

On Entropy-Preserving Stochastic Averages

Alan B. Poritz

*Center for Communications Research – Princeton
805 Bunn Drive
Princeton, New Jersey 08540
USA*

Jonathan A. Poritz*

*Department of Mathematics and Physics
Colorado State University, Pueblo
2200 Bonforte Blvd.
Pueblo, CO 81001
USA*

Abstract

When an $n \times n$ doubly stochastic matrix A acts on the left as a linear transformation on an n -long probability vector P , we refer to the new probability vector AP as the *stochastic average* of the pair (A, P) . Let Γ_n denote the set of pairs (A, P) whose stochastic average preserves the entropy of P : $H(AP) = H(P)$. Γ_n is a subset of $\mathbf{B}_n \times \Sigma_n$ where \mathbf{B}_n is the Birkhoff polytope and Σ_n is the probability simplex.

We characterize Γ_n and determine its geometry, topology, and combinatorial structure. For example, we find that $(A, P) \in \Gamma_n$ if and only if $A^t AP = P$. We show that for any n , Γ_n is piecewise-linearly contractible in $\mathbf{B}_n \times \Sigma_n$. We exhibit two finite decompositions of Γ_n . We derive the geometry of the fibers (A, \cdot) and (\cdot, P) of Γ_n . Γ_3 is worked out in detail. Our analysis exploits the convexity of $x \log x$ and the structure of an efficiently computable bipartite graph that we associate to each ds-matrix. This graph also lets us represent such a matrix in a permutation-equivalent, block-diagonal form where each block is doubly stochastic and fully indecomposable.

Keywords: entropy, stochastic average, doubly stochastic matrix, Birkhoff polytope, probability simplex
2010 MSC: 94A15, 05C50

*Corresponding author: +1 (719) 337-1210 (phone); +1 (719) 549-2962 (fax)
Email addresses: alan.poritz@gmail.com (Alan B. Poritz),
jonathan.poritz@gmail.com (Jonathan A. Poritz)
URL: www.poritz.net/jonathan (Jonathan A. Poritz)

1. Introduction

In his 1948 paper *A Mathematical Theory of Communications* [1], [2], Shannon introduced entropy as a measure of information in a probability distribution. If the column vector $P = (p_1, \dots, p_n)^t \in \mathbb{R}^n$ is a probability distribution, he defined its entropy to be the quantity

$$H = H(P) = - \sum_{i=1}^n p_i \log_2(p_i) ,$$

where by convention $p \log_2(p)$ is set to 0 whenever $p = 0$. He then went on to remark, [1, p. 395]:

“Any change toward equalization of the probabilities p_1, p_2, \dots, p_n increases H . Thus if $p_1 < p_2$ and we increase p_1 , decreasing p_2 by an equal amount so that p_1 and p_2 are more nearly equal, H increases. More generally, if we perform any “averaging” operation on the p_i of the form

$$p'_i = \sum_j a_{ij} p_j$$

where $\sum_i a_{ij} = \sum_j a_{ij} = 1$ and all $a_{ij} \geq 0$, then H increases (except in the special case where this transformation amounts to no more than a permutation of the p_j with H of course remaining the same).”

Let $A = (a_{ij})$ denote the $n \times n$ doubly stochastic matrix (*ds-matrix* for short) given above. We refer to the probability distribution $P' = AP$ as the *stochastic average of the pair* (A, P) . In these terms, what Shannon has said is that a necessary and sufficient condition for a stochastic average to preserve the entropy of P (*i.e.*, $H(AP) = H(P)$) is that there exists an $n \times n$ permutation matrix π such that $AP = \pi P$.

This viewpoint provokes several questions. What is the structure of the set of all ds-matrices A whose stochastic average with a fixed P preserves its entropy? What is the structure of the set of all P whose stochastic average with a fixed ds-matrix A preserves the entropy of P ? And, finally, what is the structure of the set of all pairs (A, P) whose stochastic averages preserve entropy? In what follows, we present a number of criteria for membership in these sets and then explore their geometry, topology and combinatorics.

We write $\mathbf{S}_n \subset \mathbb{R}^{n^2}$ for the set of $n \times n$ permutation matrices, and identify the permutation group on $\{1, \dots, n\}$ with \mathbf{S}_n : a permutation takes i into j if and only if the 1 in the i -th column of the corresponding matrix occurs in row j – hence the action of a permutation on $\{1, \dots, n\}$ corresponds to the left-multiplication action of the corresponding permutation matrix on the standard unit basis vectors $\{\mathbf{e}_1^n, \dots, \mathbf{e}_n^n\}$ of \mathbb{R}^n .

The convex hull of \mathbf{S}_n in \mathbb{R}^{n^2} is called the *Birkhoff polytope* and denoted \mathbf{B}_n . The Birkhoff-von Neumann Theorem (*e.g.*, [3]) states that \mathbf{B}_n is an $(n-1)^2$ -dimensional polytope with $n!$ vertices and consists of the set of all $n \times n$ ds-matrices.

Example 1.1. The set of 2×2 ds-matrices is $\mathbf{B}_2 = \left\{ \begin{pmatrix} t & 1-t \\ 1-t & t \end{pmatrix} \mid 0 \leq t \leq 1 \right\}$.

This set forms a line segment in \mathbb{R}^4 .

Let $\Sigma_n \subset \mathbb{R}^n$ denote the $(n-1)$ -dimensional simplex of n -long probability distributions; it contains the *uniform distribution* $U = (1/n, \dots, 1/n)^t$. U uniquely maximizes the entropy function H on Σ_n , [1].

Definition 1.2. The set of *entropy-preserving stochastic averages* is

$$\Gamma_n \triangleq \{(A, P) \in \mathbf{B}_n \times \Sigma_n \mid H(AP) = H(P)\}.$$

Γ_n contains both $\mathbf{S}_n \times \Sigma_n$ and $\mathbf{B}_n \times \{U\}$. Notice also that if $A \in \mathbf{B}_n$ is a direct sum of smaller ds-matrices and $P \in \Sigma_n$ is constant on the indices of each summand of A , then $(A, P) \in \Gamma_n$.

Further, there is a natural left action of $\mathbf{S}_n \times \mathbf{S}_n$ on $\mathbf{B}_n \times \Sigma_n$ given by

$$(\mathbf{S}_n \times \mathbf{S}_n) \times (\mathbf{B}_n \times \Sigma_n) \rightarrow \mathbf{B}_n \times \Sigma_n : ((\pi_1, \pi_2), (A, P)) \mapsto (\pi_1 A \pi_2^{-1}, \pi_2 P) \quad (1.1)$$

and we have

Proposition 1.3. Γ_n is invariant under the action of $\mathbf{S}_n \times \mathbf{S}_n$ on $\mathbf{B}_n \times \Sigma_n$.

Proof. Start with $(A, P) \in \Gamma_n$, so that $H(AP) = H(P)$, and any $\pi_1, \pi_2 \in \mathbf{S}_n$. Then

$$H((\pi_1 A \pi_2^{-1})(\pi_2 P)) = H(\pi_1 AP) = H(AP) = H(P) = H(\pi_2 P)$$

so $(\pi_1, \pi_2) \cdot (A, P) = (\pi_1 A \pi_2^{-1}, \pi_2 P) \in \Gamma_n$. □

To reach a complete understanding of Γ_n , for each pair $(A, P) \in \Gamma_n \times \Sigma_n$ we introduce in Section 2 a pair of partitions of $\{1, \dots, n\}$. One partition is induced on the domain of P by its level sets. For the second, we show that the positive entries of a ds-matrix $A \in \mathbf{B}_n$ determine a bipartite graph whose connected components induce a partition on the set of column indices of A . In Section 4 we prove that $(A, P) \in \Gamma_n$ if and only if the partition induced on the column indices of A refines the partition induced on the domain of P . We can then give a proof of Shannon's interpretation of Γ_n as consisting of those pairs (A, P) satisfying $AP = \pi P$ for some permutation π . We also prove a third characterization of Γ_n : $(A, P) \in \Gamma_n$ if and only if $A^t AP = P$.

These results allow us to study the structure of Γ_n in detail. In §5, we show that a fiber of Γ_n of the form (A, \cdot) is both a sub-simplex of Σ_n and a sub-complex of its barycentric subdivision [4]. In §6, we observe that a fiber of the form (\cdot, P) is a union of disjoint and isomorphic faces of the polytope \mathbf{B}_n . That the set of fibers of either form can be placed in bijective correspondence with the set of the partitions of $\{1, \dots, n\}$ is shown in §8. We show in Section 7 that Γ_n is a contractible and hence connected set. Finally, in Section 9, we establish a pair of partitions of Γ_n into finitely many product subsets and then explicitly compute these partitions for Γ_2 and Γ_3 .

2. Partitions and Decompositions

2.1. Set Partitions and Integer Partitions

Let \mathbb{N} denote the set of natural numbers. We make frequent use of both partitions of the set $\{1, \dots, n\} \subset \mathbb{N}$ and partitions of the integer $n \in \mathbb{N}$. A *partition of $\{1, \dots, n\}$* is a set of non-empty, pairwise disjoint subsets whose union is all of $\{1, \dots, n\}$; the collection of all partitions of $\{1, \dots, n\}$ will be denoted Φ_n . The subset of Φ_n consisting of those set partitions with exactly m subsets will be denoted $\Phi_{n,m}$. While set partitions are intrinsically unordered, if $\mathcal{J} \in \Phi_{n,m}$ is indexed as $\mathcal{J} = \{J_1, \dots, J_m\}$, where the sequence of least elements of the J_i 's increases as i goes from 1 up to m , we say that \mathcal{J} is *indexed in natural order*.

A *partition of $n \in \mathbb{N}$* is a set of positive integers whose sum is n ; the collection of all partitions of n will be denoted ϕ_n . The subset of ϕ_n consisting of all integer partitions with exactly m summands will be denoted $\phi_{n,m}$.

The two uses of the word ‘‘partition’’ are closely related.

Definition 2.1. Let $\mathcal{J} \in \Phi_{n,m}$ be a partition of the set $\{1, \dots, n\}$ and let $\mathbf{n} \in \phi_{n,m}$ be the partition of the integer n whose summands are the cardinalities of the subsets in \mathcal{J} . \mathbf{n} is referred to as the *integer partition induced by \mathcal{J}* .

If the set partition \mathcal{J} is ordered, then the integer partition \mathbf{n} inherits this ordering.

Definition 2.2. Conversely, let $\mathbf{n} = \{n_1, \dots, n_m\} \in \phi_{n,m}$ be an ordered partition of the integer n and let $\mathcal{N} = \{N_1, \dots, N_m\} \in \Phi_{n,m}$ be the ordered set partition such that $N_1 = \{1, \dots, n_1\}$ and, if $m > 1$, each successive N_i consists of the next n_i elements of $\{1, \dots, n\}$, for $i \in \{2, \dots, m\}$. \mathcal{N} is referred to as the *set partition induced by \mathbf{n}* .

The induced partition \mathcal{N} is in natural order.

Suppose that the set partition $\mathcal{J} \in \Phi_{n,m}$ is ordered; say $\mathcal{J} = \{J_1, \dots, J_m\}$. Then both the induced integer partition $\mathbf{n} = \{n_1, \dots, n_m\}$ and the subsequent induced set partition $\mathcal{N} = \{N_1, \dots, N_m\}$ are ordered. The cardinality of J_i and the cardinality of N_i are both equal to n_i . There is a permutation $\pi_{\mathcal{J}}$ uniquely defined by the requirement that it carries the k -th smallest integer in J_i to the k -th smallest integer in N_i for $k = 1 \dots, n_i$ and $i = 1 \dots, m$.

Definition 2.3. We refer to $\pi_{\mathcal{J}}$ as the *unscrambling permutation* induced by the ordered partition \mathcal{J} .

2.2. Going from Partitions to d s-Matrices and Probability Distributions

Partitions generate distinguished subsets of the Birkhoff polytope and the probability simplex.

Definition 2.4. Fix a set partition $\mathcal{J} \in \Phi_n$.

1. The \mathcal{J} -adapted permutation matrices are the members of the set

$$\mathbf{S}_n(\mathcal{J}) \triangleq \{\pi \in \mathbf{S}_n \mid \pi_{ij} \neq 0 \text{ implies there is a } J \in \mathcal{J} \text{ with both } i, j \in J\};$$

i.e., $\mathbf{S}_n(\mathcal{J})$ permutes the indices of each component of \mathcal{J} separately.

2. The \mathcal{J} -adapted ds-matrices $\mathbf{B}_n(\mathcal{J})$ are the convex hull of $\mathbf{S}_n(\mathcal{J})$ in \mathbb{R}^{n^2} .
3. Similarly, the \mathcal{J} -adapted probability distributions are the members of the set

$$\Sigma_n(\mathcal{J}) \triangleq \{P \in \Sigma_n \mid p_i \neq p_j \text{ implies there is a } J \in \mathcal{J} \text{ with } i \in J, j \notin J\};$$

i.e., $\Sigma_n(\mathcal{J})$ consists of probability distributions which are constant on the indices in each $J \in \mathcal{J}$.

4. Now fix an ordered partition $\mathbf{n} = \{n_1, \dots, n_m\} \in \Phi_{n,m}$ and set

$$\mathbf{S}_n(\mathbf{n}) \triangleq \mathbf{S}_n(\mathcal{N}), \quad \mathbf{B}_n(\mathbf{n}) \triangleq \mathbf{B}_n(\mathcal{N}), \quad \text{and} \quad \Sigma_n(\mathbf{n}) \triangleq \Sigma_n(\mathcal{N}),$$

where $\mathcal{N} \in \Phi_{n,m}$ is the set partition induced by \mathbf{n} .

$\mathbf{S}_n(\mathbf{n})$ and $\mathbf{B}_n(\mathbf{n})$ are each a set of block diagonal matrices in $M_{n \times n}(\mathbb{R})$ where block i has size $n_i \times n_i$ for $i = 1, \dots, m$. It further follows from their definitions that

$$\mathbf{S}_n(\mathbf{n}) \cong \mathbf{S}_{n_1} \times \dots \times \mathbf{S}_{n_m} \quad \text{and} \quad \mathbf{B}_n(\mathbf{n}) \cong \mathbf{B}_{n_1} \times \dots \times \mathbf{B}_{n_m}.$$

2.3. Going from Probability Distributions and ds-Matrices to Partitions

To the level sets of a probability distribution in Σ_n there is naturally associated a set partition in Φ_n :

Definition 2.5. Given $P = (p_1, \dots, p_n)^t \in \Sigma_n$, let $\mathcal{I}_P \in \Phi_n$ denote the partition of the domain of P which places i and j in the same subset if and only if $p_i = p_j$. We refer to \mathcal{I}_P as the *coincidence partition of P* .

Note $\mathcal{I}_P \in \Phi_{n,l}$, where l is the number of level sets of P . Notice also that if $\pi \in \mathbf{S}_n$, then $\mathcal{I}_{\pi P} = \pi \mathcal{I}_P$.

Next, for ds-matrices, the partition we need is defined in two stages:

Definition 2.6. Given $A = (a_{ij}) \in \mathbf{B}_n$, define its *weight graph* to be the bipartite graph W_A with vertices consisting of two disjoint copies of $\{1, \dots, n\}$ – we call these the *row indices R* and the *column indices C* , so the vertex set of W_A is $R \amalg C$ – and whose edges connect an $r \in R$ and a $c \in C$ if and only if the *weight* $a_{rc} > 0$.

The topology of the weight graph gives our next partition:

Definition 2.7. Given a ds-matrix A , if its weight graph W_A has m connected components then we call m its *component count*. If we denote these components as $\{W_1, \dots, W_m\}$, we can define the *row-index partition* $\mathcal{R}_A = \{R_1, \dots, R_m\} \in \Phi_{n,m}$ by setting R_i to be the set of all row indices that appear in W_i , for each $i \in \{1, \dots, m\}$, and likewise for the *column-index partition* $\mathcal{C}_A = \{C_1, \dots, C_m\} \in \Phi_{n,m}$.

We index the partition \mathcal{C}_A in natural order and then index the partitions \mathcal{R}_A and W_A such that R_i and C_i come both from the same component W_i , for each $i \in \{1, \dots, m\}$.

The row- and column-index partitions are closely related:

Proposition 2.8. *Let A be a ds-matrix with weight graph W_A , component count m , row-index partition $\mathcal{R}_A = \{R_1, \dots, R_m\}$ and column-index partition $\mathcal{C}_A = \{C_1, \dots, C_m\}$. Then for each $i \in \{1, \dots, m\}$,*

$$\#(R_i) = \#(C_i) \quad .$$

Proof. In any component of W_A , the sum of the weights on the edges equals the number of row indices in that component, since the sum of all weights going into any individual row vertex is 1. By the same reasoning, this sum must also equal the number of column indices. Therefore the corresponding sets in \mathcal{R}_A and \mathcal{C}_A must have the same cardinality. \square

Definition 2.9. For a ds-matrix A with weight graph W_A , component count m , row-index partition $\mathcal{R}_A = \{R_1, \dots, R_m\} \in \Phi_{n,m}$ and column-index partition $\mathcal{C}_A = \{C_1, \dots, C_m\} \in \Phi_{n,m}$, we call the partition $\mathbf{n}_A = \{n_1, \dots, n_m\} \in \phi_{n,m}$ of n defined either by $n_i = \#(R_i)$ or $n_i = \#(C_i)$, for $i \in \{1, \dots, m\}$, the *component size partition of A* . We give \mathbf{n}_A the ordering induced by the natural order of the column-index partition \mathcal{C} .

When the matrix A in question is clear from context, we suppress the subscript A in all of the objects defined above: W instead of W_A , \mathcal{C} instead of \mathcal{C}_A , etc. Likewise, if the probability distribution P is clear, we write \mathcal{I} instead of \mathcal{I}_P .

Given a ds-matrix A , there is an efficient algorithm for constructing all of the objects in definitions 2.6, 2.7, and 2.9: One connected component of the graph is produced in each round. At the start of the i -th round, R_i and E_i are the empty set and C_i has as single entry, the smallest column index not yet used in earlier rounds. Each round consists of a sequence of steps; each step consists of two moves. Move 1: for each column index c just adjoined to C_i , adjoin to R_i the row indices for which A has positive entries in column c , and adjoin to E_i the corresponding edges. Move 2: for each row index r just adjoined to R_i , adjoin to C_i the column indices for which A has positive entries in row r , and adjoin to E_i the corresponding edges. After each step, test whether the cardinality of R_i equals the cardinality of C_i . If they are not equal, step again; if they are equal, the component W_i is complete and a new round begins. The process terminates in at most n steps. Pseudocode for this is shown as Algorithm 2.1.

2.4. A Block Diagonal Equivalent Form for Doubly Stochastic Matrices

Recall that two $n \times n$ matrices A and B are called *permutation equivalent* if there exist $\pi_1, \pi_2 \in \mathbf{S}_n$ such that $B = \pi_1 A \pi_2^{-1}$. In other words, B is in the orbit of A under the left action $(\mathbf{S}_n \times \mathbf{S}_n) \times M_{n \times n}(\mathbb{R}) \rightarrow M_{n \times n}(\mathbb{R}) : ((\pi_1, \pi_2), A) \mapsto$

Algorithm 2.1 Graph- and partition-creation algorithm

```

1:  $i \leftarrow 0$ 
2:  $C^u \leftarrow \{1, \dots, n\}$ 
3: repeat
4:    $i \leftarrow i + 1$ 
5:    $R_i \leftarrow \emptyset$ 
6:    $C_i \leftarrow \min C^u$ 
7:   repeat
8:      $R_i \leftarrow R_i \cup \{r \in R \mid a_{rc} > 0 \text{ for some } c \in C_i\}$ 
9:      $C_i \leftarrow C_i \cup \{c \in C \mid a_{rc} > 0 \text{ for some } r \in R_i\}$ 
10:  until  $\#(R_i) = \#(C_i)$ 
11:  define  $W_i$  as the graph with vertices  $R_i \amalg C_i$  and edges
     $\{(r, c) \in R_i \times C_i \mid a_{rc} > 0\}$ 
12:   $n_i \leftarrow \#(R_i)$  (or  $\#(C_i)$ )
13:   $C^u \leftarrow C^u \setminus C_i$ 
14: until  $C^u$  is empty

                                COMPONENT COUNT:            $m \leftarrow i$ 
                                WEIGHT GRAPH:               $W \leftarrow \bigcup_{i=1}^m W_i$ 
15: output  ROW-INDEX PARTITION:       $\mathcal{R} = \{R_1, \dots, R_m\}$ 
                                COLUMN-INDEX PARTITION:     $\mathcal{C} = \{C_1, \dots, C_m\}$ 
                                COMPONENT SIZE PARTITION:   $\mathbf{n} = \{n_1, \dots, n_m\}$ 

```

$\pi_1 A \pi_2^{-1}$. (See, e.g., [3] for further explication of this and other common terms used in this section).

The notion of permutation equivalence plays well with weight graphs and the partitions we have defined in §2.3:

Proposition 2.10. *If the matrices $A, B \in \mathbf{B}_n$ are permutation equivalent then their weight graphs are isomorphic, $W_B \cong W_A$. In fact, if $B = \pi_1 A \pi_2^{-1}$ for $\pi_1, \pi_2 \in \mathbf{S}_n$, then the row- and column-index partitions are related by $\mathcal{R}_B = \pi_1 \mathcal{R}_A$ and $\mathcal{C}_B = \pi_2 \mathcal{C}_A$, while the component size partitions are equal.*

An $n \times n$ matrix A is said to be *partly decomposable* if it contains $m \times (n - m)$ zero submatrix for some m satisfying $0 < m < n$; that is, if there are permutations $\pi_1, \pi_2 \in \mathbf{S}_n$ such that

$$\pi_1 A \pi_2^{-1} = \begin{pmatrix} B & C \\ 0_{m \times (n-m)} & D \end{pmatrix}. \quad (2.1)$$

A is called *fully indecomposable* if it is not partly decomposable.

Here then is our equivalent form for ds-matrices:

Theorem 2.11. *Let $A \in \mathbf{B}_n$ have component size partition $\mathbf{n} = \{n_1, \dots, n_m\}$. Then A is permutation equivalent to a matrix in $\mathbf{B}_n(\mathbf{n})$, which in turn is a block diagonal matrix with blocks of sizes n_1, \dots, n_m , each of which blocks is fully indecomposable.*

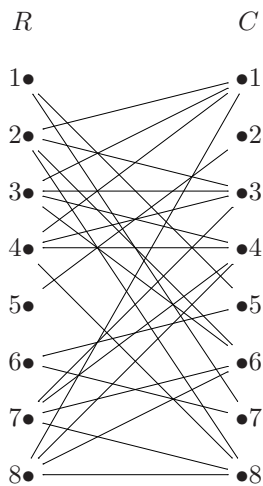
Proof. Say the row- and column-index partitions of A are $\mathcal{R} = \{R_1, \dots, R_m\}$ and $\mathcal{C} = \{C_1, \dots, C_m\}$, with \mathcal{C} indexed in its natural order and \mathcal{R} indexed so that R_i and C_i come from the same component of the weight graph W_A , for $i \in \{1, \dots, m\}$. Then Proposition 2.10 tells us immediately that the permutation equivalence using the unscrambling permutations $\pi_2 = \pi_{\mathcal{C}}$ and $\pi_1 = \pi_{\mathcal{R}}$ has exactly the desired form in $\mathbf{B}_n(\mathbf{n})$.

To show the blocks are fully indecomposable, we restrict our attention to a single block in $\mathbf{B}_n(\mathbf{n})$. Here it suffices to show that if a matrix A is partly decomposable, then its component count must be greater than 1. Let notation be as above at equation (2.1). Since $\pi_1 A \pi_2^{-1}$ is column-stochastic, the sum of all entries in the first $n - m$ columns must be $n - m$, which must then be the sum of all the elements of B as the block of zeros contributes nothing to that column sum. The same argument using now that $\pi_1 A \pi_2^{-1}$ is row-stochastic shows that C must in fact be an $(n - m) \times m$ block of zeros. Thus in fact $\pi_1 A \pi_2^{-1} \in \mathbf{B}_{n-m} \times \mathbf{B}_m$ and so has component count of at least 2. \square

Example 2.12. Consider the 8×8 doubly stochastic matrix:

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & .4 & 0 & .6 & 0 \\ .2 & 0 & .4 & 0 & 0 & .1 & 0 & .3 \\ .4 & 0 & .3 & .1 & 0 & .2 & 0 & 0 \\ .1 & 0 & .2 & .3 & 0 & 0 & 0 & .4 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .6 & 0 & .4 & 0 \\ 0 & 0 & .1 & .4 & 0 & .3 & 0 & .2 \\ .3 & 0 & 0 & .2 & 0 & .4 & 0 & .1 \end{pmatrix}.$$

The weight graph of this A is:



Algorithm 2.1 of the preceding section yields a component count of $m = 3$, column-index partition \mathcal{C} , row-index partition \mathcal{R} , and component size partition

n of the integer $n = 8$, which are displayed in the following table:

i	1	2	3
C_i	{1, 3, 4, 6, 8}	{2}	{5, 7}
R_i	{2, 3, 4, 7, 8}	{5}	{1, 6}
n_i	5	1	2

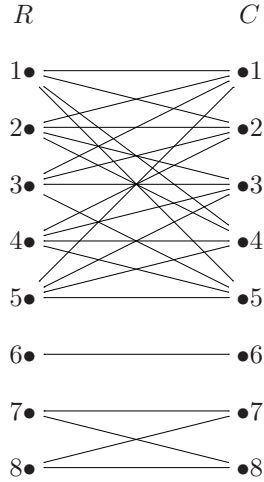
(2.2)

A is permutation equivalent to the block diagonal matrix:

$$B = \pi_{\mathcal{R}}^{-1} A \pi_{\mathcal{C}} = \begin{pmatrix} .2 & .4 & 0 & .1 & .3 & 0 & 0 & 0 \\ .4 & .3 & .1 & .2 & 0 & 0 & 0 & 0 \\ .1 & .2 & .3 & 0 & .4 & 0 & 0 & 0 \\ 0 & .1 & .4 & .3 & .2 & 0 & 0 & 0 \\ .3 & 0 & .2 & .4 & .1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .4 & .6 \\ 0 & 0 & 0 & 0 & 0 & 0 & .6 & .4 \end{pmatrix} \in \mathbf{B}_n(\mathbf{n}) = \mathbf{B}_5 \times \mathbf{B}_1 \times \mathbf{B}_2 \subset \mathbf{B}_8$$

where $\pi_{\mathcal{C}}$ and $\pi_{\mathcal{R}}$ can be read off as the permutations which send the second and third rows, *resp.*, of the above table (2.2) to $(1, \dots, 8)$ (or, in cycle notation, $\pi_{\mathcal{C}} = (3\ 2\ 6\ 4)(5\ 7\ 8)$ and $\pi_{\mathcal{R}} = (2\ 1\ 7\ 4\ 3)(5\ 6\ 8)$).

Note that the weight graph of B is isomorphic to that of A but shows the components much more clearly:



3. Stochastic Averaging Does Not Decrease Entropy

The point of this paper is to study when stochastic averages preserve entropy. Of course, typically they *increase* the entropy – see the discussion in Shannon [1]. But they certainly do not decrease the entropy:

Proposition 3.1. (Shannon) [1] Let $A \in \mathbf{B}_n$ and $P \in \Sigma_n$. Then the entropy of the stochastic average AP is not less than the entropy of P : $H(AP) \geq H(P)$.

Proof. Say $P = (p_1, \dots, p_n)^t$. The entropy of P can also be expressed as $H(P) = -\sum \varphi(p_i)$, where $\varphi(x) = x \log_2(x)$, extended (by continuity) so that $\varphi(0) = 0$. But since this $\varphi(x)$ is convex on the interval $[0, 1]$ and A is column-stochastic,

$$\varphi\left(\sum_{j=1}^n a_{ij}p_j\right) \leq \sum_{j=1}^n a_{ij}\varphi(p_j) \quad (3.1)$$

holds for any i . But then

$$\begin{aligned} H(AP) &= -\sum_{i=1}^n \varphi\left(\sum_{j=1}^n a_{ij}p_j\right) \\ &\geq -\sum_{i=1}^n \sum_{j=1}^n a_{ij}\varphi(p_j) \\ &= -\sum_{j=1}^n \left(\sum_{i=1}^n a_{ij}\right)\varphi(p_j) \\ &= -\sum_{j=1}^n \varphi(p_j) \\ &= H(P) \end{aligned} \quad (3.2)$$

□

4. Three Characterizations of Γ_n

4.1. A Critical Definition and a New Block Diagonal Form for ds-Matrices

Recall that given $\mathcal{J}, \mathcal{J}' \in \Phi_n$, we say that \mathcal{J} *refines* \mathcal{J}' if $\forall J \in \mathcal{J} \exists J' \in \mathcal{J}'$ such that $J \subseteq J'$. In such a case, we write $\mathcal{J} \leq \mathcal{J}'$, while we write $\mathcal{J} < \mathcal{J}'$ if $\mathcal{J} \leq \mathcal{J}'$ and $\mathcal{J} \neq \mathcal{J}'$. Φ_n becomes a partially ordered set with this ordering.

The following notion is crucial to the rest of this paper:

Definition 4.1. Let $A \in \mathbf{B}_n$ and $P \in \Sigma_n$. We say that A *respects* P if the column-index partition \mathcal{C}_A of A refines the coincidence partition \mathcal{I}_P of P : $\mathcal{C}_A \leq \mathcal{I}_P$. We write

$$\mathcal{R}_n = \{(A, P) \in \mathbf{B}_n \times \Sigma_n \mid A \text{ respects } P\}$$

for the set of respectful pairs.

Just as entropy preservation is invariant under the $\mathbf{S}_n \times \mathbf{S}_n$ action (defined in equation 1.1), so is the status of a ds-matrix respecting a probability distribution:

Proposition 4.2. \mathcal{R}_n is invariant under the action of $\mathbf{S}_n \times \mathbf{S}_n$ on $\mathbf{B}_n \times \Sigma_n$.

Proof. We must show that for any $\pi_1, \pi_2 \in \mathbf{S}_n$, $A \in \mathbf{B}_n$, and $P \in \Sigma_n$, A respects P if and only if $B = \pi_1 A \pi_2^{-1}$ respects $Q = \pi_2 P$. But for any permutation π_2 , \mathcal{C}_A refines \mathcal{I}_P if and only if $\pi_2 \mathcal{C}_A$ refines $\pi_2 \mathcal{I}_P$. By Proposition 2.10, $\mathcal{C}_B = \pi_2 \mathcal{C}_A$, while $\mathcal{I}_Q = \pi_2 \mathcal{I}_P$, by the note following Definition 2.5. Hence \mathcal{C}_A refines \mathcal{I}_P if and only if \mathcal{C}_B refines \mathcal{I}_Q . □

Now we show that every $\mathbf{S}_n \times \mathbf{S}_n$ -orbit in \mathcal{B}_n intersects $\mathbf{B}_n(\mathbf{n}) \times \Sigma_n(\mathbf{n})$:

Proposition 4.3. *Let $P \in \Sigma_n$ be a probability distribution. Choose an ordering of its coincidence partition $\mathcal{I} \in \Phi_{n,l}$, say as $\mathcal{I} = \{I_1, \dots, I_l\}$, and let $\mathbf{n} \in \phi_{n,l}$ be the induced integer partition. Let $\pi_{\mathcal{I}} \in \mathbf{S}_n$ be the corresponding unscrambling permutation. Then for any $A \in \mathbf{B}_n$ that respects P there is a permutation $\pi \in \mathbf{S}_n$ such that $\pi A \pi^{-1} \in \mathbf{B}_n(\mathbf{n})$ and $\pi_{\mathcal{I}} P \in \Sigma_n(\mathbf{n})$.*

Proof. The approach in the proof of Theorem 2.11 works here as well. First, $\pi_{\mathcal{I}} P \in \Sigma_n(\mathbf{n})$, from the definition of the unscrambling permutation.

Next, given an $A \in \mathbf{B}_n$ of component count m and row- and column-index partitions \mathcal{R} and \mathcal{C} which respects P , we define an ordered partition $\mathcal{J} = \{J_1, \dots, J_l\}$ of $\{1, \dots, n\}$ as follows: for $i \in \{1, \dots, l\}$, let

$$J_i = \bigcup_{j \ni C_j \subset I_i} R_j .$$

Letting π be the unscrambling permutation $\pi_{\mathcal{J}}$, a direct application of Proposition 2.10 yields that $\pi A \pi^{-1} \in \mathbf{B}_n(\mathbf{n})$. \square

4.2. Main Result – Conditions for Entropy Preserving Stochastic Averages

We shall need Theorem 90 in Hardy, Littlewood, and Polya's *Inequalities* [5, p. 74], specialized to the continuous, non-linear, convex function $\varphi(x) = x \log_2 x$ of Proposition 3.1.

Lemma 4.4. *Let $\{q_1, \dots, q_k\}$ be a set of k positive weights that sum to one and let $\{p_1, \dots, p_k\}$ be a set of k real numbers in the unit interval $[0, 1]$. Then*

$$\varphi\left(\sum_{j=1}^k q_j p_j\right) = \sum_{j=1}^k q_j \varphi(p_j)$$

if and only if the p_j s are all equal.

Now we are in the position to state and prove the main result of this paper:

Theorem 4.5. *Let $A \in \mathbf{B}_n$ be an $n \times n$ doubly stochastic matrix and $P \in \Sigma_n$ an n -long probability vector. The following statements are equivalent¹:*

1. $H(AP) = H(P)$ i.e., $(A, P) \in \Gamma_n$
2. A respects P
3. $A^t AP = P$
4. $AP = \pi P$ for some permutation matrix $\pi \in \mathbf{S}_n$.

¹The equivalence of (1) and (4) is Shannon's original result.

Proof. Let $\mathcal{C} = \{C_1, \dots, C_m\}$ and $\mathcal{R} = \{R_1, \dots, R_m\}$ be the column-index and row-index partitions induced by A , respectively, and let $\mathcal{I} = \{I_1, \dots, I_l\}$ be the coincidence partition induced by P .

$\boxed{(1) \Rightarrow (2)}$: If $H(AP) = H(P)$, then inequality (3.2) is actually an equality. In particular, the first two lines of (3.2) yield

$$\sum_{i=1}^n \varphi \left(\sum_{j=1}^n a_{ij} p_j \right) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} \varphi(p_j) .$$

By the inequality (3.1), the summands here all change in the same direction; since the sums are equal and it cannot happen that one term increases while the other decreases, in fact the summands must all individually be equal. Since $\varphi(0) = 0$, we may as well restrict our attention only to those terms which are non-zero, *i.e.*,

$$\varphi \left(\sum_{\substack{j \ni \\ a_{ij} \neq 0}} a_{ij} p_j \right) = \sum_{\substack{j \ni \\ a_{ij} \neq 0}} a_{ij} \varphi(p_j) .$$

But applying Lemma 4.4, we can conclude that the non-zero entries in the i -th row of A must form a set of indices for which the p_j are all equal – that is, the non-zero indices in any row of A are a subset of some $I_j \in \mathcal{I}$.

Refer now to lines 7-10 of the Algorithm 2.1 constructing the components $\{W_1, \dots, W_m\}$ of the weight graph W . The sets adjoined to some C_j each time through this loop are unions of sets of the form

$$C(r) = \{c \in C \mid a_{rc} > 0\}$$

for various values of r . We have just seen that such sets $C(r)$ are always each contained in a corresponding element of the partition \mathcal{I} , but it remains to see that the $C(r)$ whose union forms a C_j overlap sufficiently to ensure that they are subsets of *the same* element of \mathcal{I} .

Say some $c \in C_j$ and then that $r, r' \in R$ give edges (r, c) and (r', c) of W – in fact, the corresponding $C(r)$ and $C(r')$ are both to be adjoined to C_j . But the existence of these edges tells us that both rows r and r' of A have non-zero entries in column c (at least), *i.e.*, $c \in C(r) \cap C(r')$. This means that the constant value P takes on for $C(r)$ is the same as that value it takes on for $C(r')$, so $C(r) \cup C(r') \subseteq I_{j'} \in \mathcal{I}$ for some j' . Then continuing this way, also with new c' coming from the $C(r) \cup C(r')$ adjoined to C_j , builds all of C_j , which hence entirely lies in $I_{j'}$. Thus A respects P .

$\boxed{(2) \Rightarrow (3) \ \& \ (4)}$: Let $\pi, \pi_{\mathcal{I}}$, and \mathbf{n} be as in Proposition 4.3, so that $(B, Q) = (\pi A \pi_{\mathcal{I}}^{-1}, \pi_{\mathcal{I}} P) \in \mathcal{R}_n \cap (\mathbf{B}_n(\mathbf{n}) \times \Sigma_n(\mathbf{n}))$. But then the first $n_1 \times n_1$ block of B is itself a ds-matrix, while Q restricted to $\{1, \dots, n_1\}$ is a multiple of the uniform distribution, so this block leaves that part of Q unchanged. The same argument

applies to the remaining blocks of B and Q , so in fact $BQ = Q$; exactly the same reasoning shows also that $B^t Q = Q$. But then

$$A^t AP = (\pi_{\mathcal{I}}^t B^t \pi)(\pi^t B \pi_{\mathcal{I}}) \pi_{\mathcal{I}}^t Q = \pi_{\mathcal{I}}^t B^t B Q = \pi_{\mathcal{I}}^t B^t Q = \pi_{\mathcal{I}}^t Q = P$$

(using the fact that the transpose of a permutation matrix is its inverse), which is (3). Similarly,

$$AP = (\pi^t B \pi_{\mathcal{I}}) \pi_{\mathcal{I}}^t Q = \pi^t B Q = \pi^t Q = (\pi^t \pi_{\mathcal{I}}) P$$

which is (4), using the permutation $\pi^t \pi_{\mathcal{I}}$.

(3) \Rightarrow (1): If $A^t AP = P$, then

$$H(P) = H(A^t AP) \geq H(AP) \geq H(P),$$

where the inequalities are two applications of Proposition 3.1, one using also the fact that transpose preserves \mathbf{B}_n .

(4) \Rightarrow (1): If $AP = \pi P$ for some permutation matrix $\pi \in \mathbf{S}_n$, then $H(AP) = H(\pi P) = H(P)$. □

5. Subsets of Γ_n of the Form (A, \cdot)

Let us consider fibers of the projection map $\Gamma_n \rightarrow \mathbf{B}_n$:

Definition 5.1. We write σ_A for the set of distributions $P \in \Sigma_n$ whose stochastic averages with the fixed ds-matrix $A \in \mathbf{B}_n$ preserve the entropy of P .

Theorem 4.5 tells us that σ_A consists of the distributions respected by A , and is therefore completely determined by the column-index partition \mathcal{C}_A . In order to establish another description of σ_A , we build a map as follows: Let $\mathcal{J} = \{J_1, \dots, J_m\} \in \Phi_{n,m}$ be indexed in natural order, and for each $i \in \{1, \dots, m\}$, write $n_i = \#(J_i)$. Write

$$T_{\mathcal{J}} : \mathbb{R}^m \rightarrow \mathbb{R}^n : (x_1, \dots, x_m) \mapsto \sum_{i=1}^m \frac{x_i}{n_i} \left(\sum_{j \in J_i} \mathbf{e}_j^n \right)$$

(where, again, the \mathbf{e}_j^n are the standard unit vectors in \mathbb{R}^n)

Here are some basic properties of this map:

Proposition 5.2. *Let $\mathcal{J} \in \Phi_{n,m}$ be indexed in natural order. Then $T_{\mathcal{J}}$ is a rank m linear map which sends probability distributions to probability distributions. In fact, for any ds-matrix A whose column-index partition is \mathcal{C}_A , $T_{\mathcal{C}_A}(\Sigma_m) = \sigma_A$.*

Proof. Let the n_i be as above. The normalization factors of $\frac{1}{n_i}$ in the definition of $T_{\mathcal{J}}$ ensure that it carries Σ_m into Σ_n . Next, if $Q \in \Sigma_m$, then A respects $P = T_{\mathcal{C}_A}(Q)$, so that $T_{\mathcal{C}_A}(\Sigma_m) \subseteq \sigma_A$. Conversely, if A respects a distribution P , then P takes on at most m distinct values, and is constant on each set of indices C_i , so $\sigma_A \subseteq T_{\mathcal{C}_A}(\Sigma_m)$. □

This map allows us to determine the geometry of σ_A :

Theorem 5.3. *Let $A \in \Sigma_n$ be an $n \times n$ ds-matrix and let m be the component count of its weight graph. Then*

- $\sigma_A \subset \Sigma_n$ is an $(m - 1)$ -dimensional simplex and
- σ_A is a sub-complex of the barycentric subdivision of Σ_n .

Proof. The standard unit vectors $\{\mathbf{e}_j^m \mid j = 1, \dots, m\}$ of \mathbb{R}^m , which are the vertices of Σ_m , are each mapped by the nonsingular map T_{C_A} to the linearly independent vectors

$$\left\{ T_{C_A}(\mathbf{e}_i^m) = \frac{1}{n_i} \left(\sum_{j \in C_i} \mathbf{e}_j^n \right) \mid i = 1, \dots, m \right\} \subset \mathbb{R}^n .$$

These vectors are then the vertices of σ_A , which therefore has the stated dimension. Furthermore, the i -th of these vertices is the barycenter of the $(n_i - 1)$ -dimensional face of Σ_n with vertices $\{\mathbf{e}_j \mid j \in C_i\}$, which yields the second statement of this theorem. \square

Example 5.4. For the ds-matrix A from Example 2.12, the barycenters of the sub-simplices of Σ_8 determined by $C_A = \{C_1, C_2, C_3\} = \{\{1, 3, 4, 6, 8\}, \{2\}, \{5, 7\}\}$ are, respectively, the distributions:

$$\begin{aligned} T_{C_A}(\mathbf{e}_1^3) &= (1/5, 0, 1/5, 1/5, 0, 1/5, 0, 1/5)^t \\ T_{C_A}(\mathbf{e}_2^3) &= (0, 1, 0, 0, 0, 0, 0, 0)^t \\ T_{C_A}(\mathbf{e}_3^3) &= (0, 0, 0, 0, 1/2, 0, 1/2, 0)^t. \end{aligned}$$

The convex hull of these three distributions is σ_A . In particular, let $Q = (q_1, q_2, q_3) \in \Sigma_3$ be any distribution and write $P = T_{C_A}(Q) = \sum_{i=1}^3 q_i T_{C_A}(\mathbf{e}_i^3)$. Then

$$P = (q_1/5, q_2, q_1/5, q_1/5, q_3/2, q_1/5, q_3/2, q_1/5)^t$$

while

$$AP = (q_3/2, q_1/5, q_1/5, q_1/5, q_2, q_3/2, q_1/5, q_1/5)^t.$$

This AP is a permutation of P , hence $H(AP) = H(P)$, and so $P \in \sigma_A$.

Example 5.5. If $A \in \mathbf{S}_n$, then A has a component count of n and $\sigma_A = \Sigma_n$.

Example 5.6. If $A \in \mathbf{B}_n$ has a strictly positive row or column, then A has a component count of 1 and $\sigma_A = \{U\}$.

Example 5.7. If A is a direct sum of m fully indecomposable ds-matrices, then A has a component count of m by Theorem 2.11 and σ_A is an $(m - 1)$ -simplex.

6. Subsets of Γ_n of the Form (\cdot, P)

Now we study fibers of the projection $\Gamma_n \rightarrow \Sigma_n$:

Definition 6.1. We write \mathbf{b}_P for the set of ds-matrices $A \in \mathbf{B}_n$ whose stochastic averages with the fixed distribution $P \in \Sigma_n$ preserve its entropy.

Again, Theorem 4.5 says that \mathbf{b}_P consists of the ds-matrices that respect P . Here is a concrete description of \mathbf{b}_P :

Theorem 6.2. *Let $P \in \Sigma_n$ be a probability distribution. Choose an ordering of its coincidence partition $\mathcal{I} \in \Phi_{n,l}$ and let $\mathbf{n} = \{n_1, \dots, n_l\} \in \phi_{n,l}$ be the induced integer partition. Let $\pi_{\mathcal{I}} \in \mathbf{S}_n$ be the corresponding unscrambling permutation. Then we have*

$$\mathbf{b}_P = \bigcup_{\pi \mathbf{S}_n(\mathbf{n}) \in \mathbf{S}_n / \mathbf{S}_n(\mathbf{n})} \pi \cdot \mathbf{B}_n(\mathbf{n}) \cdot \pi_{\mathcal{I}}^{-1} \quad , \quad (6.1)$$

which expresses \mathbf{b}_P as a union of $n! / n_1! n_2! \cdots n_l!$ pairwise disjoint faces of the polytope \mathbf{B}_n , each of which is isomorphic to $\mathbf{B}_n(\mathbf{n})$. Hence, also, \mathbf{b}_P has dimension $\sum_{k=1}^l (n_k - 1)^2$.

Proof. The conclusion of Proposition 4.3 is that for every $A \in \mathbf{b}_P$, there is a $\pi \in \mathbf{S}_n$ such that $\pi A \pi_{\mathcal{I}}^{-1} \in \mathbf{B}_n(\mathbf{n})$. In other words,

$$\mathbf{b}_P \subseteq \{(\pi B \pi_{\mathcal{I}}^{-1} \mid \pi \in \mathbf{S}_n, B \in \mathbf{B}_n(\mathbf{n}))\} = \mathbf{S}_n \cdot \mathbf{B}_n(\mathbf{n}) \cdot \pi_{\mathcal{I}}^{-1} \quad .$$

Since the opposite inclusion is quite clear in light of Theorem 4.5 and the definition of respect, we actually have that $\mathbf{b}_P = \mathbf{S}_n \cdot \mathbf{B}_n(\mathbf{n}) \cdot \pi_{\mathcal{I}}^{-1}$. However, this description is not the most efficient, since the \mathbf{S}_n -action is not effective: all of the elements of $\mathbf{S}_n(\mathbf{n})$ fix $\mathbf{B}_n(\mathbf{n})$. We can divide by this subgroup to have the above decomposition (6.1).

Since $\#(\mathbf{S}_n(\mathbf{n})) = n_1! n_2! \cdots n_l!$ and $\dim(\mathbf{B}_n(\mathbf{n})) = \sum_{k=1}^l (n_k - 1)^2$, we shall be done if we can show that the different faces $\pi \cdot \mathbf{B}_n(\mathbf{n}) \cdot \pi_{\mathcal{I}}^{-1}$ for representatives π of different cosets in $\mathbf{S}_n / \mathbf{S}_n(\mathbf{n})$ are disjoint. For this, say $B_1, B_2 \in \mathbf{B}_n(\mathbf{n})$ and $\pi_1, \pi_2 \in \mathbf{S}_n$ have $\pi_1 B_1 \pi_{\mathcal{I}}^{-1} = \pi_2 B_2 \pi_{\mathcal{I}}^{-1}$. But then $(\pi_2^{-1} \pi_1) B_1 = B_2$, which is a permutation matrix $\pi_2^{-1} \pi_1$ acting on the left on a matrix in $\mathbf{B}_n(\mathbf{n})$ and yielding again a matrix in $\mathbf{B}_n(\mathbf{n})$. Look at a row in B_1 , in the a block of rows defined by j th subset in the partition which defines $\mathbf{B}_n(\mathbf{n})$. Since the entire block is a $n_j \times n_j$ ds-matrix, there must be at least one non-zero element in this row. But then if we permute the rows by left-multiplication with $\pi_2^{-1} \pi_1$ and have again an element of $\mathbf{B}_n(\mathbf{n})$, the row in question must not have been moved out of the j th block. Therefore $\pi_2^{-1} \pi_1 \in \mathbf{S}_n(\mathbf{n})$ and we are done. \square

Example 6.3. If $P = (1, 0, \dots, 0)^t \in \Sigma_n$, then $l = 2$ and \mathbf{b}_P is the disjoint union of the n faces $\pi(\mathbf{B}_1 \times \mathbf{B}_{n-1})$ of \mathbf{B}_n as π runs over the set of the permutation matrices obtained by exchanging the first and i -th rows of the identity matrix, for $i \in \{1, \dots, n\}$.

Example 6.4. For the uniform distribution $U \in \Sigma_n$, we have: $l = 1$, $\mathbf{n} = \{n\}$, $\mathbf{B}_n(\mathbf{n}) = \mathbf{B}_n$, $\mathbf{S}_n(\mathbf{n}) = \mathbf{S}_n$, and decomposition in (6.1) yields \mathbf{b}_U as the one face \mathbf{B}_n .

Example 6.5. If $P \in \Sigma_n$ takes n distinct values, then: $l = n$, $\mathbf{n} = \{1, \dots, 1\}$, $\mathbf{B}_n(\mathbf{n}) = \{\text{Id}_{n \times n}\} = \mathbf{S}_n(\mathbf{n})$, and the decomposition in (6.1) writes \mathbf{b}_P as the union of $n!$ 0-dimensional faces, $\mathbf{b}_P = \mathbf{S}_n \cdot \{\text{Id}_{n \times n}\} = \mathbf{S}_n \subset \mathbf{B}_n$.

7. The Topology of Γ_n

It is natural to think of the topology and linear structure of Γ_n as coming from its realization as a subset of $\mathbf{B}_n \times \Sigma_n \subset \mathbb{R}^{n^2+n}$. Then we can ask how the different fibers discussed in the last two sections fit together. The answer is quite simple:

Theorem 7.1. Γ_n is PL-contractible².

Proof. Each σ_A contains U and is convex. Also, as we have seen in Example 6.4, $\mathbf{b}_U = \mathbf{B}_n$, which is also convex. So for any $B \in \mathbf{B}_n$, we can contract first each of the fibers of $\Gamma_n \rightarrow \mathbf{B}_n$ to the distribution U , then contract the base of this fibration to the point B . Explicitly, define $G_B : \mathbf{B}_n \times \Sigma_n \times [0, 1] \rightarrow \mathbf{B}_n \times \Sigma_n$ by

$$G_B(A, P, t) = \begin{cases} (A, (1-2t)P + 2tU) & \text{if } 0 \leq t \leq 1/2 \\ (2-2t)A + (2t-1)B, U & \text{if } 1/2 \leq t \leq 1 \end{cases} .$$

This $G_B|_{\Gamma_n}$ is a PL-homotopy between the identity map on Γ_n and the constant map $\Gamma_n \rightarrow \{(B, U)\}$. \square

8. Enumerating the Distinct Fibers of the Maps $\mathbf{B}_n \leftarrow \Gamma_n \rightarrow \Sigma_n$

Let us now examine all the fibers from §§5 and 6 separately. It is convenient to have notation for the sets of such fibers.

Definition 8.1. We write

$$s_n = \{\sigma_A \mid A \in \mathbf{B}_n\} \subset 2^{\Sigma_n}$$

$$\beta_n = \{\mathbf{b}_P \mid P \in \Sigma_n\} \subset 2^{\mathbf{B}_n}$$

and endow each with a partial ordering coming from set inclusion.

We have

Theorem 8.2. *There exist bijections $F : \Phi_n \rightarrow s_n$, which reverses order, and $G : \Phi_n \rightarrow \beta_n$, which preserves order.*

²See [6] for a classical background on the PL category.

Proof. First, given $\mathcal{J} \in \Phi_n$, define a ds-matrix $f(\mathcal{J}) \in \mathbf{B}_n$ as follows: for $r, c \in \{1, \dots, n\}$ let

$$f(\mathcal{J})_{rc} = \begin{cases} 1/\#(J) & \text{if } r \text{ and } c \text{ are in the same } J \in \mathcal{J} \\ 0 & \text{otherwise.} \end{cases}$$

In order to understand $f : \Phi_n \rightarrow \mathbf{B}_n$, consider the case of a partition $\mathbf{n} = \{n_1, \dots, n_m\} \in \phi_{n,m}$ of n , with some chosen ordering, and then $\mathcal{N} \in \Phi_{n,m}$, the set partition induced by \mathbf{n} . Now $f(\mathcal{N})$ is a block-diagonal matrix, with successive blocks being of size $n_i \times n_i$ and filled with the entry $1/n_i$. Therefore the weight graph of $f(\mathcal{N})$ has m components, each of which is the complete bipartite graph on $2n_i$ vertices and hence the column-index partition of $f(\mathcal{N})$ is \mathcal{N} .

Any other $\mathcal{J} \in \Phi_n$ is a permutation of a partition \mathcal{N} as just examined, so $f(\mathcal{J})$ is conjugate to a block-diagonal matrix and still has an isomorphic weight graph. Thus we always have $\mathcal{C}_{f(\mathcal{J})} = \mathcal{J}$.

Also, the function f is one-to-one since following it with the function that takes a ds-matrix to its column-index partition yields the identity on Φ_n .

Out of this f , build the desired function

$$F : \Phi_n \rightarrow s_n : \mathcal{J} \mapsto \sigma_{f(\mathcal{J})} .$$

Similarly, starting with $\mathcal{J} \in \Phi_{n,m}$, define a distribution $g(\mathcal{J}) \in \Sigma_n$ as follows: index \mathcal{J} as $\{J_1, \dots, J_m\}$ in natural order. Then, for $j \in \{1, \dots, n\}$, set

$$g(\mathcal{J})_j = \frac{i}{\sum_{k=1}^m k\#(J_k)} \quad \text{if } j \in J_i .$$

The coincidence partition of $g(\mathcal{J})$ is \mathcal{J} , *i.e.*, $\mathcal{I}_{g(\mathcal{J})} = \mathcal{J}$. The function $g : \Phi_n \rightarrow \Sigma_n$ is one-to-one since following it with the function that takes a distribution into its coincidence partition is the identity on Φ_n . As above, we build

$$G : \Phi_n \rightarrow \beta_n : \mathcal{J} \mapsto \mathbf{b}_{g(\mathcal{J})}$$

out of g .

Since, for any $A \in \mathbf{B}_n$, σ_A is completely determined by its column-index partition \mathcal{C}_A and $\mathcal{C}_A = \mathcal{C}_{f(\mathcal{C}_A)}$ we have $F(\mathcal{C}_A) = \sigma_{f(\mathcal{C}_A)} = \sigma_A$ and so F is surjective. To see that it is injective, assume $F(\mathcal{J}) = F(\mathcal{J}')$. The distribution $g(\mathcal{J})$ is in $\sigma_{f(\mathcal{J})}$ since the column-index partition of $f(\mathcal{J})$ refines (indeed, equals) the coincidence partition of $g(\mathcal{J})$. But $\sigma_{f(\mathcal{J})} = \sigma_{f(\mathcal{J}')}$, so it follows that \mathcal{J}' refines \mathcal{J} . Symmetry then tells us that $\mathcal{J} = \mathcal{J}'$.

Likewise, for any $P \in \Sigma_n$, \mathbf{b}_P is completely determined by its coincidence partition \mathcal{I}_P and $\mathcal{I}_P = \mathcal{I}_{g(\mathcal{I}_P)}$, we have $G(\mathcal{I}_P) = \mathbf{b}_{g(\mathcal{I}_P)} = \mathbf{b}_P$ and so G is also surjective. To see that G is one-to-one, let $G(\mathcal{J}) = G(\mathcal{J}')$. The ds-matrix $f(\mathcal{J})$ is in $\mathbf{b}_{g(\mathcal{J})}$ since its column-index partition refines the coincidence partition of $g(\mathcal{J})$. But $\mathbf{b}_{g(\mathcal{J})} = \mathbf{b}_{g(\mathcal{J}')}$, so it follows that \mathcal{J} refines \mathcal{J}' . By symmetry again we have $\mathcal{J} = \mathcal{J}'$.

Now take $\mathcal{J}, \mathcal{J}' \in \Phi_n$ satisfying $\mathcal{J} \leq \mathcal{J}'$. Since $F(\mathcal{J}) = \sigma_{f(\mathcal{J})}$ consists of distributions P for which $\mathcal{J} = \mathcal{C}_{f(\mathcal{J})} \leq \mathcal{I}_P$, it follows that $F(\mathcal{J}') \subseteq F(\mathcal{J})$. However, $G(\mathcal{J}) = \mathbf{b}_{g(\mathcal{J})}$ consists of ds-matrices A for which $\mathcal{C}_A \leq \mathcal{I}_{g(\mathcal{J})} = \mathcal{J}$, so $G(\mathcal{J}) \subseteq G(\mathcal{J}')$. \square

9. Partitioning Γ_n into Finitely Many Product Subsets

We defined in §§5 and 6 fibers of maps from Γ_n onto \mathbf{B}_n and Σ_n , out of which we can write Γ_n as an uncountable union of product subsets in the two ways

$$\Gamma_n = \bigcup_{A \in \mathbf{B}_n} \{A\} \times \sigma_A = \bigcup_{P \in \Sigma_n} \mathbf{b}_P \times \{P\}.$$

These descriptions are highly inefficient, however, since many points of Γ_n appear many times – the key point here being that the specific fibers σ_A and \mathbf{b}_P are determined entirely by the underlying partitions, as described in previous sections. We are in fact able to show the

Theorem 9.1. *Each of the following expressions decomposes $\Gamma_n \subset \mathbf{B}_n \times \Sigma_n$ into a finite number of disjoint product subsets:*

$$\begin{aligned} \Gamma_n &= \bigcup_{\mathcal{J} \in \Phi_n} \mathbf{b}_{g(\mathcal{J})} \times \left(\sigma_{f(\mathcal{J})} \setminus \bigcup_{\substack{\mathcal{J}' \in \Phi_n \\ \ni \mathcal{J} < \mathcal{J}'}} \sigma_{f(\mathcal{J}')} \right) \\ &= \bigcup_{\mathcal{J} \in \Phi_n} \left(\mathbf{b}_{g(\mathcal{J})} \setminus \bigcup_{\substack{\mathcal{J}' \in \Phi_n \\ \ni \mathcal{J}' < \mathcal{J}}} \mathbf{b}_{g(\mathcal{J}')} \right) \times \sigma_{f(\mathcal{J})}, \end{aligned}$$

where f and g are the same functions defined in the previous section.

Proof. Consider the function

$$F_0 : \Phi_n \rightarrow 2^{\Sigma_n} : \mathcal{J} \mapsto \{P \in \Sigma_n \mid \mathcal{I}_P = \mathcal{J}\}.$$

A distribution in Σ_n has only one coincidence partition, so F_0 carries distinct \mathcal{J} into non-intersecting subsets. Also, $P \in F_0(\mathcal{I}_P)$, so F_0 surjects onto Σ_n . Therefore, $\{F_0(\mathcal{J}) \mid \mathcal{J} \in \Phi_n\}$ is a finite decomposition of Σ_n . Consider also the function

$$G_0 : \Phi_n \rightarrow 2^{\mathbf{B}_n} : \mathcal{J} \mapsto \{A \in \mathbf{B}_n \mid \mathcal{C}_A = \mathcal{J}\}.$$

Reasoning as above, we have that $\{G_0(\mathcal{J}) \mid \mathcal{J} \in \Phi_n\}$ is a finite decomposition of \mathbf{B}_n .

By Theorem 4.5, Γ_n is the union over $\mathcal{J} \in \Phi_n$ of the disjoint sets

$$\{(A, P) \in \mathbf{B}_n \times \Sigma_n \mid \mathcal{C}_A \text{ refines } \mathcal{I}_P \text{ and } \mathcal{I}_P = \mathcal{J}\}$$

i.e., the sets

$$\{A \in \mathbf{B}_n \mid \mathcal{C}_A \text{ refines } \mathcal{J}\} \times \{P \in \Sigma_n \mid \mathcal{I}_P = \mathcal{J}\} = \mathbf{b}_{g(\mathcal{J})} \times F_0(\mathcal{J}).$$

Similarly, $\mathbf{\Gamma}_n$ is also the union over $\mathcal{J} \in \mathbf{\Phi}_n$ of the disjoint sets

$$\{(A, P) \in \mathbf{B}_n \times \mathbf{\Sigma}_n \mid \mathcal{C}_A \text{ refines } \mathcal{I}_P \text{ and } \mathcal{C}_A = \mathcal{J}\}$$

i.e., the sets

$$\{A \in \mathbf{B}_n \mid \mathcal{C}_A = \mathcal{J}\} \times \{P \in \mathbf{\Sigma}_n \mid \mathcal{J} \text{ refines } \mathcal{I}_P\} = G_0(\mathcal{J}) \times \sigma_{f(\mathcal{J})} .$$

Say that we have a pair $\mathcal{J}, \mathcal{J}' \in \mathbf{\Phi}_n$ for which $\mathcal{J} < \mathcal{J}'$. The matrix $f(\mathcal{J})$ has column-index partition which refines that of $f(\mathcal{J}')$ – after left-multiplying by a suitable permutation matrix, the blocks of $f(\mathcal{J})$ are properly contained in the blocks of $f(\mathcal{J}')$. Therefore there are more probability distributions of which $f(\mathcal{J})$ leaves invariant the entropy than $f(\mathcal{J}')$, *i.e.*, $\sigma_{f(\mathcal{J}')} \subsetneq \sigma_{f(\mathcal{J})}$. In contrast, $\mathcal{I}_{g(\mathcal{J})} = \mathcal{J} < \mathcal{J}' = \mathcal{I}_{g(\mathcal{J}')}$, so the column-index partitions of ds-matrices the averages with which do not change the entropy of $g(\mathcal{J})$ must refine those for $g(\mathcal{J}')$, which means there are fewer such matrices for $g(\mathcal{J})$ than for $g(\mathcal{J}')$. That is, $\mathbf{b}_{g(\mathcal{J})} \subsetneq \mathbf{b}_{g(\mathcal{J}')}$.

In particular, for any $\mathcal{J} \in \mathbf{\Phi}_n$

$$F_0(\mathcal{J}) = \sigma_{f(\mathcal{J})} \setminus \bigcup_{\mathcal{J}' \in \mathbf{\Phi}_n \ni \mathcal{J} < \mathcal{J}'} \sigma_{f(\mathcal{J}')}$$

and

$$G_0(\mathcal{J}) = \mathbf{b}_{g(\mathcal{J})} \setminus \bigcup_{\mathcal{J}' \in \mathbf{\Phi}_n \ni \mathcal{J}' < \mathcal{J}} \mathbf{b}_{g(\mathcal{J}')}$$

from which follow the claimed finite decompositions of $\mathbf{\Gamma}_n$ □

Example 9.2. Let us construct the two decompositions of $\mathbf{\Gamma}_2$.

There are only two partitions of $\{1, 2\}$. They, and their images under f and g , are:

$$\begin{aligned} \mathbf{\Phi}_2 : \quad & \{\{1\}, \{2\}\} < \quad \{\{1, 2\}\} \\ f : \quad & \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad , \quad \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} \\ g : \quad & (1/3, 2/3)^t \quad , \quad (1/2, 1/2)^t [= U] \end{aligned}$$

Therefore, putting the terms in the unions of Theorem 9.1 in increasing order with regards to the refinement ordering “ $<$ ” of partitions,

$$\mathbf{\Gamma}_2 = [\mathbf{S}_2 \times (\mathbf{\Sigma}_2 \setminus \{U\})] \cup [\mathbf{B}_2 \times \{U\}] = [\mathbf{S}_2 \times \mathbf{\Sigma}_2] \cup [(\mathbf{B}_2 \setminus \mathbf{S}_2) \times \{U\}] , \quad (9.1)$$

using (repeatedly) the examples in §§5 and 6. Since $\mathbf{\Gamma}_2 \subset \mathbf{B}_2 \times \mathbf{\Sigma}_2 \subset \mathbb{R}^2$, it is easy to make a picture of these decompositions: see Figure 1.

Example 9.3. Now let us consider the case of $\mathbf{\Gamma}_3$. It is convenient to use some notation as follows: write $\mathbf{S}_3 = \mathbf{G} \cup \mathbf{H}$, where \mathbf{G} denotes the cyclic subgroup of order 3 with elements

$$id = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \rho = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad \rho^2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

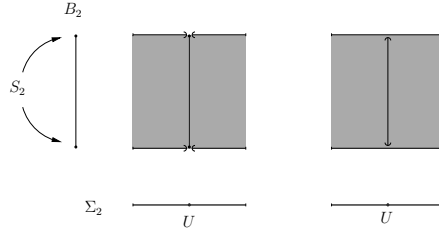


Figure 1: Decompositions of Γ_2 : the two sideways “H”s are the two versions of Γ_2 as given in (9.1), shown as sitting inside the square $\mathbf{B}_2 \times \Sigma_2$.

and $\mathbf{H} =$ denotes the set of order 2 elements of \mathbf{S}_3 , being

$$\tau_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \tau_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \text{and} \quad \tau_3 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

To construct the two decompositions of Γ_3 given by Theorem 9.1, it suffices to compute $\sigma_{f(\mathcal{J})}$ and $\mathbf{b}_{g(\mathcal{J})}$ for each of the five partitions $\mathcal{J} \in \Phi_3$. First, here are the partitions and their images under the maps f and g from §8:

$$\Phi_3 : \quad \{1\}, \{2\}, \{3\} \quad < \quad \begin{matrix} \{1\}, \{2, 3\} \\ \{1, 3\}, \{2\} \\ \{1, 2\}, \{3\} \end{matrix} \quad < \quad \{1, 2, 3\} \quad (9.2)$$

$$f : \quad id \quad , \quad \begin{matrix} 1/2(id + \tau_1) \\ 1/2(id + \tau_2) \\ 1/2(id + \tau_3) \end{matrix} \quad , \quad 1/3(id + \rho + \rho^2),$$

$$g : \quad (1/6, 1/3, 1/2)^t \quad , \quad \begin{matrix} (1/5, 2/5, 2/5)^t \\ (1/4, 1/2, 1/4)^t \\ (1/4, 1/4, 1/2)^t \end{matrix} \quad , \quad (1/3, 1/3, 1/3)^t = U,$$

According to the examples and calculations in §5, we can compute $\sigma_{f(\cdot)}$ for the above partitions:

$$\sigma_{f(\cdot)} : \quad \Sigma_3 \quad , \quad \begin{matrix} \mathbf{e}_1 \star 1/2(\mathbf{e}_2 + \mathbf{e}_3) \\ \mathbf{e}_2 \star 1/2(\mathbf{e}_1 + \mathbf{e}_3) \\ \mathbf{e}_3 \star 1/2(\mathbf{e}_1 + \mathbf{e}_2) \end{matrix} \quad , \quad \{U\}$$

(here we are suppressing the superscript 3 in these basis vectors, in the notation of §1, to avoid clutter and since Σ_3 is firmly in three dimensions) where the join (\star) of a pair of points yields the line segment connecting them. Geometrically, these values of $\sigma_{f(\cdot)}$ are: all of Σ_3 – which is an equilateral triangle; that triangle’s three medians; and the centroid of the triangle; see Figure 2.

Turning to the $\mathbf{b}_{g(\cdot)}$, we shall apply Theorem 6.2, for which we need first to record the permutations $\pi_{\mathcal{I}}$ from Proposition 4.3, again corresponding to the

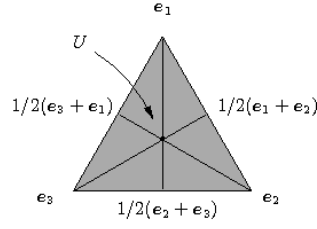


Figure 2: The probability simplex Σ_3 , with the possible values of $\sigma_{f(\cdot)}$ shown.

partitions in Φ_3 as laid out above in (9.2):

$$\pi_{\mathcal{I}} : \quad id \quad , \quad \begin{matrix} id \\ \tau_3 \\ \rho^2 \end{matrix} \quad , \quad id,$$

Note next that the partitions of 3 induced by the partitions $\mathcal{J} \in \Phi_3$ are, reading across the Φ_3 row above:

$$\mathbf{n} : \quad 1 + 1 + 1 \quad , \quad \begin{matrix} 1 + 2 \\ 2 + 1 \\ 2 + 1 \end{matrix} \quad , \quad 3$$

These partitions of 3 yield corresponding adapted permutations and ds-matrices (see Definition 2.4):

$$\mathbf{S}_3(\mathbf{n}) : \quad \{id\} \quad , \quad \begin{matrix} \{id, \tau_1\} \\ \{id, \tau_2\} \\ \{id, \tau_3\} \end{matrix} \quad , \quad \mathbf{S}_3$$

$$\mathbf{B}_3(\mathbf{n}) : \quad \{id\} \quad , \quad \begin{matrix} id \star \tau_1 \\ id \star \tau_2 \\ id \star \tau_3 \end{matrix} \quad , \quad \mathbf{B}_3 .$$

The last ingredient we need is a complete set of representatives of the coset spaces:

$$\mathbf{S}_3/\mathbf{S}_3(\mathbf{n}) : \quad \mathbf{S}_3 \quad , \quad \begin{matrix} \{id, \tau_2, \tau_3\} \\ \{id, \tau_1, \tau_3\} \\ \{id, \tau_1, \tau_2\} \end{matrix} \quad , \quad \{id\}$$

Putting all this information together into Theorem 6.2, we find:

$$\mathbf{b}_{g(\cdot)} : \quad \mathbf{S}_3 \quad , \quad \begin{matrix} (id \star \tau_1) \cup (\rho \star \tau_3) \cup (\rho^2 \star \tau_2) \\ (id \star \tau_2) \cup (\rho \star \tau_1) \cup (\rho^2 \star \tau_3) \\ (id \star \tau_3) \cup (\rho \star \tau_2) \cup (\rho^2 \star \tau_1) \end{matrix} \quad , \quad \mathbf{B}_3,$$

Picturing how these $\mathbf{b}_{g(\cdot)}$ fit together in the Birkhoff polytope \mathbf{B}_3 is quite a bit harder, since \mathbf{B}_3 is four-dimensional. It is, however, a well-studied four-dimensional polytope, whose geometric features are spelled out in [7], for example. There a similar diagram to our Figure 3 is shown, which depicts the six vertices of \mathbf{B}_3 , its 15 edges – 9 short and 6 long (here drawn with thicker lines). The two triangles consisting entirely of long sides are, in the full four-dimensional polytope, perpendicular and intersect only at the center point.

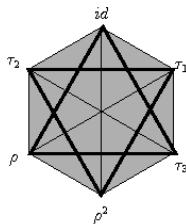


Figure 3: A two-dimensional projection of the four-dimensional Birkhoff polytope \mathbf{B}_3 .

With this four-dimensional geometry now clear, we can show in Figure 4 the subsets $\mathbf{b}_{g(\cdot)}$ inside \mathbf{B}_3 . Theorem 9.1 gives us then the following two decompo-

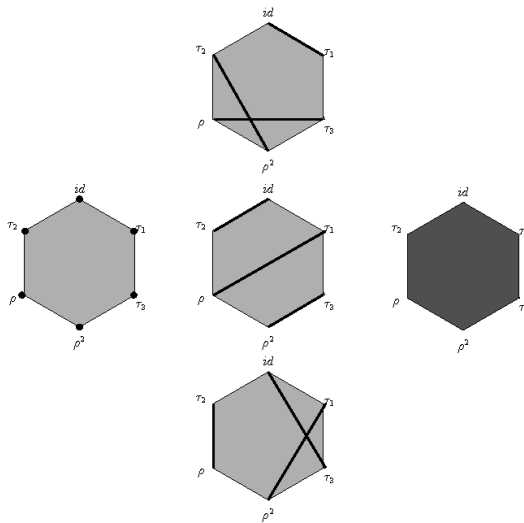


Figure 4: The subsets $\mathbf{b}_{g(\cdot)}$ of \mathbf{B}_3 , laid out corresponding to the partitions in equation (9.2).

sitions of Γ_3 :

$$\begin{aligned}
\Gamma_3 &= \mathbf{S}_3 \times \left(\Sigma_3 \setminus [(\mathbf{e}_1 \star 1/2(\mathbf{e}_2 + \mathbf{e}_3)) \cup (\mathbf{e}_2 \star 1/2(\mathbf{e}_1 + \mathbf{e}_3)) \cup (\mathbf{e}_3 \star 1/2(\mathbf{e}_1 + \mathbf{e}_2))] \right) \\
&\cup \left[(id \star \tau_1) \cup (\rho \star \tau_3) \cup (\rho^2 \star \tau_2) \right] \times \left((\mathbf{e}_1 \star 1/2(\mathbf{e}_2 + \mathbf{e}_3)) \setminus \{U\} \right) \\
&\cup \left[(id \star \tau_2) \cup (\rho \star \tau_1) \cup (\rho^2 \star \tau_3) \right] \times \left((\mathbf{e}_2 \star 1/2(\mathbf{e}_1 + \mathbf{e}_3)) \setminus \{U\} \right) \\
&\cup \left[(id \star \tau_3) \cup (\rho \star \tau_2) \cup (\rho^2 \star \tau_1) \right] \times \left((\mathbf{e}_3 \star 1/2(\mathbf{e}_1 + \mathbf{e}_2)) \setminus \{U\} \right) \\
&\cup \mathbf{B}_3 \times \{U\} \\
&= \mathbf{S}_3 \times \Sigma_3 \\
&\cup \left[((id \star \tau_1) \cup (\rho \star \tau_3) \cup (\rho^2 \star \tau_2)) \setminus \mathbf{S}_3 \right] \times \left(\mathbf{e}_1 \star 1/2(\mathbf{e}_2 + \mathbf{e}_3) \right) \\
&\cup \left[((id \star \tau_2) \cup (\rho \star \tau_1) \cup (\rho^2 \star \tau_3)) \setminus \mathbf{S}_3 \right] \times \left(\mathbf{e}_2 \star 1/2(\mathbf{e}_1 + \mathbf{e}_3) \right) \\
&\cup \left[((id \star \tau_3) \cup (\rho \star \tau_2) \cup (\rho^2 \star \tau_1)) \setminus \mathbf{S}_3 \right] \times \left(\mathbf{e}_3 \star 1/2(\mathbf{e}_1 + \mathbf{e}_2) \right) \\
&\cup \left[\mathbf{B}_3 \setminus (\{id, \rho, \rho^2\} \star \{\tau_1, \tau_2, \tau_3\}) \right] \times \{U\} ,
\end{aligned}$$

where on the last line here we use the fact that the join of those two finite point sets is the union of all line segments from a point of the first set to a point of the second set.

10. Acknowledgement

We thank David DeGeorge for several enlightening comments.

References

- [1] C. E. Shannon, A mathematical theory of communication, Bell Systems Technical Journal 27 (1948) 379–423, 623–656.
- [2] C. E. Shannon, A mathematical theory of communication, 50th anniversary edition, IEEE Symposium on Information Theory at MIT (1998).
- [3] H. Minc, Nonnegative matrices, Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, New York, 1988.
- [4] S. Eilenberg, N. Steenrod, Foundations of algebraic topology, Princeton University Press, Princeton, 1952.
- [5] G. H. Hardy, J. E. Littlewood, G. Pólya, Inequalities, Cambridge University Press, Cambridge, 1952, 2nd Edition.

- [6] C. P. Rourke, B. J. Sanderson, Introduction to Piecewise-Linear Topology, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 69, Springer-Verlag, New York, Berlin, Heidelberg, 1972.
- [7] I. Bengtsson, E. Åsa, M. Kuś, W. Tadej, K. Życzkowski, Birkhoffs polytope and unistochastic matrices, $N = 3$ and $N = 4$, Communications in Mathematical Physics 259 (2005) 307–324.