An upper bound on the number of high-dimensional permutations

Nathan Linial[*] Zur Luria[†]

Abstract

What is the higher-dimensional analog of a permutation? If we think of a permutation as given by a permutation matrix, then the following definition suggests itself: A d-dimensional permutation of order n is an $n \times n \times \ldots n = [n]^{d+1}$ array of zeros and ones in which every *line* contains a unique 1 entry. A line here is a set of entries of the form $\{(x_1, \ldots, x_{i-1}, y, x_{i+1}, \ldots, x_{d+1}) | n \geq y \geq 1\}$ for some index $d+1 \geq i \geq 1$ and some choice of $x_j \in [n]$ for all $j \neq i$. It is easy to observe that a one-dimensional permutation is simply a permutation matrix and that a two-dimensional permutation is synonymous with an order-n Latin square. We seek an estimate for the number of d-dimensional permutations. Our main result is the following upper bound on their number

$$\left((1+o(1))\frac{n}{e^d}\right)^{n^d}.$$

We tend to believe that this is actually the correct number, but the problem of proving the complementary lower bound remains open. Our main tool is an adaptation of Brègman's [1] proof of the Minc conjecture on permanents. More concretely, our approach is very close in spirit to Radhakrishnan's [10] proof of Brègman's theorem.

*Department of Computer Science, Hebrew University, Jerusalem 91904, Israel. e-mail: nati@cs.huji.ac.il . Supported by ISF and BSF grants.

†Department of Computer Science, Hebrew University, Jerusalem 91904, Israel. e-mail: zluria@cs.huji.ac.il .

1 Introduction

The permanent of an $n \times n$ matrix $A = (a_{ij})$ is defined by

$$Per(A) = \sum_{\sigma \in \mathbb{S}_n} \prod_{i=1}^{n} a_{i,\sigma_i}$$

Permanents have attracted a lot of attention [9]. They play an important role in combinatorics. Thus if A is a $0-1$ matrix, then $Per(A)$ counts perfect matchings in the bipartite graph whose adjacency matrix is A. They are also of great interest from the computational perspective. It is $\#P$-hard to calculate the permanent of a given $0-1$ matrix [11], and following a long line of research, an approximation scheme was found [6] for the permanents of nonnegative matrices. Bounds on permanents have also been studied at great depth. Van der Waerden conjectured that $Per(A) \geq \frac{n!}{n^n}$ for every $n \times n$ doubly stochastic matrix A, and this was established more than fifty years later by Falikman and Egorychev [4, 3]. More recently, Gurvitz [5] discovered a new conceptual proof for this conjecture (see [8] for a very readable presentation). What is more relevant for us here are upper bounds on permanents. These are the subject of Minc's conjecture which was proved by Brègman.

Theorem 1.1. *If A is an $n \times n$ $0-1$ matrix with r_i ones in the i-th row, then*

$$Per(A) \leq \prod_{i=1}^{n} (r_i!)^{1/r_i}.$$

In the next section we review Radhakrishnan's proof, which uses the entropy method. Our plan is to imitate this proof for a d-dimensional analogue of the permanent. To this end we need the notion of d-dimensional permutations.

Definition 1.2.

1. *Let A be an $[n]^d$ array. A line of A is vector of the form*

$$(A(i_1, ..., i_{j-1}, t, i_{j+1}, ..., i_d))_{t=1}^{n},$$

where $1 \leq j \leq d$ and $i_1, ..., i_{j-1}, i_{j+1}, ..., i_d \in [n]$.

2. *A d-dimensional permutation of order n is an $[n]^{d+1}$ array P of zeros and ones such that every line of P contains a single one and $n-1$ zeros. Denote the set of all d-dimensional permutations of order n by $S_{d,n}$.*

For example, a two dimensional array is a matrix. It has two kinds of lines, usually called rows and columns. Thus a 1-permutation is an $n \times n$ $0-1$ matrix with a single one in each row and a single one in each column, namely a permutation matrix. A 2-permutation is identical to a Latin square and $S_{2,n}$ is the same as the set $\mathcal{L}_n$, of order-n Latin squares. We now explain the correspondence between the two sets. If X is a 2-permutation of order n, then we associate with it a Latin square L, where $L(i,j)$ as the (unique) index of a 1 entry in the line $A(i,j,*)$. For more on the subject of Latin squares, see [12]. The same definition yields a one-to-one correspondence between 3-dimensional permutations and Latin cubes. In general, d-dimensional permutations are synonymous with d-dimensional Latin hypercubes. For more on d-dimensional Latin hypercubes, see [13]. To summarize, the following is an equivalent definition of a d-dimensional permutation. It is an $[n]^d$ array with entries from $[n]$ in which every line contains each $i \in [n]$ exactly once. We interchange freely between these two definitions according to context.

Our main concern here is to estimate $|S_{d,n}|$, the number of d-dimensional permutations of order n. By Stirling's formula

$$|S_{1,n}| = n! = \left((1+o(1))\frac{n}{e}\right)^n.$$

As we saw, $|S_{2,n}|$ is the number of order n Latin squares. The best known estimate [12] is

$$|S_{2,n}| = |\mathcal{L}_\mathrm{n}| = \left((1+\mathrm{o}(1))\frac{\mathrm{n}}{\mathrm{e}^2}\right)^{\mathrm{n}^2}.$$

This relation is proved using bounds on permanents. Brégman's theorem for the upper bound, and the Falikman-Egorychev theorem for the lower bound.

This suggests

Conjecture 1.3.

$$|S_{d,n}| = \left((1+o(1))\frac{n}{e^d}\right)^{n^d}.$$

In this paper we prove the upper bound

Theorem 1.4.

$$|S_{d,n}| \leq \left((1+o(1))\frac{n}{e^d}\right)^{n^d}.$$

As mentioned, our method of proof is an adaptation of [10]. We first need

Definition 1.5. *1. An $[n]^{d+1}$ $0-1$ array M_1 is said to include an array M_2 if*

$$M_2(i_1,...,i_{d+1}) = 1 \Rightarrow M_1(i_1,...,i_{d+1}) = 1.$$

2. The d-permanent of a $[n]^{d+1}$ $0-1$ array A is

$$Per_d(A) = \text{The number of } d\text{-dimensional permutations included in } A.$$

Note that in the one-dimensional case, this is indeed the usual definition of $Per(A)$. It is not hard to see that for $d = 1$ following theorem coincides with Brègman's theorem.

Theorem 1.6. *Define the function $f : \mathbb{N}_{\geq 0} \times \mathbb{N} \longrightarrow \mathbb{R}$ recursively by:*

- *$f(0, r) = \log(r)$, where the logarithm is in base e.*
- *$f(d, r) = \frac{1}{r}\sum_{k=1}^{r} f(d-1, k)$.*

Let A be an $[n]^{d+1}$ $0-1$ array with $r_{i_1,\dots i_d}$ ones in the line $A(i_1,...,i_d,*)$. Then

$$Per_d(A) \leq \prod_{i_1,\dots,i_d} e^{f(d,r_{i_1,\dots i_d})}.$$

We will derive below fairly tight bounds on the function f that appears in theorem 1.6. It is then an easy matter to prove theorem 1.4 by applying theorem 1.6 to the all-ones array.

What about proving a matching lower bound on $S_{d,n}$ (and thus proving conjecture 1.3)? In order to follow the footsteps of [12], we would need a lower bound on $Per_d A$, namely, a higher-dimensional analog of the van der Waerden conjecture. The entries of a *multi-stochastic* array are nonnegative reals and the sum of entries along every line is 1. This is the higher-dimensional counterpart of a doubly-stochastic matrix. It should be clear how to extend the notion of $Per_d(A)$ to real-valued arrays. In this approach we would need a lower bound on $Per_d(A)$ that holds for every multi-stochastic array A. However, this attempt (or at least its most simplistic version) is bound to fail. An easy consequence of Hall's theorem says that a $0-1$ matrix in which every line or column contains the same (positive) number of one-entries, has a *positive* permanent. (We still do not know exactly how small such a permanent can be, see [8] for more on this). However, the higher dimensional analog of this is simply incorrect. There exist multi-stochastic arrays whose d-permanent vanishes, as can easily be deduced e.g., from [7].

We can, however, derive a lower bound of $|S_{d,n}| \geq \exp(\Omega(n^d))$ for even n. Consider the following construction: Let n be an even integer, and let P be a d-dimensional permutation of order $\left[\frac{n}{2}\right]^d$. It is easy to see that such a P exists. Simply set

$$P(i_1, ..., i_d) = (i_1 + ... + i_d) \mod \frac{n}{2}.$$

Now we construct a d-dimensional permutation Q of order $[n]^d$ by replacing each element of P with a $[2]^d$ block. If $P(i_1, ..., i_d) = j$, then the corresponding block contains the values j and $j + n$. It is easy to see that there are exactly two ways to arrange these values in each block, and that Q is indeed a d-dimensional permutation of order $[n]^d$. There are $\left(\frac{n}{2}\right)^d$ blocks, and so the number of possible Q's is $2^{\left(\frac{n}{2}\right)^d}$. For a constant d this is $\exp(\Omega(n^d))$.

In section 2 we present Radhakrishnan's proof of the Brègman bound. In section 3 we prove theorem 1.6. In section 4 we use this bound to prove theorem 1.4.

2 Radhakrishnan's proof of Brègman's theorem

2.1 Entropy - Some basics

We review the basic material concerning entropy that is used here and refer the reader to [2] for further information on the topic.

Definition 2.1. *The entropy of a discrete random variable X is given by*

$$H(X) = \sum_x Pr(X = x) \log \left(\frac{1}{Pr(X = x)} \right).$$

For random variables X and Y, the conditional entropy of X given Y is

$$H(X|Y) = \mathbb{E}(H(X|Y = y)) = \sum_y Pr(Y = y) H(X|Y = y).$$

In this paper we will always consider the base e entropy of X which simply means that the logarithm is in base e.

Theorem 2.2. *1. If X is a discrete random variable, then*

$$H(X) \leq \log |range(X)|,$$

with equality iff X has a uniform distribution.

2. If $X_1, ..., X_n$ is a sequence of random variables, then

$$H(X_1, ..., X_n) = \sum_{i=1}^{n} H(X_i | X_1, ..., X_{i-1}).$$

3. The inequality

$$H(X|Y) \leq H(X|f(Y))$$

holds for every two discrete random variables X and Y and every real function $f(\cdot)$.

The following is a general approach using entropy that is useful for a variety of approximate counting problems. Suppose that we need to estimate the cardinality of some set S. If X is a random variable which takes values in S under the uniform distribution on S, then $H(X) = \log(|S|)$. So, a good estimate on $H(X)$ yields bounds on $|S|$.

This approach is the main idea of both Radhakrishnan's proof and our work.

2.2 Radhakrishnan's proof

Let A be an $n \times n$ $0-1$ matrix with r_i ones in the ith row. Our aim is to prove the upper bound

$$Per(A) \leq \prod_{i=1}^{n} (r_i!)^{\frac{1}{r_i}}.$$

Let $\mathcal{M}$ be the set of permutation matrices included in A, and let X be a uniformly sampled random element of $\mathcal{M}$. Our plan is to evaluate $H(X)$ using the chain rule and estimate $|\mathcal{M}|$ using the fact (theorem 2.2) that $H(X) = \log(|\mathcal{M}|)$.

Let X_i be the unique index j such that $X(i,j) = 1$. We consider a process where we scan the rows of X in sequence and estimate $H(X) = H(X_1, ..., X_n)$ using the chain rule in the corresponding order. To carry out this plan, we need to bound the contribution of the term involving X_i conditioned on the previously observed rows. That is, we write

$$H(X) = \sum_{i=1}^{n} H(X_i | X_1, ..., X_{i-1}).$$

Let R_i be the set of indices of the one-entries in A-th i-th row. That is,

$$R_i = \{j : A(i,j) = 1\}.$$

Let

$$Z_i = \{j \in R_i : X_{i'} = j \text{ for some } i' < i\}.$$

Note that $X_i \in R_i$, because X is included in A. In addition, given that we have already exposed the values $X_{i'}$ for $i' < i$, it is impossible for X_i to take any value $j \in Z_i$, or else the column $X(*, j)$ contains more than a single 1-entry. Therefore, given the variables that precede it, X_i must take a value in $R_i \setminus Z_i$. The cardinality $N_i = |R_i \setminus Z_i|$ is a function of $X_1, ..., X_{i-1}$ and so by theorem 2.2,

$$H(X) = \sum_{i=1}^{n} H(X_i | X_1, ..., X_{i-1})$$

$$= \sum_{i=1}^{n} \sum_{x_1, ..., x_{i-1}} \Pr(X_1 = x_1, ..., X_{i-1} = x_{i-1}) H(X_i | X_1 = x_1, ..., X_{i-1} = x_{i-1})$$

$$\leq \sum_{i=1}^{n} \sum_{x_1, ..., x_{i-1}} \Pr(X_1 = x_1, ..., X_{i-1} = x_{i-1}) \log(N_i)$$

$$= \sum_{i=1}^{n} \mathbb{E}_{X_1, ..., X_{i-1}} [\log(N_i)] = \sum_{i=1}^{n} \mathbb{E}_X [\log(N_i)].$$

It is not clear how we should proceed from here, for how can we bound $\log(N_i)$ for a general matrix? Moreover, different orderings of the rows will give different bounds. We use this fact to our advantage and consider the expectation of this bound over all possible orderings. Associated with a permutation $\sigma \in \mathbb{S}_n$ is an ordering of the rows where X_j is revealed before X_i if $\sigma(j) < \sigma(i)$. We redefine Z_i and N_i to take the ordering σ into account. Let

$$Z_i(\sigma) = \{j \in R_i : X_{i'} = j \text{ for some } \sigma(i') < \sigma(i)\}.$$

$$N_i(\sigma) = |R_i \setminus Z_i(\sigma)|.$$

Then $N_i(\sigma)$ is the number of available values for X_i, given all the variables X_j for j such that $\sigma(j) < \sigma(i)$. As before, using the chain rule we obtain the inequality

$$H(X) = \sum_{i=1}^{n} H(X_i | X_j : \sigma(j) < \sigma(i)) \leq \sum_{i=1}^{n} \mathbb{E}_X \left[\log(N_i(\sigma))\right].$$

The inequality remains true if we take the expected value of both sides when σ is a random permutation sampled from the uniform distribution on $\mathbb{S}_n$.

$$H(X) \leq \sum_{i=1}^{n} \mathbb{E}_\sigma \left[\mathbb{E}_X \left[\log(N_i(\sigma))\right]\right] = \sum_{i=1}^{n} \mathbb{E}_X \left[\mathbb{E}_\sigma \left[\log(N_i(\sigma))\right]\right].$$

Thus, the bound we get on $H(X)$ depends on the distribution of the random variable $N_i(\sigma)$. The final observation that we need is that the distribution of $N_i(\sigma)$ is very simple and that it does not depend on X. Consequently we can eliminate the step of taking expectation with respect to the choice of X. Let us fix a specific X.

Let W_i denote the set of $r_i - 1$ row indices $j \neq i$ for which $X_j \in R_i$. Note that N_i is equal to r_i minus the number of indices in W_i that precede i in the random ordering σ. Since σ was chosen uniformly, this number is distributed uniformly in $\{0, ..., r_i - 1\}$. Thus, N_i is uniform on the set $\{1, ..., r_i\}$. Therefore

$$\mathbb{E}_\sigma \left[\log(N_i(\sigma))\right] = \sum_{k=1}^{r_i} \frac{1}{r_i} \log(k) = \frac{1}{r_i} \log(r_i!).$$

Hence

$$H(X) \leq \sum_{i=1}^{n} \mathbb{E}_X \left[\frac{1}{r_i} \log(r_i!)\right] = \sum_{i=1}^{n} \frac{1}{r_i} \log(r_i!)$$

which implies the Brègman bound.

3 The d-dimensional case

3.1 An informal discussion

The core of the above-described proof of the Brègman bound can be viewed as follows. Let us pick first a 1-permutation X that is contained in the matrix A and consider the set R_i of the r_i 1-entries in A's i-th row. There are exactly r_i indices j for which $X_j \in R_i$. The random ordering of the rows that we select σ, determines which of these will precede the i-th row (or will *cast its shadow* on the i-th row). The random number u_i of rows that cast a shadow on the i-th row is uniformly distributed in the range $\{0, \ldots, r_i - 1\}$. The contribution of this row to the upper bound on $H(X)$ is $\log N_i$, where $N_i = r_i - u_i$. The expectation of $\log N_i$ is exactly $\frac{1}{r_i} \sum_{j=1,\ldots,r_i} \log j = \frac{1}{r_i} \log(r_i!)$. In other words, N_i is the number of 1-entries in the i-th row that are still unshaded when we reach the i-th row in the chain rule evaluation of the entropy. The contribution of this row to the estimate is $\log N_i$.

How should we modify this argument to deal with d-dimensional permutations? We fix a d-dimensional permutation X that is contained in A and consider a random ordering of all lines of the form $A(i_1, ..., i_d, *)$. Given such an ordering, we use the chain rule to derive an upper bound on $H(X)$. Each ordering yields a different bound. However, as in the one dimensional case, the key insight is that averaging over all possible orderings (in a class that we later define) gives us a simple bound on $H(X)$.

The overall structure of the argument remains the same. We consider a concrete line $A(i_1, ..., i_d, *)$. Its contribution to the estimate of the entropy is $\log N$ where N is the number of 1-entries that remain unshaded at the time (according to the chosen ordering) at which we compute the corresponding term in the chain rule for the entropy. However, now shade can fall from d different directions. The contribution of the line to the entropy will be the expected logarithm of the number of ones that remain unshaded after each of the d dimensions has cast its shade on it.

The ordering of the lines is done as follows: At the coarsest level lines are ordered according to their first coordinate i_1. This ordering is chosen uniformly from $\mathbb{S}_n$. The (random) order of lines with a fixed value of i_1 is chosen by recursion on the dimension d. To understand how many 1's remain unshaded in a given line, we first consider the shade along the first coordinate. If it initially has r 1-entries, then the number of unshaded 1-entries after this stage is uniformly distributed on $[r]$. We then recurse with the remaining 1-entries and proceed on the subcube of codimension 1 that is defined by the value of the first coordinate. It is not hard to see how the recursive expression for $f(d, r)$ reflects this calculation.

3.2 In detail

Let A be a $[n]^{d+1}$-dimensional array of zeros and ones, and X is a random d-permutation sampled uniformly from the set of d-permutations contained in A. Then $H(X) = \log(Per_d(A))$ by theorem 2.2 and again we seek an upper bound on $H(X)$.

We think of X as an $[n]^d$ array each line of which contains each member of $[n]$ exactly once. The proof does its accounting using lines of the form $A(i_1, ..., i_d, *)$, i.e., lines in which the last $((d+1)$-st) coordinate varies. Such a line is specified by $\mathbf{i} = (i_1, ..., i_d)$. The random variable $X_{\mathbf{i}}$ is defined to be the value of $X(i_1, ..., i_d)$. We think of the variables $X_{\mathbf{i}}$ as being revealed to us one by one. Thus, $X_{i_1,...,i_d}$ must belong to

$$R_{\mathbf{i}} = R_{i_1,...,i_d} = \{j : A(i_1, ..., i_d, j) = 1\}$$

the set of 1-entries in this line.

In the proof we scan these lines in a particular randomly chosen order. Let us ignore this issue for a moment and consider some fixed ordering of these lines. Initially, the number of 1-entries in this line is $r_{\mathbf{i}}$. As we proceed, some of these 1's become unavailable to $X_{\mathbf{i}}$, since choosing them would result in a conflict with the choice made in some previously revealed line. We say that these 1's are *in the shade* of previously considered lines. This shade can come from any of the d possible *directions*. Thus we denote by $Z_{\mathbf{i}} \subseteq R_{\mathbf{i}}$ the set of the indices of the 1-entries in $R_{\mathbf{i}}$ that are unavailable to $X_{\mathbf{i}}$ given the values of the preceding variables. We can express $Z_{\mathbf{i}} = \cup_{k=1,...,d} Z_{\mathbf{i}}^k$ where entries in $Z_{\mathbf{i}}^k$ are shaded from direction k. Namely, $j \in Z_{\mathbf{i}}^k$ if there is an already scanned line indexed by **i'** with $X_{\mathbf{i'}} = j$ and where **i** and **i'** coincide on all coordinates except the k-th. Thus, given the values of the previously considered variables, there are at most

$$N_{\mathbf{i}} = |R_{\mathbf{i}} \setminus Z_{\mathbf{i}}|$$

values that are available to $X_{\mathbf{i}}$.

We next turn to the random ordering of the lines. Now, however, we do not select a completely random ordering, but opt for a more structured ordering that respects the coordinate structure. Namely, we first select a random ordering σ on $[n]$. If $i_1 \neq i'_1$, then all lines of the form $A(i_1, ..., i_d, *)$ precede all those of the form $A(i'_1, ..., i'_d, *)$ provided that i_1 precedes i'_1 in the ordering σ. We next select $\sigma_1, \dots, \sigma_n$ which are also orderings of $[n]$. If $i_1 = i'_1 = i$ but $i_2 \neq i'_2$, then the precedence between lines of the form $A(i, i_2, ..., i_d, *)$ and $A(i, i'_2, ..., i'_d, *)$ is determined by the order of $\sigma_i(i_2)$ and $\sigma_i(i'_2)$. This is extended appropriately to all further coordinates where we select for each $d-1 \geq k \geq 0$ and each choice of $(i_1, \dots, i_k)$, a random ordering $\sigma_{(i_1,\dots,i_k)}$ of $[n]$. To determine which of the two lines $A(i_1, ..., i_d, *)$ and $A(i'_1, ..., i'_d, *)$ precede each other, we consult the ordering $\sigma_{(i_1,\dots,i_k)}$, where $i_1 = i'_1, \dots, i_k = i'_k$ but $i_{k+1} \neq i'_{k+1}$. The ordering between these lines is the same as the ordering between $\sigma_{(i_1,\dots,i_k)}(i_{k+1})$ and $\sigma_{(i_1,\dots,i_k)}(i'_{k+1})$. Thus a complete choice of the orderings $\sigma_{(i_1,\dots,i_k)}$ induces a total order on the lines $A(i_1, ..., i_d, *)$. Denote this order by $\prec$. That is, we write $\mathbf{i} \prec \mathbf{j}$ if $\mathbf{i}$ comes before $\mathbf{j}$. We write $\mathbf{i} \prec_k \mathbf{j}$ if $\mathbf{i} \prec \mathbf{j}$ and $\mathbf{i}$ and **i'** differ *only* in the kth coordinate.

We think of $X_{\mathbf{i}}$ as being revealed to us according to this order.

We turn to the definition of $R_{\mathbf{i}}$, $Z^k_{\mathbf{i}}$ and $N_{\mathbf{i}}$. Their definitions are affected by the chosen ordering of the lines. In addition, for reasons to be made clear later, we generalize the definition of $N_{\mathbf{i}}$. It is defined as the number of values available to $X_{\mathbf{i}}$ (given the preceding lines) from a given index set $W \subseteq R_{\mathbf{i}}$. In the discussion below, we fix X, a d-dimensional permutation that is contained in A.

Definition 3.1. *The index set of the* 1*-entries in the line* $A(i_1, ..., i_d, *)$ *is denoted by*

$$R_{\boldsymbol{i}} = R_{i_1,\dots,i_d} = \{j : A(i_1, ..., i_d, j) = 1\},$$

and its cardinality is $r_{\boldsymbol{i}} = |R_{\boldsymbol{i}}|$.

Let $W \subseteq R_{\boldsymbol{i}}$ *with* $\boldsymbol{i} = (i_1, ..., i_d)$*, and suppose that* $X_{\boldsymbol{i}}$ *belongs to* W*. For a given ordering* $\prec$*, let*

$$Z^k_{\boldsymbol{i}}(X, \prec) = \{j \in R_{\boldsymbol{i}} : X_{\boldsymbol{i'}} = j \text{ for some } \boldsymbol{i'} \prec_k \boldsymbol{i}\}.$$

$$N_{\boldsymbol{i}}(W, X, \prec) = |W \setminus \cup_{k=1}^d Z^k_{\boldsymbol{i}}(X, \prec)|.$$

Thus, $N_{\mathbf{i}}$ is a function of $W \subseteq R_{\mathbf{i}}$, X and the ordering $\prec$. Each variable $X_{\mathbf{i}}$ specifies a 1 entry of the line $A(i_1, ..., i_d, *)$. The entry thus specified must conform to the values taken by the preceding variables. Namely, no line of X can contain more than a single 1 entry. We consider the number of values that the variable $X_{\mathbf{i}}$ can take, given the values that precede it. Fix an index tuple $\mathbf{i} = (i_1, ..., i_d)$. The variable $X_{\mathbf{i}}$ should specify an index i_{d+1} with $A(i_1, ..., i_{d+1}) = 1$, i.e., an element of $R_{\mathbf{i}}$. Consider some element $j \in R_{\mathbf{i}}$. If $X_{\mathbf{i'}} = j$, for some $\mathbf{i'} \prec_k \mathbf{i}$ and $k \leq d$ then clearly $X_{\mathbf{i}} \neq j$, or else the line $X(i_1, ..., i_{k-1}, *, i_{k+1}, ..., i_d)$ contains more than a single j-entry. In other words, $X_{\mathbf{i}}$ cannot specify an element of $Z^k_{\mathbf{i}}(X, \prec)$ and is restricted to the set $R_{\mathbf{i}} \setminus \cup_{k=1}^{d} Z^k_{\mathbf{i}}(X, \prec)$. Therefore, there are at most $N_{\mathbf{i}}(R_i, X, \prec)$ possible values that $X_{\mathbf{i}}$ can take given the variables that precede it in the order $\prec$.

For a given order $\prec$, we can use the chain rule to derive

$$H(X) = \sum_{\mathbf{i}} H(X_{\mathbf{i}} | X_{\mathbf{j}} : \mathbf{j} \prec \mathbf{i}).$$

By theorem 2.2,

$$\begin{aligned} H(X_{\mathbf{i}} | X_{\mathbf{j}} : \mathbf{j} \prec \mathbf{i}) &= \mathbb{E}_{X_{\mathbf{j}}:\mathbf{j} \prec \mathbf{i}} \left[H(X_{\mathbf{i}} | X_{\mathbf{j}} = x_{\mathbf{j}} : \mathbf{j} \prec \mathbf{i}) \right] \\ &\leq \mathbb{E}_{X_{\mathbf{j}}:\mathbf{j} \prec \mathbf{i}} \left[\log(N_{\mathbf{i}}(R_{\mathbf{i}}, X, \prec)) \right] = \mathbb{E}_X \left[\log(N_{\mathbf{i}}(R_{\mathbf{i}}, X, \prec)) \right]. \end{aligned}$$

The last equality holds because $N_{\mathbf{i}}$ depends only on the lines of X that precede $X_{\mathbf{i}}$, and so taking the expectation over the rest of X doesn't change anything.

As in the one dimensional case, the next step is to take the expectation of both sides of the above inequality over $\prec$.

$$\begin{aligned} H(X) &\leq \sum_{\mathbf{i}} \mathbb{E}_{\prec} \left[\mathbb{E}_X \left[\log(N_{\mathbf{i}}(R_{\mathbf{i}}, X, \prec)) \right] \right] \\ &= \sum_{\mathbf{i}} \mathbb{E}_X \left[\mathbb{E}_{\prec} \left[\log(N_{\mathbf{i}}(R_{\mathbf{i}}, X, \prec)) \right] \right]. \end{aligned}$$

The key to unraveling this expression is the insight that the random variable $N_{\mathbf{i}}$ has a simple distribution (as a function of $\prec$), and moreover, that this distribution does not depend on X.

Recall that in the one dimensional case, we obtained the distribution of N_i as follows. Initially, the number of ones in the ith row was r_i. Then the rows preceding the ith row were revealed, and some of the ones in the i-th row became unavailable to X, because some other row had placed a one in their column. We defined $N_i = |R_i \setminus Z_i(\sigma)|$. The size of $Z_i(\sigma)$ was shown to be uniformly distributed over $\{0, ..., r_i - 1\}$, and thus the distribution of N_i was shown to be uniform over $\{1, ..., r_i\}$.

A similar argument works in the d dimensional case, but the distribution of $N_{\mathbf{i}}$ is no longer uniform. Recall that the function f is defined recursively by

$$f(0,r) = \log(r)$$
$$f(d,r) = \frac{1}{r}\sum_{k=1}^{r} f(d-1,k).$$

Claim 3.2. *Let X be a d-permutation, $\boldsymbol{i} = (i_1, ..., i_d)$ and let W be an index set such that $X_{\boldsymbol{i}} \in W$. Then $\mathbb{E}_{\prec}[\log(N_{\boldsymbol{i}}(W, X, \prec))]$ depends only on d and $r = |W|$, and*

$$\mathbb{E}_{\prec}[\log(N_{\boldsymbol{i}}(W, X, \prec))] = f(d,r).$$

Proof. The proof proceeds by induction on d.

First, note that if $|W| = r$ and $d = 0$, then $N_{\mathbf{i}}(W, X, \prec) = |W| = r$ by definition, and therefore

$$\mathbb{E}_{\prec}[\log(N_{\mathbf{i}}(W, X, \prec))] = \log(r) = f(0,r).$$

In order to proceed with the induction step, we must describe $N_{\mathbf{i}}(W, X, \prec)$ in terms of parameters of dimension $d-1$ instead of d. To this end we need the following definitions:

- $X' = X(i_1, *, ..., *)$. Note that X' is a $(d-1)$-dimensional permutation.
- $W' = W \setminus Z^1_{\mathbf{i}}(X, \prec)$. Note that $|W'|$ actually depends only on σ, the ordering of the first coordinate.
- Let **i'** $= (i'_1, ..., i'_{d-1}) = (i_2, ..., i_d)$.
- Given an ordering $\prec$, let $\prec'$ be the ordering on the index tuples $(i'_1, ..., i'_{d-1})$ defined by the orderings

$$\sigma_{i_1}, \sigma_{i_1,i_2}, ..., \sigma_{i_1,...,i_{d-1}}.$$

Note that for every X,W, $\mathbf{i}$ and $\prec$ we have $N_{\mathbf{i}}(W, X, \prec) = N_{\mathbf{i'}}(W', X', \prec')$. This equality follows directly from the definition of N. Now,

$$\mathbb{E}_{\prec}[\log(N_{\mathbf{i}}(W, X, \prec))] = \mathbb{E}_{\sigma}[\mathbb{E}_{\prec'}[\log(N_{\mathbf{i}}(W, X, \prec))]]$$

$$= \mathbb{E}_\sigma \left[\mathbb{E}_{\prec'} \left[\log(N_{\mathbf{i'}}(W', X', \prec')) \right] \right] = \mathbb{E}_\sigma \left[f(d-1, |W'|) \right]$$

The last step follows from the induction hypothesis. Consequently,

$$\mathbb{E}_\prec \left[\log(N_{\mathbf{i}}(W, X, \prec)) \right] = \sum_k \Pr(|W'| = k) f(d-1, k).$$

The only remaining question is to determine the distribution of $|W'|$ as a function of σ. Note, however, that we have already answered this question in the one dimensional proof, namely, $|W'|$ is uniformly distributed on $\{1, ..., r\}$. Indeed, $W' = |W \setminus Z^1_{\mathbf{i}}(X, \prec)|$, and $Z^1_{\mathbf{i}}(X, \prec)$ is the set of indices s such that:

- For some $j \in W$, $X(s, i_2, ..., i_d) = j$ (there are $r - 1$ such indices, one for each $j \in W$).
- The random ordering σ places s before i_1.

In a random ordering, the position of i_1 is uniformly distributed. Therefore $|Z^1_{\mathbf{i}}(X, \prec)|$ is uniformly distributed on $\{0, ..., r-1\}$, and $Pr(|W'| = k) = \frac{1}{r}$ for every $1 \leq k \leq r$.

Putting this together, we have shown that

$$\mathbb{E}_\prec \left[\log(N_{\mathbf{i}}(W, X, \prec)) \right] = \frac{1}{r} \sum_{k=1}^{r} f(d-1, k) = f(d, r).$$

□

In conclusion, we have shown that

$$H(X) \leq \sum_{\mathbf{i}} \mathbb{E}_X \left[\mathbb{E}_\prec \left[\log(N_{\mathbf{i}}(R_{\mathbf{i}}, X, \prec)) \right] \right]$$

$$= \sum_{\mathbf{i}} \mathbb{E}_X \left[f(d, r_{\mathbf{i}}) \right] = \sum_{\mathbf{i}} f(d, r_{\mathbf{i}}),$$

where $r_{\mathbf{i}} = r_{i_1, \ldots, i_d}$ is the number of ones in the vector $A(i_1, ..., i_d, *)$. Therefore,

$$Per_d(A) \leq \prod_{\mathbf{i}} e^{f(d, r_{\mathbf{i}})}.$$

4 An upper bound for the number of d-permutations

As mentioned, the upper bound on the number of d-dimensional permutations is derived by applying theorem 1.6 to the all-ones array J. The main technical step is a derivation of an upper bound on the function $f(d,r)$.

Theorem 4.1. *For every d there exist constants c_d and r_d such that for all $r \geq r_d$,*

$$f(d,r) \leq \log(r) - d + c_d \frac{\log^d(r)}{r}.$$

One possible choice that we adopt here is $r_d = e^d$ for every d, $c_1 = 5$, $c_2 = 8$, and $c_d = \frac{d^3(1.1)^d}{d!}$ for $d \geq 3$.

Proof. A straightforward induction on d yields the weaker bound $f(d,r) \leq \log(r)$ for all d, r. For $d = 0$ there is equality and the general case follows since $f(d,r) = \frac{1}{r}\sum_{k=1}^{r} f(d-1,k) \leq \frac{1}{r}\sum_{k=1}^{r}\log(k) \leq \log(r)$. This simple bound serves us to deal with the range of small r's (below r_{d-1}). We turn to the main part of the proof

$$f(d,r) = \frac{1}{r}\sum_{k=1}^{r} f(d-1,k) = \frac{1}{r}\left[\sum_{k=1}^{r_{d-1}} f(d-1,k) + \sum_{k=r_{d-1}+1}^{r} f(d-1,k)\right]$$

$$\leq \frac{1}{r}\left[r_{d-1}\log(r_{d-1}) + \sum_{k=1}^{r}\log(k) - (d-1) + c_{d-1}\frac{\log^{d-1}(k)}{k}\right]$$

$$\leq \frac{\xi}{r} + \frac{1}{r}\log(r!) - (d-1) + \frac{c_{d-1}}{r}\sum_{k=1}^{r}\frac{\log^{d-1}(k)}{k}$$

where $\xi = r_{d-1}\log(r_{d-1}) = (d-1)e^{d-1}$. It is easily verified that for $r \geq r_d \geq 3$ there holds $\log(r!) \leq r\log(r) - r + 2\log(r)$. We can proceed with

$$\leq \frac{\xi}{r} + \log(r) + \frac{2\log(r)}{r} - d + \frac{c_{d-1}}{r}\sum_{k=1}^{r}\frac{\log^{d-1}(k)}{k}.$$

We now bound the sum $\sum_{k=1}^{r}\frac{\log^{d-1}(k)}{k}$ by means of the integral $\int_1^r \frac{\log^{d-1}(x)dx}{x} = \frac{\log^d(r)}{d}$. Note that the integrand is unimodal and its maximal value is $\gamma =$

$\left(\frac{d-1}{e}\right)^{d-1}$. Thus,

$$\frac{c_{d-1}}{r}\sum_{k=1}^{r}\frac{\log^{d-1}(k)}{k} \le \frac{c_{d-1}}{r}\left(\frac{\log^d(r)}{d}+\gamma\right).$$

Putting this together, we have the inequality

$$f(d,r) \le \log(r) - d + \frac{2\log(r)+\xi+c_{d-1}\left(\gamma+\frac{\log^d(r)}{d}\right)}{r}.$$

Therefore it is sufficient to choose c_d such that for every $r \ge e^d$

$$2\log(r)+\xi+c_{d-1}\left(\gamma+\frac{\log^d(r)}{d}\right) \le c_d \log^d(r)$$

i.e.,

$$\frac{2}{\log^{d-1}(r)}+\frac{\xi}{\log^d(r)}+c_{d-1}\left(\frac{\gamma}{\log^d(r)}+\frac{1}{d}\right) \le c_d.$$

The left hand side of the above inequality is clearly a decreasing function of r. Therefore it is sufficient to verify the inequality for $r = e^d$. Plugging this and the values of the constants ξ and γ into the left hand side of the above inequality, we get

$$\frac{2}{d^{d-1}}+\frac{(d-1)e^{d-1}}{d^d}+c_{d-1}\left(\frac{(d-1)^{d-1}}{e^{d-1}d^d}+\frac{1}{d}\right)$$
$$\le \left(1+\frac{1}{e^{d-1}}\right)\frac{c_{d-1}}{d}+d\left(\frac{2}{d^d}+\left(\frac{e}{d}\right)^d\right).$$

Thus, we may take

$$c_d = \left(1+\frac{1}{e^{d-1}}\right)\frac{c_{d-1}}{d}+d\left(\frac{2}{d^d}+\left(\frac{e}{d}\right)^d\right).$$

Calculating c_d using this recursion and the fact that $c_0 = 0$, we get that $c_1 = 2+e \le 5$, $c_2 \le 8$, and $c_3 \le \frac{3^3(1.1)^3}{3!}$. Proceeding by induction,

$$c_d = \left(1+\frac{1}{e^{d-1}}\right)\frac{(d-1)^3(1.1)^{d-1}}{d!}+d\left(\frac{2}{d^d}+\left(\frac{e}{d}\right)^d\right)$$

$$\leq \frac{(1.1)^d(d-1)^3}{d!} + 2d\left(\frac{e}{d}\right)^d \leq \frac{(1.1)^d(d-1)^3 + 2d^2}{d!} \leq \frac{(1.1)^d d^3}{d!}.$$

In the inequality before the last one, we used Stirling's approximation to show that $\frac{d}{d!} \leq \left(\frac{e}{d}\right)^d$.

□

For the $[n]^{d+1}$ all ones array J, $r_{i_1,\ldots,i_d} = n$ for every tuple $(i_1, \ldots, i_d)$, and so for large enough n we have the bound

$$Per_d(J) \leq \prod_{i_1,\ldots,i_d} e^{f(d,n)} = (e^{f(d,n)})^{n^d} \leq exp\left[\log(n) - d + c_d\frac{\log^d(n)}{n}\right]^{n^d}.$$

For a constant d, letting n go to infinity, $c_d\frac{\log^d(n)}{n} = o(1)$ and therefore the number of d permutations is at most

$$\left(\frac{n}{e^d}(1+o(1))\right)^{n^d}.$$

References

[1] L. M. Brègman, *Certain properties of nonnegative matrices and their permanents*, Dokl. Akad. Nauk SSSR 211 (1973), 27-30. MR MR0327788 (48 #6130)

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.

[3] G.P. Egorichev, *Proof of the van der Waerden conjecture for permanents*, Siberian Math. J. 22 (1981), 854-859.

[4] D.I. Falikman, *A proof of the van der Waerden conjecture regarding the permanent of a doubly stochastic matrix*, Math. Notes Acad. Sci. USSR 29 (1981), 475-479.

[5] L. Gurvits , *Van der Waerden/SchrijverValiant like conjectures and stable (aka hyperbolic) homogeneous polynomials: one theorem for all. With a corrigendum*, Electron. J. Combin. 15 (2008), R66 (26 pp).

[6] M. Jerrum, A. Sinclair, and E. Vigoda, *A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries*, J. ACM, 671-697.

[7] M. Kochol, *Relatively narrow Latin parallelepipeds that cannot be extended to a Latin cube*, Ars Combin. 40 (1995), 247260.

[8] Laurent, M. and Schrijver, A., *On Leonid Gurvitss proof for permanents*, The American Mathematical Monthly 10 (2010), 903–911.

[9] H. Minc, *Permanents*, Encyclopedia of Mathematics and Its Applications Vol. 6, Addison-Wesley, Reading, Mass, 1978

[10] Jaikumar Radhakrishnan, *An entropy proof of Bregman's theorem*, J. Combinatorial Theory Ser. A 77 (1997), no. 1, 80-83. MR MR1426744 (97m:15006)

[11] L. G. Valiant, *The complexity of computing the permanent*, Theoret. Comput. Sci. 8 (1979), 189-201

[12] J. H. van Lint and R. M. Wilson, *A Course in Combinatorics*, Cambridge U.P., 1992.

[13] B.D. McKay and I.M. Wanless, *A census of small latin hypercubes*, SIAM J. Discrete Math. 22 (2008), pp. 719736.