Random walks on the vertices of transportation polytopes with constant number of sources

Citation for published version (APA):

Document status and date:
Published: 01/01/2002

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.

Download date: 26. Jun. 2020
Random walks on the vertices of transportation Polytopes with constant number of sources

M. Cryan
M. Dyer
H. Müller
L. Stougie

SPOR-Report 2002-14

Eindhoven, September 2002
The Netherlands
Random Walks on the Vertices of Transportation Polytopes with Constant Number of Sources *

Mary Cryan† Martin Dyer† Haiko Müller† Leen Stougie†
September 24, 2002

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 [20] together with ideas developed by Morris and Sinclair [15, 16] for the knapsack problem, and Cryan et al [2] for contingency tables, to establish that the random walk approaches the uniform distribution in time $O(m^2 n)$. 

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. We are able to show that this walk converges to the uniform distribution on the vertex set in time $O(m^2 n)$ 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 [11] and Ziegler [21].) 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 [18]). In fact, the only previous mixing results known are for very special, and highly symmetric polytopes, such as the $n$-cube [4] and the Birkhoff polytope [17].

Our approach is inspired by that of Cryan, Dyer, Goldberg, Jerrum and Martin [2] for sampling contingency tables. This was itself based on the “balanced permutation” ideas of Morris and Sinclair [15, 16] for the knapsack problem. However, following the line of proof given in [2], and using the $m$-dimensional balanced permutations of [15], leads inevitably to a mixing time bound of $n^{2O(m)}$. To obtain our improvement in the exponent, from exponential to polynomial, it is necessary to sharpen the tools of [15, 16] 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. (See [15] for further information.)

2 Background

The transportation problem (TP) is the combinatorial optimization problem of assigning shipments of some commodity from sources to destinations so that the transportation cost is minimized. We are given a list of $m$ sources and a list $r = (r_1, \ldots, r_m)$ of the quantities at each source ($r_i$ is the quantity at source $i$). We are given a list of $n$ destinations and a list $c = (c_1, \ldots, c_n)$ of the quantities required at each destination ($c_j$ units are required at destination $j$). Without loss of generality, we will assume that $\sum_{i=1}^m r_i = \sum_{j=1}^n c_j$, so that demand exactly matches supply.

We will let the total number of units be denoted by $N = \sum_{i=1}^m r_i$. We are also given transportation costs $t_{ij}$ for $i \in [m], j \in [n]$, where $t_{ij}$ denotes the cost of shipping one unit from source $i$ to destination $j$. (We use the notation $t_{ij}$ to denote the $i, j$ element of a matrix.)

A feasible solution to the TP is a set of real numbers $\{X_{ij} : i \in [m], j \in [n]\}$ satisfying the set of equations (1)-(3):

$$\sum_{j=1}^n X_{ij} = c_j \text{ for all } j \in [n]$$

Each feasible solution corresponds to a possible way of shipping units from suppliers to consumers subject to the given constraints on supply and demand. The set of feasible solutions, the feasible region is a convex polytope $P(r, c)$ in $\mathbb{R}^{mn}$. We call $P(r, c)$ a transportation polytope. The TP is to find $X \in P(r, c)$ minimizing $\sum_{i \in [m], j \in [n]} t_{ij} X_{ij}$.

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 [7], where it was shown to be \#P-complete to count exactly the number of vertices of a $2 \times n$ transportation polytope,\footnote{In fact [7] only claims NP-hardness, but the proof establishes \#P-completeness.} 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 $P(r,c)$ almost uniformly at random, when the number of sources $m$ is a constant. We define a Markov chain $\mathcal{W}$ on the set $\Omega$ of all vertices of $P(r,c)$ and prove it is rapidly mixing. Our chain $\mathcal{W}$ is a random walk on the edge-vertex graph $G(\mathcal{W})$ of the polytope $P(r,c)$. This graph is sometimes called the skeleton of the transportation polytope. By Lemma 2 below, we know that any vertex $Z$ of $P(r,c)$ has at most $d_m = \lceil mc^{m-1}n^m \rceil$ 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 $1 - \deg(Z)/2d_m$, which is at least $1/2$. A well-known result of Balinski [1] states that the edge-vertex graph of any convex polytope of dimension $k$ is $k$-connected. Thus, since $G(\mathcal{W})$ is connected, $\mathcal{W}$ is ergodic. Also, for any two vertices $Z, W$ of $P(r,c)$, $\Pr_{\mathcal{W}}[Z,W] = \Pr_{\mathcal{W}}[W,Z]$, so the chain converges to the uniform distribution on $\Omega$. Observe that all "null" steps at $Z$, where $\mathcal{W}$ remains at $Z$, can be simulated by updating 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 not been reached.

We will show that $\mathcal{W}$ is rapidly mixing by first showing that a "heat bath chain", which can make much larger changes, mixes rapidly. This chain, $\mathcal{M}_{HB}$, is described in section 4 below, and analysed in sections 5-6. Subsequently, in section 7, we use the comparison technique of Diaconis and Saloff-Coste [5] (see also Randall and Tetali [19]) to lift the mixing result from $\mathcal{M}_{HB}$ to $\mathcal{W}$. Finally, in section 8, we outline how sampling can be used to count approximately the number of vertices of a transportation polytope.

3 Preliminaries

For basic information about polytopes we refer the reader to Ziegler [21], and for specific details about transportation polytopes to Klee and Witzgall [12]. In particular, it is shown in [12] that $P(r,c)$ has dimension $(m - 1)(n - 1)$, so that the representation (4)-(7) is full-dimensional.

The following (see also Hadley [10]) is proved.

Lemma 1 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 no vertex of $P(r,c)$ has more than $n + m - 1$ non-zero coordinates. A non-degenerate vertex has exactly $n + m - 1$ non-zero coordinates. Any $(m - 1)(n - 1)$-dimensional transportation polytope has at most $m!n!$ vertices. Another upper bound for the number of vertices is $(cm)^{n+m-1}$.

We note that any point of $P(r,c)$ must have at least $n$ non-zero coordinates, and therefore any vertex has between $n$ and $n + m - 1$ (inclusive) non-zero coordinates. If $P(r,c)$ is non-degenerate, then every vertex will have $n + m - 1$ non-zero coordinates. If $Z$ is any vertex of $P(r,c)$, let $T_Z = \{ j : Z_j$ has more than one non-zero$\}$. Then $|T_Z| \leq m - 1$. If $Z$ has $n + m - 1 - q$ non-zero coordinates in total ($0 \leq q \leq m - 1$), we will say it has degeneracy $q$. We will sometimes refer to co-ordinates as cells.
Lemma 2 A vertex of $P(r,c)$ has at least $(m-1)(n-1)$, and less than $me^{m-1}n^m$, incident edges.

Proof: The lower bound comes from the non-degenerate case, noting that such a vertex is the intersection of exactly $(m-1)(n-1)$ facets.

For the upper bound, suppose we perturb $P(r,c)$, using the procedure described in Hadley [10]. We obtain the non-degenerate transportation polytope $P(r+c,m+emc)$, where $c$ is the $m$-vector of $1's$, $c_n$ is the $n$th unit vector and $c$ is small. Any vertex $v$ of $P(r,c)$ with degeneracy $q$ is perturbed to a set of at most $(m-1)(n-1)+q$ vertices, since some set of $q$ zero coordinates of $v$ will become non-zero in the perturbation. Since $(m-1)(n-1)+q < (en)^{m-1}$, each perturbed vertex is incident to exactly $(m-1)(n-1)$ edges. Hence $v$ is incident to at most $(en)^{m-1} \times (m-1)(n-1) < me^{m-1}n^m$ edges. □

4 The heat-bath chain

We now define an auxiliary "heat-bath" Markov chain $M_{HB}$, which operates on a $b_m$-sized window of the table representing the current vertex $Z$, where $b_m = 94m^2$. A single step of $M_{HB}$ is performed as follows: a set of columns $B \subseteq [n]$, with $|B| = b_m$, is chosen uniformly at random, subject to $T_Z \subseteq B$. 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 $Z_j$ $(j \in B)$.

We know that $M_{HB}$ is ergodic, since it includes all moves of $\mathcal{W}$, and therefore it converges to a stationary distribution $\pi$ on $\Omega$. By definition, $Pr_{M_{HB}}[Z \rightarrow W] = Pr_{M_{HB}}[W \rightarrow Z]$ for any two vertices $Z$, $W$. Therefore the stationary distribution $\pi$ is the uniform distribution on $\Omega$.

To show rapid mixing of $M_{HB}$ in section 6, we will use the multicommodity flow approach of Sinclair [20] (see also Diaconis and Stroock [6]), together with a construction based on ideas of Morris and Sinclair [16] which we develop in section 5 below. For any Markov chain $M$ on the state space $\Omega$, a multicommodity flow is defined on the underlying graph $G(M)$ of the chain $M$. The vertex set is $\Omega$, and there is an edge $(u \rightarrow v)$ for every pair of states such that $Pr_{M}(u \rightarrow v) > 0$ in $M$. For $x,y \in \Omega$, a unit flow from $x$ to $y$ is a set $\mathcal{P}_{x,y}$ of simple directed paths in $G(M)$ from $x$ to $y$, such that each path $p \in \mathcal{P}_{x,y}$ has positive weight $\alpha_p$, and the sum of the $\alpha_p$ over all $p \in \mathcal{P}_{x,y}$ is 1. A multicommodity flow is a family of unit flows $F = \{P_{x,y} : x,y \in \Omega\}$ containing a unit flow for every pair of states from $\Omega$.

The length $L(F)$ of the multi-commodity flow $F$ is $L(F) = \max_{p \in \mathcal{P}_{x,y}} \{p : p \in \mathcal{P}_{x,y}\}$, where $p$ denotes the edge length of $p$. For any edge $e$ of $G(M)$, we define $F(e)$ to be the sum of the $\alpha_p$ weights over all $e_p$ such that $e \in e_p$ and $e_p \in \mathcal{P}_{x,y}$ for some $x,y \in \Omega$. Then we will use the following.

Theorem 3 (Sinclair [20]) Let $P$ be the transition matrix of an ergodic, reversible Markov chain $M$ on $\Omega$ whose stationary distribution is the uniform distribution. Let $F$ be a multicommodity flow on the graph $G(M)$. Then the mixing time of the chain is bounded above by

$$\tau(\epsilon) \leq 2|\Omega|^{-1} L(F) \max_{e \in \mathcal{P}_M} F(e) (\log |\Omega| + \log \epsilon^{-1})$$

Finally, in section 7, we apply a comparison technique of Diaconis and Stroock [6] to extend our analysis to the random walk $\mathcal{W}$.

5 Balanced permutations

Suppose we are given two vertices $X$, $Y$ of $P(r,c)$, so $|T_X \cup T_Y| \leq 2(m-1)$. Let $T = \{j : X_j = Y_j\}$, $L = [n] \setminus (T_X \cup T_Y \cup T)$ and $\ell = |L|$. Let $r_i = r_i - \sum_{j \in T} X_j = r_i - \sum_{j \in T} Y_j$ for $i \in [n]$.

For $j \in L$, define the "weight vectors" $w_j = Y_j - X_j \in \mathbb{R}^n$, and let $\mu = \sum_{j \in L} w_j / \ell$ with coordi-
nates \( \mu' \) (\( i \in [m] \)). 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

\[
S_{i,i'} = S_{i',i} = \{ j \in L : \{ w^i_j, w^{i'}_j \} = \{ -c_j, +c_j \} \},
\]

Let \( \ell_{i,i'} = |S_{i,i'}| \), and let \( \mu_{i,i'} = \frac{\sum_{j \in S_{i,i'}} w^i_j / \ell_{i,i'}}{\ell_{i,i'}} \) 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 use a result from [15] (see also [16]) to help us define a suitable random permutation \( \sigma_{i,i'} \) on each of the \( S_{i,i'} \) sets.

**Lemma 4 (Morris [15])** Suppose we are given real weights \( \{ w_j \}_{j=1}^{s} \) with total \( W = \sum_{j=1}^{s} w_j \). Let \( M = \max_{j=1}^{s} |w_j| \). Suppose that \( |W| \geq 21M \).

Then there is a random permutation \( \pi_1 \) of \( [s] \) that satisfies the following two conditions: For some universal constant \( C > 1 \), and each \( 1 \leq k \leq s \),

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

(ii) for every \( U \subseteq [s] \) with \( |U| = k \),

\[
\Pr[\pi_1(1, \ldots, k) = U] \leq Cs^{23} (\frac{s}{k})^{-1}.
\]

We say \( \pi_1 \) is a \( k \)-balanced \( Cs^{23} \)-uniform permutation.

From this we will deduce a statement more convenient for our application (cf Morris [15, Ch. 3]). Let \( C \) be the constant from Lemma 4.2

**Lemma 5** Let \( \{ w_j \}_{j=1}^{s} \) be a set of real numbers with mean \( \mu = \sum_{j=1}^{s} w_j / s \). Then there exists a random permutation \( \pi \) of \( [s] \) such that, for each \( 1 \leq k \leq s \), there are sets \( D_1, D_2 \) with \( |D_1|, |D_2| \leq 42 \) satisfying

(i) \( \sum_{j \in [s] \cap D_1} w_{\pi(j)} \leq k \mu \), \( \sum_{j \in [s] \cap D_2} w_{\pi(j)} \geq k \mu \).

(ii) for every \( U \subseteq [s] \) with \( |U| = k \),

\[
\Pr[\pi(1, \ldots, k) = U] \leq Cs^{23} (\frac{s}{k})^{-1}.
\]

We call \( \pi \) a strongly \( 42 \)-balanced \( Cs^{23} \)-uniform permutation. (Strong balance means that the sign of \( \sum_{j \in [k]} (w_{\pi(j)} - \mu) \) can be altered by exchanging constantly many weights.)

**Proof:** Assume, by symmetry, that \( \mu \geq 0 \).

(i) If \( s \leq 42 \) let \( \pi \) be a random permutation of \( [s] \), \( D_1 = [k] \) and \( D_2 = [s] \setminus [k] \). Otherwise, let \( Q \) contain the indices of the 21 columns for which \( (w_j - \mu) \) is greatest and \( R \) contain the indices of the 21 columns for which \( (w_j - \mu) \) is smallest. There are two cases:

(a) first suppose \( -\sum_{j \in R} (w_j - \mu) \geq \sum_{j \in Q} (w_j - \mu) \). Then let the set \( \{ w_{\pi(j)} : j \in [s - 20, s] \} \) contain the values \( w_j \) for \( j \in R \).

We will apply Lemma 4 to the set of weights \( \{ w_j - \mu \}_{j \notin R} \) to construct our permutation \( \pi \). Note that \( W = \sum_{j \in R} (w_j - \mu) = -\sum_{j \in R} (w_j - \mu) \geq \sum_{j \in Q} (w_j - \mu) \). For every \( j \notin Q \cup R \), we have \( 21 |w_j - \mu| \leq W \). For now, assume that \( 21 |w_j - \mu| \leq W \) for \( j \in Q \), so that we have \( W \geq 21M \). (We will show how to remove this assumption below).

We have already constructed \( \pi \) for \( j \in [s - 20, s] \). Let \( \pi \) be the permutation of Lemma 4 on \( [s - 21] \). If \( k \leq 21 \), take \( D_1 = [k] \), \( D_2 = \emptyset \). If \( 21 < k \leq s - 21 \), property (i) of \( \pi_1 \) gives

\[
0 \leq \sum_{j=1}^{k} (w_{\pi(j)} - \mu) \leq \sum_{j=1}^{s-21} (w_{\pi(j)} - \mu) = -\sum_{j=s-20}^{s} (w_{\pi(j)} - \mu).
\]

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 \( k \leq s - 21 \), we have

\[
\sum_{j=1}^{k-21} (w_{\pi(j)} - \mu) + \sum_{j=s-20}^{s} (w_{\pi(j)} - \mu) \leq 0.
\]
Then, setting \( D_1 = [k-20,k] \cup [s-20,s] \), we have
\[
\sum_{j \in [k] \cup D_1} w_\pi(j) \leq k\mu.
\]
If \( k > s - 21 \), the conclusion follows easily from
\[
\sum_{j=1}^{s-21} (w_\pi(j) - \mu) \geq 0, \quad \sum_{j=1}^{s} (w_\pi(j) - \mu) = 0.
\]
Note that \( |D_1|, |D_2| \leq 21 \), expect for \( D_1 \) when \( 21 < k \leq s - 21 \) (then \( |D_1| \leq 42 \)). We now show how to deal with the possibility that there is some \( j \in Q \) such that \( 21|w_j - \mu| > W \). When we construct the permutation \( \pi_1 \), we replace the weights \( \{w_j - \mu\}_{j \in Q} \) by \( \{w_j' - \mu\}_{j \in Q} \), where \( w_j' = \sum_{j \in Q} w_j/21 \) for all \( j \in Q \). Then \( W \) does not change and the condition \( 21M \leq W \) is satisfied. When \( \pi_1 \) 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 \( Q \) for others) to obtain \( D_1, D_2 \) sets satisfying condition (i) for the original weights. Moreover, for \( 21 < k \leq s - 21 \), we can define \( D_1 = [s-20,s] \cup (Q \cap [k]) \cup [k-20+|\{j : j \in Q \cap [k]\}], k] \) to ensure (i) holds. Therefore we still have \( |D_1|, |D_2| \leq 42 \), as claimed.

(b) Suppose \( -\sum_{j \in R} (w_j - \mu) < \sum_{j \in Q} (w_j - \mu) \). Let \( \{w_\pi(j) : j \in [s-20,s]\} \) be the set of weights \( \{w_j : j \in Q\} \). Then, when we apply Lemma 4 to the set of weights \( \{w_j - \mu\}_{j \in [s] \setminus Q} \), the total of the weights \( W \) is negative. Again, assuming for now that \( |W| \geq 21 \max_{j \in [s] \setminus Q} |w_j - \mu| \), we let \( \pi \) be the permutation \( \pi_1 \) of Lemma 4 on \([s-21]\).

For \( k \leq 21 \), we take \( D_1 = \emptyset \) and \( D_2 = [s-20,s] \). For \( 21 < k \leq s - 21 \), we take \( D_1 = \emptyset \) and \( D_2 = [s-20,s] \). The case \( k > s - 21 \) is similar to case (a).

Finally, we treat the possibility that there exists \( j \in R \) with \( 21|w_j - \mu| > |W| \) in a similar way to case (a).

(ii) If \( s \leq 42 \), the property follows from the fact that \( \pi \) is a random permutation. Otherwise, if \( k \leq 21 \) or \( k > s - 21 \), the statement is trivially true. In all other cases, property (ii) of \( \pi_1 \) implies \( \Pr[\pi(1, \ldots, k) = U] \leq C(s-21)^2(s-21)^{-1} \leq C s^{23} t^{-1} \).

Remark: It would be possible to improve the constants in Lemma 5 by proving it directly, rather than starting from Lemma 4. However, we are not aiming to optimize the constants.

We apply the construction of Lemma 5 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 \( \sum_{j \in J} w_j' = -\sum_{j \in J} w_j' \). Hence \( \sum_{j \in J} w_j' \geq k\mu_{i,i'} \iff \sum_{j \in J} w_j' \leq k\nu_{i,i'} \). 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 \( \sigma \) 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** \( i^* = \arg\max_{i=1}^{q} (\alpha_i k_i - k_i) \)

**then** \( k_{i^*} \leftarrow k_{i^*} + 1 \).

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

**Lemma 6** For all \( k \in [0,\nu] \), \( k_i \leq [\alpha_i k] \leq \nu_i \), \( i \in [q] \), and \( \sum_{i=1}^{q} |k_i - \alpha_i k| < 2(q-1) \).

**Proof:** First note that \( [\alpha_i k] \leq \nu_i \). Otherwise \( [\alpha_i k] = \lfloor \nu_i k/\nu \rfloor > \nu_i \), giving \( k > \nu \), a contradiction.

Let \( \gamma_i(k) = \alpha_i k - k_i \). Note that \( \sum_{i=1}^{q} \gamma_i = 0 \), so \( \gamma_i \geq 0 \). Let primes denote quantities at step \((k+1)\), so \( \gamma_i' = \gamma_i(k+1) \). Then \( \gamma_i' = \gamma_i + \alpha_i > \gamma_i \) \((i \neq i^*)\), but \( \gamma_{i^*}' = \gamma_{i^*} - (1 - \alpha_{i^*}) > -1 \). Since \( \gamma_i(0) = 0 \) for all \( i \), it follows by induction that
\( \gamma_i(k) > -1 \) for all \( i, k \). Now let \( k_i \leq [\alpha_i k] \) follows immediately. Also, since \( \sum_{i=1}^{q} \gamma_i = 0 \) and \( \gamma_i \geq 0 \),

\[
\sum_{i=1}^{q} |k_i - \alpha_i k| = \sum_{i=1}^{q} |\gamma_i| = 2 \sum_{\gamma_i < 0} |\gamma_i|
\]

This 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 \( \sigma \).

**Lemma 7** The random permutation \( \sigma \) has the following properties.

(i) For all \( k \in [\ell] \) there exist sets \( D^i_k \) \((s = 1, 2; i \in [m])\) such that \( |\bigcup_{i=1}^{m} D^i_k| < 23m^2 \) \((s = 1, 2)\) and

\[
\sum_{j \in [k] \cap D^i_1} w_{i,j} \leq k\mu^i, \quad \sum_{j \in [k] \cap D^i_2} w_{i,j} \geq k\mu^i.
\]

(ii) For any \( U \subseteq [\ell] \) with \( |U| = k \), \( \Pr_r[\sigma \{1, \ldots , k\} \subseteq U] \leq 4m^{14m^2}(\ell^{-1})^{-1} \), for some constant \( C_m \).

We then say that \( \sigma \) is a strongly \( 23m^2 \)-balanced \( C_m^{14m^2} \)-uniform permutation.

**Proof:** (i): We prove only the first inequality, the other being entirely similar. Suppose the values at step \( k \) in \textsc{interleave} are \( k_{i,i'} \) and \( \alpha_{i,i'} = \ell_{i,i'}/l \). Define \( k_{i,i'}^* \) to be \( [k\alpha_{i,i'}] \) if \( \mu_{i,i'} \geq 0 \), and \( [k\alpha_{i,i'}] \) otherwise. Using Lemma 6, observe that \( \sum_{i,i'} (k_{i,i'}^* - k_{i,i'}) \) is at most

\[
\sum_{i,i'} (k_{i,i'}^* - k_{i,i'}) \leq \frac{\binom{m}{2}}{2} + \binom{m}{2} = 3\binom{m}{2}.
\]

Let \( D^i_{1,\ell} \) be the set associated with \( \sigma_{i,i'} \), \( k_{i,i'}^* \) such that \( \sum_{j \in [k_{i,i'}^*] \cap D^i_{1,\ell}} w_{i,j} \leq k_{i,i'}^* \mu_{i,i'} \), and let \( I_{1,\ell} \) 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 \( D^i_1 = \bigcup_{\ell} (D^i_{1,\ell} \cup I_{1,\ell}) \). Then, using Lemma 5, \( |\bigcup_{i=1}^{m} D^i_1| < 42\binom{m}{2} < 3\binom{m}{2} < 45m^2/2 \). Also

\[
\sum_{j \in [k] \cap D^i_1} w_{i,j} \leq \sum_{i,i'} k_{i,i'} \mu_{i,i'} = \sum_{i,i'} k\ell_{i,i'} \mu_{i,i'}/\ell = k\mu^i.
\]

(ii): Let \( \tau^* \) be the random permutation we get when we apply \textsc{interleave} to the collection of uniform distributions \( \tau_{i,i'} \) on \( S_{i,i'} \) for every \( i,i' \). Let \( \tau \) represent the uniform distribution on \([\ell] \). We will first bound \( \Pr_\tau[\tau \{1, \ldots , k\} = U] \) in terms of \( \Pr_\tau[\tau \{1, \ldots , k\} = U] = \binom{k}{\ell} \), 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

\[
\Pr_\tau \left[ |K_{i,i'} - \alpha_{i,i'} k| \geq \sqrt{k \log(\ell)} \right] \leq 2e^{-2k\mu^2 \binom{m}{2}} = 2\ell^{-2}
\]

by a single application of the Chernoff bound (see McDiarmid [13]. Summing over all \( k \) and all \( i,i' \) (\( \binom{m}{2} \) in total), we find that under the uniform distribution \( \tau \),

\[
|K_{i,i'} - \alpha_{i,i'} k| \leq \sqrt{k \log(\ell)} \quad (8)
\]

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 (8) holds with probability at least \( 1/2 \).

Let \( \tau^* \) be the uniform distribution on the permutations that satisfy (8) (for all \( k \), all \( i,i' \)). Note the probability of any event in \( \tau^* \) is at most twice its probability in the uniform distribution \( \tau \). Also, since the integer variable \( K_{i,i'} \) has maximum probability of taking values \( \{\alpha_{i,i'} k, [\alpha_{i,i'} k]\} \), we have

(a) \( \Pr_{\tau^*}[K_{i,i'} = q_{i,i'}] \geq (\sqrt{k \log(\ell)})^{-1} \) for \( q_{i,i'} \in \{[\alpha_{i,i'} k], [\alpha_{i,i'} k]\} \)

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 6, we find that
\[
\Pr_{r}(K_i,i' = k_{i,i'}, \forall i,i') \geq (k \log \epsilon)^{-m^2}/4 \left( \prod_{i,i'} \epsilon^{-k_{i,i'} - a_{i,i'}(i' - 1)/2} \right)
\geq \epsilon^{-m^2/4} \epsilon^{-3m^2/2} / 2 \geq \epsilon^{-2m^2/2}
\]
So \( \Pr_\sigma \{1, \ldots, k\} = \cup \sim \sim 2^m \epsilon 2m^2 \).

Then applying Lemma 4 to each of the \( S_i,i' \), we have

\[
\Pr_\sigma \{1, \ldots, k\} = \cup \sim \sim 2m^2 \epsilon 2m^2 \epsilon 2m^2 / 2 \sim \sim 2m^2 / 2
\]
and we have \( C_m \epsilon 14m^2 \)-uniformity.

6 Analysis of the heat bath

In a similar manner to [2] (see also [16]), we use the permutation \( \sigma \) constructed by \textit{interleave} to route flow from \( X \) to \( Y \), by changing \( X_{\sigma(k)} \) to \( Y_{\tau(k)} \) for \( k = 1, 2, \ldots, \ell \). If \( Z_k \) is any intermediate matrix obtained in this way, in general it cannot be completed to a vertex of \( P(r,c) \), and our “encoding”, \( \hat{Z}_k \), for \( Z_k \) has column \( X \) and vice versa. Again, \( \hat{Z}_k \) cannot be completed to a vertex in general. But we will show that both \( Z_k \) and \( \hat{Z}_k \) are close to vertices of \( P(r,c) \) in the following sense. If we delete only a constant number of columns from either \( Z_k \) or \( \hat{Z}_k \), then it can be completed to a vertex of \( P(r,c) \), \( Z_k \) and \( \hat{Z}_k \) respectively. Moreover, both \( X \) and \( Y \) can be reconstructed from \( Z_k \) and \( \hat{Z}_k \) using a suitably small amount of information.

Let \( D_s = \bigcup_{i=1}^m D_s^i (s = 1, 2) \). Since \( X_j^i, Y_j^i \geq 0 \) for all \( i,j \), for each \( i \) in \( m \) we have
\[
\sum_{j \in L \setminus \{k\}} X_j^i + \sum_{j \in \{k\} \cap D_s^i} Y_j^i \leq \sum_{j \in L \setminus \{k\} \cap D_s^i} X_j^i + \sum_{j \in \{k\} \cap D_s^i} Y_j^i = \sum_{j \in L} X_j^i + \sum_{j \in \{k\} \cap D_s^i} w_s^j(j).
\]
By Lemma 7, we also have
\[
\sum_{j \in L} X_j^i + \sum_{j \in \{k\} \cap D_s^i} w_s^j(j) \leq \sum_{j \in L} X_j^i + k \mu^i
\]
which is at most \( r_i^k \).

Hence, if we delete the columns in \( D_1 \), \( Z_k \) can be completed to a vertex of \( P(r,c) \) by setting \( Z_i = X_j = Y_j (j \in T) \), and filling in the columns of \( D_1 \cup T_X \cup T_Y \) according (say) to the “northwestern corner rule” [10]. The proof for \( \hat{Z}_k \) is identical, by interchanging \( X_j^i \) with \( Y_j^i \), \( w_s^j \) with \( -w_s^j \), \( D_1 \) with \( D_2 \), and using the lower bound in (i) of Lemma 7.

Now suppose we are given \( Z_k^i, \hat{Z}_k \) and we wish to recover \( X, Y \). Let us assume, using the uniformity property of \( \sigma \), that we are given \( U = \sigma[k] \). We still need to know the “deleted” columns \( D_1, D_2, T_X, T_Y \), but there are at most \( (s_{m}^n)^2 \left( \binom{n}{m-1} \right)^2 < n^{4m^2} \) ways of selecting these sets. We can easily reconstruct both \( X \) and \( Y \) except for the deleted columns. However, there are at most \( 4m^2 \) such columns, and \( X \) and \( Y \) are both vertices. Thus there are at most \( (s_{m}^{4m^2})^2 < n^{4m^2} \) ways of completing \( X \) and \( Y \), i.e. a constant number. So there are at most \( n^{4m^2} \) ways of augmenting the encoding so that we can uniquely identify \( X \) and \( Y \) from \( Z_k^i, \hat{Z}_k \). Note that \( Z_k^i \) and \( \hat{Z}_k \) differ in at most \( 4m^2 \) columns, justifying the choice of \( b_m \) above.

We can now bound the flow through any state \( Z \in \Omega \). There are \( |\Omega| \) ways of choosing \( Z \), \( \left( \binom{n}{k} \right)^2 \) ways of choosing \( |U| \) and \( n^{4m^2} n^{4m^2} \) ways of specifying the additional information needed to uniquely identify \( X \) and \( Y \). However, by the uniformity of \( \sigma \), \( \Pr(\sigma[k] = U) \leq C_m n^{14m^2} \left( \binom{n}{k} \right)^{-1} \). Hence the flow through any state may be bounded by
\[
|\Omega| \times \left( \binom{n}{k} \right)^2 \times n^{4m^2} n^{4m^2} \times C_m n^{14m^2} \left( \binom{n}{k} \right)^{-1}
= O(n^{61m^2}) |\Omega|.
\]
We can now use the “flow spreading” device from [2] and Theorem 3 to conclude that
\[
\tau(\epsilon) \leq 4|\Omega|^{-1} L(F) O(n^{61m^2}) |\Omega| \log |\Omega| + \log \epsilon^{-1})
= O(n^{52m^2}) \log \epsilon^{-1}, \quad (9)
\]
on noting that \( \mathcal{L}(\mathcal{F}) \leq n \), and \( |\Omega| \leq \left( \frac{mn}{n+m-1} \right) < \left( \frac{m}{e} \right)^{n+m-1} \).

7 Analysis of the random walk

We now show that the natural random walk \( \mathcal{W} \) defined in section 2 is rapidly mixing. We prove this using the comparison theorem of Diaconis and Saloff-Coste [5]. For a Markov chain \( \mathcal{M} \) on a state space \( \Omega \), let \( \ker(\mathcal{M}) \) denote the set of pairs \( (X, Y) \in \Omega^2 \) such that \( \Pr_{\mathcal{M}}[X \rightarrow Y] > 0 \).

**Theorem 8 (Diaconis and Saloff-Coste [5])**

Let \( \Omega \) 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 \( \Omega \). 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

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

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

\[
O\left( \tau_{\mathcal{M}}(r) \log(|\Omega|) \max_{(Z,W) \in \ker(\mathcal{M}')} A_{Z,W} \right).
\]

We now use Theorem 8 to bound the mixing time of \( \mathcal{W} \) in terms of the mixing time of \( \mathcal{M}_{HB} \).

We construct a canonical path \( p_{X,Y} \) on \( G(\mathcal{W}) \) for every pair of vertices \( (X, Y) \in \ker(\mathcal{M}_{HB}) \). Recall that by our definition of \( \mathcal{M}_{HB} \) in Section 4, for any pair \( (X, Y) \in \ker(\mathcal{M}_{HB}) \), 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 \( j \in T_X \cup T_Y \). Let \( b = |J_{X,Y}| \). Let \( \hat{X} \) be the table consisting of the columns \( X_j \) for \( j \in J_{X,Y} \), and let \( \hat{Y} \) be the table 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 \( i \)th row of \( \hat{X} \). By definition of \( J_{X,Y} \), \( s_i \) is also the source quantity for the \( i \)th row of \( \hat{Y} \). Let \( P(s,c) \) be the \( (m-1)(b-1) \)-dimensional transportation polytope with source quantities \( s_i \) for \( i \in [m] \) and destination quantities \( c_j \) for \( j \in J_{X,Y} \). \( \hat{X} \) and \( \hat{Y} \) are both vertices of \( P(s,c) \).

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

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

is a vertex of the original transportation polytope \( P(r,c) \) (see, for example, Hadley [10]).

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 \( \hat{X}(0) = \hat{X}, \hat{X}(1), \ldots, \hat{X}(\ell - 1), \hat{X}(\ell) = \hat{Y} \) connecting \( \hat{X} \) to \( \hat{Y} \) 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 \( P(r,c) \). For every \( i \in [\ell] \), \( X(i) \) is the table consisting of the columns \( X_j \) for \( j \notin J_{X,Y} \) and the columns \( \hat{X}(i)_j \) for \( j \in J_{X,Y} \). Also, \( X(i) \) is a vertex of \( P(r,c) \) for every \( i \in [\ell] \) and also \( (X(i-1), X(i)) \) is an edge of \( P(r,c) \) for every \( i \in [\ell] \) (see Hadley [10]). Therefore the path \( p_{X,Y} \) given by \( X(0) = X, X(1), \ldots, X(\ell) = Y \) is a path of length at most \( m!b! \) (see Lemma 1) in the edge-vertex graph \( G(\mathcal{W}) \).

Let \( \mathcal{P} = \{ p_{X,Y} : X, Y \in \ker(\mathcal{M}_{HB}) \} \). 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 \( \ker(\mathcal{M}_{HB}) \) according to the set \( B \) of \( b_m \) columns used to move from \( X \) to \( Y \). We will write \((X, Y) \in \mathcal{M}_{HB}(B) \) if \((X, Y) \) is an element of \( \ker(\mathcal{M}_{HB}) \) and \( X \) and \( Y \) differ only
on the columns in \( B \). Then we find that \( A_{Z,W} \) is at most

\[
\Pr_{W}(Z,W) \sum_{B \subseteq [n], |B| = b_m} \Pr_{M}(X,Y) \sum_{(X,Y) \in \text{ker}(M_{MB}(B))} |p_{X,Y}|Pr_{M}(X,Y)
\]

which is at most

\[
\Pr_{W}(Z,W) \sum_{B \subseteq [n], |B| = b_m} \Pr_{M}(X,Y) \sum_{(X,Y) \in \text{ker}(M_{MB}(B))} (m!b_m!)Pr_{M}(X,Y)
\]

However, once we fix a set of columns \( B \), we know that there are at most \( m!b_m \) different vertices of \( P(r,c) \) which agree with \( Z \) (and \( W \)) on all columns \( j \not\in B \). Using this, and the fact that \( \Pr_{M}(X,Y) \leq 1 \), we find

\[
A_{Z,W} \leq \frac{1}{\Pr_{W}(Z,W)} \sum_{B \subseteq [n], |B| = b_m} (m!b_m!)^3
\]

\[
\leq 2m e^{m-1}m^{m} \left( \frac{n}{b_m} \right) (m!b_m!)^3
\]

for any \((Z,W) \in \text{ker}(W)\). Using \( b_m = 94m^2 \), we have

\[
A_{Z,W} \leq 2m e^{m-1}m^{m}n^{94m^2}(m!b_m!)^3
\]

Then by Theorem 8 and by (9), we find that \( \tau_W(\varepsilon) \) is

\[
= O\left( n^{156m^2+m+2} \log(\varepsilon^{-1}) \right)
\]

using the fact that \(|\Omega| \leq (em)^{n+m-1} \), and therefore \( \log |\Omega| \) is \( O(n) \).

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 \( P(r,c) \). We will briefly sketch the method.

If \( n < 2(m+1) \), determine \(|\Omega|\) by complete enumeration. (See, for example, [7].) Otherwise, at least \( n-m+1 \) columns \( j \) have the single entry \( c_j \) at any vertex, and each column has only \( n \) cells. Therefore some cell \((s,t)\) 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 tables in which it contains \( c_j \), by sampling. But \( p = |\Omega'|/|\Omega| \), where \(|\Omega'|\) is the number of vertices of a transportation polytope \( P(r',c') \).

Here \( c' = (c_1, \ldots, c_{t-1}, c_{t+1}, \ldots, c_m) \), \( r'_s = r_s - c_t \), and \( r'_i = r_i, i = [m]\setminus\{s\} \). We estimate \(|\Omega'|\) recursively, and hence \(|\Omega|\) by \(|\Omega'|/p\).

References


