האוניברסיטה העברית בירושלים
The Hebrew University of Jerusalem

HIGH DIMENSIONAL COMBINATORICS

A thesis submitted for the degree of
Doctor of Philosophy

By

Zur Luria

Submitted to the senate of the Hebrew University

July 2014

This work was carried out under the supervision of

Prof. Nathan Linial

Acknowledgements

First and foremost I would like to thank my advisor, Prof. Nathan (Nati) Linial for his endless patience and encouragement. He taught me, both consciously and unconsciously, how good mathematics is done. I appreciate all of his contributions of time, ideas, and funding to make my Ph.D. productive and stimulating. Nati's enthusiastic and optimistic approach to research was contagious and motivational for me, even during tough times.

This thesis would not have been possible without the support of my parents. I am grateful for their ongoing encouragement and their help and support.

Last but not least, I would like to thank all of the people in the theory lab for making the university a fun place to which I was happy to come every day.

Abstract

In many areas of science and mathematics we are interested in modeling large systems that are defined through *pairwise* interactions. The most natural object to model such interactions is a graph, which is just a collection of pairs. A permutation is equivalent to a perfect matching in a bipartite graph, so it also encodes information about pairs. Even such ubiquitous objects as matrices fall into this framework, because a matrix is a mapping from pairs of indices to some set.

However, in many cases the interesting interactions involve more than two entities, and there is still not enough theoretical machinery to deal with this difficulty. High dimensional combinatorics involves the development of combinatorial objects relevant to multi-way interactions, which we call high dimensional objects.

Often there are several different ways to generalize a classical object such as a perfect matching or a permutation to higher dimensions. Usually, these different higher dimensional objects correspond to different answers to the questions: What is a perfect matching? What is a permutation?

For example, a matching in a graph is a set of disjoint edges. This definition points to one generalization of matchings - a d-matching can be defined to be a set of disjoint edges in a $(d+1)$-uniform hypergraph. However, a matching is also a collection of 2-sets (that is, sets of size 2) where every 1-set is contained in at most one 2-set. This point of view leads the definition of what are known as *designs*. We can also view perfect matchings in bipartite graphs as permutation matrices, or one to one mappings, and these definitions also lead to a different types of high dimensional matchings.

We are interested in finding generalizations that are useful and natural, and studying their basic combinatorial properties in light of what is known about their lower dimensional counterparts. We ask questions like:

- What are the best high dimensional analogs for a permutation (or a perfect matching, or a coloring of a graph, etc.)?
- Given such an analog, *how many* such objects of a given size exist?
- Is there an efficient algorithm to construct such an object?
- What are the properties of a typical object?

In addition to the introduction, this thesis contains the following chapters.

Chapter 2: An upper bound on the number of high-dimensional permutations: This paper introduces the notion of a high dimensional permutation as a gen-

eralization of a permutation matrix. These objects are also equivalent to Latin squares, cubes and hypercubes, which are important and well studied objects in their own right.

We define a high dimensional Permanent function on arrays of arbitrary dimension, and prove an upper bound on the permanent of 0-1 arrays. Our upper bound is a generalization of Brègman's upper bound on the Permanent. Using this bound we prove an upper bound on the number of order n d dimensional permutations.

Our proof uses the entropy method, a type of probabilistic method that has proven very useful for enumeration problems in combinatorics.

Chapter 3: An upper bound on the number of Steiner triple systems: This paper considers Steiner triple systems, which are a basic type of combinatorial design. Steiner triple systems have a long history, and have been extensively studied. In particular, in 1974 Richard Wilson proved asymptotic upper and lower bounds on their number [12]. He conjectured that in fact the number of order n Steiner triple systems is equal to $\left((1+o(1))\frac{n}{e^2}\right)^{n^2/6}$. We proved that this is an upper bound, resolving half of Wilson's conjecture.

Chapter 4: On the vertices of the d-dimensional Birkhoff polytope: This paper introduces the notion of d-stochastic arrays as a high dimensional analog of doubly stochastic matrices. An important basic fact about doubly stochastic matrices is Birkhoff's theorem, which states that every doubly stochastic matrix is a convex combination of permutation matrices. In geometric terms, this means that the vertices of the polytope of doubly stochastic matrices are precisely the permutation matrices.

We ask whether the same thing happens in higher dimensions. It turns out that the answer is no. As we show, even for tristochastic arrays the number of vertices is asymptotically much larger than the number of 2-permutations.

Contents

Chapter 1

Introduction

1.1 Classical objects

1.1.1 Permutations

A permutation of order n is an ordering of the numbers $1, ..., n$. Permutations are ubiquitous objects that appear in many different contexts and have many different interpretations.

The set of order-n permutations is denoted by S_n. A permutation $\sigma \in S_n$ can be represented by a vector of length n over $\{1, ..., n\}$ whose entries are all different. Alternatively, σ can be represented by a permutation matrix.

A permutation matrix is an $n \times n$ 0-1 matrix with a single 1 in each row and and a single 1 in each column. Given a permutation $\sigma = (\sigma(1), ..., \sigma(n))$, the corresponding permutation matrix A is given by $A(i, \sigma(i)) = 1$ for $i \in \{1, ..., n\}$ and $A = 0$ elsewhere.

$$\begin{pmatrix} 5 \\ 3 \\ 1 \\ 2 \\ 4 \end{pmatrix} \leftrightarrow \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

It is also possible to view an order-n permutation as a perfect matching in a bipartite graph with n vertices on each side. Given a permutation σ, the corresponding perfect matching on $G = \langle U \cup V, E \rangle$ is given by $M = \{\{u_j, v_{\sigma(j)}\}\}_{j=1}^{n}$.

The number of order n permutations is $n!$, which by Stirling's approximation is equal to $\left((1 + o(1))\frac{n}{e}\right)^n$. It is very easy to sample an order-n permutation at random, and the properties of random permutations have been extensively studied and are well understood.

1.1.2 Regular Graphs

A graph is defined by set of vertices V and a set of edges $E \subseteq \binom{V}{2}$. The degree of a vertex $v \in V$ is the number of edges that contain it. A graph is *d-regular* if all of the vertices have the same degree, which means that every vertex is contained in exactly d edges.

One basic fact about regular graphs is that there exists a d-regular graph on n vertices iff $n > d$ and $d \cdot n$ is even. This is a necessary condition because the number of edges in a graph is equal to half of the sum of the degrees. In a regular graph the sum of the degrees is $d \cdot n$, so this number must be even. It is less immediate, but also easy, to show that there exists a d-regular graph on n vertices whenever $n > d$ and $d \cdot n$ is even.

However, sampling an n vertex d-regular graph at random is a difficult problem which remains an active topic of research. For small values of d, this problem was solved in the 80's in [2], [1] via the configuration model.

The study of random regular graphs started to take a central position in combinatorics and computer science when it became clear that such graphs were *expanders* with high probability. These are graphs that have properties that make them very useful for many practical and theoretical applications. For more on this, see [8].

The enumeration of regular graphs is also of great interest. There is no known closed formula for the number of d-regular graphs on n vertices. It is, however, possible to develop asymptotic formulas for this number [4].

1.1.3 The Permanent

Let A be an $n \times n$ matrix. The *permanent* of A is defined by

$$Per(A) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} A(i, \sigma(i)).$$

Unlike its cousin the determinant, calculating the permanent of a matrix is a very difficult computational problem. In fact, it is known to be $\#P$-complete, which implies that the existence of a polynomial time algorithm to compute the permanent of an $n \times n$ matrix would imply that $P = NP$.

Permanents are connected to perfect matchings in bipartite graphs. If A is a 0-1 matrix, its permanent is the number of permutation matrices whose 1-entries are contained by A's support. Let $G = \langle U \cup V, E \rangle$ be a bipartite graph with $|U| = |V| = n$. We define G's adjacency matrix to be a 0-1 $n \times n$ matrix with $A(i, j) = 1$ iff $\{u_i, v_j\} \in E$. This is a one to one correspondence between 0-1 $n \times n$ matrices and bipartite graphs with n vertices on each side. A perfect matching in G corresponds to a permutation matrix contained by A, so $Per(A)$ is the number of perfect matchings in G. The problem of computing the permanent restricted to 0-1 matrices remains $\#P$-complete, however.

Since the permanent is so difficult to compute exactly, much attention has focused on the problems of bounding and approximating the permanent of a matrix. Minc conjectured an upper bound for the permanent, which was proved by Brègman in 1973 [2], and in the 1920's van der Waerden conjectured a lower bound, which was proved by Falikman [6] and Egorichev [5] in 1981. These two bounds have been of great importance to us in our research.

1.1.4 Doubly Stochastic Matrices

A doubly stochastic matrix is a nonnegative $n \times n$ matrix whose rows and columns all sum to one. The following is an example of an order-3 permutation matrix.

$$\begin{pmatrix} 0 & 0.4 & 0.6 \\ 0 & 0.6 & 0.4 \\ 1 & 0 & 0 \end{pmatrix}$$

Van der Waerden's conjecture (= The Falikman-Egorychev theorem) is a lower bound on the permanent of doubly stochastic matrices.

The following theorem is an important fact about doubly stochastic matrices that connects them to permutations.

Theorem 1.1.1 (Birkhoff-von Neumann)**.** *Every doubly stochastic matrix is a convex combination of permutation matrices.*

It is possible to state this result in geometric terms. The set of order-n doubly stochastic matrices is a polytope in $\mathbb{R}^{n^2}$. Birkhoff's theorem means that the vertices of this polytope are precisely the permutation matrices.

1.2 The Search for the Right Generalization

1.2.1 High Dimensional Permutations

There are several natural ways to generalize permutations to high dimensions. The generalization that we define here stems from the representation of permutations as permutation matrices.

As mentioned, a permutation matrix is an $n \times n$ 0-1 matrix with a single 1 in every row and column. We define a d-permutation to be an $\underbrace{n \times ... \times n}_{d+1}$ 0-1 array with a single 1 in each *line*. Here a line is the set of entries we get by fixing all but one of the indices and allowing the last index to vary over $\{1, ..., n\}$. Thus, an order-n d permutation has $(d+1)n^d$ lines.

A *Latin square* is an $n \times n$ matrix over $\{1, ..., n\}$ such that the entries in every row and column are all different. Latin squares are well known since ancient times, and they have been extensively studied.

Just as permutation matrices are equivalent permutations, 2-permutations are equivalent to Latin squares. In general, d-permutations are equivalent to d-dimensional Latin hypercubes. These are $\underbrace{n \times ... \times n}_{d}$ arrays over $\{1, ..., n\}$ such that the entries in every line are all different.

High dimensional permutations are considered in chapter 2. Our main result is an upper bound on their number. Namely, we prove that for any constant d, the number of order-n d-permutations is at most $\left((1 + o(1))\frac{n}{e^d})\right)^{n^d}$. We conjecture that this is indeed the correct asymptotic formula, and note that, if true, this would constitute a generalization of Stirling's approximation.

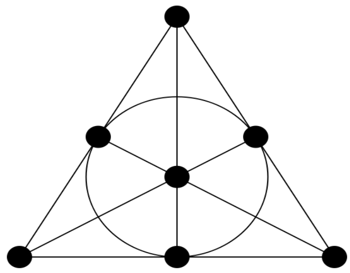

Figure 1.1: An order 7 Steiner triple system. Every line represents a triple.

Our proof proceeds by first establishing an upper bound on a high dimensional analog of the permanent. This bound is an analog of Brègman's theorem. It turns out that the natural analog of the van der Waerden conjecture is false for higher dimensions.

1.2.2 High Dimensional Perfect Matchings

A perfect matching is a collection M of pairs (or edges) from a set V such that every vertex in V appears in exactly one pair $e \in E$.

A Steiner Triple System (STS) is a collection X of *triples* from a set S such that every *pair* of objects in S appears in exactly one triple $x \in X$.

Steiner triple systems have an interesting history. They first appeared as a math puzzle in a magazine in 1844. We are interested in them because they are a higher dimensional analog of a perfect matching.

A perfect matching of a graph G is a subset of its edges such that every vertex is contained in a unique edge. As mentioned, a perfect matching in a bipartite graph corresponds to a permutation matrix.

There is another way in which Steiner triple systems can be viewed as higher dimensional perfect matchings. For a (not necessarily bipartite) graph, a perfect matching corresponds to A symmetric permutation matrix with zeros on the main diagonal. Given a perfect matching on n vertices, we construct such a permutation matrix A by setting $A(i,j) = 1$ iff $\{v_i, v_j\}$ is an edge in the matching. A is symmetric because a perfect matching is a symmetric object. If v_i is matched to v_j then v_j is matched to v_i.

It is interesting to note that Steiner triple systems are equivalent to symmetric 2-permutations. That is, if X is a Steiner triple system, we define a corresponding 2-permutation A by $A(i,j,k) = 1$ iff $\{i,j,k\} \in X$, $A(i,i,i) = 1$ for every i and $A = 0$ elsewhere. It is easy to see that A is indeed a 2-permutation. Moreover, A is symmetric in the sense that it is invariant to a permutation on the indices: $A(i,j,k) = A(j,k,i) = ...$

Let $STS(n)$ denote the number of Steiner triple systems of order n. We would like to find an asymptotic formula for $STS(n)$.

In 1974, Richard Wilson proved that

$$\left(\frac{n}{e^2 3^{3/2}}\right)^{\frac{n^2}{6}} \leq STS(n) \leq \left(\frac{n}{e^{1/2}}\right)^{\frac{n^2}{6}}.$$

Wilson conjectured that, in fact,

$$STS(n) = \left((1+o(1))\frac{n}{e^2}\right)^{\frac{n^2}{6}}.$$

We proved that $STS(n) \leq \left((1+o(1))\frac{n}{e^2})\right)^{n^2/6}$, resolving half of Wilson's conjecture. This result is presented in chapter 3.

1.2.3 High Dimensional Regularity

The natural high dimensional analog of a graph is a d-uniform hypergraph, which is a pair $H = \langle V, E\rangle$ such that $E \subseteq \binom{V}{d}$.

For every $1 \leq l \leq d-1$, we can extend the notion of regularity to d-regular graphs in the following way. We say that H is regular if every l-element subset of V is contained in exactly k d-element subsets in E. Note that when $l = 1$ and $d = 2$ we get the usual notion of regularity in graphs.

Just as a perfect matching is a 1-regular graph, we can think of Steiner triple systems as 1-regular 3-uniform hypergraphs (with $l = 2$). Part of our interest in Steiner triple systems stems from the fact that understanding them seems to be the first step on the way to understanding this notion of high dimensional regularity.

1.2.4 d-Stochastic Arrays

We define a d-stochastic array to be a $\underbrace{n \times ... \times n}_{d+1}$ nonnegative array such that every line sums to one. Note that a 1-stochastic array is a doubly stochastic matrix.

In the spirit of Birkhoff's theorem, it is natural to wonder whether the convex hull of the set of order n d-permutations is equivalent to the set of d-stochastic arrays. Birkhoff's theorem tells us that for $d = 1$ the answer is yes.

It turns out that in higher dimensions the answer is no. Indeed, as we show in chapter 4, even for $d = 2$ the number of vertices of the polytope of order-n 2-stochastic arrays is asymptotically much greater than the number of order-n 2-permutations.

1.3 The Entropy Method

Claude Shannon founded the field of information theory with a landmark paper that was published in 1948 [9]. Shannon's main motivation was to prove fundamental bounds on data processing tasks such as communication and compression. Since then, information theory has found applications in many diverse areas in mathematics, physics and computer science.

The central concept in information theory is information entropy, which measures the amount of uncertainty inherent in a random process. The entropy of a discrete random variable X is

$$H(X) = \sum_x \Pr(X = x)\log_2\left(\frac{1}{\Pr(X = x)}\right).$$

If a random variable can take n possible values, then its entropy is at most $\log_2 n$, and the entropy of a random variable that is uniformly distributed over n possible values is equal to $\log_2 n$. In general, the number of bits needed on average to encode X's value (given that we know X's distribution) is $H(X)$.

Recently, combinatorists have started to make use of information entropy for enumeration problems. For a survey of recent works that use these methods, see [7]. The basic idea is as follows.

We want to count the number of objects in a set S. Instead of addressing the problem directly, we use the many tools that have been developed to approximate the entropy of random variables.

- We define a random variable X that is uniformly distributed on S. In other words, we sample a member of the set S uniformly at random.
- Use information theoretic methods to bound $H(X)$. This, of course, is the crucial step.
- Since $H(X) = \log(|S|)$, this yields bounds on the number of objects in S.

This method was a central tool in our works on the asymptotic enumeration of high dimensional permutations and Steiner triple systems.

Bibliography

[1] E. A. Bender and E. R. Canfield, *The asymptotic number of non-negative integer matrices with given row and column sums*, Journal of Combinatorial Theory, Series A, 24 (1978), 296307.

[2] B. Bollobás, *A probabilistic proof of an asymptotic formula for the number of labelled regular graphs*, European Journal of Combinatorics, 1 (1980), 311316.

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

[4] E. R. Canfield and B. McKay, *Asymptotic enumeration of dense 0-1 matrices with equal row sums and equal column sums*, Electronic Journal of Combinatorics, 12 (2005), 31.

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

[6] 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.

[7] Galvin, David, *Three tutorial lectures on entropy and counting*, arXiv preprint arXiv:1406.7872 (2014).

[8] S. Hoory, N. Linial and A. Wigderson, *Expander graphs and their applications*, Bulletin of the American mathematical society, 43 (2006), 439-561.

[9] C. E. Shannon, *A mathematical theory of communication*, ACM SIGMOBILE Mobile Computing and Communications Review 5.1 (2001): 3-55.

[10] Richard M. Wilson*Nonisomorphic Steiner Triple Systems*, Math. Z. 135 (1974), 303–313.

Chapter 2

High Dimensional Permutations

This chapter includes the following publication.

N. Linial and Z. Luria, *An upper bound on the number of high-dimensional permutations*, Combinatorica, to appear.

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 [2] proof of the Minc conjecture on permanents. More concretely, our approach is very close in spirit to Schrijver's [11] and Radhakrishnan's [10] proofs of Brègman's theorem.

2.1 Introduction

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

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

*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 .

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 [12], 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 by Egorychev [5, 4]. 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 2.1.1. *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 2.1.2. 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 (or d-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 [10]. 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 [14]. To summarize, the following is an equivalent definition of a d-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-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 [10] is

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

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 2.1.3.

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

In this paper we prove the upper bound

Theorem 2.1.4.

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

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

Definition 2.1.5. *1. An $[n]^{d+1}$ 0-1 array M_1 is said to support 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

$$Per_d(A) = \text{The number of } d\text{-permutations supported by } A.$$

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

Theorem 2.1.6. *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 $r_{i_1,\ldots i_d}$ ones in the line $A(i_1, ..., i_d, *)$. Then

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

We will derive below fairly tight bounds on the function f that appears in theorem 2.1.6. It is then an easy matter to prove theorem 2.1.4 by applying theorem 2.1.6 to the all-ones array.

What about proving a matching lower bound on $S_{d,n}$ (and thus proving conjecture 2.1.3)? In order to follow the footsteps of [10], 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 1-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-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-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 $j + \frac{n}{2}$. 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-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 2.1.6. In section 4 we use this bound to prove theorem 2.1.4.

2.2 Radhakrishnan's proof of Brègman's theorem

2.2.1 Entropy - Some basics

We review the basic material concerning entropy that is used here and refer the reader to [5] for further information on the topic.

Definition 2.2.1. *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 2.2.2. *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.

2.2.2 Radhakrishnan's proof

Let A be an $n \times n$ 0-1 matrix with r_i ones in the i-th 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 supported by 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 2.2.2) 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 1-entries in A's 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 supported by 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 2.2.2,

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

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

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

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

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

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

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 $\mathbb{S}_n$.

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

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.

2.3 The d-dimensional case

2.3.1 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 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 $\mathbb{E}_\sigma[\log N_i]$, where $N_i = r_i - u_i$ is the number of 1-entries in the i-th row that are still unshaded. The expectation of $\log N_i$ is exactly $\frac{1}{r_i} \sum_{j=1}^{r_i} \log j = \frac{1}{r_i} \log(r_i!)$.

How should we modify this argument to deal with d-dimensional permutations? We fix a d-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 lines are ordered by a random lexicographic ordering. At the coarsest level lines are ordered according to their first coordinate i_1. This ordering is chosen uniformly from $\mathbb{S}_n$. 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.

2.3.2 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 2.2.2 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 $(d+1)$-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, a member j of $R_{\mathbf{i}}$ belongs to $Z_{\mathbf{i}}^k$ 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 random lexicographic ordering. Namely, we select d random permutations $\sigma_1, ..., \sigma_d \in \mathbb{S}_n$. The line $A(i_1, ..., i_d, *)$ precedes $A(i'_1, ..., i'_d, *)$ if there is a $k \in [n]$ such that $\sigma_k(i_k) < \sigma_k(i'_k)$ and $i_j = i'_j$ for all $j < k$. Thus a choice of the orderings σ_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 **i** comes before **j**. We write $\mathbf{i} \prec_k \mathbf{j}$ if $\mathbf{i} \prec \mathbf{j}$ and **i** and **j** differ *only* in the k-th coordinate.

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

We turn to the definition of $R_{\mathbf{i}}$, $Z_{\mathbf{i}}^k$ 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-permutation that is contained in A.

Definition 2.3.1. *The index set of the* 1*-entries in the line* $A(i_1, ..., i_d, *)$ *is denoted by*

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

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* $X_{\boldsymbol{i}} \in W$*. For a given ordering* $\prec$*, let*

$$Z_{\boldsymbol{i}}^k(X, \prec) = \{j \in R_{\boldsymbol{i}} : X_{\boldsymbol{i'}} = j \text{ for some } \boldsymbol{i'} \prec_k \boldsymbol{i}\}.$$
$$N_{\boldsymbol{i}}(W, X, \prec) = |W \setminus \cup_{k=1}^d Z_{\boldsymbol{i}}^k(X, \prec)|.$$

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 $X_{\mathbf{i}}$ must 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 $k \le d$ then clearly $X_{\mathbf{i}} \ne j$, 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 $Z_{\mathbf{i}}^k(X, \prec)$ and is restricted to the set $R_{\mathbf{i}} \setminus \cup_{k=1}^d Z_{\mathbf{i}}^k(X, \prec)$. 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 2.2.2,

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

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 $\prec$.

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

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

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 i-th row was r_i. Then the rows preceding the i-th 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 2.3.2. *Let X be a d-permutation, $\boldsymbol{i} = (i_1, ..., i_d)$ and let $W \subseteq R_{\boldsymbol{i}}$ be an index set such that $X_{\boldsymbol{i}} \in W$. Then $\mathbb{E}_{\prec}\left[\log(N_{\boldsymbol{i}}(W, X, \prec))\right]$ depends only on d and $r = |W|$, and*

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

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

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

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_1, 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 $\sigma_2, \sigma_3, ..., \sigma_d$.

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,

$$\mathbb{E}_{\prec}\left[\log(N_{\mathbf{i}}(W, X, \prec))\right] = \mathbb{E}_{\sigma_1}\left[\mathbb{E}_{\prec'}\left[\log(N_{\mathbf{i}}(W, X, \prec))\right]\right]$$

$$= \mathbb{E}_{\sigma_1}\left[\mathbb{E}_{\prec'}\left[\log(N_{\mathbf{i'}}(W', X', \prec'))\right]\right] = \mathbb{E}_{\sigma_1}\left[f(d-1, |W'|)\right]$$

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_1. 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_1 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 $r_{\mathbf{i}} = r_{i_1,...,i_d}$ is the number of ones in the vector $A(i_1, ..., i_d, *)$. Therefore,

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

2.4 The number of d-permutations – An upper bound

As mentioned, the upper bound on the number of d-dimensional permutations is derived by applying theorem 2.1.6 to the all-ones array J. The main technical step is a derivation of an upper bound on the function $f(d, r)$.

Theorem 2.4.1. *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.

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

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

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

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

where $\xi = r_{d-1}\log(r_{d-1}) + r_{d-1}(d-1) = 2(d-1)e^{d-1}$. 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,

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

Putting this together, we have the inequality

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

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

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

i.e.,

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

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

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

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

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 $c_1 = 2 + e \leq 5$, $c_2 \leq 8$, and $c_d \leq \frac{d^3(1.1)^d}{d!}$ for $3 \leq d \leq 10$. 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 the fact that for $d \geq 10$, $\left(\frac{e}{d}\right)^d \leq \frac{d}{d!}$.

$\square$

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

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

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

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

Bibliography

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

[11] A. Schrijver, *A short proof of Minc's conjecture*, J. Comb. Theory Ser. A 25 (1978), 80-83.

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

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

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

Chapter 3

Steiner Triple Systems

This chapter includes the following publication.

N. Linial and Z. Luria, *An upper bound on the number of Steiner triple systems*, Random Structures and Algorithms 43.4 (2013): 399-406.

An upper bound on the number of Steiner triple systems

Nathan Linial Zur Luria

Abstract

Richard Wilson conjectured in 1974 the following asymptotic formula for the number of n-vertex Steiner triple systems:
$STS(n) = \left((1+o(1))\frac{n}{e^2}\right)^{\frac{n^2}{6}}$. Our main result is that

$$STS(n) \le \left((1+o(1))\frac{n}{e^2}\right)^{\frac{n^2}{6}}.$$

The proof is based on the entropy method.

As a prelude to this proof we consider the number $F(n)$ of 1-factorizations of the complete graph on n vertices. Using the Kahn-Lovász theorem it can be shown that

$$F(n) \le \left((1+o(1))\frac{n}{e^2}\right)^{\frac{n^2}{2}}.$$

We show how to derive this bound using the entropy method. Both bounds are conjectured to be sharp.

3.1 Introduction

A Steiner triple system on a vertex set V is a collection of triples $T \subseteq \binom{V}{3}$ such that each pair of vertices is contained in exactly one triple from T. It is well known that a Steiner triple system on $n \ge 1$ vertices exists if and only if $n \equiv 1$ or 3 (mod 6). We denote the number of Steiner triple systems on the vertex set $[n] := \{1, ..., n\}$ by $STS(n)$.

A 1-factorization of the complete graph on n vertices K_n is a partition of the edges of K_n into $n-1$ perfect matchings, or in other words, a proper edge coloring of K_n using

*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 .

$n-1$ colors. Let $F(n)$ denote the number of 1-factorizations of K_n. It is well known that a 1-factorization of K_n exists if and only if n is even.

The main results of this paper are a new upper bound on $STS(n)$ and a new proof of a known upper bound on $F(n)$.

It has been observed (e.g., [3]) that 1-factorizations and Steiner triple systems are special types of Latin squares. We view a Latin square as an $n \times n \times n$ array A with $0-1$ entries in which each *line* has exactly one element that equals 1. To see that this description of Latin squares is equivalent to the usual definition, we associate to the array A a matrix L, that is defined via $L(i,j) = k$ where k is the unique index for which $A(i,j,k) = 1$. A 1-factorization is a Latin square A such that $A(i,j,k) = 1 \Leftrightarrow A(j,i,k) = 1$ and $A(i,i,n) = 1$ for all i. Thus, L is a symmetric matrix in which all diagonal terms equal n. A Steiner triple system is a Latin square A where $A(i,j,k) = 1$ implies that $A(\sigma(i), \sigma(j), \sigma(k)) = 1$ for every permutation $\sigma \in S_3$ on i,j,k, and $A(i,i,i) = 1$ for all i. This can also be expressed in terms of L, though it's a bit more complicated to formulate.

These relations suggest that there might be deeper analogies to reveal among Latin squares, STS's and 1-factorizations. Indeed, we have recently proved an asymptotic upper bound on the number of Latin hypercubes [9], and here we prove analogous statements for $STS(n)$ and $F(n)$.

The best previously known estimates for the number of n-point Steiner triple systems are due to Richard Wilson [12].

$$\left(\frac{n}{e^2 3^{3/2}}\right)^{\frac{n^2}{6}} \leq STS(n) \leq \left(\frac{n}{e^{1/2}}\right)^{\frac{n^2}{6}}.$$

Wilson also conjectured that, in fact, $STS(n) = \left((1+o(1))\frac{n}{e^2}\right)^{\frac{n^2}{6}}$. We show that this is an upper bound on the number of Steiner triple systems.

Theorem 3.1.1.

$$STS(n) \leq \left((1+o(1))\frac{n}{e^2}\right)^{\frac{n^2}{6}}.$$

The Kahn-Lovász theorem shows that a graph with degree sequence $r_1, ..., r_n$ has at most $\prod_{i=1}^{n} (r_i!)^{\frac{1}{2r_i}}$ perfect matchings. In particular a d-regular graph has at most $(d!)^{\frac{n}{2d}}$ perfect matchings. For a proof see Alon and Friedland [1]. These results are inspired by Brégman's proof [2] of Minc's conjecture on the permanent. For a very recent proof of this result that uses the entropy method, see [6].

This theorem easily yields an upper bound on $F(n)$ as follows: Choose first a perfect matching of K_n. The remaining edges constitute an $n-2$ regular graph in which we again choose a perfect matching. We proceed to choose perfect matchings until we exhaust all of $E(K_n)$. The theorem implies that we have at most $((n-k)!)^{\frac{n}{2(n-k)}}$ choices for the k-th step, so that $F(n) \leq \prod_{d=1}^{n-1} (d!)^{\frac{n}{2d}}$. An application of Stirling's formula gives:

Theorem 3.1.2.

$$F(n) \leq \left((1+o(1))\frac{n}{e^2}\right)^{\frac{n^2}{2}}.$$

One of the results of the present paper is a new proof of this bound.

It is an interesting question to seek lower bounds to complement these upper bounds. We have already mentioned Wilson's lower bound on $STS(n)$. Cameron gave a lower bound for $F(n)$ in [4]. His argument yields

$$F(n) \geq \left((1+o(1))\frac{n}{4e^2}\right)^{\frac{n^2}{2}}.$$

For the sake of completeness we repeat his argument. It starts with the inequality $F(n) \geq L(\frac{n}{2})(F(n/2))^2$, where $L(n)$ is the number of order-n Latin squares. This inequality is shown as follows: Partition the vertex set $[n]$ into two equal parts, and select an arbitrary 1-factor on each. It is well-known and easy to prove that a 1-factorization of $K_{r,r}$ is equivalent to an order-r Latin square. It follows easily from the Van der Waerden conjecture that $L(n) \geq (\frac{(1+o(1))n}{e^2})^{n^2}$ (see [10]). The derivation of Cameron's lower bound is a simple matter now. We note that this argument works when n is divisible by 4. When $n = 4r + 2$ some additional care is required.

For the record, we complement Wilson's conjecture with a conjecture on the number of 1-factorizations:

Conjecture 3.1.3.

$$F(n) = \left((1+o(1))\frac{n}{e^2}\right)^{\frac{n^2}{2}}.$$

Our proofs are based on the entropy method, a useful tool for a variety of counting problems. The basic idea is this: In order to estimate the size of a finite set $\mathcal{F}$, we introduce a random variable X that is uniformly distributed on the elements of $\mathcal{F}$. Since $H(X) = \log(|\mathcal{F}|)$, bounds on $H(X)$ readily translate into bounds on $|\mathcal{F}|$. The bounds we derive on $H(X)$ are based on several elementary properties of the entropy function. Namely, if a random variable takes values in a finite set S then its entropy does not exceed $\log |S|$ with equality iff the distribution is uniform over S. Also, if X can be expressed as $X = (Y_1, \ldots, Y_k)$, then $H(X) = \sum_j H(Y_j|Y_1, \ldots, Y_{j-1})$. The expression $X = (Y_1, \ldots, Y_k)$ can be viewed as a way of gradually revealing the value of the random variable X. It is a key ingredient of our proofs to randomly select the order in which the variables Y_i are revealed and average over the resulting identities $H(X) = \sum_j H(Y_j|Y_i \text{ s.t. } i \text{ precedes } j)$. Similar ideas can be found in the literature, but to the best of our knowledge this method of proof is mostly due to Radhakrishnan [10]. We deviate somewhat from the standard notation in that our logarithms are always natural, rather than binary. Formally, we should use the notation H_e for the entropy function, but to simplify matters, we stick to the standard notation $H(X)$. We refer the reader to [5] for a thorough discussion of entropy. For an example of the entropy method, see [10].

In section 2, we give an entropy proof of theorem 3.1.2. Using similar methods, in section 3 we give an entropy proof of theorem 3.1.1.

3.2 An upper bound on 1-factorizations

Let n be an even integer, and let X be a random, uniformly chosen 1-factorization of K_n. Define the random variable $X_{\{i,j\}}$ to be the color of the edge $\{i, j\}$ in X. In order to analyze these random variables we first fix an ordering of the edges and we seek to bound the number of colors which are available for the edge $\{i,j\}$, given the colors of the preceding edges.

A color c is unavailable for $X_{\{i,j\}}$ if there is a previously seen variable of the form $X_{\{i,k\}}$ or $X_{\{k,j\}}$ which is equal to c. Let $N_{\{i,j\}}$ denote the number of available colors. It is really an upper bound on the number of values that $X_{\{i,j\}}$ can take given values for previously seen variables. Note that $N_{\{i,j\}}$ depends both on X and on the ordering.

We first apply the chain rule for the entropy function.

$$\log(F(n)) = H(X) = \sum_{\{i,j\}} H(X_{\{i,j\}}|X_{\{k,l\}} : \{k,l\} \text{ precedes } \{i,j\}) \tag{3.1}$$

$$= \sum_{\{i,j\}} \mathbb{E}_X[H(X_{\{i,j\}}|X_{\{k,l\}} = x_{\{k,l\}} : \{k,l\} \text{ precedes } \{i,j\})].$$

We now use the bound $H(X) \le \log(|Range(X)|)$ and conclude that

$$\log(F(n)) \le \sum_{\{i,j\}} \mathbb{E}_X[\log(N_{\{i,j\}})].$$

This bound holds for any ordering of the edges. We choose a random ordering by selecting a random mapping $\lambda : \binom{[n]}{2} \to [0,1]$. Edges are scanned according to the order of the real numbers $\lambda(\{i,j\})$ starting from the largest values. Of course we may assume that λ is $1:1$. This description of the ordering turns out to simplify matters in the discussion below.

We now take the expectation with respect to the random choice of the ordering, i.e., the choice of the mapping λ.

$$\log(F(n)) \le \mathbb{E}_\lambda[\sum_{\{i,j\}} \mathbb{E}_X[\log(N_{\{i,j\}})]] = \sum_{\{i,j\}} \mathbb{E}_X[\mathbb{E}_\lambda[\log(N_{\{i,j\}})]].$$

We bound the expectation $\mathbb{E}_\lambda[\log(N_{\{i,j\}})]$ using Jensen's inequality. If we do this right away, the resulting upper bound is not optimal. Therefore, we first condition on the value of $\lambda(\{i,j\})$ and only then use Jensen's inequality.

$$\mathbb{E}_\lambda[\log(N_{\{i,j\}})] = \mathbb{E}_{\lambda(\{i,j\})}[\mathbb{E}_\lambda[\log(N_{\{i,j\}})|\lambda(\{i,j\})]] \le$$

$$\mathbb{E}_{\lambda(\{i,j\})}[\log(\mathbb{E}_\lambda[N_{\{i,j\}}|\lambda(\{i,j\})])]$$

In order to evaluate this expression it is necessary to compute the expectation of $N_{\{i,j\}}$ given $\lambda(\{i,j\})$.

Lemma 3.2.1. $\mathbb{E}_\lambda[N_{\{i,j\}}|\lambda(\{i,j\})] = 1 + (n-2)\lambda(\{i,j\})^2$.

Proof. The true color of the edge $\{i,j\}$ in X is obviously always available to $X_{\{i,j\}}$. For each remaining color c, there is an edge of the form $\{i,a\}$ and an edge of the form $\{b,j\}$ that take the color c in X. If either of these edges λ-precedes $\{i,j\}$ then c is unavailable.

The edge $\{i,j\}$ precedes any edge of smaller λ value. Since these values are chosen independently, the probability that c is available is $\lambda(\{i,j\})^2$, and the result follows from the linearity of the expectation. $\square$

Using lemma 3.2.1, we have

$$\mathbb{E}_{\lambda(\{i,j\})}[\log(\mathbb{E}_\lambda[N_{\{i,j\}}|\lambda(\{i,j\})])] = \int_0^1 \log(1+(n-2)t^2)dt$$

$$= \log(n-1) - 2 + \frac{2\arctan(\sqrt{n-2})}{\sqrt{n-2}}$$

$$= \log(n) - 2 + O\left(\frac{1}{\sqrt{n}}\right).$$

Consequently,

$$\log(F(n)) \leq \sum_{\{i,j\}} \log(n) - 2 + O\left(\frac{1}{\sqrt{n}}\right)$$

$$= \binom{n}{2}\left(\log(n) - 2 + O\left(\frac{1}{\sqrt{n}}\right)\right)$$

which yields the bound

$$F(n) \leq \left(\left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right)\frac{n}{e^2}\right)^{\frac{n^2}{2}}.$$

3.3 An upper bound on the number of Steiner triple systems

The ideas here are similar to those in section 2, but the details are different.

Let X be a uniformly chosen random Steiner triple system on n vertices. Define $X_{\{i,j\}}$ to be the unique vertex k such that $\{i,j,k\}$ is a triple in X.

As above, for a given order on the pairs we define a random variable $N_{\{i,j\}}$.

Let $X_{\{i,j\}} = k$, and let $F_{\{i,j\}}$ denote the event that $\{i,j\}$ precedes both $\{j,k\}$ and $\{i,k\}$. If $F_{\{i,j\}}$ doesn't occur, set $N_{\{i,j\}} := 1$.

Let $t \in [n] \setminus \{i,j,k\}$ be a vertex. Since $\{i,j,t\} \notin X$, there are vertices a and b such that $\{i,a,t\}, \{j,b,t\} \in X$. We say that the vertex t is unavailable for $X_{\{i,j\}}$ if any of the six pairs in these triples precede $\{i,j\}$. If $F_{\{i,j\}}$ does occur, define $N_{\{i,j\}}$ to be the number of available vertices.

Now, If $F_{\{i,j\}}$ doesn't occur, then $X_{\{i,j\}}$ is uniquely determined by the preceding variables. Otherwise, the unavailable vertices are ruled out as possible values for $X_{\{i,j\}}$. If, for instance, $\{a,t\}$ is revealed before $\{i,j\}$, then by the time that $X_{\{i,j\}}$ is revealed to us we already know that $\{i,a,t\} \in X$, and therefore $\{i,j,t\} \notin X$ and $X_{\{i,j\}} \neq t$.

Thus, $N_{\{i,j\}}$ is an upper bound on the number of vertices that are available for $X_{\{i,j\}}$, given the values of the preceding variables.

For a given ordering of the pairs, as in Equation (3.1) we derive:

$$\log(STS(n)) = H(X) \leq \sum_{\{i,j\}} \mathbb{E}_X[\log(N_{\{i,j\}})].$$

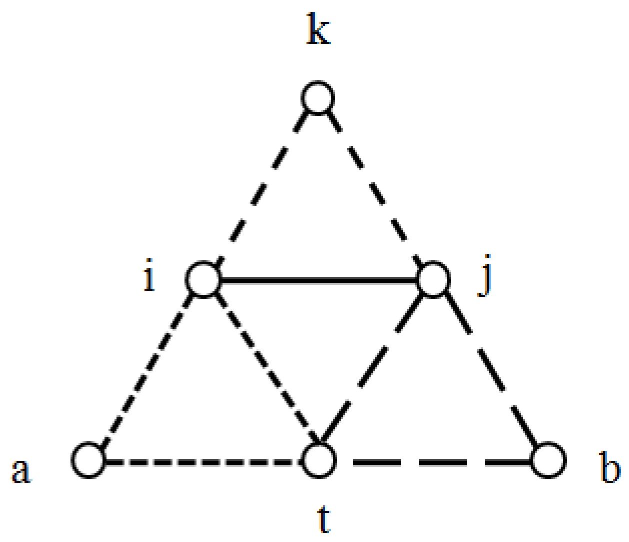

Figure 3.1: If either of the triangles $\{t, i, a\}$ or $\{t, j, b\}$ are revealed before $X_{\{i,j\}}$, then t is unavailable.

As before, we choose a random ordering by selecting a random mapping $\lambda : \binom{[n]}{2} \to [0, 1]$. Pairs are considered by decreasing order of their λ values. We take the expectation over the choice of λ to obtain

$$\log(STS(n)) \le \sum_{\{i,j\}} \mathbb{E}_X[\mathbb{E}_\lambda[\log(N_{\{i,j\}})]].$$

Let us fix X and an edge $\{i, j\}$ and turn to bound $\mathbb{E}_\lambda[\log(N_{\{i,j\}})]$. The next step is to condition over $\lambda(\{i, j\})$.

$$\mathbb{E}_\lambda[\log(N_{\{i,j\}})] = \mathbb{E}_{\lambda(\{i,j\})}[\mathbb{E}_\lambda[\log(N_{\{i,j\}})|\lambda(\{i, j\})]].$$

The event $F_{\{i,j\}}$ occurs iff $\lambda(\{i, j\}) > \lambda(\{i, k\})$ and $\lambda(\{i, j\}) > \lambda(\{k, j\})$ so that $\Pr(F_{\{i,j\}}|\lambda(\{i, j\})) = \lambda(\{i, j\})^2$. Therefore

$$\mathbb{E}_\lambda[\log(N_{\{i,j\}})|\lambda(\{i, j\})] = \lambda(\{i, j\})^2 \mathbb{E}_\lambda[\log(N_{\{i,j\}})|\lambda(\{i, j\}), F_{\{i,j\}}]$$

$$\le \lambda(\{i, j\})^2 \log(\mathbb{E}_\lambda[N_{\{i,j\}}|\lambda(\{i, j\}), F_{\{i,j\}}]), \tag{3.2}$$

where the final inequality follows from Jensen's inequality.

Lemma 3.3.1. $\mathbb{E}_\lambda[N_{\{i,j\}}|\lambda(\{i, j\}), F_{\{i,j\}}] = 1 + (n - 3)\lambda(\{i, j\})^6$.

Proof. The vertex k that participates in a triple with i, j is obviously always available to $X_{\{i,j\}}$. As mentioned, for each remaining vertex t, there are six pairs that $\{i, j\}$ must λ-precede for t to be available, and this occurs with probability $\lambda(\{i, j\})^6$. The result follows from the linearity of the expectation. □

Using 3.2 and lemma 3.3.1, we have

$$\mathbb{E}_{\lambda(\{i,j\})}[\log(\mathbb{E}_\lambda[N_{\{i,j\}}|\lambda(\{i,j\})])] = \int_0^1 t^2 \log(1+(n-3)t^6)dt$$

$$= \frac{1}{3}\left((\log((n-3)x^6+1) - 2x^3) + \frac{2\arctan(\sqrt{n-3}x^3)}{\sqrt{n-3}}\right)\Big|_0^1$$

$$= \frac{1}{3}\left((\log(n-2)-2) + \frac{2\arctan(\sqrt{n-3})}{\sqrt{n-3}}\right)$$

$$= \frac{1}{3}\left(\log(n) - 2 + O\left(\frac{1}{\sqrt{n}}\right)\right).$$

Consequently,

$$\log(STS(n)) \leq \sum_{\{i,j\}} \frac{1}{3}\left(\log(n) - 2 + O\left(\frac{1}{\sqrt{n}}\right)\right)$$

$$= \frac{n^2}{6}\left(\log(n) - 2 + O\left(\frac{1}{\sqrt{n}}\right)\right)$$

which yields the bound

$$STS(n) \leq \left(\left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right)\frac{n}{e^2}\right)^{\frac{n^2}{6}}.$$

Acknowledgement

We are grateful to the anonymous referees for useful remarks. In particular, we acknowledge comments made by one of the referees that helped us to shorten and simplify our proofs.

Bibliography

[1] Alon, N. and Friedland, S., *The maximum number of perfect matchings in graphs with a given degree sequence*, the electronic journal of combinatorics, 2008 .

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

[3] P. J. Cameron, *A generalization of t-designs*, Discrete Math., Discrete Math. 309 (2009), 4835–4842.

[4] P. J. Cameron, *Parallelisms of Complete Designs*, London Math. SOC. Lecture Note Ser. 23. Cambridge Univ. Press, Cambridge (1976) 144 pp. MR 54#7269.

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

[6] Cutler, J. and Radcliffe, AJ, *An entropy proof of the Kahn-Lovász theorem*, the electronic journal of combinatorics, 2011 .

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

[8] 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.

[9] N. Linial and Z. Luria, *An upper bound on the number of higher dimensional permutations*, http://arxiv.org/abs/1106.0649.

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

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

[12] Richard M. Wilson*Nonisomorphic Steiner Triple Systems*, Math. Z. 135 (1974), 303–313.

Chapter 4

A high dimensional Birkhoff Polytope

This chapter includes the following publication.

N. Linial and Z. Luria, *On the vertices of the d-dimensional Birkhoff polytope*, Discrete and Computational Geometry 51.1 (2014): 161-170.

On the vertices of the d-dimensional Birkhoff polytope

Nathan Linial Zur Luria

Abstract

Let us denote by Ω_n the *Birkhoff polytope* of $n \times n$ doubly-stochastic matrices. As the Birkhoff-von Neumann theorem famously states, the vertex set of Ω_n coincides with the set of all $n \times n$ permutation matrices. Here we consider a higher-dimensional analog of this basic fact. Let $\Omega_n^{(2)}$ be the polytope which consists of all *tristochastic* arrays of order n. These are $n \times n \times n$ arrays with nonnegative entries in which every *line* sums to 1. What can be said about $\Omega_n^{(2)}$'s vertex set? It is well-known that an order-n Latin square may be viewed as a tristochastic array where every line contains $n-1$ zeros and a single 1 entry. Indeed, every Latin square of order n is a vertex of $\Omega_n^{(2)}$, but as we show, such vertices constitute only a vanishingly small subset of $\Omega_n^{(2)}$'s vertex set. More concretely, we show that the number of vertices of $\Omega_n^{(2)}$ is at least $(L_n)^{\frac{3}{2}-o(1)}$, where L_n is the number of order-n Latin squares.

We also briefly consider similar problems concerning the polytope of $n \times n \times n$ arrays where the entries in every *coordinate hyperplane* sum to 1, improving a result from [8]. Several open questions are presented as well.

4.1 Introduction

Let $\Omega_n \subset \mathbb{R}^{n^2}$ be the Birkhoff polytope, namely the set of order-n doubly stochastic matrices. The defining equations and inequalities of Ω_n are

$$\sum_{i=1}^{n} x_{i,j} = 1 \text{ for all } 1 \leq j \leq n,$$

$$\sum_{j=1}^{n} x_{i,j} = 1 \text{ for all } 1 \leq i \leq n,$$

and

$$x_{i,j} \geq 0 \text{ for all } 1 \leq i,j \leq n.$$

The vertex set of Ω_n is determined by the Birkhoff-von Neumann theorem [1, 11].

*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 .

Theorem 4.1.1. *The vertex set of Ω_n coincides with the set of permutation matrices of order n.*

We consider here some higher-dimensional analogs of the polytope Ω_n and ask about their vertex sets in light of Theorem 4.1.1.

A *line* in an $n \times n \times n$ array A is the set of entries obtained by fixing two indices and letting the third vary from 1 to n. A line of the form $A(\cdot, j, k)$ is called a column, a line of the form $A(i, \cdot, k)$ is a row and a line of the form $A(i, j, \cdot)$ is a shaft. A *coordinate hyperplane* in A is the $n \times n$ matrix obtained by fixing one index and letting the other two vary. Such a hyperplane of the form $A(\cdot, \cdot, k)$ is called a *layer* of A. We denote the k-th layer of A by A_k. We denote the support of an array A by $\mathrm{supp}(A)$.

Let $\Omega_n^{(2)}$ be the polytope of all *tristochastic* arrays of order n. Namely, $n \times n \times n$ arrays with nonnegative entries in which every *line* sums to 1. Latin squares of order n can be viewed as two-dimensional permutations and it is easily verified that every Latin square of order n is a vertex of $\Omega_n^{(2)}$. Does the natural analog of Theorem 4.1.1 hold true? As we show (Theorem 4.1.5), this is far from the truth. Of the $v = v_n$ vertices of $\Omega_n^{(2)}$ only fewer than $v^{2/3+o(1)}$ correspond to Latin squares.

In Section 4.3 we establish a similar phenomenon for a related polytope. Namely, now we consider $n \times n \times n$ arrays of nonnegative reals in which every coordinate hyperplane sums to 1. It is shown in [8] that a natural, combinatorially defined set of vertices, comprise no more than $v^{\frac{2}{3}+o(1)}$ of the v vertices of this polytope. We improve this bound to $v^{\frac{1}{2}+o(1)}$.

The polytopes that we consider here have been studied in various contexts. For an extensive coverage, see [12],[3]. These earlier studies were motivated mostly by interest in optimization problems. The (original) Birkhoff polytope plays an important role in assignment problems. Likewise, its higher dimensional analogs are of interest in the study of transportation polytopes, multi-index assignment problems and other classical optimization problems.

4.1.1 Background material

A Latin square L of order n is an $n \times n$ matrix with entries from $[n] := \{1, ..., n\}$ such that each symbol appears exactly once in every row and column. Equivalently, it is an $n \times n \times n$ array A of zeros and ones in which every *line* has exactly one 1 entry. The correspondence between the two definitions is this: $A(i, j, k) = 1 \Leftrightarrow L(i, j) = k$. We denote the number of order-n Latin squares by L_n.

The permanent of an $n \times n$ matrix A is defined as

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

A lower bound on permanents of doubly stochastic matrices was conjectured by van der Waerden and proved by Falikman and by Egorychev [5, 4].

Theorem 4.1.2. *If A is an $n \times n$ doubly stochastic matrix, then*

$$Per(A) \geq \frac{n!}{n^n}.$$

An upper bound on the permanent of zero/one matrices was conjectured by Minc and proved by Brègman [2].

Theorem 4.1.3. *Let A be an $n \times n$ matrix of zeros and ones with r_i ones in the i-th row. Then*

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

The following argument of van Lint and Wilson [10] utilizes these two bounds to derive an estimate for L_n by constructing a Latin square A and bounding the number of ways to do this. Consider the $n \times n \times n$ zero-one array representation of a Latin square layer by layer. Each layer is a permutation matrix, so that there are $n!$ choices for the first layer. Having already specified $k-1$ layers, the number of choices for the k-th layer can be expressed as the permanent of B, a zero/one matrix where $b_{ij} = 1$ iff $a_{ijt} = 0$ for all $k > t$. Using the above upper and lower bounds on $per(B)$ it follows that

Theorem 4.1.4.

$$L_n = \left((1 + o(1))\frac{n}{e^2}\right)^{n^2}.$$

4.1.2 A higher dimensional Birkhoff polytope

Definitions

Let $\Omega_n^{(d)}$ be the set of $[n]^{d+1}$ nonnegative arrays such that the sum of each line is 1. Thus, $\Omega_n^{(1)} = \Omega_n$, the set of order-$n$ doubly stochastic matrices. Likewise, we call a member of $\Omega_n^{(d)}$ a $(d+1)$-stochastic array. Maintaining the analogy, we let $S_n^{(d)}$ be the set of $[n]^{d+1}$ arrays of zeros and ones with a single one in each line. In other words, $S_n^{(d)}$ consists of all $(d+1)$-stochastic arrays all of whose entries are zero or one. Thus, $S_n^{(1)}$ is the set of order n permutation matrices and $S_n^{(2)}$ coincides with the set of order-n Latin squares. Members of $S_n^{(d)}$ are called d-permutations.

In the literature, the set of all nonnegative d-dimensional arrays with line sums equal to 1, which we denote by $\Omega_n^{(d-1)}$, is called the *d-index planar assignment polytope*. In Section 4.3 we consider the polytope of all nonnegative d-dimensional arrays with hyperplane sums equal to 1, and denote it by $\Sigma_n^{(d-1)}$. In the literature this polytope is called the *d-index axial assignment polytope*. These polytopes are instances of *multi-way transportation polytopes*. See [9] for a survey of what is known about these more general objects.

We turn to investigate the vertex set of $\Omega_n^{(d)}$. It is easily verified that every member of $S_n^{(d)}$ is a vertex of $\Omega_n^{(d)}$. However, as we show here $\Omega_n^{(d)}$ can have numerous additional vertices.

4.1.3 A motivating example

It is instructive to consider the smallest such example. Namely, the following array A is a vertex of $\Omega_3^{(2)}$.

$$A_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} A_2 = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix} A_3 = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

To see that A is indeed a vertex, assume to the contrary that $A = \alpha B + (1-\alpha)C$ for some $0 < \alpha < 1$ and $B \neq C$ in $\Omega_3^{(2)}$. If $A(i,j,k)$ is 0 or 1, then necessarily $A(i,j,k) = B(i,j,k)$. So wherever $A(i,j,k) \neq B(i,j,k)$, there holds $A(i,j,k) = \frac{1}{2}$.

Consider the graph $G = G(A)$ whose vertices are the $\frac{1}{2}$ entries of A, where two vertices are adjacent iff they are on the same line. Since A is tristochastic, it follows that $B(i,j,k)+B(i',j',k') = 1$ for every two neighbors (i,j,k) and (i',j',k') in G. Specifically, if $B(i,j,k) = \frac{1}{2}+\epsilon$, then $B(i',j',k') = \frac{1}{2}-\epsilon$. Consequently, the connected component of G which contains the vertices (i,j,k) and (i',j',k') is bipartite. The color of a vertex is determined according to whether the B entry is $\frac{1}{2} \pm \epsilon$. However, it is easy to verify that G is connected and not bipartite, which proves our claim.

4.1.4 A scheme for constructing vertices

The above example suggests a construction for vertices of $\Omega_n^{(2)}$. Let A be an order-n tristochastic array whose support consists of exactly two $\frac{1}{2}$ entries in each line. The graph $G = G(A)$ defined as above is 3-regular and has $2n^2$ vertices. As we now show, A is a vertex of $\Omega_n^{(2)}$ iff no connected component of G is bipartite.

Indeed, suppose that G has a bipartite connected component with parts P and Q. Let Δ be the $[n]^3$ array with ± 1 entries at the elements of P, Q respectively and 0 everywhere else. Note that every line of Δ sums to zero. To see that A is not a vertex, note that $A = \frac{X+Y}{2}$, where $X, Y = A \pm \frac{1}{2}\Delta$ are clearly tristochastic.

Conversely, suppose that $A = \alpha B + (1-\alpha)C$ with $1 > \alpha > 0$ and $B \neq C$ in $\Omega_n^{(2)}$ is not a vertex. The same consideration that worked for the above example shows that the relevant component of G is bipartite.

This discussion suggests that we construct A so that no connected component of $G(A)$ is bipartite. This shouldn't be too hard, since G is 3-regular. Indeed, we suspect (but we still cannot show) that with high probability a randomly chosen tristochastic array with two $\frac{1}{2}$'s in each line is a vertex. This idea still yields the following lower bound on the number of vertices of $\Omega_n^{(2)}$.

Theorem 4.1.5. *The polytope $\Omega_n^{(2)}$ has at least $L_n^{\frac{3}{2}-o(1)}$ vertices.*

4.2 Proof of theorem 4.1.5

4.2.1 The construction

Let $I = (i_1, i_2, \ldots, i_n)$ and $J = (j_1, j_2, \ldots, j_n)$ be two permutations of $[n]$. We let

$$H(I, J) := \{(i_1, j_1), (i_2, j_1), (i_2, j_2), (i_3, j_2), \ldots, (i_n, j_n), (i_1, j_n)\},$$

and call such a collection of index pairs an *H-cycle*. If the elements of an H-cycle are interpreted as the indices of entries in an $n \times n$ matrix, there are exactly two such entries in every row and column. Note that $H(I, J) = H(I', J')$ where $I' = (i_2, i_3, \ldots, i_n, i_1)$ and $J' = (j_2, j_3, \ldots, j_n, j_1)$. Likewise, if we "reverse" the order of the i_ν and the j_ν by replacing I with $I' = (i_1, i_n, i_{n-1}, ..., i_2)$ and J with $J' = (j_n, n_{n-1}, ...j_1)$, $H(I, J)$ remains unchanged. Consequently, the number of H-cycles is $\frac{1}{2}n!(n-1)!$.

Definition 4.2.1. *Let n be an even integer. An order n double Latin square is an $n \times n$ matrix with entries from $\{1, ..., \frac{n}{2}\}$ where each symbol appears exactly* ***twice*** *in each row and column.*

We say that a double Latin square X is *Hamiltonian* if the indices of the k-entries of X constitute an H-cycle for every $k \in \{1, ..., \frac{n}{2}\}$. (This explains the choice of the term H-cycle).

Let A be a $t \times t$ matrix and let $\sigma \in S_t$ be a permutation. We denote by $\sigma(A)$ the matrix obtained from A by applying σ to its rows. We need the following result from [7]:

Proposition 4.2.2. *Let A, B be two order $\frac{n}{2}$ Latin squares and let $\sigma \in \mathbb{S}_{\frac{n}{2}}$ be a cyclic permutation. Then the block matrix*

$$X = \left(\begin{array}{c|c} A & B \\ \hline \sigma(A) & B \end{array}\right).$$

is an order n Hamiltonian double Latin square.

It follows that the number of Hamiltonian order-n double Latin squares is at least $(\frac{n}{2} - 1)! \cdot L_{\frac{n}{2}}^2 = ((1 + o(1))\frac{n}{2e^2})^{\frac{n^2}{2}}$.

We want to construct a tristochastic array A with exactly two $\frac{1}{2}$'s in each line, in such a way that $G(A)$ is non-bipartite and connected (and therefore A is a vertex).

The idea is to use a Hamiltonian double Latin square X to define the top $\frac{n}{2}$ layers of A. We use the fact that X is Hamiltonian to complete A in such a way that $G(A)$ is connected, and then "plant" an odd cycle in $G(A)$ to ensure that $G(A)$ isn't bipartite.

Given a Hamiltonian double Latin square X, we use it as the "topographical map" of the top $\frac{n}{2}$ layers of A. Namely, $A(i, j, k) = \frac{1}{2} \Leftrightarrow X(i, j) = k$. Let us observe the subgraph of $G(A)$ spanned by the entries of A that reside in these top layers. Every positive entry $A(i, j, k) = \frac{1}{2}$ comes from $X(i, j) = k$, and X has exactly two k entries in each line. Therefore this subgraph of $G(A)$ is 2-regular. Moreover, since X is Hamiltonian, for every $1 \leq k \leq \frac{n}{2}$ the vertices of $G(A)$ that correspond to $\mathrm{supp}(A_k)$ constitute a cycle of length $2n$. In other words, the subgraph of $G(A)$ corresponding to the entries of the top half of A is the disjoint union of $\frac{n}{2}$ cycles of length $2n$.

At this point, there are two $\frac{1}{2}$ entries in every line that resides in one of the top $\frac{n}{2}$ layers of A, and a single $\frac{1}{2}$ entry in every shaft.

We turn to define the next layer, $A_{\frac{n}{2}+1}$. Our purpose is to choose the $\frac{1}{2}$ entries in this layer so as to form a single cycle of length $2n$. The vertices of this subgraph should also be connected to each of the cycles in the top $\frac{n}{2}$ layers. Clearly, if we manage to accomplish this task, then the part of $G(A)$ that is already revealed is connected. Furthermore, note that every shaft contains a positive entry in the top half of A. Therefore, $G(A)$ will remain connected regardless of our choices in the lower layers of A.

In order to achieve our goals concerning $A_{\frac{n}{2}+1}$, we want to find $\frac{n}{2}$ index pairs $(i_1, j_1), ..., (i_{\frac{n}{2}}, j_{\frac{n}{2}})$ such that $X(i_l, j_l) = l$ for all $1 \leq l \leq \frac{n}{2}$ and no two of them share a row or a column. We find such pairs successively as follows: Suppose that, for some $k < \frac{n}{2}$, we already have k pairs $(i_1, j_1), ..., (i_k, j_k)$ with $X(i_1, j_1) = 1, ..., X(i_k, j_k) = k$ and no two pairs share a row or column. We claim that there is an additional pair (i_{k+1}, j_{k+1}) that does not share a row or column with any of the above index pairs, and $X(i_{k+1}, j_{k+1}) = k + 1$. Since X is a double Latin square, every row and column of X has exactly two elements that equal $k+1$. Therefore at most $4k$ of these entries share a row or column with a previous pair. But $2n > 4k$, so that such an index pair (i_{k+1}, j_{k+1}) must exist.

We choose $\frac{n}{2}$ more pairs of indices $(i_{\frac{n}{2}+1}, j_{\frac{n}{2}+1}), ..., (i_n, j_n)$ in such a way that no two pairs of $(i_1, j_1), ..., (i_n, j_n)$ share a row or a column.

It is possible to rename, if necessary, the set of chosen pairs $\{(i_\alpha, j_\alpha)|\alpha = 1, \ldots, n\}$ as $\{(\nu, \tau_\nu)|\nu = 1, \ldots, n\}$ for some permutation $\tau \in S_n$. Let P be the permutation matrix of τ. We next select a permutation $\sigma \in S_n$ whose permutation matrix P' is such that $P + P'$ consists of a single cycle. (We note that given τ, there are exactly $(n-1)!$ possible choices for σ). We achieve our aim by setting $A_{\frac{n}{2}+1} := \frac{1}{2}(P + P')$.

The purpose of our choices for $A_{\frac{n}{2}+2}$ is to introduce an odd cycle into $G(A)$. This odd cycle must use elements from the top half of A. Additionally, the indices of the $\frac{1}{2}$ entries in $A_{\frac{n}{2}+2}$ must avoid all index pairs used in $A_{\frac{n}{2}+1}$, so as not to create a shaft with three $\frac{1}{2}$ entries.

To this end, we seek two vertices $x = (x_1, x_2, k)$ and $y = (y_1, y_2, k)$ with $x_1 \neq y_1$ and $x_2 \neq y_2$ that are connected by a path of odd length in the part of $G(A)$ constructed so far. The construction of $A_{\frac{n}{2}+2}$ will yield a length four path between x and y, ensuring that $G(A)$ is not bipartite. This path will have the form x, x', w, y', y where $x' = (x_1, x_2, \frac{n}{2} + 2), y' = (y_1, y_2, \frac{n}{2} + 2)$ and w is either $(x_1, y_2, \frac{n}{2} + 2)$ or $(y_1, x_2, \frac{n}{2} + 2)$.

A simple counting argument shows the feasibility of this construction. Two vertices from the same layer can serve as x and y if their distance in that layer is odd and ≥ 3. There are $\Omega(n^2)$ such pairs in every layer with a total of $\Omega(n^3)$ such candidate pairs. On the other hand, as we show below, only $O(n^2)$ such pairs are ruled out, so at least for large n a good choice of such x, y must exist.

The reason that an entry cannot play the role of x is that its shaft meets $\text{supp}(A_{\frac{n}{2}+1})$. There are $O(n)$ vertices in x's layer which might serve as y, and $\text{supp}(A_{\frac{n}{2}+1})$ has cardinality $2n$, so only $O(n^2)$ pairs x, y get ruled out for this reason. It remains to see how the pair $x = (x_1, x_2, k)$ and $y = (y_1, y_2, k)$ can be disqualified when both x's and y's shaft do not meet $\text{supp}(A_{\frac{n}{2}+1})$. This can happen only if both $(x_1, y_2, \frac{n}{2} + 2)$ and $(y_1, x_2, \frac{n}{2} + 2)$ are unavailable to us, namely $A(x_1, y_2, \frac{n}{2} + 1) = A(y_1, x_2, \frac{n}{2} + 1) = \frac{1}{2}$. There are only $O(n^2)$

such instances, one per each pair of vertices in the $2n$-cycle residing in $A_{\frac{n}{2}+1}$.

By doing these computations carefully, one shows that already for $n \geq 10$ there must exist a good pair for the above argument.

Next we need to complete $\mathrm{supp}(A_{\frac{n}{2}+2})$. We are currently committed to three elements and $2n-3$ more $\frac{1}{2}$ entries need to be chosen, so that altogether there are exactly two in each row and column. The locations that must not be chosen are those in the "shadow" of $\mathrm{supp}(A_{\frac{n}{2}+1})$. It is easily seen that we need the following simple graph-theoretic claim.

Proposition 4.2.3. *Let $G = (L, R, E)$ be a $(n-2)$-regular bipartite graph with $|R| = |L| = n \geq 6$ and let M be a path of length 3 in G. Then there is a 2-factor in G which contains the three edges of M.*

Proof. Let $M = x_1, x_2, x_3, x_4$. A bipartite graph with sides of size k and degrees $\geq k/2$ has a perfect matching. Let Φ be a perfect matching in $G \setminus \{x_1, x_2, x_3, x_4\}$. Next let Ψ be a perfect matching in $G \setminus \{x_2, x_3\} \setminus \Phi$. The desired 2-factor is $\Phi \cup \Psi \cup \{(x_1, x_2), (x_2, x_3), (x_3, x_4)\}$. □

To recap, the graph $G(A)$ is connected, it contains an odd cycle, and these properties are retained regardless of how the remaining $\frac{n}{2} - 2$ layers are completed.

The remaining layers are constructed as follows. Let K be an $n \times n$ matrix where $K(i, j) = 1$ or 0 according to whether the shaft $A(i, j, \cdot)$ has one or two $\frac{1}{2}$ entries. Each row and column of K has $n - 4$ one-entries. In other words, K is the adjacency matrix of an $(n-4)$-regular bipartite graph which, therefore, has a 2-factor. This process can be completed layer by layer. This is just an existential argument and we next turn to estimate the number of ways in which our construction can be realized.

To this end we will multiply the number of ways to construct the top half and the appropriate number for the bottom half. As stated above, there are $L_n^{\frac{1}{2}+o(1)}$ ways to construct the top half. The estimate for the bottom $\frac{n}{2} - 2$ layers is a slight variation on van Lint and Wilson's [10] approximate enumeration of Latin squares. By the van der Waerden bound [5, 4], a k-regular (n, n) bipartite graph H has at least $\left((1 + o(1)\frac{k}{e}\right)^n$ perfect matchings. By the same argument, there are at least $\left((1 + o(1)\frac{k-1}{e}\right)^n$ ways to complete a perfect matching in H to a 2-factor. The product of these two numbers is an overcount, since every cycle in the 2-factor can be split in two ways between the first and second 1-factors. Consequently, H has at least

$$\left((1 + o(1))\frac{k(k-1)}{e^2\sqrt{2}}\right)^n$$

2-factors.

We think of K as the adjacency matrix of such an H, and each layer is just a 2-factor supported by K. With each choice, the edges of the chosen 2-factor are removed from H, which goes from being d-regular to $(d-2)$-regular. This yields the following lower bound on the number of choices:

$$\prod_{2 \leq k \leq n-4,\ k \text{ is even}} \left((1 + o(1))\frac{k(k-1)}{e^2\sqrt{2}}\right)^n =$$

$$(n-4)!^n \cdot \left(\frac{1+o(1)}{e^2\sqrt{2}}\right)^{n(n-4)/2} = \left((1+o(1))\frac{n}{2^{\frac{1}{4}}e^2}\right)^{n^2} = L_n^{1-o(1)}.$$

The product of the bound for the top half and the bound for the bottom half yields a total of $L_n^{\frac{3}{2}-o(1)}$.

4.3 A variation on the theme

Here is another natural extension of the notion of doubly stochastic matrices. Namely, let $\Sigma_n^{(d)}$ be the set of all $[n]^{d+1}$ arrays of nonnegative reals such that the entries in each coordinate hyperplane sum to one. The collection of such arrays clearly constitutes a convex polytope. Our goal is to investigate the vertex set of this polytope.

The vertices of $\Sigma_n^{(d)}$ have been studied before ([8], [6]). It is a well known fact that these polytopes have some noninteger vertices. V. M. Kravtsov [8] has considered the problem of enumerating the vertices, and gave a lower bound of $(n!)^{3+o(1)}$. Our construction improves this result and yields a lower bound of $(n!)^{4+o(1)}$.

The arithmetic properties of vertices of $\Sigma_n^{(d)}$ have also been studied. Kravtsov wrote a series of papers discussing constructions of vertices, and in particular showed that the denominator of the fractional elements of a vertex can grow exponentially with n. Gromova [6] gave a characterization of the sets of numbers that can appear as entries in a vertex.

Let us define $T_n^{(d)}$ as the collection of all $[n]^{d+1}$ arrays of zeros and ones with a single one in each coordinate hyperplane. It is clear that $T_n^{(d)}$ is included in the vertex set of $\Sigma_n^{(d)}$. There is a natural bijection between tuples $(\sigma_1, ..., \sigma_d) \in \mathbb{S}_n^d$ and members $A \in T_n^{(d)}$ which is given by $A(i, \sigma_1(i), ..., \sigma_d(i)) = 1$ for all $1 \le i \le n$. In particular $|T_n^{(d)}| = (n!)^d$.

As it happens, noninteger vertices are easy to construct. Here is the smallest example:

$$A_1 = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}, A_2 = \begin{bmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{bmatrix}$$

Clearly $A \in \Sigma_2^{(2)}$. We now consider the graph $\bar{G}(A)$ with vertex set $\mathrm{supp}(A)$ with an edge between every two vertices that lie in the same coordinate hyperplane. As in Section 4.1.2, we show that A is a vertex by observing that $\bar{G}(A)$ has no bipartite connected component. In the present case, $\bar{G} = K_4$.

Our general construction is similar in nature to this example. We first construct an $n \times n$ matrix M with entries from $[n]$ in which every row and column contains exactly two nonzero entries and where each integer in $[n]$ appears exactly twice in M. We view M as a way to encode A as follows: $M(i,j) = k$ for some $k \neq 0$ says that $A(i,j,k) = 1/2$ and $A(i,j,k') = 0$ for all $k' \neq k$. Also $M(i,j) = 0$ means that $A(i,j,l) = 0$ for all l. It is not hard to verify that if the graph $\bar{G}$ corresponding to M is connected and non-bipartite, then A is a vertex of $\Sigma_n^{(2)}$.

We now turn to construct many such matrices M and thus generate many vertices for $\Sigma_n^{(2)}$ which are not in $T_n^{(2)}$. Let

$$H = \{(i_1, j_1), (i_2, j_1), (i_2, j_2), \ldots, (i_n, j_n), (i_1, j_n)\}$$

be an H-cycle and let

$$M(i_1, j_1) = M(i_2, j_2) = 1 \text{ and } M(i_2, j_1) = 2.$$

The remaining entries of the H-cycle $M(i_\alpha, j_\alpha)$ and $M(i_{\alpha+1}, j_\alpha)$ are filled arbitrarily with the elements of the multiset $\{2, 3, 3, 4, 4, \ldots, n, n\}$. Note that the resulting graph $\bar{G}$ is connected, due to the fact that the first two indices of the $\frac{1}{2}$ elements of A form an H-cycle, and non-bipartite, since it contains the triangle $\{(i_1, j_1, 1), (i_2, j_2, 1), (i_2, j_1, 2)\}$. There are $\frac{1}{2}n!(n-1)!$ choices for H and and $\frac{(2n-3)!}{2^{n-2}}$ ways to map the multiset to the nonzero entries of M. Altogether, this construction yields more than $(n!)^4 = \left(T_n^{(2)}\right)^2$ vertices of $\Sigma_n^{(2)}$.

4.4 Conjectures and some experimental results

This paper raises many open questions. Here are several of them:

- Get a better estimate for the number of vertices of $\Omega_n^{(2)}$.
- The analogous question for $\Omega_n^{(d)}$ with $d > 2$ seems completely open at this writing.
- The polytope $\Omega_n^{(2)}$ is defined by requiring that one-dimensional subsets of the array sum to one. In the definition of $\Sigma_n^{(2)}$ this is required of two-dimensional subsets. For larger d there are a whole range of possible polytopes to consider, depending on which sets of entries sum to 1.

If we knew the support size of vertices in $\Omega_n^{(d)}$, we could make progress on these questions. By standard linear programming arguments, every vertex of $\Omega_n^{(d)}$ has at least aff-dim$(\Omega_n^{(d)})$ zero coordinates. Since aff-dim$(\Omega_n^{(d)}) = (n-1)^{d+1}$, every vertex has support size at most $n^{d+1} - (n-1)^{d+1} \leq (d+1) \cdot n^d$.

It follows that $\Omega_n^{(d)}$ has at most $\binom{n^{d+1}}{(d+1)n^d} \leq \left(\frac{ne}{d+1}\right)^{(d+1)n^d}$ vertices. In particular, $\Omega_n^{(2)}$ has fewer than n^{3n^2} vertices. If we knew, say, that a typical vertex of $\Omega_n^{(2)}$ has support size $\leq \alpha n^2$ vertices, we could conclude that it has at most $n^{(1+o(1))\alpha n^2}$ vertices.

We have conducted some numerical experiments to get a sense of the numbers. Using linear programming tools, it is possible to find the vertex that maximizes a randomly chosen linear objective function. Needless to say, this distribution on the vertices is by no means uniform. We nevertheless hope that our experiments do tell us something meaningful about the properties of typical vertices. We selected the coordinates in the objective function independently from normal distribution. The average value of α in these experiments seems to increase slowly with n. We don't know whether the typical support size of a vertex converges to $3n^2$ or to αn^2 for some $\alpha < 3$.

Bibliography

[1] G. Birkhoff, *Tres observaciones sobre el álgebra lineal*, Univ. nac. Tucumán Revista, Ser. A, 5:147-150, 1946.

[2] L. M. Brègman, *Certain properties of nonnegative matrices and their permanents*, Soviet Math. Dokl. 14 (1973), 945-949.

[3] R. Burkard, M. Dell'Amico and S. Martello, *Assignment Problems*, SIAM, Philadelphia, 2009.

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

[5] 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.

[6] M. B. Gromova, *The Birkhoff-von Neumann theorem for polystochastic matrices*, Selecta Math. Soviet. 11 (1992), 145-158.

[7] A. J. W. Hilton, M. Mays, C. St. J. Nash-Williams, C. A. Rodger, *Hamiltonian double latin squares*, J. Combin. Theory B 87 (2003), 81-129.

[8] V. M. Kravtsov, *Combinatorial properties of noninteger vertices of a polytope in a three-index axial assignment problem*, Cybernetics and Systems Analysis 43.1 (2007), 25-33.

[9] J. A. De Loera and E. D. Kim, *Combinatorics and Geometry of Transportation Polytopes: An Update*, Preprint, http://arxiv.org/abs/1307.0124.

[10] J. H. van Lint and R. M. Wilson, *A Course in Combinatorics*, Cambridge Univ. Press, 1992.

[11] J. von Neumann, *A certain zero-sum two-person game equivalent to the optimal assignment problem* Contributions to the Theory of Games, 2:5-12, Princeton, 1953.

[12] V. A. Yemelichev, M. M. Kovalov and M. K. Kravtsov, *Polytopes, Graphs and Optimization*, Cambridge Univ. Press, 1984.

Chapter 5

Discussion and Open Questions

The research presented in this thesis is part of a large scale effort in the combinatorics world to understand the properties of high dimensional objects. In many cases these objects exhibit properties that are richer and more complicated than those of their lower dimensional counterparts. Here are some examples of this from our research.

- For $d = 2$ van der Waerden's theorem gives us a simple way to construct 2-permutations (or Latin squares) and also to obtain a lower bound on their number. However, the natural extension of van der Waerden's conjecture does not hold true for higher dimensions. There are d-stochastic arrays whose d-dimensional permanent is zero.

- There are simple algorithms for constructing perfect matchings and sampling them at random. However, for higher dimensions it seems that these problems are extremely difficult.

- As Birkhoff's theorem states, the vertices of the polytope of doubly stochastic matrices are the permutation matrices. As we have shown, the vertices of the polytope of tristochastic arrays are much more varied. We do not yet have a good description for them, and we only have weak upper and lower bounds on their number.

There are many basic tantalizing questions about high dimensional objects that we do not know the answer to. For one thing, we would very much like to find matching lower bounds for high dimensional permutations and Steiner systems. One approach that seems promising is to find a lower bound on the d-Permanent. We know that the natural analog of van der Waerden's conjecture fails in higher dimensions, but perhaps there is a bound with a different flavor that could be of use.

A Steiner system with parameters (n, t, k) is an n element set S together with a collection X of k-element subsets of S with the property that each t-element subset of S is contained in a unique set $x \in X$. Using our methods, we can prove an upper bound on the number of order n Steiner systems with $k = t + 1$. However, finding a lower bound on Steiner systems in general seemed until recently to be all but impossible, since the question of the existence of Steiner systems for $t \geq 6$ had been open for almost 200 years. This was one of the central open problems in combinatorics, until Peter Keevash solved it earlier this year [2], showing that there exist infinitely many Steiner systems for any parameters t, k such that $1 \leq t < k$. It is our hope that using his methods it may be possible to get better lower bounds on high dimensional permutations and Steiner systems.

Another problem that we would like to make progress on is understanding random high dimensional objects. Is there an efficient algorithm to sample a high dimensional permutation uniformly at random? Matthews and Jacobson proposed a Random walk over the set of order-n Latin squares and showed that it is connected [1]. If this random walk could be shown to be rapidly mixing, it would yield an efficient algorithm for sampling Latin squares uniformly at random. This walk can easily be adapted to Steiner triple systems, but it is not known whether this yields a connected Markov chain.

Random regular graphs have proved immensely important both for theoretical and practical applications. What are the properties of random regular hypergraphs? For example, it is well known that a random regular graph is with high probability a good expander. Recently, much work has been done on generalizing the concept of expansion in graphs to higher dimensions (see for example [3]). Is it true that a random regular hypergraph is with high probability a good high dimensional expander? This seems to be a good topic for further research.

Bibliography

[1] M. T. Jacobson and P. Matthews, *Generating uniformly distributed random Latin squares*, Journal of Combinatorial Designs 4.6 (1996): 405-437.

[2] P. Keevash, *The existence of designs*,arXiv preprint arXiv:1401.3665 (2014).

[3] A. Lubotzky, *Ramanujan complexes and high dimensional expanders*, arXiv preprint arXiv:1301.1028 (2013).