and } \pi(k+1) = l]}{\Pr_{\text{unif}}[\pi\{1, \ldots, k\} = U \text{ and } \pi(k+1) = l]} \leq Cm^{1/2+\epsilon},$$

where we write $\Pr_{\text{unif}}$ for the probability when π_I is uniform and Pr for the probability when π_I is $Cm^{1/2+\epsilon}$–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, only the interval containing l can contain elements from both U and U^c (the complement of U). Additionally, in the interval containing l, all the elements before l must be in U and all those after l must be in U^c. Let A_1 be the collection of intervals in $\{I_i\}_{i=1}^{q-1}$ containing only elements of U, and A_2 the collection of intervals containing only elements of U^c. Then $|A_1|+|A_2|$ must have value either $q-1$ or $q-2$. 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 above likelihood ratio is

$$\frac{\Pr[\mathcal{E}_1 \text{ and } \mathcal{E}_2]}{\Pr_{\text{unif}}[\mathcal{E}_1 \text{ and } \mathcal{E}_2]} \leq Cq^{1/2+\epsilon} \leq Cm^{1/2+\epsilon}.$$

In the case where $|A_1|+|A_2| = q-1$ this is just the $Cq^{1/2+\epsilon}$–uniformity property; when $|A_1|+|A_2| = q-2$ it is the strong $Cq^{1/2+\epsilon}$–uniformity property. Thus π is strongly $Cm^{1/2+\epsilon}$–uniform, and the proof is complete. $\square$

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 4 for constructing a good flow in the general case. Our goal will be to obtain a flow f of cost $\mathcal{C}(f) = \mathrm{O}(n^{3/2+\epsilon})$ for any $\epsilon > 0$. So we assume from now on that ϵ is arbitrary but fixed.

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 4: we essentially followed a balanced almost uniform permutation of $X \oplus Y$, except that we removed a constant number of items 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 basically the same, except that we will now remove a fixed number of items from $X \cup Y$ and add/delete these repeatedly along the path to maintain fine balance. Moreover, before applying the random permutation, we first need to "pre-process" the pair (X, Y) so that neither X nor Y is too close to the hyperplane: in contrast to the bounded ratio case, this is not guaranteed by the removal of a fixed number of items because of the possibly large variations in weights. However, we can overcome this obstacle by randomly switching items between X and Y to roughly balance their weights. The resulting flow-carrying paths will not in general be geodesics, as before, though they will have length only $\mathrm{O}(n)$.

In preparation for describing the flow, we first describe the pre-processing operation. We assume that $a(X) + a(Y) \le 2b - 6\Delta M$, where $M = \max_{i \in X \oplus Y} a_i$ and $\Delta = \Delta(\epsilon)$ is the constant appearing in Theorem 5.2. (In our application, we will reduce to this case by deleting a fixed number of items from $X \cup Y$.) Call a *pair* of vertices (X', Y') *full* if either $a(X') > b - \Delta M$ or $a(Y') > b - \Delta M$. Our goal is to shift items randomly between X and Y and thereby reach a pair (X', Y') that is not full.

Consider the following random walk on $\{(X', Y') : X' \cup Y' = X \cup Y,\ X' \cap Y' = X \cap Y,\ a(X'), a(Y') \le b\}$. If the current state is (X', Y'), choose an index $i \in X' \oplus Y'$ u.a.r. With probability $\frac{1}{2}$, do nothing; else move i from X' to Y' or Y' to X' if possible. We call this the "pre-processing random walk" (PRW). We claim in the following lemma that, if we run the PRW for a number of steps chosen randomly between 1 and $\mathrm{O}(n)$, we will with reasonable probability arrive at a pair (X', Y') that is not full. The proof uses a martingale argument and is deferred to the Appendix.

Lemma 5.3 *Let (X, Y) be a full pair of vertices in G_Ω with $a(X) + a(Y) \le 2b - 6\Delta M$, where $M = \max_{i \in X \oplus Y} a_i$. Pick T u.a.r. from $\{1, 2, \ldots, C_1 m\}$, where $m = |X \oplus Y|$ and C_1 is a suitable constant (which depends only on Δ), and let (X', Y') be the result of running the PRW for T steps starting from (X, Y). Then $\Pr[(X', Y') \text{ is not full}] \ge 1/C_2$ for a positive constant C_2 (again depending only on Δ).*

We are now ready to construct and analyze the flow in the general case.

Lemma 5.4 *For arbitrary weights and any $\epsilon > 0$, it is possible to construct a multicommodity flow f in G_Ω with $\mathcal{C}(f) = \mathrm{O}(n^{3/2+\epsilon})$ and $\mathcal{L}(f) = \mathrm{O}(n)$.*

Proof: Let X, Y be arbitrary vertices of G_Ω. Viewing X and Y as subsets of $\{1, \ldots, n\}$, let H be the $h = \lceil 6\Delta \rceil$ elements of $X \oplus Y$ having largest weight (or let $H = X \oplus Y$ if $|X \oplus Y| \le h$), with ties broken according to index order. Define $X' = X - H$, $Y' = Y - H$ and $S = X' \oplus Y'$. Let $m = |S|$, and let $\{w_i\}_{i=1}^m$ be an arbitrary enumeration of the weights of the items in S, with the weights of items in X', Y' appearing with negative and positive signs respectively. Let $M = \max_i |w_i|$.

We will say that a set of indices Z is *good* if $Z - H \in \Omega$ and $(Z \oplus X \oplus Y) - H \in \Omega$. For a set of indices Z and an index i, define

$$Zi = \begin{cases} Z \oplus \{i\} & \text{if } Z \oplus \{i\} \text{ is good;} \\ Z & \text{otherwise.} \end{cases}$$

Define $Zi_1 i_2 = ((Zi_1)i_2)$ and so on. Note that if $Y = Xi_1 \cdots i_l$ then the sequence $i_1, \ldots, i_l$ defines a path from X to Y in the unit hypercube of length at most l. This path need not in general lie within G_Ω; however, it is "close to" G_Ω in the sense that for every point Z of the path, $Z - H \in \Omega$.

If (X', Y') is not full then set $T = 0$, otherwise choose T u.a.r. from $\{1, \ldots, C_1 m\}$, where C_1 is the constant in Lemma 5.3. Next, let $i_1, \ldots, i_T$ be i.i.d. uniform over S. Define $X'' = X' i_1 \cdots i_T$ and $Y'' = X'' \oplus X' \oplus Y' = Y' i_1 \cdots i_T$. Thus (X'', Y'') is the result of running the PRW for T steps starting from (X', Y'). Note that $a(X') + a(Y') \leq 2b - a(H) \leq 2b - 6\Delta M$. So, by Lemma 5.3, we can condition on the event that the pair (X'', Y'') is not full and thus increase the probability of any path by a factor of at most C_2.

Now let π be a Δ-balanced, strongly $Cm^{1/2+\epsilon}$–uniform permutation on the weights $\{w_i\}$, whose existence is guaranteed by Theorem 5.2. We claim that the sequence

$$i_1, \ldots, i_T, \pi(1), \ldots, \pi(m), i_T \ldots, i_1 \tag{19}$$

defines a path from X to Y in the hypercube. This is true because the condition that (X'', Y'') be not full, together with the fact that π is balanced, guarantees that all of the transitions indicated by π will actually take place.

Set

$$j_k = \begin{cases} i_k & \text{if } 1 \leq k \leq T; \\ \pi(k - T) & \text{if } T < k \leq T + m; \\ i_{2T+m-k-1} & \text{if } T + m < k \leq 2T + m, \end{cases}$$

and let $l = 2T + m$. Then $j_1, \ldots, j_l$ is the sequence in (19). Our flow from X to Y will essentially follow the sequence j_k, 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 new items j_k to be added. Thus each intermediate state Z will be of the form $\overline{H} \oplus Xj_1 \cdots j_k$, for some $\overline{H} \subseteq H$ and $k \leq l$.

Suppose that, after processing the first $k \leq l$ elements of the sequence in (19), we have $Z = \overline{H} \oplus Xj_1 \cdots j_k$ for some $\overline{H} \subseteq H$. The transition rule will be as follows.

1. If $k < l$ and $j_{k+1} \notin Z$ then move to $Z\,j_{k+1}$ if possible (i.e., if the result is an element of Ω); otherwise delete an element from H.

2. If $k < l$ and $j_{k+1} \in Z$ then add an element from H if possible; otherwise move to $Z\,j_{k+1}$.

3. If $k = l$ then add an element from $H \cap Y$ if possible; otherwise delete an element from $H \cap X$.

The fact that all of the sets $X\,j_1 \cdots j_k$ are good ensures that sufficient elements of H can always be removed so as to make room to add the next element j_{k+1} when necessary; hence the above rule defines a feasible random path from X to Y. Similarly, goodness also implies that $a(Z \oplus X \oplus Y - H) \leq b$ for every intermediate state Z; since our rule keeps the weight as large as possible this implies that, at any intermediate edge (Z, W) along the path, there exists (at most) one element $u \in H$ such that

$$a(Z \oplus X \oplus Y - \{u, z\}) \leq b, \tag{20}$$

where z is the index such that $\{z\} = Z \oplus W$. Then $(Z' - \{u, z\}) \in \Omega$, where exactly as in the analysis in section 4, we define the "encoding" Z' by

$$Z' = X \oplus Y \oplus Z.$$

Thus, for any given edge (Z, W), the number of encodings Z' is at most $n|\Omega|$.

Note that the path from X to Y can be naturally divided into three *stages*, corresponding to the three parts of the sequence j_k. We will write the flow through any given edge $(Z, W) \in G_\Omega$ as $f(Z, W) = f_1(Z, W) + f_2(Z, W) + f_3(Z, W)$, where $f_i(Z, W)$ is the contribution of stage i paths. We will bound f by bounding each of the three contributions f_i separately.

Consider stage 1 first, and focus on a particular edge (Z, W). For any pair (X, Y) that sends flow through (Z, W) in stage 1, we can write $Z = \overline{H} \oplus X j_1 \ldots j_k$, where $j_1, \ldots, j_k$ are the first k elements processed along the path. Thus the pair (X, Y) is completely specified by $k, j_1, \ldots, j_k, Z'$ and $\overline{H}$, via the easily verified relations

$$X = \overline{H} \oplus Z j_k \cdots j_1\,; \qquad Y = \overline{H} \oplus Z' j_k \cdots j_1.$$

The amount of flow corresponding to a given sequence $j_1, \ldots, j_k$ is bounded above by the probability that $j_1, \ldots, j_k, z$ were the first $k+1$ indices chosen in the pre-processing random walk, which is at most $C_2 m^{-(k+1)}$. (The factor C_2 here arises from our earlier conditioning on the event that (X'', Y'') is not full.) Thus we can bound the stage-1 flow $f_1(Z, W)$ as in section 4. We have

$$\begin{aligned} f_1(Z, W) &\le \sum_{Z'} \sum_k \sum_{j_1, \ldots, j_k} \sum_{\overline{H}} C_2 m^{-(k+1)} \\ &\le \sum_{Z'} \sum_k 2^h C_2 m^{-1} \\ &\le \sum_{Z'} C_1 m 2^h C_2 m^{-1} \\ &\le C_1 C_2 2^h n |\Omega|, \end{aligned}$$

where the factors $C_1 m$ and 2^h arise from summing over k and $\overline{H}$ respectively.

The flow $f_3(Z, W)$ from stage-3 paths can be handled symmetrically, so consider now the stage-2 paths. For a given edge (Z, W), every pair (X, Y) that sends flow through (Z, W) in stage 2 can be completely specified by $Z', T, k, j_1, \ldots, j_T, U$ and $\overline{H}$, where k is the number of elements of the sequence in (19) processed along the path from X to Z and $U = \{\pi(1), \ldots, \pi(k-T)\}$, via

$$X = \overline{H} \oplus (Z \oplus U) j_T \cdots j_1\,; \qquad Y = \overline{H} \oplus (Z' \oplus U) j_T \cdots j_1.$$

Let $k' = k - T$. The amount of flow corresponding to a given $j_1, \ldots, j_T$ and U is bounded above by

$$(C_2 m^{-T})(C_1 m)^{-1} \left[C m^{1/2+\epsilon} \binom{m}{k', m-k'-1, 1}^{-1} \right],$$

where the first factor comes from the pre-processing random walk, the second factor is the probability of choosing a particular T, and the third factor is an upper bound on the probability $\Pr[\pi\{1, \ldots, k'\} = U \text{ and } \pi(k'+1) = z]$, which comes from the strong almost uniformity of π. Thus we can again bound the flow $f_2(Z, W)$ as in section 4. We have

$$f_2(Z, W) \le \sum_{Z'} \sum_T \sum_k \sum_{j_1, \ldots, j_T} \sum_U \sum_{\overline{H}} (C_2 m^{-T})(C_1 m)^{-1} \left[C m^{1/2+\epsilon} \binom{m}{k', m-k'-1, 1}^{-1} \right]$$

$$
\begin{aligned}
&\le \sum_{Z'} (C_1 m) m\, m^T \binom{m-1}{k'} 2^h (C_2 m^{-T})(C_1 m)^{-1} \left[C m^{1/2+\epsilon} \binom{m}{k', m-k'-1, 1}^{-1} \right] \\
&= \sum_{Z'} 2^h C_2 \left[m \binom{m-1}{k'} \binom{m}{k', m-k'-1, 1}^{-1} \right] C m^{1/2+\epsilon} \\
&= \sum_{Z'} 2^h C_2 C m^{1/2+\epsilon} \\
&\le 2^h C C_2 n^{3/2+\epsilon} |\Omega|,
\end{aligned}
$$

where the factors in the second line are written in the same order as the sums they arise from.

Adding the contributions f_1, f_2 and f_3, we see that the above flow satisfies $\mathcal{C}(f) = \mathrm{O}(n^{3/2+\epsilon})$, while plainly $\mathcal{L}(f) = \mathrm{O}(n)$. Since $\epsilon > 0$ was arbitrary, this completes the proof. □

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

Theorem 5.5 *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 $\tau_{\mathrm{mix}} = \mathrm{O}(n^{9/2+\epsilon})$ for any $\epsilon > 0$.*

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

Remark: The mixing time bound of $\mathrm{O}(n^{9/2+\epsilon})$ in Theorem 5.5 is reasonably tight for this type of analysis. If we apply Theorem 2.1 to analyze random walk on the entire cube $\{0,1\}^n$, we get a bound of $\mathrm{O}(n^3)$ even with an optimal flow. Thus the truncation introduces an extra factor of only $\mathrm{O}(n^{3/2+\epsilon})$ into the bound. It is instructive to see where this extra factor comes from: $\mathrm{O}(n^{1/2+\epsilon})$ is due to the balanced almost uniform permutation construction (Theorem 5.2, which is tight), while $\mathrm{O}(n)$ comes from the fact that the "encoding" Z' may lie just outside Ω. □

6 Multiple hyperplanes

6.1 Introduction

In this section, we will extend our earlier results to handle multiple hyperplanes. For a non-negative 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 $A\mathbf{x} \le \mathbf{b}$. The vertices in Ω constitute the set of feasible solutions to the multidimensional knapsack problem with the d simultaneous constraints

$$
\mathbf{a}^j \cdot \mathbf{x} \equiv \sum_{i=1}^{n} a_i^j x_i \le b^j \quad \text{for } 1 \le j \le d, \tag{21}
$$

where $a_i^j \equiv a_{ji}$. (In equation (21) the superscript j indexes the jth linear constraint; we will follow this convention throughout.)

Geometrically, Ω is obtained by truncating the unit cube 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 section will easily extend to any collection of hyperplanes with this property[5].

[5]However, we cannot allow the hyperplanes to be arbitrary. If arbitrary truncations were allowed, then it would be possible to use just two hyperplanes to cause exponential bottlenecks or even disconnect the graph G_Ω.

Following our earlier notation, we identify a 0-1 vector $\mathbf{x} = (x_i)_{i=1}^n$ with the set of indices $X = \{i : x_i = 1\}$, and write $a(X) = (a^1(X), \ldots, a^d(X))$ for the (now d-dimensional) weight of X. As before we denote by G_Ω the subgraph of the hypercube $\{0,1\}^n$ induced by the vertices in Ω, and we again study symmetric random walk on G_Ω; i.e., transitions from a given state $X \subseteq \{1, \ldots, n\}$ are made as follows:

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

Again, to avoid issues involving periodicity, we add to every state a holding probability of $\frac{1}{2}$.

In this section we will prove that this random walk on G_Ω has mixing time that is polynomially bounded in n, for any fixed dimension d. Just as in the one-dimensional case, this immediately gives a polynomial time algorithm for sampling (almost) uniformly at random from Ω, and a fpras for computing $|\Omega|$.

We note that the degree of our polynomial upper bound for the mixing time will depend on the dimension d, but this is unavoidable as the following simple example shows. Consider a d-dimensional knapsack problem in which there are $\frac{n}{2d}$ items having each of the d weight vectors $(n, 0, \ldots, 0), (0, n, \ldots, 0), \ldots, (0, \ldots, 0, n)$, and the remaining $\frac{n}{2}$ items have weight vector $(1, 1, \ldots, 1)$; the knapsack capacity is $\mathbf{b} = (n, \ldots, n)$. Let S be the set of feasible solutions in Ω which do not 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 easily that the mixing time is $n^{\Omega(d)}$.

In fact, for arbitrary d there can be no uniform polynomial upper bound for the running time of *any* sampling algorithm unless RP = NP. This follows immediately by reduction from the problem of sampling independent sets in a graph. By theorem 1.17 of [17], there is no algorithm for (almost) uniformly sampling independent sets in a graph unless RP = NP. Now if $G = (V, E)$ is an arbitrary (undirected) graph, 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 $x_u + x_v \leq 1$ for all $\{u, v\} \in E$.

To prove rapid mixing of the random walk on G_Ω for any fixed d, we use the multicommodity flow technique as before. Recall that Theorem 2.1, which bounds the mixing time in terms of the cost of a flow f, holds for symmetric random walk on any connected subset of the hypercube, so it again suffices to come up with a flow of small cost. As before, the idea is to spread each $X \to Y$ 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.

Definition 6.1 *Fix an integer $d > 0$, and let $\{w_i\}_{i=1}^m$ be a set of weights in $\mathbf{R}^d$ satisfying $\sum_{i=1}^m w_i = 0$. For a positive real number Δ, a permutation $\pi \in \mathcal{S}_m$ is* Δ-balanced *if*

$$\max_k \Big| \sum_{i=1}^k w^j_{\pi(i)} \Big| \leq \Delta M^j \quad \text{for } 1 \leq j \leq d, \tag{22}$$

where $w_i = (w_i^1, \ldots, w_i^d)$ and $M^j = \max_{1 \leq i \leq m} |w_i^j|$

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 (22). Note that this generalizes our earlier Definition 5.1 for the one-dimensional case (except that, for simplicity, we have assumed that $\sum_i w_i = 0$).

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 for an arbitrary set of vector weights that even a *single* balanced permutation exists. (For $d = 1$ of course this is trivial.) The existence of such a permutation follows at once from a lemma due to Grinberg and Sevast'yanov [6], which was proved in an entirely different context:

Lemma 6.2 [6] *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*

$$\sum_{i=1}^{k} x_{\nu(i)} \in d \times \mathrm{conv}\{x_1, \ldots, x_n\} \quad \text{for } 1 \leq k \leq n,$$

where conv *denotes the convex hull.*

Of course, we need something much stronger than this, namely *almost uniform* permutations with a similar balance property. Perhaps surprisingly, we will show that balanced almost uniform permutations exist in arbitrary dimension d. To illustrate the main ideas involved in extending from one to higher dimensions, we 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.

In this setting, let $I_1 = \{i : w_i^2 \geq 0\}$, $I_2 = \{i : w_i^2 < 0\}$, and define $v = \sum_{i \in I_1} w_i$. For every $i \leq m$, let y_i be the projection of w_i onto $v^\perp$. 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 $\{w_i^2\}_{i=1}^m$ (the projections of the w_i onto the second coordinate axis). Since π_1 and π_2 are both almost uniform, so is π, by an argument similar to that in the proof of Theorem 3.3.

Furthermore, since π_1 and π_2 are each balanced with respect to projections onto $v^\perp$, 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 the second coordinate axis and onto $v^\perp$ are both bounded, and since the w_i^j are all in $[1, 2]$, the angle between the coordinate axis and $v^\perp$ is bounded away from zero. Thus, the partial sums $\sum_{i=1}^k w_{\pi(i)}$ stay inside a parallelogram of bounded diameter. Hence π is balanced with respect to the weights $\{w_i\}_{i=1}^m$.

This concludes the sketch proof for the above special case with $d = 2$. Note that it is a straightforward reduction to the one-dimensional result. Unfortunately, in general the reduction from d to $d-1$ dimensions is not quite so straightforward; we deal with the extra technical difficulties in the next subsection.

6.2 Balanced almost uniform permutations in arbitrary dimensions

The following theorem says that one can always construct balanced almost uniform permutations when the dimension d is fixed.

Theorem 6.3 *Let d be any positive integer. There is a constant c_d and a polynomial function p_d such that, for any set of weights $\{w_i\}_{i=1}^m$ in $\mathbf{R}^d$ with $\sum_i w_i = 0$, 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 5.2, with $c_1 = 15$ and $p_1(m) = Cm$. Now let $d > 1$ be arbitrary, and suppose that the result holds for dimensions up to $d-1$. Let $\{w_i\}_{i=1}^m$ be a set of weights in $\mathbf{R}^d$. Suppose first that the weights satisfy

$$\begin{aligned} M^j &= 2 \quad \text{for all } j; & (23)\\ 1 \leq \max_{1\leq j\leq d} |w_i^j| &\leq 2 \quad \text{for all } i. & (24)\end{aligned}$$

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

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

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

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

Let $I_1 = \{i : w_i^d \geq 0\}$, $I_2 = \{i : w_i^d < 0\}$, $m_1 = |I_1|$, 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 $\mu_1, \mu_2 \geq \frac{1}{2d}$. For $1 \leq j < d$, 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 $i \leq m$ 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 $|\gamma^j| \leq 1$, $|y_i^j| \leq 4$, 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 $1 \leq k_1 \leq m_1$ we have

$$\Big|\sum_{i=1}^{k_1} w_{\pi_1(i)}^d - k_1\mu_1\Big| \leq 2\alpha\sqrt{k_1^*} \tag{25}$$

where $k_1^* = \min\{k_1, m_1 - k_1\}$. In similar fashion to the proof of Lemma A.1.1, Hoeffding's bounds [7] imply that for a particular value of k_1 we have

$$\Pr_{\text{unif}}\left(\pi_1 \text{ does not satisfy (25)}\right) \leq 2\exp(-2\alpha^2),$$

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

$$\begin{aligned}\Pr(\pi_1 \text{ does not satisfy (25)}) &\leq p_{d-1}(m)\cdot \Pr_{\text{unif}}(\pi_1 \text{ does not satisfy (25)})\\ &\leq p_{d-1}(m)\cdot 2\exp(-2\alpha^2).\end{aligned}$$

Hence

$$\Pr[\pi_1 \text{ is not } \alpha\text{-good}] \leq m p_{d-1}(m)\cdot 2\exp(-2\alpha^2). \tag{26}$$

Suppose that for some constants C and r, the polynomial p_{d-1} satisfies $p_d(k) \leq Ck^r$ for all k. If we let $\alpha = \sqrt{(r+1)\ln(m)}$, then the RHS of (26) is at most $2Cm^{r+1-2(r+1)} \leq \frac{1}{2}$, 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 $|\sum_{i=1}^k w_{\pi(i)}^d| \le 2$, and

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

for all $j < d$ and k. Hence π is c_d'–balanced for $c_d' = 4c_{d-1}+1$ by assumption (23) .

To verify almost uniformity, we follow the proof of Theorem 3.3. 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

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

Now we can bound the quantity $\binom{m_1}{k_1}\binom{m_2}{k_2}$ by mimicking (with minor modifications) the calculations from equation (7) to equation (10) in the proof of Theorem 3.3. In our current setting, we have $\mu_1, \mu_2 \ge \frac{1}{2d}$, and the $|w_i^d|$ are in $[0,2]$. Because we have changed the definition of α-good and the value of α, we also have to make the substitutions $(B-1)^2 \to 2^2$ and $\ln m \to (r+1)\ln m$, respectively. Thus the bound on the exponent given in equation (9) becomes

$$\frac{10\cdot 2^2}{(\frac{1}{2d})^2}(r+1)\ln m = 160d^2(r+1)\ln m. \tag{27}$$

Hence π is $p_d(m)$-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 (23) and (24). To generalize to arbitrary weights $\{w_i\}_{i=1}^m$, we use the interval trick introduced in section 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. Now use the T_i to divide β into intervals $I_1, \ldots, I_q$. Let $\{\alpha_i\}_{i=1}^{q-1}$ be the aggregated (d-dimensional) 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 (23) and (24). Hence these weights admit a c_d'–balanced, $p_d(q)$–uniform permution (though see the remark immediately following this proof). Rearranging the intervals $\{I_i\}_{i=1}^{q-1}$ according to such a permutation gives a permutation on $\{1,\ldots,m\}$ which is $p_d(m)$–uniform and c_d–balanced for $c_d = 2c_d'+1$. □

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 zero. 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 (27) 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 on $\{\alpha_i\}_{i=1}^{q-1}$ which is (c'_d+1)–balanced and $cp_d(q)$–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 specify our flow, we need one more definition.

Definition 6.4 *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 positive 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 $|S \oplus \pi\{1, \ldots, k\}| \le l$, 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 some fixed number l of items, and *vice versa*. As the name suggests, the strong balance condition is stricter than the usual balance condition. Nonetheless, the following lemma says that strongly balanced permutations always exist.

Lemma 6.5 *For any sequence $\{w_i\}_{i=1}^m$ in $\mathbf{R}^d$, there exists a strongly $16d^2$–balanced permutation.*

Note that this lemma claims only that a *single* strongly balanced permutation exists; unlike Theorem 6.3, it makes no claims regarding almost uniformity. The proof of the lemma relies heavily on the result of Grinberg and Sevast'yanov quoted earlier (Lemma 6.2); the proof is straightforward but rather technical, so we defer it to the Appendix.

6.3 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.

Lemma 6.6 *Fix any number of knapsack constraints d. For arbitrary item weights, it is possible to construct a multicommodity flow f in G_Ω with $\mathcal{C}(f)$ bounded by a polynomial in n and $\mathcal{L}(f) = \mathrm{O}(n)$.*

Proof: Recall that we identify each vertex $\mathbf{x} \in \Omega$ with the index set $X = \{i : x_i = 1\}$. Let $\widehat{\Omega} = \{X \in \Omega : a^j(X) \le b^j - 3c_d \max_{i \in X} a_i^j\}$, where c_d is the constant in the construction of balanced almost uniform permutations as in Theorem 6.3. Our main goal will be to construct a flow $\widehat{f}$ which, simultaneously for every $X, Y \in \widehat{\Omega}$, sends one unit of flow from X to Y. This flow will satisfy $\mathcal{C}(\widehat{f}) \le \mathrm{poly}(n)$ and $\mathcal{L}(\widehat{f}) = \mathrm{O}(n)$.

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

It remains to define the flow $\widehat{f}$ and show that it has the properties claimed. Fix $X, Y \in \widehat{\Omega}$. 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 could 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" items 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 $L = \{i \in X \oplus Y : a_i^j > M^j$ for some $j\}$ and $S = (X \oplus Y) - L$. (L and S are the "large" and "small" items respectively.) 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 $p_d(|S|)$–uniform permutation which is c_d–balanced with respect to the weights $\{w_i - \mu_2\}_{i\in S}$. The existence of π_1 and π_2 is guaranteed by Lemma 6.5 and Theorem 6.3 respectively. 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

$$\sum_{i\in V_1} w^j_{\pi(i)} \le (k-1)\mu^j + 3c_d M^j; \tag{28}$$

$$\sum_{i\in V_2} w^j_{\pi(i)} \ge (k-1)\mu^j - 3c_d M^j. \tag{29}$$

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. Again, 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

$$\frac{k_1-1}{k-1} \le \frac{|L|}{|L|+|S|};$$
$$\frac{k_2-1}{k-1} \le \frac{|S|}{|L|+|S|}.$$

Therefore, since ${\mu_1}^j \ge {\mu_2}^j$, we have

$$(k_1-1)\mu_1^j + k_2\mu_2^j \le (k-1)\mu^j; \tag{30}$$
$$k_1\mu_1^j + (k_2-1)\mu_2^j \ge (k-1)\mu^j. \tag{31}$$

Clearly, the strong balance condition on π_1 implies that there exist A, A', with $|A \oplus \{1, \ldots, k_1\}| \le 16d^2 + 1$ and similarly for A', such that

$$\sum_{i\in A} w^j_{\pi_1(i)} \le (k_1-1)\mu_1^j; \tag{32}$$

$$\sum_{i\in A'} w^j_{\pi_1(i)} \ge k_1\mu_1^j. \tag{33}$$

Also, by the balance condition on π_2 we have

$$\sum_{i=1}^{k_2} w^j_{\pi_2(i)} \leq k_2\mu^j_2 + c_d \max_{i\in S}\{|w^j_i - \mu^j_2|\} \leq k_2\mu^j_2 + 3c_d M^j; \tag{34}$$

$$\sum_{i=1}^{k_2} w^j_{\pi_2(i)} \geq k_2\mu^j_2 - c_d \max_{i\in S}\{|w^j_i - \mu^j_2|\} \geq (k_2-1)\mu^j_2 - 3c_d M^j. \tag{35}$$

Now, let $B = \pi^{-1}(\pi_1(A) \cup \pi_2\{1,\ldots,k_2\})$ and $B' = \pi^{-1}(\pi_1(A') \cup \pi_2\{1,\ldots,k_2\})$. Then we have $|B \oplus \{1,\ldots,k\}| \leq 16d^2 + 1 \leq 17d^2$, and

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

Exactly analogous relations hold with B, A replaced by B', A'. Combining this with equations (30)–(35) gives (28) and (29).

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 $1 \leq i \leq |X \oplus Y|$. This path might not stay in Ω, but we can alter it slightly so that it does. Equations (28) and (29) 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

$$a^j(Z_k - W^j_k) \leq \max\{a^j(X), a^j(Y)\} + 3c_d M^j \leq b^j; \tag{36}$$

$$a^j(Z_k \cup W^j_k) \geq \min\{a^j(X), a^j(Y)\} - 3c_d M^j. \tag{37}$$

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 $0 \leq k \leq |X \oplus Y|$, define

$$\overline{Z}_k = Z_k - W_k.$$

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

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

2. Add each item in $\overline{Z}_{k+1} - (\overline{Z}_k \cap \overline{Z}_{k+1})$ 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 4 and 5, for each intermediate point Z along the path define the "encoding" Z' by

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

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

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

where the second inequality follows from (37). Hence $Z' \in \Omega$. We can therefore bound the flow $\widehat{f}(Z)$ through Z by

$$\widehat{f}(Z) \quad \le \quad \sum_{Z' \in \Omega} \; \sum_{W_x, W_y, U} \Pr[\{\pi(1), \ldots, \pi(|U|)\} = U]. \tag{38}$$

Finally, 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 (38). Therefore, we have

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

where in the second line we have appealed to the almost uniformity of permutation π_2. This completes the proof. $\square$

Given such a flow, we can appeal to Theorem 2.1 to derive the main result of this section.

Theorem 6.7 *Fix any dimension $d > 0$, and let Ω be the set of solutions to an arbitrary instance of the d-dimensional 0-1 knapsack problem. 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 $|\Omega|$ in this more general setting.

Acknowledgments

We thank Persi Diaconis for bringing reference [6] to our attention.

References

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

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

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

[4] 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** (1993), pp. 271–284.

[5] G. R. Grimmett and D. R. Stirzaker. *Probability and Random Processes (3rd Ed.).* Oxford University Press, 1992.

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

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

[8] M. R. Jerrum. Mathematical foundations of the Markov chain Monte Carlo method. In *Probabilistic Methods for Algorithmic Discrete Mathematics* (M. Habib, C. McDiarmid, J. Ramirez-Alfonsin & B. Reed, eds), Algorithms and Combinatorics **16**, Springer-Verlag, 1998, pp. 116–165.

[9] M. R. Jerrum and A. J. Sinclair. Approximating the permanent. *SIAM Journal on Computing* **18** (1989), pp. 1149–1178.

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

[11] M. R. Jerrum and A.J. Sinclair. The Markov chain Monte Carlo method: an approach to approximate counting and integration. In *Approximation algorithms for NP-hard problems*, D.S. Hochbaum ed., PWS Publishing, Boston, 1997, pp. 482–520.

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

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

[14] B. Morris. *Random walks in convex sets.* PhD thesis, University of California at Berkeley, July 2000.

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

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

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

Appendix

A.1 Appendix for section 3

This section contains two technical lemmas that were used in the proof of Theorem 3.3.

Lemma A.1.1 *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 3.3. 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 [7]. □

Lemma A.1.2 *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

$$\frac{\binom{m_1}{\alpha m_1}\binom{m_2}{\alpha m_2}}{\binom{m}{\alpha m}} \quad \geq \quad C_1 m^{-1/2} \tag{39}$$

and

$$\frac{\binom{m_1}{\alpha m_1 - l}\binom{m_2}{\alpha m_2 + l}}{\binom{m_1}{\alpha m_1}\binom{m_2}{\alpha m_2}} \quad \geq \quad C_2 \exp\left\{-\frac{l^2 + \frac{1}{2}|l|}{\alpha(1-\alpha)}\Big(\frac{1}{m_1} + \frac{1}{m_2}\Big)\right\} \tag{40}$$

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 (40):

$$\begin{gathered}\left[\frac{\mathrm{P}(\alpha m_1)\mathrm{P}((1-\alpha)m_1)\mathrm{P}(\alpha m_2)\mathrm{P}((1-\alpha)m_2)}{\mathrm{P}(\alpha m_1 - l)\mathrm{P}((1-\alpha)m_1 + l)\mathrm{P}(\alpha m_2 + l)\mathrm{P}((1-\alpha)m_2 - l)}\right] \times \\ \left[\frac{\alpha m_1 (1-\alpha) m_1 \alpha m_2 (1-\alpha) m_2}{(\alpha m_1 - l)((1-\alpha)m_1 + l)(\alpha m_2 + l)((1-\alpha)m_2 - l)}\right]^{1/2},\end{gathered} \tag{41}$$

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

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

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 (41) is bounded below by

$$\exp\left\{-\frac{l^2}{\alpha(1-\alpha)}\Big(\frac{1}{m_1} + \frac{1}{m_2}\Big)\right\}. \tag{42}$$

A similar calculation bounds the second parenthesis in (41) by

$$\exp\left\{-\frac{|l|}{\alpha(1-\alpha)}\Big(\frac{1}{m_1} + \frac{1}{m_2}\Big)\right\}. \tag{43}$$

Combining (42) and (43) completes the verification of inequality (40) above, and hence the proof of the lemma. $\square$

A.2 Appendix for section 5.2

Here we provide the proof of Lemma 5.3, the analysis of the pre-processing random walk (PRW), which was omitted from the main text.

Proof of Lemma 5.3: By removing $X \cap Y$ from both X and Y and replacing b by $b - a(X \cap Y)$, we may assume that $X \cap Y = \emptyset$. Moreover, by scaling all the a_i and b we may assume that $M = \max_i a_i = 1$. Finally, we may assume that $b \geq 3\Delta$ since otherwise there are no pairs (X, Y) satisfying the hypothesis of the lemma.

Define

$$\mathcal{F} = \{(X', Y') : a(X') \geq b - \Delta \text{ or } a(Y') \geq b - \Delta\};$$
$$\mathcal{E} = \{(X', Y') : a(X') \leq b - 2\Delta \text{ and } a(Y') \leq b - 2\Delta\}.$$

Note that $\mathcal{F}$ contains all full pairs (X', Y'), and $\mathcal{E}, \mathcal{F}$ are disjoint. Also, define the hitting times

$$\mathcal{T} = \max_{(X',Y') \in \mathcal{F}} \mathrm{E}(\text{number of PRW steps to hit } \mathcal{E} \text{ starting at } (X', Y'));$$
$$\mathcal{U} = \min_{(X',Y') \in \mathcal{E}} \mathrm{E}(\text{number of PRW steps to hit } \mathcal{F} \text{ starting at } (X', Y')).$$

Now we claim that the lemma will follow if we can show:

(i) $\mathcal{T} \leq \alpha m$ for some constant $\alpha > 0$;

(ii) $\mathcal{U}/\mathcal{T} \geq \beta$ for some constant $\beta > 0$.

To see this, set the length of the PRW to be $C_1 m = 4\alpha m$, and let (X_t, Y_t) denote the sequence of pairs visited by the PRW, with $(X_0, Y_0) = (X, Y) \in \mathcal{F}$. Let T_0 be the first time t at which $(X_t, Y_t) \in \mathcal{E}$ (or $T_0 = C_1 m$ if the walk ends before this occurs); then let U_1 be the first t for which $(X_{T_0+t}, Y_{T_0+t}) \in \mathcal{F}$, and T_1 the first t for which $(X_{T_0+U_1+t}, Y_{T_0+U_1+t}) \in \mathcal{E}$. Continue defining a sequence of hitting times $U_2, T_2, U_3, T_3, \ldots$ in this way until the end of the walk is reached. Note that the PRW is not full for at least $\sum_i U_i$ steps, and that $\sum_{i \geq 0} T_i + \sum_{i \geq 1} U_i = 4\alpha m$ is the total walk length. Now from facts (i) and (ii) we have

$$\mathrm{E}\Big(\sum_{i \geq 0} T_i - \tfrac{1}{\beta} \sum_{i \geq 1} U_i\Big) = \mathrm{E}\Big(T_0 + \sum_{i \geq 1} (T_i - \tfrac{1}{\beta} U_i)\Big) \leq \alpha m.$$

An application of Markov's inequality then ensures that $\sum_{i \geq 0} T_i - \frac{1}{\beta} \sum_{i \geq 1} U_i \leq 2\alpha m$ with probability at least $\frac{1}{2}$. Conditioning on this event we have $(1 + \frac{1}{\beta}) \sum U_i \geq 2\alpha m$, and thus the proportion of steps during which the PRW is not full is at least $1/2(1 + \frac{1}{\beta})$, a constant. The lemma now follows easily.

It remains to verify facts (i) and (ii) above: these are immediate consequences of the following two claims. Let $\sigma^2 = \frac{1}{m} \sum_{i \in X \cup Y} a_i^2$ be the second moment of the item weights, and note that $\sigma^2 \geq 1/m$ since $\max_i a_i = 1$.

Claim 1: $\mathcal{T} \leq \gamma_1/\sigma^2$ for a constant $\gamma_1 > 0$.

Claim 2: $\mathcal{U} \geq \gamma_2/\sigma^2$ for a constant $\gamma_2 > 0$.

Proof of Claim 1: Choose an initial pair $(X_0, Y_0) \in \mathcal{F}$ that maximizes the expected time until the PRW hits $\mathcal{E}$, and let (X_t, Y_t) denote the PRW starting at (X_0, Y_0). We may assume w.l.o.g. that $a(X_0) > a(Y_0)$, so that $a(X_0) \in [b - \Delta, b]$ and $\mathcal{T}$ is the expected time until $a(X_t) \leq b - 2\Delta$. We estimate $\mathcal{T}$ by coupling the PRW with the *unconstrained* random walk, which behaves exactly like the PRW except that the constraint $\sum a_i \leq b$ is ignored. (Thus it is just simple random walk on an m-dimensional hypercube with holding probability $\frac{1}{2}$ at every step.) Write $(\widehat{X}_t, \widehat{Y}_t)$ for the unconstrained random walk, with $(\widehat{X}_0, \widehat{Y}_0) = (X_0, Y_0)$, and consider the first time $t = \widehat{T}$ at which $|a(\widehat{X}_t) - a(X_0)| \geq 2\Delta$. Now $a(\widehat{X}_t)$ is a supermartingale up to time $\widehat{T}$, since $\mathrm{E}(a(\widehat{X}_{t+1}) - a(\widehat{X}_t)|\widehat{X}_t) = \frac{1}{2m}(a(\widehat{Y}_t) - a(\widehat{X}_t)) < 0$. Thus with constant probability $a(\widehat{X}_{\widehat{T}}) \leq a(\widehat{X}_0) - 2\Delta$, and so $(X_{\widehat{T}}, Y_{\widehat{T}}) \in \mathcal{E}$. Hence $\mathcal{T}$ is bounded above by a constant times $\mathrm{E}(\widehat{T})$. But we also have $\mathrm{E}((a(\widehat{X}_{t+1}) - a(\widehat{X}_t))^2|\widehat{X}_t) = \frac{1}{2m}\sum_i a_i^2 = \frac{\sigma^2}{2}$. So $\mathrm{E}(\widehat{T})$ is the expected time for a supermartingale with increments bounded by ± 1 and with second moment $\sigma^2/2$ to move a distance $\pm 2\Delta$ from its initial value. A standard application of the martingale Optional Stopping Theorem (see, e.g., [5, Section 12.5]), now yields that $\mathrm{E}(\widehat{T}) \leq \frac{(4\Delta+1)^2}{\sigma^2/2} = \gamma_1/\sigma^2$ for a positive constant γ_1. This completes the proof of Claim 1.

Proof of Claim 2: As above let (X_t, Y_t) denote the PRW, but now with $(X_0, Y_0) \in \mathcal{E}$. We follow the random variable $Z_t = a(X_t) - a(Y_t)$, which always has a drift towards 0 (i.e., $\mathrm{E}(Z_{t+1} - Z_t|X_t) \times Z_t \leq 0$ for all t). Note that initially $|Z_0| \leq 2(b - 2\Delta) - V$, where $V = a(X_t) + a(Y_t) = \sum_i a_i$ is independent of t. And when $(X_t, Y_t) \in \mathcal{F}$ we have $|Z_t| \geq 2(b - \Delta) - V$. Thus $\mathcal{U}$ is bounded below by the minimum expected time for $|Z_t|$ to increase by 2Δ from its initial value. But the second moment is $\mathrm{E}((Z_{t+1} - Z_t)^2|X_t) = \frac{1}{2m}\sum_i (2a_i)^2 = 2\sigma^2$, so by a similar application of the Optional Stopping Theorem we conclude that $\mathcal{U} \geq \gamma_2/\sigma^2$, as claimed.

This completes the verification of Claims 1 and 2, and hence the proof of the lemma. □

A.3 Appendix for section 6.2

Here we prove the existence of strongly balanced permutations, as claimed in Lemma 6.5.

Proof of Lemma 6.5: 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 with the largest values of w_i^j, and the $4d$ indices i with the largest values of $-w_i^j$, 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 $k \leq m - |L|$, we have

$$-s^{j+} \leq \sum_{i=1}^{k} w_{\pi(i)}^j \leq s^{j-}, \tag{44}$$

where

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

We will need the Grinberg-Sevast'yanov result (Lemma 6.2), which states that for any set of vectors $x_1, \ldots, x_n$ in $\mathbf{R}^d$ with $\sum_i x_i = 0$, there exists a permutation $\nu \in \mathcal{S}_n$ such that

$$\sum_{i=1}^{k} x_{\nu(i)} \in d \times \mathrm{conv}\{x_1, \ldots, x_n\} \quad \text{for } 1 \leq k \leq n.$$

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

$$\sum_{i=1}^{k} x_{\nu'(i)} \in -d \times \mathrm{conv}\{x_1, \ldots, x_n\} \quad \text{for } 1 \leq k \leq n.$$

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

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

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

$$\sum_{i=1}^{k} (w_{\pi_2(i)} - \mu_2) \in d \times \mathrm{conv}\{w_i - \mu_2 : i \in S_2\}, \tag{45}$$

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 (44). 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

$$\begin{aligned}
\sum_{i=1}^{k} w_{\pi(i)}^j &= \sum_{i=1}^{k} w_{\pi_1(i)}^j \\
&= k\mu_1^j + \sum_{i=1}^{k} (w_{\pi_1(i)} - \mu_1^j) \\
&\le r\mu_1^j + d \max_{1 \le i \le m'} \{-(w_i^j - \mu_1^j)\} \\
&= \tfrac{1}{2}(s^{j-} - s^{j+}) + d \max_{1 \le i \le m'} \{-w_i^j\} + d\mu_1^j \\
&\le \tfrac{1}{2}(s^{j-} - s^{j+}) + (\tfrac{1}{4}s^{j-}) + (\tfrac{1}{4}s^{j+}) \\
&= \tfrac{3}{4}s^{j-} - \tfrac{1}{4}s^{j+} \\
&\le s^{j-} - \tfrac{1}{4}s^{j+}.
\end{aligned} \tag{46}$$

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

$$\sum_{i=1}^{k} w_{\pi(i)}^j = \sum_{i=1}^{r} w_{\pi_1(i)}^j + \sum_{i=1}^{k-r} w_{\pi_2(i)}^j. \tag{47}$$

Now if $\mu_2^j < 0$ then the conditional expectation of $\sum_{i=1}^k w_{\pi(i)}^j$ given π_1 is at least $-s^{j+} + s^{j-}$. Hence we must have

$$\begin{aligned}
\sum_{i=1}^{k} w_{\pi(i)}^j &\geq -s^{j+} + s^{j-} - d\max_{i\in S_2}\{-(w_i^j - \mu_2^j)\} \\
&\geq -s^{j+} + s^{j-} - d\max_{i\in S_2}\{-w_i^j\} \\
&\geq -s^{j+} + s^{j-} - s^{j-}/4 \\
&\geq -s^{j+}.
\end{aligned}$$

On the other hand, if $\mu_2^j \geq 0$ the right-hand side of (47) can be bounded below as follows:

$$\begin{aligned}
\sum_{i=1}^{r} w_{\pi_1(i)}^j + \sum_{i=1}^{k-r} w_{\pi_2(i)}^j &= \left[\sum_{i=1}^{r} w_{\pi_1(i)}^j + (k-r)\mu_2^j\right] + \sum_{i=1}^{k-r}(w_{\pi_2(i)}^j - \mu_2^j) \\
&\geq \left[\tfrac{1}{2}(s^{j-} - s^{j+}) - d\max_{1\leq i\leq m'}\{w_i^j - \mu_1^j\}\right] - d\max_{i\in S_2}\{-(w_i^j - \mu_2^j)\} \\
&\geq \left[\tfrac{1}{2}(s^{j-} - s^{j+}) - d\max_{1\leq i\leq m'}\{w_i^j\}\right] - d\max_{i\in S_2}\{-w_i^j\} - d\mu_2^j \\
&\geq \left[\tfrac{1}{2}(s^{j-} - s^{j+}) - s^{j+}/4\right] - s^{j-}/4 - s^{j-}/4 \\
&= -\frac{3}{4}s^{j+} \\
&\geq -s^{j+}.
\end{aligned}$$

For a corresponding upper bound, we can write

$$\sum_{i=1}^{k} w_{\pi(i)}^j \leq \sum_{i=1}^{r} w_{\pi_1(i)}^j + (k-r)\mu_2^j + \min\{d, k-r\}\max_{i\in S_2}\{w_i^j - \mu_2^j\}. \tag{48}$$

If $\mu_2^j \geq 0$, the right-hand side of (48) is bounded above by

$$(s^{j-} - s^{j+}) + d\max_{i\in S_2}\{w_i^j\} \leq (s^{j-} - s^{j+}) + s^{j+}/4 \leq s^{j-}.$$

If $\mu_2^j < 0$, the right-hand side of (48) is bounded above by

$$\sum_{i=1}^{r} w_{\pi_1(i)}^j + \min\{d, k-r\}\max_{i\in S_2}\{w_i^j\} \leq (s^{j-} - s^{j+}/4) + s^{j+}/4 \leq s^{j-},$$

where in the first inequality we have used equation (46).

Putting all the above together we see that π is strongly $8d^2$–balanced. Furthermore, in light of (44) we know also that, for all k and j, there exists a set of indices $S \supseteq \{1, \ldots, k\}$, with $|S| \leq k + 8d^2$, such that $\sum_{i=1}^k w_{\pi(i)}^j$ and $\sum_{i\in S} w_{\pi(i)}^j$ have opposite signs.

Now let $\{w_i\}_{i=1}^m$ be arbitrary. From the 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, and define $\mu^j = \frac{1}{m}\sum_{i=1}^m w_i^j$. W.l.o.g. $\mu^j \geq 0$. Then there exists $S \supseteq \{1, \ldots, k\}$ with $|S| \leq k + 8d^2$ such that

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

In addition, there exists $S \supseteq \{1,\ldots,k\}$, 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. $\square$