Random Walks on the Vertices of Transportation Polytopes with Constant Number of Sources *

Mary Cryan† Martin Dyer‡ Haiko Müller‡ Leen Stougie§

Abstract

We consider the problem of uniformly sampling a vertex of a transportation polytope with m sources and n destinations, where m is a constant. We analyse a natural random walk on the edge-vertex graph of the polytope. The analysis makes use of the multi-commodity flow technique of Sinclair [30] together with ideas developed by Morris and Sinclair [24, 25] for the knapsack problem, and Cryan *et al.* [3] for contingency tables, to establish that the random walk approaches the uniform distribution in time $n^{O(m^2)}$.

1 Introduction

In this paper we study the mixing time behaviour of a natural random walk on the edge-vertex graph of a *transportation polytope* with m sources and n destinations. We are able to show that this walk converges to the uniform distribution on the vertex set in time $n^{O(m^2)}$. Therefore the random walk mixes rapidly whenever the number of sources m is a constant. As far as we are aware, this is the first result proving rapid mixing of a random walk on the graph of any non-trivial class of polytopes. Very little is known about the mixing times of random walks on polytope graphs in general. In fact, it is not even known whether the diameter of the graph is polynomially bounded in the dimension and number of facets of the polytope. (See Kalai [19] and Ziegler [32].) In consequence, Markov chain Monte Carlo (MCMC) has not been well explored as a means of sampling, or approximately counting, vertices of general polytopes. Even for special classes of polytopes, such as arbitrary transportation polytopes, approximate counting algorithms are not known to exist, either by MCMC or by other means (see, for example, Pak [28]). This is despite the fact that the diameter of any transportation polytope is bounded above by a linear function in $m + n$ (see Brightwell et al. [2]; an earlier paper by Dyer and Frieze [13] gave a polynomial upper bound). In fact, the only previous mixing results known for random walks on the edge-vertex graph of a polytope are for very special, and highly symmetric polytopes, such as the n-cube [7] and the Birkhoff polytope [27].

Our approach to proving rapid mixing for our random walk on the transportation polytope is inspired by that of Cryan, Dyer, Goldberg, Jerrum and Martin [3] for sampling contingency tables. This was itself based on the "balanced permutation" ideas of Morris and Sinclair [24, 25] for the

*Supported by the EPSRC grant "Sharper Analysis of Randomised Algorithms: a Computational Approach", by the EC IST project RAND-APX, by the MRT Network ADONET of the European Community (MRTN-CT-2003-504438) and the Dutch BSIK-BRICKS project.

†School of Informatics, University of Edinburgh, King's Buildings, Mayfield Road, Edinburgh EH9 3JZ, Scotland.

‡(dyer|hm)@comp.leeds.ac.uk. School of Computing, University of Leeds, Leeds LS2 9JT, England.

§leen@win.tue.nl. Dept. of Mathematics and Computer Science, Eindhoven University of Technology, the Netherlands and CWI Amsterdam, the Netherlands.

knapsack problem. However, following the line of proof given in [3], and using the m-dimensional balanced permutations of [24], would lead inevitably to a mixing time bound of $n^{2^{O(m)}}$ for our random walk. To obtain our improvement in the exponent, from exponential to polynomial, it is necessary to sharpen the tools of [24, 25] using the special structure of the problem at hand. Our improvement then results principally from the fact that we can prove that a *strongly* $O(m^2)$-balanced $n^{O(m^2)}$-uniform permutation exists for this problem. Note that it is unknown whether a strongly-balanced almost-uniform permutation exists for an arbitrary set of m-dimensional vectors, when m is variable. (See [24] for further information.)

Our paper is organised as follows: In §2 we give basic information and background concerning the transportation polytope and the natural random walk on that polytope. In §3 we list results about the structure of vertices and edges of the transportation polytope, and prove upper and lower bounds on the number of adjacent edges of any vertex (the ratio of the upper bound to the lower bound is a key parameter in our definition of the random walk). §4 introduces a new Markov chain called a heat-bath chain, which can make larger moves on the edge-vertex graph than the natural random walk, but which also converges to the uniform distribution on the vertices of the transportation polytope. This heat-bath chain is then analysed in the next two sections. In §5, we present our improved balanced almost-uniform permutations (based on the permutations of Morris and Sinclair [25]), which will be used in the analysis of the heat-bath chain. In §6 we prove that the heat-bath chain mixes rapidly, when the number of sources is constant. In §7 we prove, by comparison to the heat-bath chain, that the natural random walk also mixes rapidly in this case. In §8 we show how to use our sampling algorithm to obtain a polynomial-time algorithm to approximately count vertices of the transportation polytope when m is constant.

2 Background

The transportation problem (TP) is a classic problem in operations research. The problem was posed for the first time by Hitchcock in 1941 [18] and independently by Koopmans in 1947 [21], and appears in any standard introductory course on operations research. It is the combinatorial optimization problem of assigning shipments of some commodity from sources to destinations so that the total transportation cost is minimized. We are given m sources and a list $r = (r_1, \ldots, r_m)$ of supplies for these sources (r_i is the supply at source i). We are given n destinations and a list $c = (c_1, \ldots, c_n)$ of demands for these destinations (c_j is the demand at destination j). Without loss of generality, we assume that $\sum_{i=1}^{m} r_i = \sum_{j=1}^{n} c_j$, and define $N = \sum_{i=1}^{m} r_i$. Let t_j^i denote the cost of transporting one unit from source i to destination j, for $1 \leq i \leq m, 1 \leq j \leq n$. We use the somewhat uncommon notation t_j^i to denote the i, j element of a matrix.

We will represent an assignment to the variables of the transportation problem by a $m \times n$-dimensional matrix X, and write X_j to denote the j-th column of X (X_j^i denotes the i-th entry of column X_j). For integers $p \leq q$, let $[p, q]$ denote the set of integers $\{p, \ldots, q\}$. Similarly $(p, q]$ denotes the set $\{p + 1, \ldots, q\}$ etc. Also $[p]$ denotes $[1, p]$ for $p > 0$. The TP, satisfying all supplies and demands at minimum total transportation costs, is formulated by the following linear program:

$$\min \sum_{i=1}^{m} \sum_{j=1}^{n} t_j^i X_j^i$$

$$X_j^i \geq 0 \qquad \text{for all } i \in [m], j \in [n] \tag{1}$$

$$\sum_{j=1}^{n} X_j^i = r_i \qquad \text{for all } i \in [m] \tag{2}$$

$$\sum_{i=1}^{m} X_j^i = c_j \qquad \text{for all } j \in [n] \tag{3}$$

The set of feasible solutions of the TP, the *feasible region*, is a convex polytope $\mathrm{P}(r, c)$ in $\mathbb{R}^{mn}$ called the *transportation polytope*. The existence of a strongly polynomial time algorithm for the TP follows directly from the seminal work of Tardos [31]. Orlin [26] gave a strongly polynomial time primal simplex algorithm for the more general minimum cost flow problem.

The integer feasible solutions for the TP arise in another context. Given a list of non-negative integer row sums and column sums, a *contingency table* is defined to be any $m \times n$ matrix of non-negative integers with the given row and column sums. Therefore the set of integer feasible solutions to the TP corresponds exactly to the set of contingency tables with row sums $(r_1, \ldots, r_m)$ and column sums $(c_1, \ldots, c_n)$ [6]. The problem of generating contingency tables almost uniformly at random has been widely studied, for example, by Dyer, Kannan and Mount [15], Diaconis and Saloff-Coste [9], Dyer and Greenhill [14], Morris [23], Cryan et al. [3] and Dyer [12]. In particular, it was shown in [3] that a 2×2 "heat-bath" Markov chain is rapidly mixing when the number of rows is constant.

The minimum cost for a TP is always attained at a vertex. Therefore counting and enumerating the vertices of transportation polytopes is of interest. Some results on the complexity of enumerating the vertices of a polytope appeared in Dyer [11], where it was shown to be #P-complete to count exactly the number of vertices of a $2 \times n$ transportation polytope,[1] and that it is NP-complete to decide if a $2 \times n$ transportation polytope is degenerate.

In this paper we consider the problem of sampling the vertices of $\mathrm{P}(r, c)$ almost uniformly at random, when the number of sources m is a constant. We define a Markov chain $\mathcal{W}$ on the set Ω of all vertices of $\mathrm{P}(r, c)$ and prove it is rapidly mixing. Our chain $\mathcal{W}$ is a random walk on the *edge-vertex graph* of the polytope $\mathrm{P}(r, c)$. This graph, also called the *skeleton* of the transportation polytope, contains a vertex Z for every vertex of $\mathrm{P}(r, c)$, and an edge (Z, W) for every pair of vertices Z, W that form an edge of $\mathrm{P}(r, c)$. We denote the edge-vertex graph (see Definition 1 below) by $G(\mathcal{W})$.

By Lemma 4 of §3, we know that any vertex Z of $\mathrm{P}(r, c)$ has at most d_m incident edges, where $d_m = n^m$ is polynomially bounded in n. A single step of our Markov chain is performed as follows: if Z is the current vertex, we walk along any incident edge of Z with probability $1/2d_m$. If $\deg(Z)$ denotes the vertex degree of Z in $G(\mathcal{W})$, then the probability of remaining at Z is $1 - \deg(Z)/2d_m$. A well-known result of Balinski [1] states that the edge-vertex graph of any convex polytope of dimension k is k-connected. Therefore, $G(\mathcal{W})$ is connected, and the Markov chain $\mathcal{W}$ is *irreducible*. Also, at any given step, the probability of remaining at the current vertex is at least $1/2$, so $\mathcal{W}$ is *aperiodic*. Hence $\mathcal{W}$ is ergodic and therefore has a unique stationary distribution. Also, for any two vertices Z, W of $\mathrm{P}(r, c)$ such that $Z \neq W$, $\mathrm{Pr}_{\mathcal{W}}[Z, W] = \mathrm{Pr}_{\mathcal{W}}[W, Z]$ (this has the value $1/2d_m$ if Z is adjacent to W, and 0 otherwise), which implies that the unique stationary distribution for $\mathcal{W}$ is the uniform distribution. Observe that all "null" steps at Z, where $\mathcal{W}$ remains at Z, can be simulated by updating

[1] In fact, [11] only claims NP-hardness, but the proof establishes #P-completeness.

the clock with a single geometrically distributed random variable, and then moving to a neighbour of Z chosen uniformly at random, provided that the end time has been reached.

We will show that $\mathcal{W}$ is rapidly mixing by first showing that a "heat bath chain", which can make much larger moves in the edge-vertex graph, mixes rapidly. This chain, $\mathcal{M}_{\mathrm{HB}}$, is described in §4, and analysed in §5–6. Subsequently, in §7, we use the comparison technique of Diaconis and Saloff-Coste [8] (see also Randall and Tetali [29]) to lift the mixing result from $\mathcal{M}_{\mathrm{HB}}$ to $\mathcal{W}$. Finally, in §8, we outline how sampling can be used to count approximately the number of vertices of a transportation polytope. However, first of all in §3 we present structural results concerning the vertices and edges of $\mathrm{P}(r, c)$, and justify our definition of d_m.

3 Vertices and Edges

For basic information about polytopes we refer the reader to Ziegler [32], and for specific details about transportation polytopes to Klee and Witzgall [20]. For basic information on the linear programming formulation and a simplex algorithm for the transportation problem we refer to any introductory textbook on operations research, e.g. [17]. We mention a few highlights here. It is shown in [20] that $\mathrm{P}(r, c)$ has dimension $(m-1)(n-1)$.

Now we give a formal definition of the edge-vertex graph of a polytope:

Definition 1 *Let $P \subseteq \mathbb{R}^r$ be any polytope. $F \subseteq P$ is a* face *of P iff*

$$F \quad = \quad P \cap \{Z : cZ = c_0\},$$

for some $c \in \mathbb{R}^r$ and some $c_0 \in \mathbb{R}$ such that $cZ \leq c_0$ holds for all $Z \in P$.
The face is non-trivial *if $F \neq \emptyset$ and $F \neq P$.*
A facet *of P is a non-trivial face F such that $dim(F) = dim(P) - 1 = r - 1$.*
An edge *of P is a non-trivial face F such that $dim(F) = 1$.*
A vertex *of P is a non-trivial face F such that $dim(F) = 0$ ($F = \{X\}$ for some $X \in P$).*
The edge-vertex graph *of P is the graph G which contains a vertex X for every vertex of P and an edge connecting X to Y iff $\{\alpha X + (1-\alpha)Y : \alpha \in [0,1]\}$ is an edge of the polytope P.*

All facets of the polytope $\mathrm{P}(r, c)$ correspond to $X^i_j \geq 0$ for some $i \in [m], j \in [n]$, and therefore every face of $\mathrm{P}(r, c)$ corresponds to setting $X^i_j = 0$ (when X is represented in tabular format) for some number of (i, j)-pairs. The following lemma is due to Dantzig [5] and others (see Klee and Witzgall [20] for a history).

Lemma 2 *If $(r_1, \ldots, r_m)$ and $(c_1, \ldots, c_n)$ are lists of positive values such that $\sum_{i=1}^m r_i = \sum_{j=1}^n c_j$, then for every vertex of $\mathrm{P}(r, c)$, the (i, j)-pairs corresponding to non-zero coordinates of that vertex form a spanning forest $\mathcal{F}$ on the bipartite graph $[m] \uplus [n]$. This implies that each vertex of $\mathrm{P}(r, c)$ has no more than $n + m - 1$ non-zero coordinates.*

A non-degenerate *vertex has* exactly $n+m-1$ *non-zero coordinates, corresponding to a spanning tree on $[m] \uplus [n]$.*
Any $(m-1)(n-1)$-dimensional transportation polytope has at most $m^{n-1}n^{m-1} \leq (em)^{n+m-1}$ vertices (for $n \geq m \geq 2$). □

We note that any vertex of $\mathrm{P}(r, c)$ must have at least n non-zero coordinates, and therefore any vertex has between n and $n + m - 1$ (inclusive) non-zero coordinates. If $\mathrm{P}(r, c)$ is non-degenerate, then every vertex will have $n + m - 1$ non-zero coordinates in one-to-one correspondence with the basic

variables of a *basic feasible solution* of the linear program (1)-(3) (see e.g. [17]). If Z has $n+m-1-q$ non-zero coordinates in total $(0 \leq q \leq m-1)$, we say it has *degeneracy* q. We sometimes refer to co-ordinates as *cells* (of the tabular representation of Z). The spanning forest $\mathcal{F}_Z$ of $[m] \uplus [n]$ of a vertex Z with degeneracy q consists of $q+1$ vertex disjoint trees. A spanning forest $\mathcal{F}_Z$ together with any set of q edges which creates a spanning tree of $[m] \uplus [n]$, corresponds uniquely to a basic feasible solution of (1)-(3), where the $m+n-1$ basic variables of this solution are the cells of Z corresponding to the edges of this spanning tree (the q added edges correspond to basic variables with value 0). Thus, any degenerate vertex Z of the polytope corresponds to a number of basic feasible solutions of (1)-(3), and each such basic feasible solution corresponds to a unique spanning tree of $[m] \uplus [n]$.

We define a pivot operation from one basic feasible solution to another one as an operation on the corresponding spanning trees. It can be found in any elementary textbook on operations research (see e.g. [17]), though it is usually described in terms of the tabular representation of basic feasible solutions. Consider any basic feasible solution Z with spanning tree $\mathcal{T}_Z$, and consider any edge $(a,b) \notin E(\mathcal{T}_Z)$. Then $E(\mathcal{T}_Z) \cup \{(a,b)\}$ contains a single unique simple cycle $\mathcal{C}$. Since $\mathcal{C}$ is an even cycle we can label its edges alternately $+$ and $-$, giving (a,b) the label $+$. Let $E^+(\mathcal{C})$ and $E^-(\mathcal{C})$ be the edges of $\mathcal{C}$ with label $+$ and $-$ respectively, and let $(c,d) = \operatorname{argmin}\{Z^i_j : (i,j) \in E^-(\mathcal{C})\}$. (if (c,d) is not unique, any choice will give a pivot). A *pivot operation (on (a,b))* then consists of increasing the value of all Z^i_j for $(i,j) \in E^+(\mathcal{C})$ by Z^c_d and decreasing the value of all Z^i_j with $(i,j) \in E^-(\mathcal{C})$ by Z^c_d. Observe that in particular, the (a,b) cell of the new table now has the value Z^c_d (and becomes a basic variable), while the (c,d) cell obtains the value 0 (and becomes a non-basic variable). The new spanning tree is then $(\mathcal{T}_Z \cup \{(a,b)\}) \setminus \{(c,d)\}$. In the case where Z^c_d is originally 0, the only effect of the pivot operation is that Z^a_b becomes a basic variable instead of Z^c_d. The vertex of the polytope does not change in this case.

A pivot on any edge (a,b) satisfying $(a,b) \notin E(\mathcal{T}_Z)$ corresponds to an edge of the edge-vertex graph of the transportation polytope if and only if $Z^i_j > 0$ for all $(i,j) \in E^-(\mathcal{C})$. Formulated in terms of the vertices of the polytope this gives the following Lemma:

Lemma 3 *Let Z be a vertex of $\mathrm{P}(r,c)$, and let $\mathcal{F}_Z$ denote the forest on $[m] \uplus [n]$ given by the non-zero cells of Z (see Lemma 2). Let $\mathcal{T}^1_Z, \ldots, \mathcal{T}^{q+1}_Z$ be the maximal trees constituting $\mathcal{F}_Z$. Let W be another vertex of $\mathrm{P}(r,c)$ and let $\mathcal{C} = \{(i,j) : Z^i_j \neq W^i_j\}$.*

Then Z and W are joined by an edge of $\mathrm{P}(r,c)$ if and only if $\mathcal{C}$ is a simple cycle of the form

$$(i_1,j_1), p_1, (i_2,j_2), \ldots, (i_\kappa, j_\kappa), p_\kappa$$

where cell $Z^{i_k}_{j_k} = 0$ for every $k \in [\kappa]$, where p_k is a path in some $T^h_Z \in \mathcal{F}_Z$ from $j_k \in [n]$ to $i_{k+1} \in [m]$ for every $k \in [\kappa]$ (identifying $i_{\kappa+1}$ with i_1), and where the $\mathcal{T}^h_Z$ are all distinct.

Moreover, for every vertex Z of $\mathrm{P}(r,c)$, and every set of cells $\mathcal{C}$ of Z forming this type of cycle in Z, there is exactly one vertex W in $\mathrm{P}(r,c)$ which is adjacent to Z such that $\{(i,j) : Z^i_j \neq W^i_j\} = \mathcal{C}$.
□

Lemma 4 *Any vertex of the polytope $\mathrm{P}(r,c)$ has at least $(m-1)(n-1)$ and at most n^m incident edges.*

Proof: We first prove the Lemma in the non-degenerate case.

First consider any non-degenerate vertex $Z \in \mathrm{P}(r,c)$, corresponding to a spanning tree $\mathcal{T}_Z$ in $[m] \uplus [n]$. In this case, if we perform a pivot operation on any of the $nm-n-m+1$ edges of $[m] \uplus [n] \setminus \mathcal{T}_Z$ (each of these is a non-basic variable), we create another vertex of $\mathrm{P}(r,c)$ adjacent to

Z. Each of these vertices is distinct. By Lemma 3 these are the only vertices adjacent to Z. Therefore the Lemma holds in the non-degenerate case (with $deg(Z) = (m-1)(n-1)$).

Now consider a vertex $Z \in \mathrm{P}(r,c)$ of degeneracy $q > 0$. We first prove the lower bound.

Suppose the non-zero cells of Z correspond to a forest $\mathcal{F}_Z$ with $q+1$ maximal trees $\mathcal{T}_Z^1, \ldots, \mathcal{T}_Z^{q+1}$. We create two basic feasible solutions X and Y by extending $\mathcal{F}_Z$ to spanning trees $\mathcal{T}_X$ and $\mathcal{T}_Y$ of $[m] \uplus [n]$ as follows: For every $1 \leq i \leq q+1$, we arbitrarily fix $u_Z^i \in [m] \cap \mathcal{T}_Z^i$ and $v_Z^i \in [n] \cap \mathcal{T}_Z^i$. We then define $\mathcal{T}_X$ and $\mathcal{T}_Y$ as follows:

$$\begin{aligned} \mathcal{T}_X &= \mathcal{F}_Z \cup M_X \qquad \text{with } M_X = \{(u_Z^{i+1}, v_Z^i) \mid i = 1, \ldots, q\}, \\ \mathcal{T}_Y &= \mathcal{F}_Z \cup M_Y \qquad \text{with } M_Y = \{(u_Z^i, v_Z^{i+1}) \mid i = 1, \ldots, q\}. \end{aligned}$$

Now consider $\mathcal{T}_X$ and consider a pivot on any of the $nm - m - n + 1$ edges $(u,v) \in [m] \uplus [n] \setminus E(\mathcal{T}_X)$. If $u \in \mathcal{T}_Z^i$ and $v \in \mathcal{T}_Z^i$ for some $1 \leq i \leq q+1$, then there is a unique path $p_{u,v}$ between u and v consisting entirely of edges of $\mathcal{T}_Z^i$ (and of non-zero edges of $\mathcal{T}_X$). Hence $\{(u,v)\} \cup p_{u,v}$ forms a cycle where all cells except (u,v) are strictly positive in Z. Therefore in this case the pivot corresponds to an edge of the polytope, and performing the pivot operation creates a neighbouring vertex W with $W_v^u > 0$. Alternatively $u \in \mathcal{T}_Z^h$ and $v \in \mathcal{T}_Z^k$ for some h, k such that $h \neq k$. First assume that $h < k$. In this case $\mathcal{T}_X \cup (u,v)$ contains a unique cycle $C_X(u,v)$ in which the only edges from $E(\mathcal{T}_X) \setminus E(\mathcal{F}_Z)$ are the edges (u_Z^h, v_Z^{h+1}), $h = i, \ldots, k-1$. Moreover, these are all in $E^+(C_X(u,v))$. Hence, $E^-(C_X(u,v)) \setminus E(\mathcal{F}_Z) = \emptyset$, and therefore $\min\{Z_j^i : (i,j) \in E^-(C_X(u,v))\} > 0$. Therefore the pivot on (u,v) corresponds to an edge of the polytope $\mathrm{P}(r,c)$, leading to a vertex W with $W_v^u > 0$. In particular this holds for the q edges $(u_Z^i, v_Z^{i+1}) \in M_Y$ (in which case $k = h+1$), leading to a vertex W which has a non-zero value for (u_Z^i, v_Z^{i+1}) and (u_Z^{i+1}, v_Z^i), and which satisfies $W_v^u = 0$ for all other $(u,v) \in [m] \uplus [n] \setminus E(\mathcal{F}_Z)$. Similarly, if $u \in \mathcal{T}_Z^h$ and $v \in \mathcal{T}_Z^k$ and $h > k$ then $\mathcal{T}_Y \cup (u,v)$ contains a unique cycle. Then applying the pivot operation for (u,v) on the spanning tree $\mathcal{T}_Y$ generates a neighbour vertex with $W_v^u > 0$.

Now recall that $E(\mathcal{T}_X) = E(\mathcal{F}_Z) \cup E(M_X)$ and $E(\mathcal{T}_Y) = E(\mathcal{F}_Z) \cup E(M_Y)$. In total, there are $(m-1)(n-1)$ (zero)-cells in $[m] \times [n] \setminus E(\mathcal{T}_X)$. Define the following three subsets of $[m] \times [n]$:

$$\begin{aligned} E_1 &= \{(u,v) : (u,v) \notin \mathcal{T}_X, u, v \in \mathcal{T}_Z^h \text{ for some } h.\} \\ E_2 &= \{(u,v) : (u,v) \notin \mathcal{T}_X, u \in \mathcal{T}_Z^h, v \in \mathcal{T}_Z^k, \text{ for some } h, k \text{ such that } h < k.\} \\ E_3 &= \{(u,v) : (u,v) \notin \mathcal{T}_Y, u \in \mathcal{T}_Z^h, v \in \mathcal{T}_Z^k, \text{ for some } h, k \text{ such that } h > k.\} \end{aligned}$$

Observe that the sets E_1, E_2 and E_3 are disjoint and $|E_1 \cup E_2 \cup E_3| = |E_1 \cup E_2 \cup E_3 \cup M_X \cup M_Y| = (m-1)(n-1) + q$ (using $M_X \subseteq E_3$ and $M_Y \subseteq E_2$). For $i = 1, 2$, let $\mathcal{V}_i$ be the neighbouring vertices of Z that can be obtained by a pivot operation on $\mathcal{T}_X$ for some cell in E_i. Let $\mathcal{V}_3$ be the neighbouring vertices of Z obtained by a pivot operation on $\mathcal{T}_Y$ for some cell in E_3. Recall that for every $(u,v) \in E_1$, the neighbour vertex of Z constructed by a pivot operation on (u,v) is unique among all pivots on cells of E_1 (no other pivot for $(u,v) \in E_1$ induces a non-zero value for cell (u,v)). Hence $|\mathcal{V}_1| = |E_1|$. Moreover, (u,v) is the only zero-cell of Z which becomes positive as a result of this pivot, hence $V_1 \cap (\mathcal{V}_2 \cup \mathcal{V}_3) = \emptyset$. For every cell $(u,v) \in E_2$, the pivot operation on (u,v) with respect to the spanning tree $\mathcal{T}_X$ constructs a neighbouring vertex to Z with a non-zero value for (u,v). This is unique among all pivot operations on cells of E_2, hence $|\mathcal{V}_2| = |E_2|$. Similarly, we know that $|\mathcal{V}_3| = |E_3|$. We now show that $|\mathcal{V}_2 \cup \mathcal{V}_3| \geq (m-1)(n-1) - |\mathcal{V}_1|$. Suppose $(u,v) \in E_3 \setminus M_X$. Then $(u,v) \notin E(\mathcal{T}_X)$, and therefore the neighbour obtained by a pivot operation on (u,v) with respect to $\mathcal{T}_Y$ is not an element of $\mathcal{V}_2$. Hence $|\mathcal{V}_2 \cap \mathcal{V}_3| \leq |M_X| = q$. Therefore $|\mathcal{V}_1 \cup \mathcal{V}_2 \cup \mathcal{V}_3| \geq |\mathcal{V}_1| + |\mathcal{V}_2| + |\mathcal{V}_3| - q = |E_1| + |E_2| + |E_3| - q \geq (m-1)(n-1)$, as required.

Next we prove the upper bound for the degenerate case.

Assume again that Z is a vertex of $\mathrm{P}(r,c)$ of degeneracy $q \geq 1$, and the the non-zero cells of Z correspond to the forest $\mathcal{F}_Z$ on $[m] \uplus [n]$, where we write $\mathcal{F}_Z = \{\mathcal{T}_Z^1, \ldots, \mathcal{T}_Z^{q+1}\}$. For every $h \in [q+1]$, let $I_h \subseteq [m]$ denote the set of source indices in $\mathcal{T}_Z^h$, and $J_h \subseteq [n]$ denote the set of destination indices in $\mathcal{T}_Z^h$; $\{I_h : h \in [q+1]\}$ is a partition of $[m]$ and $\{J_h : h \in [q+1]\}$ is a partition of $[n]$. Let $m_k = |I_k|$ and $n_k = |J_k|$, for $k \in [q+1]$.

By Lemma 3, a vertex W with $W_j^i \neq 0$ is a neighbour of Z iff the differing edges of $\mathcal{F}_Z$ and $\mathcal{F}_W$ form a simple cycle $\mathcal{C} = (i,j), p_1, (i_2,j_2), p_2, \ldots, p_\kappa$ such that

- The cells $(i,j), (i_2,j_2), \ldots (i_\kappa, j_\kappa)$ are the cells which are zero in Z and non-zero in W.
- For every $k \in [\kappa]$, p_k is a path of odd length from destination $j_k \in [n]$ to source $i_{k+1} \in [m]$ in some tree $\mathcal{T}_Z^k \in \mathcal{F}_Z$ (assuming $(i_1, j_1) = (i,j)$ and $i_{\kappa+1} = i_1$).
- The $\mathcal{T}_Z^k$ are all distinct trees of $\mathcal{F}_Z$.

Also by Lemma 3, there is exactly one neighbouring vertex W to Z for this cycle $\mathcal{C}$. Each such cycle is completely characterised by an ordered list of paths $p_1, \ldots, p_\kappa$, where each path is from some destination $j_k \in [n]$ to some source $i_{k+1} \in [m]$ in some tree $\mathcal{T}_k \in \mathcal{F}$, and the $\mathcal{T}_k$ are distinct trees. Two different ordered lists of paths only correspond to the same set of zero cells of Z if one ordered list is a cyclic rotation of the other list.

Therefore the number of neighbouring vertices of Z can be expressed as the number of simple cycles consisting of κ simple paths from κ different trees of $\mathcal{F}_Z$, summed over $\kappa \in [q+1]$. If $\kappa = 1$, the number of zero cells with both endpoints in a tree $\mathcal{T}_Z^k \in \mathcal{F}_Z$ is $(m_k - 1) \times (n_k - 1)$. If $\kappa > 1$, then we must count the cycles defined of κ simple paths from κ different trees of $\mathcal{F}_Z$. Hence the number of vertices adjacent to Z is given by the following expression:

$$deg(Z) \;=\; \sum_{S \subseteq [q+1], |S| \geq 2} (|S|-1)! \prod_{k \in S} m_k \times n_k \;+\; \sum_{k \in [q+1]} (m_k - 1) \times (n_k - 1), \tag{4}$$

which depends only on the values of m_k and n_k for $k \in [q+1]$. Let a *pair of partitions (of size $q+1$)* be any two lists of numbers $\vec{m} = m_1, \ldots, m_{q+1}$ and $\vec{n} = n_1, \ldots, n_{q+1}$ such that $m_k \geq 1, n_k \geq 1$ for all $k \in [q+1]$, and such that $\sum_{k=1}^{q+1} m_k = m$ and $\sum_{k=1}^{q+1} n_k = n$. Thus, $deg(Z)$ is a function of $(\vec{m}_Z, \vec{n}_Z)$. We bound $deg(Z)$ by bounding the maximum of the righthand side of (4) over all possible pairs of partitions.

Observe that for any pair of partitions $\vec{m}, \vec{n}$ and any $S \subseteq [q+1]$, $\prod_{k \in S} m_k n_k \leq \prod_{k \in [q+1]} m_k n_k$. Also for any value $\kappa < q+1$, there are exactly $\binom{q+1}{\kappa}$ sets $S \subseteq [q+1]$ such that $|S| = \kappa$. Therefore by (4),

$$\begin{aligned}
deg(\vec{m}, \vec{n}) \;&\leq \sum_{\kappa=1}^{q+1} (\kappa-1)! \sum_{S \subseteq [q+1], |S| = \kappa} \prod_{k \in [q+1]} m_k n_k \\
&= \sum_{\kappa=1}^{q+1} ((q+1) \cdot \ldots \cdot (q-\kappa+2))/\kappa \prod_{k \in [q+1]} m_k n_k \\
&\leq (q+1)^{q+1} \Big(\prod_{k \in [q+1]} m_k \Big) \Big(\prod_{k \in [q+1]} n_k \Big) \\
&\leq (q+1)^{q+1} \Big(\frac{m}{q+1}\Big)^{q+1} \Big(\frac{n}{q+1}\Big)^{q+1} \;=\; m^{q+1} n^{q+1} / (q+1)^{q+1}.
\end{aligned}$$

For every $m \geq 2$ (always the case in the context of transportation polytopes) and every $1 \leq q+1 \leq m$, $m^{q+1}n^{q+1}/(q+1)^{q+1} \leq n^m$: the case for $q+1 \leq m/2$ follows from the fact that if $q+1 \geq m/2$, then $m^{q+1}n^{q+1}/(q+1)^{q+1} \leq m^{q+1}n^{q+1} \leq n^{2(q+1)} \leq n^m$; the case for $q+1 = m$ is simple to check; and the case for $m/2 \leq q+1 \leq m-1$ follows by $(m/(q+1))^{q+1}n^{q+1} = (1+(m-q-1)/(q+1))^{q+1}n^{q+1} = ((1+(m-q-1)/(q+1))^{(q+1)/(m-q-1)})^{m-q-1}n^{q+1} < e^{m-q-1}n^{q+1}$. This value is at most $m^{m-q-1}n^{q+1} \leq n^m$ if $m \geq 3$, and if $m = 2$, we can check correctness directly. □

4 The heat-bath chain

We now define our auxiliary "heat-bath" Markov chain $\mathcal{M}_{\mathrm{HB}}$, which operates on a $m \times b_m$-sized window of the matrix representing the current vertex Z, where $b_m = 47m^2$. Define

$$\Gamma_Z = \{j : Z_j \text{ has more than one non-zero}\}.$$

Then $|\Gamma_Z| \leq m-1$. A single step of $\mathcal{M}_{\mathrm{HB}}$ is performed as follows: a set of columns $B \subseteq [n]$, with $|B| = b_m$, is chosen uniformly at random from the columns of the matrix representing Z, subject to $\Gamma_Z \subseteq B$. Then Z is replaced by a vertex W chosen uniformly at random from all vertices which can be obtained from Z by modifying only the columns Z_j $(j \in B)$.

The $\mathcal{M}_{\mathrm{HB}}$ chain is ergodic because it includes all moves of $\mathcal{W}$. To see this we only need to observe that by Lemma 3, any pair of vertices which are connected by an edge in $\mathrm{P}(r,c)$ can differ in at most m columns (of the matrix representation of the vertices). Clearly $b_m \geq m$. Therefore $\mathcal{M}_{\mathrm{HB}}$ is ergodic and converges to a stationary distribution ϖ on Ω. By definition, $\Pr_{\mathcal{M}_{\mathrm{HB}}}[Z,W] = \Pr_{\mathcal{M}_{\mathrm{HB}}}[W,Z]$ for any two vertices Z, W. Therefore the stationary distribution ϖ must be the uniform distribution on Ω.

To show rapid mixing of $\mathcal{M}_{\mathrm{HB}}$ in §6, we will use the multicommodity flow approach of Sinclair [30] (see also Diaconis and Stroock [10]), together with a construction based on ideas of Morris and Sinclair [25] which we develop in §5 below. Some definitions are necessary at this point.

For any ergodic Markov chain $\mathcal{M}$ on state space Ω, a multicommodity flow can be defined on the underlying graph $G(\mathcal{M})$ of the chain $\mathcal{M}$. The vertex set of $G(\mathcal{M})$ is Ω, and there is an edge $(u \to v)$ for every pair of states such that $\Pr_{\mathcal{M}}[u,v] > 0$ in $\mathcal{M}$ (observe that for our original chain $\mathcal{W}$, this "underlying graph" of the chain is exactly $G(\mathcal{W})$). For $x, y \in \Omega$, a *unit flow* from x to y is a set $\mathcal{P}_{x,y}$ of simple directed paths in $G(\mathcal{M})$ from x to y, such that each path $p \in \mathcal{P}_{x,y}$ has positive weight α_p, and the sum of the α_p over all $p \in \mathcal{P}_{x,y}$ is 1. A *multicommodity flow* is a family of unit flows $\mathcal{F} = \{\mathcal{P}_{x,y} : x, y \in \Omega\}$ containing a unit flow for every pair of states from Ω. The *length* $\mathcal{L}(\mathcal{F})$ of the multi-commodity flow $\mathcal{F}$ is $\mathcal{L}(\mathcal{F}) = \max_{x,y} \max\{|p| : p \in \mathcal{P}_{x,y}\}$, where $|p|$ denotes the edge length of p. For any edge e of $G(\mathcal{M})$, we define $\mathcal{F}(e)$ to be the sum of the α_p weights over all p such that $e \in p$ and $p \in \mathcal{P}_{x,y}$ for some $x, y \in \Omega$. Then the following theorem holds:

Theorem 5 (Sinclair [30]) *Let P be the transition matrix of an ergodic, reversible Markov chain $\mathcal{M}$ on Ω whose stationary distribution is the uniform distribution. Let $\mathcal{F}$ be a multicommodity flow on the graph $G(\mathcal{M})$. Then the mixing time of the chain is bounded above by*

$$\tau(\varepsilon) \leq 2|\Omega|^{-1}\mathcal{L}(\mathcal{F}) \max_e \frac{\mathcal{F}(e)}{\Pr_{\mathcal{M}}[e]}(\ln|\Omega| + \ln \varepsilon^{-1}) \qquad \square$$

In §5 we will present some techniques which we will use to define a multicommodity flow on the graph $G(\mathcal{M}_{\mathrm{HB}})$. In §6 we will prove that our construction does not overload any edge of the graph, and then prove that $\mathcal{M}_{\mathrm{HB}}$ mixes rapidly. Finally, in §7, we apply a comparison technique of Diaconis and Stroock [10] to extend our analysis to the random walk $\mathcal{W}$.

5 Balanced permutations

In order to construct a multicommodity flow on the graph $G(\mathcal{M}_{\mathrm{HB}})$, we follow the example of Morris and Sinclair [24, 25] for multidimensional knapsack and of Cryan *et al.* [3] for contingency tables and think of defining a path from a vertex X to a vertex Y by changing the value of a single column j (of the matrix representing the current vertex) from X_j to Y_j at each step. The procedure of changing columns of X to columns of Y will not ensure that the points along the path are vertices of $\mathrm{P}(r, c)$, or even that they lie inside $\mathrm{P}(r, c)$. However, in §6 we will show that if we define the path appropriately (using balanced almost-uniform permutations), each interim point on our path can be transformed to a vertex of $\mathrm{P}(r, c)$ by changing the values of a constant, but large, number of columns. This is why we originally analyse the heat-bath chain, which can modify b_m columns in one step.

To spread out the flow from X to Y, we will use a random permutation σ of the columns of the vertex, to determine the (random) order in which we change the columns of the vertex. We will spread flow along a particular path according to the probability with which a particular permutation of the columns is generated. Before we construct the particular (random) permutation which we will use to define the multicommodity flow for $\mathcal{M}_{\mathrm{HB}}$, we list some relevant definitions from the work of Morris and Sinclair [25, 24]. One of the properties that we will require of our random permutation is that it should approximate the uniform permutation in the following way:

Definition 6 (Morris & Sinclair[25]) *Let σ be a random permutation on $[n]$. Let $\lambda \in \mathbb{R}$ be such that $\lambda > 0$. We say that σ is λ-uniform if for every $k \in [n]$ and every $U \subseteq [n]$ with $|U| = k$,*

$$\Pr[\sigma\{1, \ldots, k\} = U] \leq \lambda \cdot \binom{n}{k}^{-1}.$$

The second property that will be important for our random permutation is that of *balance*:

Definition 7 (Morris & Sinclair[25]) *Let $w_1, \ldots, w_n$ be real m-dimensional weights (columns) with the mean $\mu \in \mathbb{R}^m$. Let $W = \sum_{j=1}^{n} w_j$. We say that a permutation σ on the set of columns is* ℓ-balanced *for some $\ell \in \mathbb{R}$, $\ell \geq 1$, if for every $k \in [n]$, and for every $i \in [m]$,*

$$\left| \sum_{j \in [k]} w^i_{\sigma(j)} - k\mu^i \right| \;\leq\; \ell \max_{j \in [n]} |w^i_j - \mu^i|.$$

This in turn implies the following:

$$\min\{W^i, 0\} - 2\ell \max_j |w^i_j| \;\leq\; \sum_{j=1}^{k} w^i_{\sigma(j)} \;\leq\; \max\{W^i, 0\} + 2\ell \max_j |w^i_j|.$$

A variant of balance is strong balance:

Definition 8 (Morris & Sinclair[25]) *Let $w_1, \ldots, w_n \in \mathbb{R}^m$ and let $\mu \in \mathbb{R}^m$ be the mean of these weights. A permutation σ is* strongly ℓ-balanced *for $\ell \in \mathbb{R}$, if for every $k \in [n]$, and for every $i \in [m]$, there is some set $S \subseteq [n]$ with $|S \oplus \sigma\{1, \ldots, k\}| \leq \ell$ such that the following two quantities have different signs (or either is 0):*

$$\sum_{j=1}^{k} w^i_{\pi(j)} - k\mu^i \qquad \sum_{j \in S} w^i_j - k\mu^i$$

In the work of Morris & Sinclair [25, 24], an explicit distinction is made between ℓ-balance and strong ℓ-balance. This distinction is highlighted because strong ℓ-balance is a constructive property, which allows the sign of $\sum_{j=1}^{k} w_j^i - k\mu^i$ to be altered by adding or deleting a fixed number of weights. We will see in Lemma 10 that in the case of one-dimensional weights, we can always convert a 0-balanced λ-uniform permutation into a strongly-balanced almost-uniform permutation, at the cost of making some constants worse. We will then construct a strongly-balanced almost-uniform permutation σ for the m-dimensional weights which appear in the vertices of the transportation polytope, by interleaving $m(m-1)/2$ of these strongly balanced one-dimensional permutations.

Let X and Y be any two vertices of $\mathrm{P}(r, c)$, so $|\Gamma_X \cup \Gamma_Y| \leq 2(m-1)$.

Let $\Gamma = \{j : X_j = Y_j\}$, $L = [n] \setminus (\Gamma_X \cup \Gamma_Y \cup \Gamma)$, and $\ell = |L|$.

In Lemma 12 we will construct a permutation σ on the m-dimensional columns of $X - Y$ for the indices in L. Lemma 12 builds on the work of Morris and Sinclair [25, 24]. Our construction will rely on the fact that each of the columns to be permuted will only contain two non-zero entries, as seen below.

Let $r'_i = r_i - \sum_{j\in\Gamma} X_j^i = r_i - \sum_{j\in\Gamma} Y_j^i$ for $i \in [m]$.

For $j \in L$, define the "weight vectors" $w_j = Y_j - X_j \in \mathbb{R}^m$, and let $\mu = \sum_{j\in L} w_j/\ell$ with coordinates μ^i ($i \in [m]$). By definition, for all $j \in L$, we know that both X_j and Y_j have exactly one non-zero and it is equal to c_j. Thus each w_j ($j \in L$) contains exactly two non-zeros, and these are of equal modulus but opposite sign. We partition L according to the location of these two non-zeros. For each pair of rows $i \neq i'$, define

$$S_{i,i'} \;=\; S_{i',i} \;=\; \big\{j \in L : \{w_j^i, w_j^{i'}\} = \{-c_j, +c_j\}\big\},$$

Let $\ell_{i,i'} = |S_{i,i'}|$, and let $\mu_{i,i'} = \sum_{j\in S_{i,i'}} w_j^i/\ell_{i,i'}$ be the mean over $S_{i,i'}$ of the weights in row i. Note that $\mu_{i,i'} = -\mu_{i',i}$ for all $i, i' \in [m]$.

We will use results of Morris and Sinclair [24, 25] to help us define a suitable random permutation $\sigma_{i,i'}$ on each of the $S_{i,i'}$ sets. The first lemma that we need is:

Lemma 9 (Morris [24]) *Suppose we are given real weights $\{w_j\}_{j=1}^h$ with total $W = \sum_{j=1}^h w_j$. Let $M = \max_{j=1}^h |w_j|$. Suppose that $|W| \geq 21M$. Then there is a random permutation π_1 of $[h]$ that satisfies the following two conditions: For some universal constant $C > 1$, and each $1 \leq k \leq h$,*

(i) $\min\{0, W\} \;\leq\; \sum_{j=1}^{k} w_{\pi_1(j)} \;\leq\; \max\{W, 0\}$*;*

(ii) for every $U \subseteq [h]$ with $|U| = k$, $\Pr[\pi_1\{1, \ldots, k\} = U] \;\leq\; Ch^2\binom{h}{k}^{-1}$.

We say π_1 is a 0-balanced Ch^2-uniform permutation. □

From this we will deduce a statement more convenient for our application (*cf* Morris [24, Ch. 3]).

Lemma 10 *Let $\{w_j\}_{j=1}^h$ be a set of real numbers with mean $\mu = \sum_{j=1}^h w_j/h$. Let C be the constant from Lemma 9. Then there exists a random permutation π of $[h]$ such that, for each $1 \leq k \leq h$, both of the following properties hold:*

(i) there are sets $D_1, D_2 \subseteq [h]$ with $|D_1|, |D_2| \leq 42$ such that

$$\sum_{j\in[k]\oplus D_1} w_{\pi(j)} \leq k\mu, \qquad \sum_{j\in[k]\oplus D_2} w_{\pi(j)} \geq k\mu.$$

(ii) for every $U \subseteq [h]$ with $|U| = k$, $\Pr[\pi\{1, \ldots, k\} = U] \leq Ch^{23}\binom{h}{k}^{-1}$.

We call π a strongly 42-balanced Ch^{23}-uniform permutation.

Proof: Assume, by symmetry, that $\mu \geq 0$. We first show how to construct the permutation π so that property (i) is satisfied.
(i) If $h \leq 42$ we will let π be a random permutation of $[h]$. Let $D_1 = [k]$ and $D_2 = [h] \setminus [k]$. Clearly property (i) is satisfied.

Otherwise $h > 42$. Let Q_1 contain the indices of the 21 elements for which $(w_j - \mu)$ is greatest and Q_2 contain the indices of the 21 elements for which $(w_j - \mu)$ is smallest. There are two subcases:

(a) The first case is when $-\sum_{j\in Q_2}(w_j - \mu) \geq \sum_{j\in Q_1}(w_j - \mu)$. In this case we assume wlog that the indices of Q_2 are the indices $[h-20, h]$, and we let π be the identity permutation on these 21 elements (the weights w_j for $j \in Q_2$ will be the last 21 elements of our permutation).

We will apply Lemma 9 to the set of weights $\{w_j - \mu\}_{j\in[h]\setminus Q_2}$ to construct our permutation π on the $\{w_j\}_{j\in[h]\setminus Q_2}$ weights. Note that $W = \sum_{j\in[h]\setminus Q_2}(w_j - \mu) = -\sum_{j\in Q_2}(w_j - \mu) \geq \sum_{j\in Q_1}(w_j - \mu)$. Also note that we are guaranteed that $W \geq 0$. For every $j \in [h]\setminus(Q_1 \cup Q_2)$, we have $21|w_j - \mu| \leq W$. For now, assume that $21|w_j - \mu| \leq W$ for $j \in Q_1$, so that we have $W \geq 21M$, where $M = \max_{j\in[h]\setminus Q_2} |w_j - \mu|$. (We will show how to remove this assumption below).

We have already constructed π for $j \in Q_2$. Let π be the permutation π_1 of Lemma 9 on the weights $\{w_j - \mu\}$ for $j \in [h] \setminus Q_2 = [h - 21]$. If $k \leq 21$, take $D_1 = [k]$, $D_2 = \emptyset$. Observe that property (i) is satisfied. If $21 < k \leq h - 21$, property (i) of π_1 gives

$$\begin{aligned} 0 \;\leq\; \sum_{j=1}^{k}(w_{\pi(j)} - \mu) \;&\leq\; \sum_{j=1}^{h-21}(w_{\pi(j)} - \mu) \\ &=\; -\sum_{j=h-20}^{h}(w_{\pi(j)} - \mu). \end{aligned}$$

We immediately have $\sum_{j=1}^{k} w_{\pi(j)} \geq k\mu$, so we can take $D_2 = \emptyset$. Also, since the above inequalities are true for all $k \leq h - 21$, we have

$$\sum_{j=1}^{k-21}(w_{\pi(j)} - \mu) + \sum_{j=h-20}^{h}(w_{\pi(j)} - \mu) \leq 0.$$

Then, setting $D_1 = [k-20, k] \cup [h-20, h]$, we have $\sum_{j\in[k]\oplus D_1} w_{\pi(j)} \leq k\mu$. Observe that in this case we also have property (i).

If $k > h - 21$, the conclusion follows easily from

$$\sum_{j=1}^{h-21}(w_{\pi(j)} - \mu) \geq 0, \quad \sum_{j=1}^{h}(w_{\pi(j)} - \mu) = 0.$$

Note that in all the cases above, in fact we have $|D_1 \cup D_2| \leq 21$, except for D_1 when $21 < k \leq h-21$ (then $|D_1| \leq 42$). We now show how to deal with the possibility that there is some $j \in Q_1$ such that $21|w_j - \mu| > W$. When we construct the permutation π_1, we replace the weights $\{w_j - \mu\}_{j\in Q_1}$ by $\{w'_j - \mu\}_{j\in Q_1}$, where $w'_j = \sum_{j\in Q_1} w_j/21$ for all $j \in Q_1$. Then W does not change and the condition $21M \leq W$ is satisfied. When π_1 has been constructed we replace the dummy weights by the original weights in random order. Then we need to exchange at most 11 weights (exchanging some elements of Q_1 for others) to obtain D_1, D_2 sets satisfying condition (i) for the original weights. Moreover, for

$21 < k \leq h - 21$, we can define $D_1 = [h-20, h] \cup (Q_1 \cap [k]) \cup [k - 20 + |\{j : j \in Q_1 \cap [k]\}|, k]$ to ensure (i) holds. Therefore we still have $|D_1|, |D_2| \leq 42$, as claimed.

(b) The second case occurs if $-\sum_{j \in Q_2}(w_j - \mu) < \sum_{j \in Q_1}(w_j - \mu)$. In this case we assume wlog that the set of indices Q_1 is the set $[h-20, h]$, and we let π on $[h-20, h]$ be the identity permutation. Then, when we apply Lemma 9 to the set of weights $\{w_j - \mu\}_{j \in [h] \setminus Q_1}$, the total of the weights W is negative. Again, assuming for now that $|W| \geq 21 \max_{j \in [h] \setminus Q_1} |w_j - \mu|$, we let π be the permutation π_1 of Lemma 9 on $[h-21]$.

For $k \leq 21$, we take $D_1 = \emptyset$ and $D_2 = [h-20, h]$. For $21 < k \leq h - 21$, we use the fact that condition (i) of Lemma 9 holds for $k - 21$ in a similar way to that described above, and we take $D_1 = \emptyset$ and $D_2 = [k-20, k] \cup [h-20, h]$. The case $k > h - 21$ is similar to case (a). Finally, we treat the possibility that there exists $j \in Q_2$ with $21|w_j - \mu| > |W|$ in a similar way to case (a).
(ii) We now show that property (ii) holds for π. If $h \leq 42$, the property follows from the fact that π is a random permutation. Otherwise, if $k \leq 21$ or $k > h - 21$, the statement is trivially true. In all other cases, property (ii) of π_1 implies $\Pr[\pi\{1, \ldots, k\} = U] \leq C(h-21)^2 \binom{h-21}{k}^{-1} \leq Ch^{23} \binom{h}{k}^{-1}$. □

We remark that it would be possible to improve the constants in Lemma 10 by proving it directly, rather than starting from Lemma 9. Moreover, even without doing this, there are many improvements we could make if we were slightly more careful with the constants in our proofs (for example, in Lemma 10, for every k, we have $D_1 = \emptyset$ or $D_2 = \emptyset$, even though we never use this fact). However, we are not aiming to optimize the constants, so we have not made use of these observations.

We apply the construction of Lemma 10 to each of the non-empty sets $S_{i,i'}$ separately to produce permutations $\sigma_{i,i'}$. Since the entries in rows i, i' are equal and opposite, for any $J \subseteq S_{i,i'}$, we have $\sum_{j \in J} w_j^i = -\sum_{j \in J} w_j^{i'}$. Hence $\sum_{j \in J} w_j^i \geq k\mu_{i,i'}$ iff $\sum_{j \in J} w_j^{i'} \leq k\mu_{i',i}$. Therefore, to have both inequalities in the same direction, we need at most 42 "corrections" in exactly one of the rows.

We now consider how to interleave the $\sigma_{i,i'}$ to produce an overall permutation σ of L. For notational simplicity, suppose we are interleaving q sets of size $\nu_i > 0$, $i \in [q]$, with $\nu = \sum_{i=1}^q \nu_i$. Let $\alpha_i = \nu_i / \nu$, so $\sum_{i=1}^q \alpha_i = 1$. Consider the following algorithm.

interleave
$k_1, k_2, \ldots, k_q \leftarrow 0$.
while $k = \sum_{i=1}^q k_i < \nu$ **do**
  if $i^* = \arg\max_{i=1}^q (\alpha_i k - k_i)$
  then $k_{i^*} \leftarrow k_{i^*} + 1$.

We now prove some useful properties of **interleave**.

Lemma 11 *For all $k \in [0, \nu]$, $k_i \leq \lceil \alpha_i k \rceil \leq \nu_i$, $i \in [q]$, and $\sum_{i=1}^q |k_i - \alpha_i k| < 2(q-1)$.*

Proof: First note that $\lceil \alpha_i k \rceil \leq \nu_i$. Otherwise $\lceil \alpha_i k \rceil = \lceil \nu_i k / \nu \rceil > \nu_i$, giving $k > \nu$, a contradiction.

Let $\gamma_i(k) = \alpha_i k - k_i$. Note that $\sum_{i=1}^q \gamma_i(k) = 0$, so $\gamma_{i^*}(k) \geq 0$. Then $\gamma_i(k+1) = \gamma_i(k) + \alpha_i > \gamma_i(k)$ ($i \neq i^*$), but $\gamma_{i^*}(k+1) = \gamma_{i^*}(k) - (1 - \alpha_{i^*}) > -1$. Since $\gamma_i(0) = 0$ for all i, it follows by induction that $\gamma_i(k) > -1$ for all i, k. Now $k_i \leq \lceil \alpha_i k \rceil$ follows immediately. Also, since $\sum_{i=1}^q \gamma_i(k) = 0$ and $\gamma_{i^*}(k) \geq 0$,

$$\sum_{i=1}^q |k_i - \alpha_i k| \;=\; \sum_{i=1}^q |\gamma_i(k)| \;=\; 2 \sum_{\gamma_i < 0} |\gamma_i(k)|$$

This is at most $2 \sum_{i \neq i^*} 1 = 2(q-1)$. □

We interleave the $\sigma_{i,i'}$ according to the procedure above to produce the permutation σ. The following Lemma proves that σ is a strongly $23m^2$-balanced $C_m\ell^{14m^2}$-uniform permutation. The q_m-balanced p_m-uniform permutations of [24, 25] for a set of ℓ general m-dimensional weights only have bounds of the form $q_m \in 2^{O(m)}$ and $p_m \in \ell^{O(m^2)}$. Therefore by exploiting the special structure of the vertices of the transportation polytope, we are able to prove better bounds on our "balance" constant. This will positively influence the asymptotic bounds that we prove on the mixing time of $\mathcal{M}_{\mathrm{HB}}$ in §6.

Lemma 12 *For every m, there is some constant C_m such that σ has the following properties.*

(i) For all $i \in [m]$, $k \in [\ell]$, there exist two sets D_1^i and D_2^i satisfying $|\bigcup_{i=1}^m D_1^i| < 23m^2$, $|\bigcup_{i=1}^m D_2^i| < 23m^2$ such that

$$\sum_{j \in [k] \oplus D_1^i} w^i_{\sigma(j)} \leq k\mu^i, \qquad \sum_{j \in [k] \oplus D_2^i} w^i_{\sigma(j)} \geq k\mu^i.$$

(ii) For any $U \subseteq [\ell]$ with $|U| = k$, $\Pr[\sigma\{1,\ldots,k\} = U] \leq C_m \ell^{14m^2} \binom{\ell}{k}^{-1}$.

Proof: (i): We prove only the first inequality, the other being entirely similar. Suppose the values at step k in **interleave** are $k_{i,i'}$, and $\alpha_{i,i'} = \ell_{i,i'}/\ell$, for each $i' \neq i$. Define $k^*_{i,i'}$ to be $\lfloor k\alpha_{i,i'} \rfloor$ if $\mu_{i,i'} \geq 0$, and $\lceil k\alpha_{i,i'} \rceil$ otherwise. Using Lemma 11, observe that $\sum_{i,i'} |k^*_{i,i'} - k_{i,i'}|$ is at most

$$\sum_{i,i'} \left(|k^*_{i,i'} - k\alpha_{i,i'}| + |k_{i,i'} - k\alpha_{i,i'}| \right)$$

which is less than $\binom{m}{2} + 2\binom{m}{2} = 3\binom{m}{2}$.

Let $D_1^{i,i'}$ be the set associated with $\sigma_{i,i'}$, $k^*_{i,i'}$ such that $\sum_{j \in [k^*_{i,i'}] \oplus D_1^{i,i'}} w_{\sigma_{i,i'}(j)} \leq k^*_{i,i'}\,\mu_{i,i'}$, and let $I_1^{i,i'}$ be the interval $[k^*_{i,i'}+1, k_{i,i'}]$, if $k^*_{i,i'} < k_{i,i'}$, or $[k_{i,i'}+1, k^*_{i,i'}]$ otherwise. Let $D_1^i = \bigcup_{i'} (D_1^{i,i'} \cup I_1^{i,i'})$. Then, using Lemma 10, $|\bigcup_{i=1}^m D_1^i| < 42\binom{m}{2} + 3\binom{m}{2} < 45m^2/2$.

Also

$$\begin{aligned} \sum_{j \in [k] \oplus D_2^i} w^i_{\sigma(j)} &\leq \sum_{i'} k^*_{i,i'}\,\mu_{i,i'} \\ &\leq \sum_{i'} k\,\ell_{i,i'}\,\mu_{i,i'}/\ell \;=\; k\mu^i. \end{aligned}$$

(ii): Let τ^* be the random permutation we get when we apply **interleave** to the collection of uniform distributions $\tau_{i,i'}$ on $S_{i,i'}$ for every i, i'. Let τ represent the uniform distribution on $[\ell]$. We will first bound $\Pr[\tau^*\{1,\ldots,k\} = U]$ in terms of $\Pr[\tau\{1,\ldots,k\} = U]$ $(= \binom{\ell}{k})$, and then use the almost-uniformity of the $\sigma_{i,i'}$ to give the result.

Let $K_{i,i'}$ be a random variable equal to the number of elements of $S_{i,i'}$ in the prefix $\tau\{1,\ldots,k\}$. We will show that with high probability $K_{i,i'}$ is not too far from $\alpha_{i,i'}k$. Precisely, we have

$$\begin{aligned} \Pr_\tau \left[|K_{i,i'} - \alpha_{i,i'}k| \geq \sqrt{k\ln(\ell)} \right] &\leq 2e^{\frac{-2k(\ln \ell)}{k}} \\ &= 2\ell^{-2} \end{aligned}$$

by a single application of the Chernoff bound (see McDiarmid [22]). Summing over all k and all i, i' ($\binom{m}{2}$ in total), we find that under the uniform distribution τ,

$$|K_{i,i'} - \alpha_{i,i'}k| \;\leq\; \sqrt{k\ln(\ell)} \tag{5}$$

holds for all k, and all i, i' with probability at least $1 - m(m-1)/\ell$. Assume wlog that $\ell \geq 14m^2$, therefore (5) holds with probability at least $1/2$.

Let τ' be the uniform distribution on the permutations that satisfy (5) (for all k, all i, i'). Note the probability of any event in τ' is at most twice its probability in the uniform distribution τ. Also, since the integer variable $K_{i,i'}$ has maximum probability of taking values $\{\lfloor \alpha_{i,i'} k \rfloor, \lceil \alpha_{i,i'} k \rceil\}$, we have

(a) $\Pr_{\tau'}[K_{i,i'} = q_{i,i'}] \geq (\sqrt{k \ln \ell})^{-1}$ for $q_{i,i'} \in \{\lfloor \alpha_{i,i'} k \rfloor, \lceil \alpha_{i,i'} k \rceil\}$

Now we are ready to bound $\Pr[\sigma^*\{1, \ldots, k\} = U]$, where U decomposes into $U_{i,i'}$ with $|U_{i,i'}| = k_{i,i'}$. We only need the following (with the binomial coefficient defined (by continuation) for non-integer arguments):

(b) $\binom{\ell_{i,i'}}{\alpha_{i,i'} k} \leq \ell^{|k_{i,i'} - \alpha_{i,i'} k| + 1} \binom{\ell_{i,i'}}{k_{i,i'}}$

Using (a) and (b) with an application of Lemma 11, we find that $\Pr_\tau[K_{i,i'} = k_{i,i'} \ \forall i, i']$ is

$$\begin{aligned} &\geq \quad (k \ln \ell)^{-m^2/4} \Big(\prod_{i,i'} \ell^{-|k_{i,i'} - \alpha_{i,i'} k| - 1}\Big)/2 \\ &\geq \quad \ell^{-m^2/2} \ell^{-3m^2/2}/2 \ \geq \ \ell^{-2m^2}/2 \end{aligned}$$

So $\Pr[\sigma^*\{1, \ldots, k\} = U] \leq 2\ell^{2m^2} \binom{\ell}{k}$. Then applying Lemma 9 to each of the $S_{i,i'}$, we have

$$\Pr[\sigma\{1, \ldots, k\} = U] \leq 2\ell^{2m^2} C^{m^2} \ell^{23m^2/2} \binom{\ell}{k}^{-1}$$

and we have $C_m \ell^{14m^2}$-uniformity. □

6 Analysis of the heat bath

We now apply Theorem 5 to prove that $\mathcal{M}_{\mathrm{HB}}$ is rapidly mixing.

In a similar manner to [3] (see also [25]), we use the permutation σ constructed by **interleave** to route flow from X to Y. We apply σ to the columns in L and for every $k \in [\ell]$, we define the matrix $Z(k)$ as the $m \times n$ matrix where we set

$$Z(k)_j = \begin{cases} Y_j & j \in \sigma\{1, \ldots, k\} \\ X_j & j \in \sigma\{k+1, \ldots, \ell\} \cup ([n] \setminus L) \end{cases}$$

Conceptually, we think of the sequence of matrices $X = Z(0), Z(1), \ldots, Z(k), \ldots, Z(\ell)$, as defining a random path from X to Y in $G(\mathcal{M}_{\mathrm{HB}})$, along which we assign some fraction of flow determined by σ. However, if $Z(k)$ is any intermediate matrix obtained in this way, in general it will not be a vertex of $\mathrm{P}(r, c)$ (or even a point inside $\mathrm{P}(r, c)$). We will presently show how to modify the $Z(k)$ matrices to obtain $Z(k)'$ matrices which are vertices of $\mathrm{P}(r, c)$. For every $k \in [\ell]$, we also define a mirror image $\bar{Z}(k)$ of $Z(k)$, called an "encoding", in the following way:

$$\bar{Z}(k)_j = \begin{cases} X_j & j \in \sigma\{1, \ldots, k\} \\ Y_j & j \in \sigma\{k+1, \ldots, \ell\} \cup ([n] \setminus L) \end{cases}$$

This matrix $\bar{Z}(k)$ is not used in constructing the multicommodity flow for $\mathcal{M}_{\mathrm{HB}}$, but is a useful concept when we come to bound the amount of multicommodity flow that can lie on an edge of $G(\mathcal{M}_{\mathrm{HB}})$.

The $\bar{Z}(k)$ matrices do not necessarily correspond to vertices of $\mathrm{P}(r,c)$. We now show that if we delete only a constant number of columns from either $Z(k)$ or $\bar{Z}(k)$, then the matrix can be completed to a vertex of $\mathrm{P}(r,c)$, denoted by $Z'(k)$ and $\bar{Z}'(k)$ respectively. Moreover, both X and Y can be reconstructed from $Z'(k)$ and $\bar{Z}'(k)$ using a suitably small amount of information.

Let $D_1 = \bigcup_{i=1}^{m} D_1^i$ and $D_2 = \bigcup_{i=1}^{m} D_2^i$. Since $X_j^i, Y_j^i \geq 0$ for all i, j, for each $i \in [m]$ we have

$$\begin{aligned}\sum_{j \in (L \setminus [k]) \setminus D_1} X^i_{\sigma(j)} + \sum_{j \in [k] \setminus D_1} Y^i_{\sigma(j)} &\leq \\ \sum_{j \in L \setminus ([k] \oplus D_1^i)} X^i_{\sigma(j)} + \sum_{j \in [k] \oplus D_1^i} Y^i_{\sigma(j)} &= \\ \sum_{j \in L} X_j^i + \sum_{j \in [k] \oplus D_1^i} w^i_{\sigma(j)}.&\end{aligned}$$

By Lemma 12, we also have

$$\begin{aligned}\sum_{j \in L} X_j^i + \sum_{j \in [k] \oplus D_1^i} w^i_{\sigma(j)} &\leq \sum_{j \in L} X_j^i + k\mu^i \\ &= \tfrac{\ell - k}{\ell} \sum_{j \in L} X_j^i + \tfrac{k}{\ell} \sum_{j \in L} Y_j^i\end{aligned}$$

which is at most r'_i, as defined in §5.

Hence, if we "delete" the columns in $D_1 \cup \Gamma_X \cup \Gamma_Y$, we obtain partial row sums for the deleted columns, where each partial row sum has size at least $r_i - r'_i$, which is non-negative for every i. Thus $Z(k)$ can be completed to a vertex of $\mathrm{P}(r,c)$ by redefining the columns of $D_1 \cup \Gamma_X \cup \Gamma_Y$ according to the "northwestern corner rule" [16]. Hence we can map $Z(k)$, for every $k \in [\ell]$, to a vertex $Z(k)'$ of $\mathrm{P}(r,c)$. This necessitates changing the values of some of the columns in $D_1 \cup \Gamma_X \cup \Gamma_Y$. By Lemma 12, $|D_1| \leq 23m^2$, hence $Z(k)'$ may differ from $Z(k)$ in at most $23m^2 + 2(m-1)$ columns in total, for any $k \in [\ell]$. Recall that for every $k \in [\ell - 1]$, $Z(k)$ and $Z(k+1)$ differ in one column. Hence $Z(k)'$ and $Z(k+1)'$ differ in at most $46m^2 + 2(m-1) + 1$ columns, which for $m \geq 2$ is at most $47m^2$. So $Z(k)' \to Z(k+1)'$ is a transition of $\mathcal{M}_{\mathrm{HB}}$ for every $k \in [\ell - 1]$ (this justifies our choice of b_m in §4). Also $X \to Z(0)'$ and $Z(\ell)' \to Y$ are transitions of $\mathcal{M}_{\mathrm{HB}}$. Hence we obtain a path $X = Z(0), Z(1)', \ldots, Z(\ell)', Y$ in $G(\mathcal{M}_{\mathrm{HB}})$ between X and Y. The proof for $\bar{Z}(k)$ is identical, by interchanging X_j^i with Y_j^i, w_j^i with $-w_j^i$, D_1 with D_2, and using the lower bound in (i) of Lemma 12.

Now suppose we are given $Z'(k)$, $\bar{Z}'(k)$ and we wish to recover X, Y. Let us assume, using the uniformity property of σ, that we are given $U = \sigma\{1, \ldots, k\}$ (we will incorporate this into our analysis later). We still need to know the "deleted" columns D_1, D_2, Γ_X, Γ_Y, but there are at most $\binom{n}{23m^2}^2 \binom{n}{m-1}^2 < n^{47m^2}$ ways of selecting these sets. We can easily reconstruct both X and Y except for the deleted columns. However, there are at most $47m^2$ such columns, and X and Y are both vertices. Moreover, since the deleted columns are the only columns which may contain more than one non-zero cell, therefore we can complete X to a vertex iff the values we choose for the deleted columns $D_1 \cup \Gamma_X \cup \Gamma_Y$ define a vertex on the induced transportation polytope (of dimension at most $(m-1)(24m^2 - 1)$) on the deleted columns. By Lemma 2, there are at most $(em)^{24m^2+m-1}$ possible ways of completing these columns for X. Similarly, there are at most $(em)^{24m^2+m-1}$ ways of completing the deleted columns $D_2 \cup \Gamma_X \cup \Gamma_Y$ for Y. So there are at most $(em)^{49m^2} n^{47m^2}$ ways of augmenting the encoding so that we can uniquely identify X and Y from $Z'(k)$, $\bar{Z}'(k)$ (assuming we have been given $U = \sigma\{1, \ldots, k\}$).

We can now bound the flow through any state $Z \in \Omega$. There are $|\Omega|$ ways of choosing $\bar{Z}$, $\binom{n}{k}$ ways of choosing $|U|$ and $(em)^{49m^2}n^{47m^2}$ ways of specifying the additional information needed to uniquely identify X and Y. However, by the uniformity of σ, $\Pr(\sigma[k] = U) \leq C_m n^{14m^2}\binom{n}{k}^{-1}$. Hence the flow through any state may be bounded by

$$|\Omega| \times \binom{n}{k} \times (em)^{49m^2}n^{47m^2} \times C_m n^{14m^2}\binom{n}{k}^{-1} = \tag{6}$$

$$O(n^{61m^2})|\Omega|. \tag{7}$$

Observe that in this analysis of the maximum flow through any state of Ω, we have obtained a term n^{47m^2} which derives from the constant $23m^2$ of the strongly $23m^2$-balanced $C_m\ell^{14m^2}$-uniform permutation σ. If we had used the permutations of Morris & Sinclair[24, 25] for general m-dimensional weights (which are strongly $2^{O(m)}$-balanced), we would have obtained a $n^{2^{O(m)}}$ term instead. This was our motivation for exploiting the structure of the vertices to obtain an improved strongly-balanced almost-uniform permutation in §5.

In order to apply Theorem 5, we must bound the flow through any *edge* of $G(\mathcal{M}_{\mathrm{HB}})$. We observe that for the flow $\mathcal{F}$ which we have constructed, for any edge $e = (Z, W)$, (7) implies:

$$\mathcal{F}(e) \quad \leq \quad O(n^{61m^2}).$$

By construction of our multi-commodity flow, $\mathcal{L}(\mathcal{F}) \leq n$. Therefore, by Theorem 5,

$$\tau_{\mathcal{M}_{\mathrm{HB}}}(\varepsilon) \quad \leq \quad 2|\Omega|^{-1} \cdot n \cdot \max_e \left\{ \frac{O(n^{61m^2})|\Omega|}{\Pr_{\mathcal{M}_{\mathrm{HB}}}[e]} \right\} (\ln|\Omega| + \ln\epsilon^{-1}) \tag{8}$$

$$= \quad O(n^{61m^2+1}) \cdot (\min_e \Pr_{\mathcal{M}_{\mathrm{HB}}}[e])^{-1}(\ln|\Omega| + \ln\epsilon^{-1}). \tag{9}$$

Now observe that $e = (Z \to W)$ is an edge of $G(\mathcal{M}_{\mathrm{HB}})$ if and only if Z and W are vertices of $\mathrm{P}(r, c)$ and there is some set B of destinations such that $\Gamma_Z \subseteq B$, $|B| \leq b_m$ and Z and W only differ on the set B. The definition of $\mathcal{M}_{\mathrm{HB}}$ implies that this particular set of destinations B is chosen from B with probability at least $\binom{n}{b_m}^{-1}$. Also, by definition of $\mathcal{M}_{\mathrm{HB}}$, once the window B has been chosen, we choose the next state uniformly at random, by choosing from all possible assignments to B which give a vertex of $\mathrm{P}(r, c)$. It is not difficult to show that this is the case if and only if the assignment to the destinations of B is a vertex of the $(m-1) \times (b_m - 1)$-dimensional polytope $\mathrm{P}(s, d)$ induced by the set of values of Z_j for the destinations $j \in B$ (see, for example, Hadley [16]). By Lemma 2 there are at most $(e \cdot m)^{b_m+m-1}$ vertices of this smaller polytope. Therefore we can bound the probability of a transition from Z to any W in $\mathcal{M}_{\mathrm{HB}}$ as follows:

$$\Pr_{\mathcal{M}_{\mathrm{HB}}}[Z, W] \quad \geq \quad \binom{n}{b_m}^{-1} (em)^{-b_m-m+1}.$$

Therefore, substituting into (9), we have the following bound on the mixing time of $\mathcal{M}_{\mathrm{HB}}$:

$$\tau_{\mathcal{M}_{\mathrm{HB}}}(\varepsilon) \quad = \quad O(n^{b_m+61m^2+1})(\ln|\Omega| + \ln\epsilon^{-1}) \quad = \quad O(n^{109m^2})\ln\epsilon^{-1}, \tag{10}$$

where in the last step, we use the facts that $\binom{n}{b_m} \leq n^{b_m} = n^{47m^2}$ and $|\Omega| \leq (em)^{n+m-1}$.

Remark: In the conference version of this paper [4], we omitted the $\Pr_{\mathcal{M}_{\mathrm{HB}}}[Z, W]$ term when bounding $\tau_{\mathcal{M}_{\mathrm{HB}}}$. Hence we erroneously claimed a bound of $O(n^{62m^2})\ln\epsilon^{-1}$ for the mixing time of $\mathcal{M}_{\mathrm{HB}}$. However, because we are able to define $b_m = 47m^2$ in this paper (we carelessly used $b_m = 94m^2$ in [4]), the bound we derive for the random walk in §7 is the same as in [4]. We believe that the mixing time of both chains is far better than our bounds, but we have not attempted to optimize the constants.

7 Analysis of the random walk

We now show that the natural random walk $\mathcal{W}$ defined in §2 is rapidly mixing. We prove this using the comparison theorem of Diaconis and Saloff-Coste [8]. For a Markov chain $\mathcal{M}$ on a state space Ω, let $\ker(\mathcal{M})$ denote the set of pairs $(X, Y) \in \Omega^2$ such that $\Pr_{\mathcal{M}}[X, Y] > 0$.

Theorem 13 (Diaconis and Saloff-Coste [8]) *Let Ω be a set of discrete structures. Let $\mathcal{M}$ and $\mathcal{M}'$ be two ergodic and reversible Markov chains which both converge to the uniform distribution on Ω. Suppose the mixing time of $\mathcal{M}$ is bounded above by $\tau_{\mathcal{M}}(\epsilon)$.*

Suppose we are given a set $\mathcal{P} = \{p_{X,Y} : (X, Y) \in \ker(\mathcal{M})\}$ containing a canonical path *$p_{X,Y}$ connecting X to Y on $G(\mathcal{M}')$, for every pair of states $(X, Y) \in \ker(\mathcal{M})$. For $(Z, W) \in \ker(\mathcal{M}')$, define*

$$A_{Z,W} = \frac{1}{\Pr_{\mathcal{M}'}[Z,W]} \sum_{\substack{(X,Y)\in\ker(\mathcal{M}) \\ (Z,W)\in p_{X,Y}}} |p_{X,Y}| \Pr_{\mathcal{M}}[X, Y].$$

Then the mixing time $\tau_{\mathcal{M}'}(\epsilon)$ is

$$O\Big(\tau_{\mathcal{M}}(\epsilon) \ln(|\Omega|) \max_{(Z,W)\in\ker(\mathcal{M}')} A_{Z,W}\Big).$$

We now use Theorem 13 to bound the mixing time of $\mathcal{W}$ in terms of the mixing time of $\mathcal{M}_{\mathrm{HB}}$.

We construct a canonical path $p_{X,Y}$ on $G(\mathcal{W})$ for every pair of vertices $(X, Y) \in \ker(\mathcal{M}_{\mathrm{HB}})$. Recall that by our definition of $\mathcal{M}_{\mathrm{HB}}$ in §4, for any pair $(X, Y) \in \ker(\mathcal{M}_{\mathrm{HB}})$, there exists a set $J_{X,Y}$ of at most b_m columns such that $j \in J_{X,Y}$ iff either $X_j \neq Y_j$ or $j \in \Gamma_X \cup \Gamma_Y$. Let $b = |J_{X,Y}|$. Let $\widehat{X}$ be the matrix consisting of the columns X_j for $j \in J_{X,Y}$, and let $\widehat{Y}$ be the matrix consisting of the columns Y_j for $j \in J_{X,Y}$. For every $i \in [m]$, let s_i be the source quantity for the ith row of $\widehat{X}$. By definition of $J_{X,Y}$, s_i is also the source quantity for the ith row of $\widehat{Y}$. Let $\mathrm{P}(s, c)$ be the $(m-1)(b-1)$-dimensional transportation polytope with source quantities s_i for $i \in [m]$ and destination quantities c_j for $j \in J_{X,Y}$. $\widehat{X}$ and $\widehat{Y}$ are both vertices of $\mathrm{P}(s, c)$.

By Lemma 2, there are at most $(em)^{b_m+m-1}$ vertices of the $(m-1)(b-1)$-dimensional transportation polytope $\mathrm{P}(s, c)$. Also by definition of $J_{X,Y}$ (if $j \notin J_{X,Y}$, then X_j has exactly one non-zero cell) any point $\widehat{Z}$ inside $\mathrm{P}(s, c)$ is a vertex of $\mathrm{P}(s, c)$ iff the point Z defined by

$$Z_j = \begin{cases} \widehat{Z}_j & \text{if } j \in J_{X,Y} \\ X_j & \text{if } j \notin J_{X,Y} \end{cases}$$

is a vertex of the original transportation polytope $\mathrm{P}(r, c)$ (see, for example, Hadley [16]).

It is a result of Balinski [1] that the connectivity of the edge-vertex graph of a polytope is equal to its dimension. Therefore there is a path $\widehat{X}(0) = \widehat{X}, \widehat{X}(1), \ldots, \widehat{X}(\ell-1), \widehat{X}(\ell) = \widehat{Y}$ connecting $\widehat{X}$ to $\widehat{Y}$ on the edge-vertex graph of the $(m-1)(b-1)$-dimensional transportation polytope. We use this path to define a sequence of points $X(0) = X, X(1), \ldots, X(i), \ldots, X(\ell) = Y$ in the original polytope $\mathrm{P}(r, c)$. For every $i \in [\ell]$, $X(i)$ is the matrix consisting of the columns X_j for $j \notin J_{X,Y}$ and the columns $\widehat{X}(i)_j$ for $j \in J_{X,Y}$. Also, $X(i)$ is a vertex of $\mathrm{P}(r, c)$ for every $i \in [\ell]$ and also $(X(i-1), X(i))$ is an edge of $\mathrm{P}(r, c)$ for every $i \in [\ell]$ (see Hadley [16]). Therefore the path $p_{X,Y}$ given by $X(0) = X, X(1), \ldots, X(\ell) = Y$ is a path of length at most $(em)^{b_m+m-1}$ (see Lemma 2) in the edge-vertex graph $G(\mathcal{W})$.

Let $\mathcal{P} = \{p_{X,Y} : X, Y \in \ker(\mathcal{M}_{\mathrm{HB}})\}$. Now we show that this set of canonical paths does not overload any edge (Z, W) of $G(\mathcal{W})$. Partition the elements (X, Y) of $\ker(\mathcal{M}_{\mathrm{HB}})$ according to the set B of b_m columns used to move from X to Y. We will write $(X, Y) \in \mathcal{M}_{\mathrm{HB}}(B)$ if (X, Y) is an element of $\ker(\mathcal{M}_{\mathrm{HB}})$ and X and Y differ only on the columns in B. Then we find that $A_{Z,W}$ is at most

$$\frac{1}{\mathrm{Pr}_{\mathcal{W}}[Z,W]} \sum_{\substack{B\subset[n],\\ |B|=b_m}} \sum_{\substack{(X,Y)\in\ker(\mathcal{M}_{\mathrm{HB}}(B))\\ (Z,W)\in p_{X,Y}}} |p_{X,Y}|\mathrm{Pr}_{\mathcal{M}}[X,Y]$$

which is at most

$$\frac{1}{\mathrm{Pr}_{\mathcal{W}}[Z,W]} \sum_{\substack{B\subset[n],\\ |B|=b_m}} \sum_{\substack{(X,Y)\in\ker(\mathcal{M}_{\mathrm{HB}}(B))\\ (Z,W)\in p_{X,Y}}} (em)^{b_m+m-1}\mathrm{Pr}_{\mathcal{M}}[X,Y].$$

However, once we fix a set of columns B, we know that there are at most $(em)^{b_m+m-1}$ different vertices of $\mathrm{P}(r, c)$ which agree with Z (and W) on all columns $j \notin B$. Using this, and the fact that $\mathrm{Pr}_{\mathcal{M}}[X, Y] \leq 1$, we find

$$\begin{aligned} A_{Z,W} &\leq \frac{1}{\mathrm{Pr}_{\mathcal{W}}[Z,W]} \sum_{B\subset[n],|B|=b_m} (em)^{3(b_m+m-1)} \\ A_{Z,W} &\leq 2n^m \binom{n}{b_m} (em)^{3(b_m+m-1)} \end{aligned}$$

for any $(Z, W) \in \ker(\mathcal{W})$, using $\mathrm{Pr}_{\mathcal{W}}[Z, W] = 1/2d_m = 1/2n^m$. Using $b_m = 47m^2$, we have

$$A_{Z,W} \leq 2n^m n^{47m^2} (em)^{3(b_m+m-1)}.$$

Applying Theorem 13 and (10), and using $|\Omega| \leq (em)^{n+m-1}$ (Lemma 2), we find that

$$\begin{aligned} \tau_{\mathcal{W}}(\varepsilon) &\in O\big(\tau_{\mathcal{M}}(\varepsilon)\ln(|\Omega|) \max_{(Z,W)\in\ker(\mathcal{M}')} A_{Z,W}\big). \\ &\in O\big(n^{156m^2+m+1}\ln(\varepsilon^{-1})\big) = O(n^{157m^2})\ln(\varepsilon^{-1}). \end{aligned}$$

8 Approximate counting

It is not difficult to turn our sampling algorithm into a fully polynomial randomized approximation scheme (*fpras*) for counting the number of vertices $|\Omega|$ of $\mathrm{P}(r, c)$. We will briefly sketch the method.

If $n < 2(m + 1)$, determine $|\Omega|$ by complete enumeration. (See, for example, [11].) Otherwise, at least $n - m + 1$ columns j have the single entry c_j at any vertex, and each column has only m cells. Therefore some particular cell (i^*, j^*) contains c_{j^*} with probability at least $(n-m+1)/mn \geq 1/(2m)$. Identify such a cell, and estimate the proportion p of all vertices in which it contains c_{j^*}, by sampling. But $p = |\Omega'|/|\Omega|$, where $|\Omega'|$ is the number of vertices of the transportation polytope $\mathrm{P}(r', c')$, when we define $c' = (c_1, \ldots, c_{j^*-1}, c_{j^*+1}, \ldots, c_n)$, $r'_{i^*} = r_{i^*} - c_{j^*}$, and $r'_i = r_i$, $i = [m] \setminus \{i^*\}$. We estimate $|\Omega'|$ recursively, and estimate $|\Omega|$ by $|\Omega'|/p$.

References

[1] M. Balinski, On the graph structure of convex polyhedra in n-space. *Pacific Journal of Mathematics*, **11**, pp. 431–434, 1961.

[2] G. Brightwell, J. van den Heuvel, L. Stougie, A linear bound on the diameter of the transportation polytope, to appear in *Combinatorica*.

[3] M. Cryan, M. Dyer, L.A. Goldberg, M. Jerrum and R. Martin, Rapidly mixing Markov chains for sampling contingency tables with a constant number of rows. Accepted to appear in *SIAM Journal on Computing*. A short version appeared in *Proceedings of the 43rd Annual Symposium on Foundations of Computer Science*, pp. 711–720, 2002.

[4] M. Cryan, M. Dyer, H. Müller and L. Stougie, Random walks on the vertices of transportation polytopes with constant number of sources. *Proceedings of the 14th ACM-SIAM Annual Symposium on Discrete Algorithms* (SODA '03), pp. 330-339, 2003.

[5] G.B. Dantzig, Application of the simplex method to a transportation problem, In Tj. C. Koopmans (Ed.), *Activity Analysis of Production and Allocation*, Cowles Commission for Research in Economics, pp. 359–374, 1951.

[6] P. Diaconis and A. Gangolli, Rectangular arrays with fixed margins, in: D. Aldous, P.P. Varaiya, J. Spencer and J.M. Steele (Eds.), *Discrete Probability and Algorithms*, IMA Volumes on Mathematics and its Applications, **72**, Springer, New York, pp. 15–41, 1995.

[7] P. Diaconis, R.L. Graham, and J.A. Morrison, Asymptotic analysis of a random walk on a hypercube with many dimensions. *Random Structures and Algorithms*, **1**, pp. 51–72, 1990.

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

[9] P. Diaconis and L. Saloff-Coste, Random walk on contingency tables with fixed row and column sums. Technical Report, Department of Mathematics, Harvard University, 1995.

[10] P. Diaconis and D. Stroock, Geometric bounds for eigenvalues of Markov chains. *Annals of Applied Probability*, **1**, pp. 36–61, 19991.

[11] M.E. Dyer, The complexity of vertex enumeration methods. *Mathematics of Operations Research*, **8**(3), pp. 381–402, 1983.

[12] M. Dyer, Approximate counting by dynamic programming. *Proceedings of the 35th Annual ACM Symposium on Theory of Computing (STOC 2003)*, pp. 693–699, 2003.

[13] M. Dyer and A. Frieze, Random walks, totally unimodular matrices and a randomised dual simplex algorithm. *Mathematical Programming*, **64**, pp. 1–16, 1994.

[14] M. Dyer and C. Greenhill, Polynomial-time counting and sampling of two-rowed contingency tables. *Theoretical Computer Science*, **246**, pp. 265–278, 2000.

[15] M. Dyer, R. Kannan and J. Mount, Sampling contingency tables. *Random Structures & Algorithms*, **10**(4), 1997, pp. 487–506, 1997.

[16] G. Hadley, Transportation Problems, In *Linear Programming* (Chapter 9), Addison-Wesley, Massachusetts, pp. 273–330, 1962.

[17] F.S. Hillier, G.J. Lieberman, *Introduction to Operations Research* (8th Ed), McGraw Hill, New York, 2005.

[18] F. L. Hitchcock, The distribution of a product from several sources to numerous localities. *Journal of Mathematics and Physics 20*, pp. 224–230, 1941.

[19] G. Kalai, Polytope skeletons and paths, In *Handbook of Discrete and Computational Geometry* (1st Ed), J. E. Goodman and J. O'Rourke (Eds), CRC Press, Boca Raton, pp. 331–345, 1997.

[20] V. Klee and C. Witzgall, Facets and Vertices of Transportation Polytopes, In *Mathematics of the Decision Sciences, Part 1*, G.B Dantzig and A.F. Veinott (Eds), American Mathematical Society, pp 257–282, 1968.

[21] Tj. C. Koopmans, Optimum utilization of the transportation system, In D. H. Leavens (Ed), *The Econometric Society Meeting, Washington DC, 1947*, pp. 136–146, 1948.

[22] C. McDiarmid, On the method of bounded differences, *London Math. Soc. Lecture Note Series* **141**, Cambridge University Press, pp. 148–188, 1989.

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

[24] B. Morris, *Random Walks in Convex Sets*, PhD thesis, Department of Statistics, University of California, Berkeley, 2000.

[25] B. Morris and A.J. Sinclair, Random walks on truncated cubes and sampling 0-1 knapsack solutions, *SIAM Journal on Computing*, **34**(1), pp. 195–226, 2004.

[26] J. B. Orlin, A polynomial time primal network simplex algorithm for minimum cost flows. *Mathematical Programming*, **8**, pp. 109–129, 1997.

[27] I. Pak, Four questions on Birkhoff polytope. *Annals of Combinatorics*, **4**, pp. 83–90, 2000.

[28] I. Pak, On the Number of Faces of Certain Transportation Polytopes. *European Journal of Combinatorics*, **21**, pp. 689–694, 2002.

[29] D. Randall and P. Tetali, Analyzing Glauber dynamics by comparison of Markov chains, *Journal of Mathematical Physics*, **41**, pp. 1598–1615, 2000.

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

[31] E. Tardos, A strongly polynomial minimum cost circulation algorithm. *Combinatorica*, **5**, 247–255, 1985.

[32] G. M. Ziegler, *Lectures on polytopes*, Graduate Texts in Mathematics **152**, Springer-Verlag, 1995.