The alternating sign matrix polytope

Jessica Striker
School of Mathematics
University of Minnesota
Minneapolis, MN 55455
jessica@math.umn.edu

Submitted: Jan 4, 2008; Accepted: Mar 16, 2009; Published: Mar 25, 2009
Mathematics Subject Classification: 05C50, 52B05

Abstract

We define the alternating sign matrix polytope as the convex hull of $n \times n$ alternating sign matrices and prove its equivalent description in terms of inequalities. This is analogous to the well known result of Birkhoff and von Neumann that the convex hull of the permutation matrices equals the set of all nonnegative doubly stochastic matrices. We count the facets and vertices of the alternating sign matrix polytope and describe its projection to the permutohedron as well as give a complete characterization of its face lattice in terms of modified square ice configurations. Furthermore we prove that the dimension of any face can be easily determined from this characterization.

1 Introduction and background

The Birkhoff polytope, which we will denote as B_n, has been extensively studied and generalized. It is defined as the convex hull of the $n \times n$ permutation matrices as vectors in $\mathbb{R}^{n^2}$. Many analogous polytopes have been studied which are subsets of B_n (see e.g. [8]). In contrast, we study a polytope containing B_n. We begin with the following definitions.

Definition 1.1. Alternating sign matrices (ASMs) are square matrices with the following properties:

- entries $\in \{0, 1, -1\}$
- the entries in each row and column sum to 1
- nonzero entries in each row and column alternate in sign

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 1 & -1 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

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

Figure 1: The 3×3 ASMs

The total number of $n \times n$ alternating sign matrices is given by the expression

$$\prod_{j=0}^{n-1} \frac{(3j+1)!}{(n+j)!}. \tag{1}$$

Mills, Robbins, and Rumsey conjectured this formula [16], and then over a decade later Doron Zeilberger proved it [21]. Shortly thereafter, Kuperberg gave a shorter proof using a bijection between ASMs and configurations of the statistical physics model of square ice with domain wall boundary conditions (which are very similar to the *simple flow grids* defined later in this paper) [13]. For a detailed exposition of the conjecture and proof of the enumeration of ASMs, see [7]. See Figure 1 for the seven 3×3 ASMs.

Definition 1.2. The nth alternating sign matrix polytope, which we will denote as ASM_n, is the convex hull in $\mathbb{R}^{n^2}$ of the $n \times n$ alternating sign matrices.

From Definition 1.1 we see that permutation matrices are the alternating sign matrices whose entries are nonnegative. Thus B_n is contained in ASM_n. The connection between permutation matrices and ASMs is much deeper than simply containment. There exists a partial ordering on alternating sign matrices that is a distributive lattice. This lattice contains as a subposet the Bruhat order on the symmetric group, and in fact, it is the smallest lattice that does so (i.e. it is the MacNeille completion of the Bruhat order) [14]. Given this close relationship between permutations and ASMs it is natural to hope for theorems for ASM_n analogous to those known for B_n.

In this paper we find analogues for ASM_n of the following theorems about the Birkhoff polytope (see the discussion in [8] and [20]).

- B_n consists of the $n \times n$ nonnegative doubly stochastic matrices (square matrices with nonnegative real entries whose rows and columns sum to 1).
- The dimension of B_n is $(n-1)^2$.
- B_n has $n!$ vertices.
- B_n has n^2 facets (for $n \geq 3$) where each facet is made up of all nonnegative doubly stochastic matrices with a 0 in a specified entry.

- B_n projects onto the permutohedron.
- There exists a nice characterization of its face lattice in terms of elementary bipartite graphs [4].

As we shall see in Theorem 2.1, the row and column sums of every matrix in ASM_n must equal 1. Thus the dimension of ASM_n is $(n-1)^2$ because, just as for the Birkhoff polytope, the last entry in each row and column is determined to be precisely what is needed to make that row or column sum equal 1. In Section 3 we prove that ASM_n has $4[(n-2)^2+1]$ facets and its vertices are the alternating sign matrices. We also prove analogous theorems about the inequality description of ASM_n (Section 2), the face lattice (Section 4), and the projection to the permutohedron (Section 3). See [22] for background and terminology on polytopes.

The alternating sign matrix polytope was independently defined in [3] in which the authors also study the integer points in the rth dilate of ASM_n calling them *higher spin alternating sign matrices*.

2 The inequality description of the ASM polytope

The main theorem about the Birkhoff polytope is the theorem of Birkhoff [6] and von Neumann [19] which says that the Birkhoff polytope can be described not only as the convex hull of the permutation matrices but equivalently as the set of all nonnegative doubly stochastic matrices (real square matrices with row and column sums equaling 1 whose entries are nonnegative).

The inequality description of the alternating sign matrix polytope is similar to that of the Birkhoff polytope. It consists of the subset of doubly stochastic matrices (now allowed to have negative entries) whose partial sums in each row and column are between 0 and 1. The proof uses the idea of von Neumann's proof of the inequality description of the Birkhoff polytope [19].

Note that in [3] Behrend and Knight approach the equivalence of the convex hull definition and the inequality description of ASM_n in the opposite manner, defining the alternating sign matrix polytope in terms of inequalities and then proving that the vertices are the alternating sign matrices.

Theorem 2.1. *The convex hull of $n \times n$ alternating sign matrices consists of all $n \times n$ real matrices $X = \{x_{ij}\}$ such that:*

$$0 \leq \sum_{i=1}^{i'} x_{ij} \leq 1 \qquad \forall\ 1 \leq i' \leq n, 1 \leq j \leq n. \tag{2}$$

$$0 \leq \sum_{j=1}^{j'} x_{ij} \leq 1 \qquad \forall\ 1 \leq j' \leq n, 1 \leq i \leq n. \tag{3}$$

$$\sum_{i=1}^{n} x_{ij} = 1 \qquad \forall\ 1 \le j \le n. \tag{4}$$

$$\sum_{j=1}^{n} x_{ij} = 1 \qquad \forall\ 1 \le i \le n. \tag{5}$$

Proof. Call the subset of $\mathbb{R}^{n^2}$ given by the above inequalities $P(n)$. It is easy to check that the convex hull of the alternating sign matrices is contained in the set $P(n)$. It remains to show that any $X \in P(n)$ can be written as a convex combination of alternating sign matrices.

Let $X \in P(n)$. Let $r_{ij} = \sum_{j'=1}^{j} x_{ij'}$ and $c_{ij} = \sum_{i'=1}^{i} x_{i'j}$. Thus the r_{ij} are the row partial sums and the c_{ij} are the column partial sums. It follows from (2) and (3) that $0 \le r_{ij}, c_{ij} \le 1$ for all $1 \le i,j \le n$. Also, from (4) and (5) we see that $r_{in} = c_{nj} = 1$ for all $1 \le i,j \le n$. If we set $r_{i0} = c_{0j} = 0$ we see that every entry x_{ij} of X satisfies $x_{ij} = r_{ij} - r_{i,j-1} = c_{ij} - c_{i-1,j}$. Thus

$$r_{ij} + c_{i-1,j} = c_{ij} + r_{i,j-1}. \tag{6}$$

Using von Neumann's terminology, we call a real number α *inner* if $0 < \alpha < 1$. If X has no inner partial sums, then X is an alternating sign matrix and the required conclusion immediately holds. So we now assume X has at least one inner partial sum. We construct a circuit in X such that all of the partial sums between adjacent matrix entries in the circuit are inner. So we rewrite the matrix X with the partial sums between entries as shown below.

$$\left(\begin{array}{ccccccccc}
 & c_{01} & & c_{02} & & & c_{0,n-1} & & c_{0n} & \\
r_{10} & x_{11} & r_{11} & x_{12} & r_{12} & & x_{1,n-1} & r_{1,n-1} & x_{1n} & r_{1n} \\
 & c_{11} & & c_{12} & & \cdots & c_{1,n-1} & & c_{1n} & \\
r_{20} & x_{21} & r_{21} & x_{22} & r_{22} & & x_{2,n-1} & r_{2,n-1} & x_{2n} & r_{2n} \\
 & & \vdots & & & & & \vdots & & \\
 & c_{n-1,1} & & c_{n-1,2} & & & c_{n-1,n-1} & & c_{n-1,n} & \\
r_{n0} & x_{n1} & r_{n1} & x_{n2} & r_{n2} & \cdots & x_{n,n-1} & r_{n,n-1} & x_{nn} & r_{nn} \\
 & c_{n1} & & c_{n2} & & & c_{n,n-1} & & c_{nn} &
\end{array}\right)$$

Begin at the entry to the left or above any inner partial sum. Then there exists an adjacent inner partial sum by (6). By repeated application of (6) to each new inner partial sum, a path can then be formed by moving from entry to entry of X along inner partial sums. Since X is of finite size and all the boundary partial sums are 0 or 1 (i.e. non–inner), the path eventually reaches an entry in the same row or column as a previous entry yielding a circuit in X whose partial sums are all inner. See Figure 2 for an example of such a circuit. Using this circuit we can write X as a convex combination of two matrices in $P(n)$, each with at least one more non–inner partial sum, in the following way.

Label the corner matrix entries in the circuit alternately $(+)$ and $(-)$. Define

$$k' = \min(r_{ij}, 1 - r_{i'j'}, c_{i''j''}, 1 - c_{i''',j'''})$$

$$
\begin{pmatrix}
0 & .4 & .5 & .1 & 0 \\
.4 & -.4 & .5 & 0 & .5 \\
.6 & .4 & -.3 & -.1 & .4 \\
0 & .3 & -.3 & .9 & .1 \\
0 & .3 & .6 & .1 & 0
\end{pmatrix}
\Rightarrow
\begin{pmatrix}
 & 0 & & 0 & & 0 & & 0 & & 0 & \\
0 & 0 & 0 & .4 & \mathbf{.4} & .5 & \mathbf{.9} & .1 & 1 & 0 & 1 \\
 & 0 & & \mathbf{.4} & & .5 & & \mathbf{.1} & & 0 & \\
0 & .4 & \mathbf{.4} & -.4 & 0 & .5 & .5 & 0 & .5 & .5 & 1 \\
 & \mathbf{.4} & & 0 & & 1 & & \mathbf{.1} & & .5 & \\
0 & .6 & \mathbf{.6} & .4 & 1 & -.3 & \mathbf{.7} & -.1 & .6 & .4 & 1 \\
 & 1 & & \mathbf{.4} & & \mathbf{.7} & & 0 & & .9 & \\
0 & 0 & 0 & .3 & \mathbf{.3} & -.3 & 0 & .9 & .9 & .1 & 1 \\
 & 1 & & .7 & & .4 & & .9 & & 1 & \\
0 & 0 & 0 & .3 & .3 & .6 & .9 & .1 & 1 & 0 & 1 \\
 & 1 & & 1 & & 1 & & 1 & & 1 &
\end{pmatrix}
$$

Figure 2: A matrix in $P(n)$ along with the matrix rewritten with the partial sums between the entries and a circuit of inner partial sums shown in boldface red

where r_{ij}, $r_{i'j'}$, $c_{i''j''}$, and $c_{i''',j'''}$ are taken over respectively the row partial sums to the right of a $(-)$ corner along the circuit, the row partial sums to the right of a $(+)$ corner, the column partial sums below a $(-)$ corner, and the column partial sums below a $(+)$ corner. Subtract k' from the entries labeled $(-)$ and add k' to the entries labeled $(+)$. Subtracting and adding k' in this way preserves the row and column sums and keeps all the partial sums weakly between 0 and 1 (satisfying (2)–(5)), so the result is another matrix X' in $P(n)$ with at least one more non–inner partial sum than X.

Now give opposite labels to the corners in the circuit in X and subtract and add another constant k'' in a similar way to obtain another matrix X'' in $P(n)$ with at least one more non–inner partial sum than X. Then X is a convex combination of X' and X'', namely $X = \frac{k''}{k'+k''}X' + \frac{k'}{k'+k''}X''$. Therefore, by repeatedly applying this procedure, X can be written as a convex combination of alternating sign matrices (i.e. matrices of $P(n)$ with no inner partial sums). $\square$

3 Properties of the ASM polytope

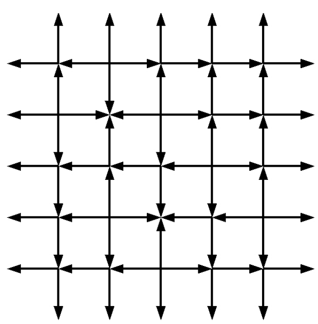
Now that we can describe the alternating sign matrix polytope in terms of inequalities, let us use this inequality description to examine some of the properties of ASM_n, namely, its facets, its vertices, and its projection to the permutohedron.

To make the proofs of the next two theorems more transparent, we introduce modified square ice configurations called *simple flow grids* which will be used more extensively in Section 4. Consider a directed graph with $n^2 + 4n$ vertices: n^2 'internal' vertices (i, j) and $4n$ 'boundary' vertices $(i, 0)$, $(0, j)$, $(i, n + 1)$, and $(n + 1, j)$ where $i, j = 1, \ldots, n$. These vertices are naturally depicted in a grid in which vertex (i, j) appears in row i and column j. Define the *complete flow grid* C_n to be the directed graph on these vertices with edge set $\{((i, j), (i, j \pm 1)), ((i, j), (i \pm 1, j)) \mid i, j = 1, \ldots, n\}$. So C_n has directed edges pointing in both directions between neighboring internal vertices in the grid, and also directed edges from internal vertices to neighboring border vertices.

Definition 3.1. A *simple flow grid* of order n is a subgraph of C_n consisting of all the vertices of C_n, and in which four edges are incident to each internal vertex: either four edges directed inward, four edges directed outward, or two horizontal edges pointing in the same direction and two vertical edges pointing in the same direction.

Proposition 3.2. *There exists an explicit bijection between simple flow grids of order n and $n \times n$ alternating sign matrices.*

Proof. Given an ASM A, we will define a corresponding directed graph $g(A)$ on the n^2 internal vertices and $4n$ boundary vertices arranged on a grid as described above. Let each entry a_{ij} of A correspond to the internal vertex (i, j) of $g(A)$. For neighboring vertices v and w in $g(A)$ let there be a directed edge from v to w if the partial sum from the border of the matrix to the entry corresponding to v in the direction pointing toward w equals 1. By the definition of alternating sign matrices, there will be exactly one directed edge between each pair of neighboring internal vertices and also a directed edge from an internal vertex to each neighboring border vertex. Vertices of $g(A)$ corresponding to 1's are sources and vertices corresponding to -1's are sinks. The directions of the rest of the edges in $g(A)$ are determined by the placement of the 1's and -1's, in that there is a series of directed edges emanating from the 1's and continuing until they reach a sink or a border vertex. Thus $g(A)$ is a simple flow grid. Also, given a simple flow grid we can easily find the corresponding ASM by replacing all the sources with 1's and all the sinks with -1's. Thus simple flow grids are in one-to-one correspondence with ASMs (see Figure 3). $\square$



$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

Figure 3: The simple flow grid—ASM correspondence

Simple flow grids are, in fact, almost the same as configurations of the six-vertex model of square ice with domain wall boundary conditions (see the discussion in [7]), the only difference being that the horizontal arrows point in the opposite direction.

Recall that for $n \geq 3$ the Birkhoff polytope has n^2 facets (faces of dimension one less than the polytope itself). ($B_2 = ASM_2$ is simply a line segment, so the number of facets equals the number of vertices which is 2.) Each facet of the Birkhoff polytope consists of all nonnegative doubly stochastic matrices with a zero in a fixed entry, that is, where one of the defining inequalities is made into an equality. The analogous theorem for ASM_n is the following.

Theorem 3.3. *ASM_n has $4[(n-2)^2+1]$ facets, for $n \geq 3$.*

Proof. Note that the $4n^2$ defining inequalities for $X \in ASM_n$ given in (2) and (3) can be restated as

$$\sum_{i'=1}^{i} x_{i'j} \geq 0 \qquad \sum_{j'=1}^{j} x_{ij'} \geq 0$$

$$\sum_{i'=i}^{n} x_{i'j} \geq 0 \qquad \sum_{j'=j}^{n} x_{ij'} \geq 0$$

for $i, j = 1, \ldots, n$. We have rewritten the statement that the row and column partial sums from the left or top must be less than or equal to 1 as the row and column partial sums from the right and bottom must be greater than or equal to 0. By counting these defining inequalities, one sees that there could be at most $4n^2$ facets, each determined by making one of the above inequalities an equality. It is left to determine how many of these equalities determine a face of dimension less than $(n-1)^2-1$.

By symmetry we can determine the number of facets coming from the inequalities $\sum_{i'=1}^{i} x_{i'j} \geq 0$ for $i, j = 1, \ldots n$ and then multiply by 4. Since the full row and column sums always equal 1, the equalities such as $\sum_{i'=1}^{n} x_{i'j} = 0$ yield the empty face ($i = n$). Also, $\sum_{i'=1}^{n-1} x_{i'j} \geq 0$ is implied by $\sum_{i'=1}^{n} x_{i'j} = 1$ and $x_{nj} \leq 1$, where $x_{nj} \leq 1$ is implied by $\sum_{j'=1}^{n} x_{nj'} = 1$ and $x_{nj'} \geq 0$ for each j'. The inequalities $x_{i'1} \geq 0$ for all i', that is, the entries in the first column are nonnegative, imply that $\sum_{i'=1}^{i} x_{i'1} \geq 0$ for $2 \leq i \leq n-1$ ($j = 1$). Thus each of these sets is a face of dimension less than $(n-1)^2-1$, and similarly for the partial sums of the last column, $\sum_{i'=1}^{i} x_{i'n} \geq 0$ for $2 \leq i \leq n-1$ ($j = n$).

So we are left with the $(n-2)^2$ inequalities $\sum_{i'=1}^{i} x_{i'j} \geq 0$ for $i = 1, \ldots, n-2$ and $j = 2, \ldots, n-1$ along with the inequality $x_{11} \geq 0$. For our symmetry argument to work, we do not include $x_{n1} \geq 0$ in our count since $x_{n1} \geq 0$ is also an inequality of the form $\sum_{j'=1}^{j} x_{ij'} \geq 0$.

Thus ASM_n has at most $4[(n-2)^2+1]$ facets, given explicitly by the $4(n-2)^2+4$ sets of all $X \in ASM_n$ which satisfy one of the following:

$$\sum_{i'=1}^{i-1} x_{i'j} = 0, \sum_{j'=1}^{j-1} x_{ij'} = 0, \sum_{i'=i+1}^{n} x_{i'j} = 0, \sum_{j'=j+1}^{n} x_{ij'} = 0,\ i, j \in \{2, \ldots, n-1\}, \tag{7}$$

$$x_{11} = 0,\ x_{1n} = 0,\ x_{n1} = 0, \text{ or } x_{nn} = 0. \tag{8}$$

They are facets (not just faces) since each equality determines exactly one more entry of the matrix, decreasing the dimension by one.

Recall that a directed edge in a simple flow grid $g(A)$ represents a location in the corresponding ASM A where the partial sum equals 1, thus a directed edge missing from $g(A)$ represents a location in A where the partial sum equals 0. Thus we can represent each of the $4(n-2)^2$ facets of (7) as subgraphs of the complete flow grid C_n from which a single directed edge has been removed: $((i\pm1,j),(i,j))$ or $((i,j\pm1),(i,j))$ with $i,j\in\{2,\ldots,n-1\}$. We can represent the facets of (8) as subgraphs of C_n from which two directed edges have been removed: $((1,1),(1,2))$ and $((1,1),(2,1))$, $((1,n),(1,n-1))$ and $((1,n),(2,n))$, $((n,1),(n-1,1))$ and $((n,1),(n,2))$, or $((n,n),(n,n-1))$ and $((n,n),(n-1,n))$.

Now given any two facets F_1 and F_2, it is easy to exhibit a pair of ASMs $\{X_1,X_2\}$ such that X_1 lies on F_1 and not on F_2. Include the directed edge(s) corresponding to F_2 but not the directed edge(s) corresponding to F_1 in $g(X_1)$, then do the opposite for X_2. Thus each of the $4[(n-2)^2+1]$ equalities gives rise to a unique facet. □

Corollary 3.4. *For $n\geq 3$, the number of facets of ASM_n on which an ASM A lies is given by $2(n-1)(n-2)+$ (number of corner 1's in A).*

Proof. Each 0 around the border of A represents one facet. Thus the number of facets corresponding to border zeros of A equals $4(n-1)-(\#$ 1's around the border of $A)$. Then there are $2(n-2)(n-3)$ facets represented by directed edges pointing in the opposite directions to the directed edges in the $(n-2)\times(n-2)$ interior array of $g(A)$. The sum of these numbers gives the above count. □

Even though ASM_n is defined as the convex hull of the ASMs, it requires some proof that each ASM is actually an extreme point of ASM_n.

Theorem 3.5. *The vertices of ASM_n are the $n\times n$ alternating sign matrices.*

Proof. Fix an $n\times n$ ASM A. In order to show that A is a vertex of ASM_n, we need to find a hyperplane with A on one side and all the other ASMs on the other side. Then since ASM_n is the convex hull of $n\times n$ ASMs, A would necessarily be a vertex.

Consider the simple flow grid corresponding to A. In any simple flow grid there are, by definition, $2n(n+1)$ directed edges, where for each entry of the corresponding ASM there is a directed edge whenever the partial sum in that direction up to that point equals 1. Since the total number of directed edges in a simple flow grid is fixed, A is the only ASM with all of those partial sums equaling 1. Thus the hyperplane where the sum of those partial sums equals $2n(n+1)-\frac{1}{2}$ will have A on one side and all the other ASMs on the other. Thus the $n\times n$ ASMs are the vertices of ASM_n. □

Another interesting property of the ASM polytope is its relationship to the permutohedron. For a vector $z=(z_1,z_2,\ldots,z_n)\in\mathbb{R}^n$, define the permutohedron P_z as the convex hull of all vectors obtained by permuting the entries of z. That is,

$$P_z=\operatorname{conv}\{(z_{\omega(1)},z_{\omega(2)},\ldots,z_{\omega(n)})\mid\omega\in S_n\}.\tag{9}$$

Also, for such a vector z, let ϕ_z be the mapping from the set of $n\times n$ real matrices to $\mathbb{R}^n$ defined by

$$\phi_z(X)=zX,\text{for any } n\times n \text{ real matrix } X.$$

It is well known, and follows immediately from the definitions, that P_z is the image of the Birkhoff polytope under the projection ϕ_z.

Proposition 3.6. *Let B_n be the Birkhoff polytope and z be a vector in $\mathbb{R}^n$. Then*

$$\phi_z(B_n) = P_z. \tag{10}$$

This result is one of many classical results about the Birkhoff polytope dating back to Hardy, Littlewood, and Pólya [11] [12]. See [17] for a nice summary of relevant results.

The next theorem states that when the same projection map is applied to ASM_n, the image is the same permutohedron whenever z is a decreasing vector. For the proof of this theorem we will need the concept of majorization [15].

Definition 3.7. Let u and v be vectors of length n. Then $u \preceq v$ (that is u is *majorized* by v) if

$$\begin{cases} \sum_{i=1}^{k} u_{[i]} \leq \sum_{i=1}^{k} v_{[i]}, & \text{for } 1 \leq k \leq n-1 \\ \sum_{i=1}^{n} u_i = \sum_{i=1}^{n} v_i \end{cases} \tag{11}$$

where the vector $(u_{[1]}, u_{[2]}, \ldots, u_{[n]})$ is obtained from u by rearranging its components so that they are in decreasing order, and similarly for v.

Theorem 3.8. *Let z be a decreasing vector in $\mathbb{R}^n$. Then*

$$\phi_z(ASM_n) = P_z. \tag{12}$$

Proof. It follows from Proposition 3.6 and $B_n \subseteq ASM_n$ that $P_z \subseteq \phi_z(ASM_n)$. Thus it only remains to be shown that $\phi_z(ASM_n) \subseteq P_z$.

Let $X = \{x_{ij}\}$ be an $n \times n$ ASM. Then there is a proposition of Rado which states that for vectors u and v of length n, $u \preceq v$ if and only if $u \in P_v$ [18]. Therefore the proof will be completed by showing $zX \preceq z$. By Definition 3.7 we need to show

$$\sum_{j=1}^{k} (zX)_{[j]} \leq \sum_{j=1}^{k} z_j, \qquad 1 \leq k \leq n-1 \tag{13}$$

$$\sum_{j=1}^{n} (zX)_j = \sum_{j=1}^{n} z_j \tag{14}$$

where the jth component $(zX)_j$ of zX is given by $\sum_{i=1}^{n} z_i x_{ij}$.

To verify (14) note that since $\sum_{j=1}^{n} x_{ij} = 1$,

$$\sum_{j=1}^{n} (zX)_j = \sum_{j=1}^{n} \sum_{i=1}^{n} z_i x_{ij} = \sum_{i=1}^{n} z_i \sum_{j=1}^{n} x_{ij} = \sum_{i=1}^{n} z_i.$$

To prove (13) we will show that $\sum_{j \in J} (zX)_j \leq \sum_{j=1}^{|J|} z_j$ given any $J \subseteq \{1, \ldots, n\}$, so that in particular $\sum_{j=1}^{|J|} (zX)_{[j]} \leq \sum_{j=1}^{|J|} z_j$.

We will need to verify the following:

$$\sum_{i=1}^{m} \sum_{j\in J} x_{ij} \leq \min(m, |J|) \qquad \forall\, m \in \{1, \ldots, n\}. \tag{15}$$

$$\sum_{i=1}^{n} \sum_{j\in J} x_{ij} = |J|. \tag{16}$$

To prove (15) note that

$$\sum_{i=1}^{m} \sum_{j\in J} x_{ij} = \sum_{j\in J} \sum_{i=1}^{m} x_{ij} \leq |J|$$

since $\sum_{i=1}^{m} x_{ij} \leq 1$. But also, since $\sum_{i=1}^{m} x_{ij} \geq 0$ and $\sum_{j=1}^{n} x_{ij} = 1$ we have that

$$\sum_{j\in J} \sum_{i=1}^{m} x_{ij} \leq \sum_{j=1}^{n} \sum_{i=1}^{m} x_{ij} = \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} = m.$$

To prove (16) observe,

$$\sum_{i=1}^{n} \sum_{j\in J} x_{ij} = \sum_{j\in J} \sum_{i=1}^{n} x_{ij} = \sum_{j\in J} 1 = |J|$$

since the columns of X sum to 1. Therefore using (15) and (16) we see that

$$\begin{aligned}
\sum_{j\in J} (zX)_j &= \sum_{j\in J} \sum_{i=1}^{n} z_i x_{ij} = \sum_{i=1}^{n} z_i \sum_{j\in J} x_{ij} = \sum_{k=1}^{n-1} (z_k - z_{k+1}) \sum_{i=1}^{k} \sum_{j\in J} x_{ij} + z_n \sum_{i=1}^{n} \sum_{j\in J} x_{ij} \\
&= \sum_{k=1}^{n-1} (z_k - z_{k+1}) \sum_{i=1}^{k} \sum_{j\in J} x_{ij} + z_n |J| && \text{by (16)} \\
&= \sum_{k=1}^{|J|-1} (z_k - z_{k+1}) \sum_{i=1}^{k} \sum_{j\in J} x_{ij} + \sum_{k=|J|}^{n-1} (z_k - z_{k+1}) \sum_{i=1}^{k} \sum_{j\in J} x_{ij} + z_n |J| \\
&\leq \sum_{k=1}^{|J|-1} (z_k - z_{k+1}) k + \sum_{k=|J|}^{n-1} (z_k - z_{k+1}) |J| + z_n |J| && \text{by (15) and since } z_k \geq z_{k+1} \\
&= \sum_{k=1}^{|J|-1} z_k - z_{|J|}(|J| - 1) + (z_{|J|} - z_n)|J| + z_n |J| \\
&= \sum_{k=1}^{|J|} z_k.
\end{aligned}$$

Thus $zX \preceq z$ and so zX is contained in the convex hull of the permutations of z. Therefore $\phi_z(ASM_n) = P_z$. □

4 The face lattice of the ASM polytope

Another nice result about the Birkhoff polytope is the structure of its face lattice [4]. Associate to each permutation matrix X a bipartite graph with vertices $u_1, u_2, \ldots, u_n$ and $v_1, v_2, \ldots, v_n$ where there is an edge connecting u_i and v_j if and only if there is a 1 in the (i, j) position of X. Such a graph will be a perfect matching on the complete bipartite graph $K_{n,n}$. A graph G is called elementary if every edge is a member of some perfect matching of G.

Theorem 4.1 (Billera–Sarangarajan)**.** *The face lattice of the Birkhoff polytope is isomorphic to the lattice of elementary subgraphs of $K_{n,n}$ ordered by inclusion.*

This lattice structure was first identified by Billera and Sarangarajan in [4] and [5], but the set of faces itself was first characterized and studied extensively by Brualdi and Gibson in [9] and [10] using certain 0-1 matrices which correspond trivially to elementary subgraphs of $K_{n,n}$. Other relevant results were also obtained by Balinski and Russakoff in [1] and [2].

A similar statement can be made about the face lattice of the ASM polytope using simple flow grids (see Definition 3.1) in place of perfect matchings, the complete flow grid C_n instead of the complete bipartite graph $K_{n,n}$, and *elementary flow grids* in place of elementary graphs.

Definition 4.2. An *elementary flow grid G* is a subgraph of the complete flow grid C_n such that the edge set of G is the union of the edge sets of simple flow grids.

Now for any face F of ASM_n define the grid corresponding to the face, $g(F)$, to be the union over all the vertices of F of the simple flow grids corresponding to the vertices. That is,

$$g(F) = \bigcup_{vertices\ A \in F} g(A).$$

Thus $g(F)$ is an elementary flow grid since its edge set is the union of the edge sets of simple flow grids.

Now we wish to define the converse, that is, given an elementary flow grid G we would like to know the corresponding face $f(G)$ of ASM_n. Define $f(G)$ to be the convex hull of the vertices of ASM_n whose corresponding simple flow grids are contained in the elementary flow grid G. So let

$$f(G) = \operatorname{conv}\{\text{vertices } A \in ASM_n \mid g(A) \subseteq G\}.$$

Recall that we can represent each of the facets of ASM_n either as subgraphs of the complete flow grid C_n from which one of the directed edges in the set $S = \{((i \pm 1, j), (i, j)), ((i, j \pm 1), (i, j)) \mid i, j \in \{2, \ldots, n-1\}\}$ has been removed or from which one of the pairs of directed edges in the set $T = \{\{((1,1)(1,2)), ((1,1),(2,1))\}, \{((1,n),(1,n-1)), ((1,n),(2,n))\}, \{((n,1),(n-1,1)), ((n,1),(n,2))\}, \{((n,n),(n,n-1)), ((n,n),(n-1,n))\}\}$ has been removed.

Thus each of the directed edges in S and the first of each pair of directed edges in T that are not in G represent facets that contain $f(G)$. Let the collection of these directed edges be called $\{e_1, e_2, \ldots, e_k\}$ and their corresponding facets $\{F_1, F_2, \ldots, F_k\}$. Let $I = \bigcap_{j=1}^{k} F_j$ be the intersection of these facets. Thus I is a face of ASM_n and $f(G) \subseteq I$.

We wish to show that $f(G)$ equals I. So suppose $f(G) \subsetneq I$. Then since I is a face of ASM_n and $f(G)$ is defined as the convex hull of vertices of ASM_n there exists an additional vertex $B \in I$ of ASM_n such that $B \notin f(G)$. But $g(B)$ must be missing the directed edges $e_1, e_2, \ldots, e_k$ since $B \in I$, thus all the directed edges of $g(B)$ must be in G. Therefore $g(B) \subseteq G$ so that $B \in f(G)$ which is a contradiction. So $f(G) = I$. Thus $f(G)$ is a face of ASM_n since it is the intersection of faces of ASM_n.

Given facets F_1 and F_2 associated with two distinct equalities in 7 and 8, it is easy to find an ASM A in $F_1 - F_2$. A must simply satisfy the equality associated with F_1, and the left–hand side of the equality associated with F_2 must equal 1 in A.

Thus elementary flow grids are in bijection with the faces of ASM_n (if we also regard the empty grid as an elementary flow grid). Elementary flow grids can be made into a lattice by inclusion, where the join is the union of the edge sets and the meet is the largest elementary flow grid made up of the directed edges from the intersection of the edges sets.

This discussion yields the following theorem:

Theorem 4.3. *The face lattice of ASM_n is isomorphic to the lattice of all $n \times n$ elementary flow grids ordered by inclusion.*

The dimension of any face of ASM_n can be determined by looking at $g(F)$ as in the following theorem. The characterization of edges of ASM_n is analogous to the result for the Birkhoff polytope which states that the graphs representing edges of B_n are the elementary subgraphs of $K_{n,n}$ which have exactly one cycle [2] [4] [10].

Definition 4.4. A *doubly directed region* of an elementary flow grid G is a connected collection of cells in G completely bounded by double directed edges but containing no double directed edges in the interior (see Figure 4). Let $\alpha(G)$ denote the number of doubly directed regions in G.

Theorem 4.5. *The dimension of a face F of ASM_n is the number of doubly directed regions in the corresponding elementary flow grid $g(F)$. In particular, the edges of ASM_n are represented by elementary flow grids containing exactly one cycle of double directed edges.*

Proof. We proceed by induction on the dimension of the face of ASM_n. The simple flow grid corresponding to any ASM A has no double directed edges, thus $\alpha(g(A)) = 0$. Now suppose for every m–dimensional face of ASM_n, the number of doubly directed regions of the elementary flow grid corresponding to the face equals m. Let F be an $(m+1)$–dimensional face of ASM_n and F' an m–dimensional subface of F. We assume $\alpha(g(F')) = m$ and wish to show that $\alpha(g(F)) = m + 1$.

Now $g(F)$ is the elementary flow grid whose edge set is the union of the edge sets of $g(F')$ and $g(A)$ over all ASMs A in $F - F'$. Every vertex in a simple flow grid must

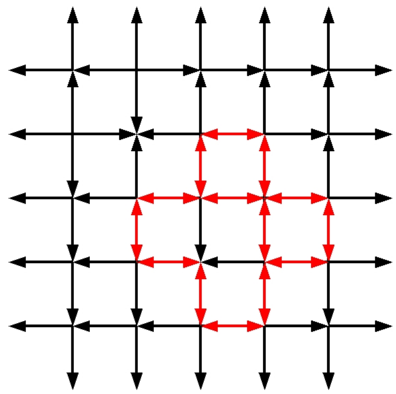


Figure 4: An elementary flow grid containing 3 doubly directed regions which corresponds by Theorem 4.5 to a 3-dimensional face of ASM_5

have even indegree and even outdegree. Therefore, if we wish to obtain $g(A)$ from $g(A')$, where A' is an ASM in F', by reversing some directed edges, the number of directed edges reversed at each vertex must be even. Thus taking the union of the directed edges of $g(A')$ with the directed edges of $g(A)$ forms one or more circuits of double directed edges, where at least one of the double directed edges is not in $g(F')$. Therefore $g(F)$ has at least one more doubly directed region than $g(F')$, so $\alpha(g(F)) \geq m+1$. Then since $g(ASM_n)$ equals the complete flow grid C_n, we have that $\alpha(g(ASM_n)) = \alpha(C_n) = (n-1)^2 = \dim(ASM_n)$. Therefore moving up the face lattice one rank increases the number of doubly directed regions by exactly one, so $\alpha(g(F)) = m + 1$. □

See Figure 5 for the elementary flow grid representing the edge in ASM_5 between

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \text{ and } \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Acknowledgments

This work is based on research which is a part of the author's doctoral thesis at the University of Minnesota under the direction of Dennis Stanton. The author would like to thank Professor Stanton for the many helpful discussions and encouragement. The author is also very grateful to the referee for the detailed report which helped to improve the clarity of the manuscript.

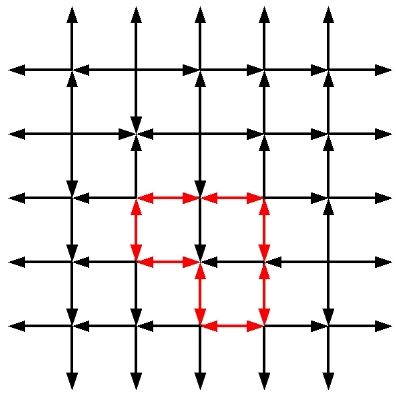


Figure 5: The elementary flow grid representing an edge in ASM_5

References

[1] M. L. Balinski and A. Russakoff, Some properties of the assignment polytope, 1972. Math. Programming, **3**, 257–258.

[2] M. L. Balinski and A. Russakoff, On the assignment polytope, 1974. SIAM Rev., **16**, 516–525.

[3] R. E. Behrend and V. A. Knight, Higher spin alternating sign matrices, 2007. Electron. J. Combin., **14**, no. 1.

[4] L. J. Billera and A. Sarangarajan, All 0-1 polytopes are traveling salesman polytopes, 1996. Combinatorica, **16**, no. 2, 175–188.

[5] L. J. Billera and A. Sarangarajan, The combinatorics of permutation polytopes, 1996. DIMACS Ser. Discrete Mat. Theoret. Comput. Sci., Vol. 24, Amer. Math. Soc., 1–23.

[6] G. Birkhoff, Three observations on linear algebra, 1946. Univ. Nac. Tucumán. Revista A., **5**, 147–151.

[7] D. M. Bressoud, Proofs and confirmations; The story of the alternating sign matrix conjecture, MAA Spectrum, Mathematical Association of America, Washington, DC, 1999.

[8] R. A. Brualdi, Combinatorial Matrix Classes, Cambridge University Press, 2006.

[9] R. A. Brualdi and P. M. Gibson, The assignment polytope, 1976. Math. Programming, **11**, 97–101.

[10] R. A. Brualdi and P. M. Gibson, Convex polyhedra of doubly stochastic matrices. I. Applications of the permanent function, 1977. J. Combin. Theory Ser. A, **22**, 194–230.

[11] G. H. Hardy, J. E. Littlewood, and G. Pólya, Some simple inequalities satisfied by convex functions, 1929. Messenger Math. **58**, 145–152.

[12] G. H. Hardy, J. E. Littlewood, and G. Pólya, Inequalities (Second Edition), Cambridge University Press, 1952.

[13] G. Kuperberg, Another proof of the ASM conjecture, 1996. Inter. Math. Not., **3**, 139-150.

[14] A. Lascoux and M. Schützenberger, Treillis et bases des groupes de Coxeter, 1996. Electron. J. Combin., **3**, no. 2.

[15] A. W. Marshall and I. Olkin, Inequalities: theory of majorization and its applications, Mathematics in Science and Engineering, **143**, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1979.

[16] W. H. Mills, D. P. Robbins, and H. Rumsey, Jr., Alternating sign matrices and descending plane partitions, 1983. J. Combin. Theory Ser. A, **34**, no. 3, 340–359.

[17] L. Mirsky, Results and problems in the theory of doubly–stochastic matrices, 1963. Z. Wahrsch. Verw. Gebiete, **1**, 319–334.

[18] R. Rado, An inequality, 1952. J. London Math. Soc., **27**, 1–6.

[19] J. von Neumann, A certain zero–sum two person game equivalent to the optimal assignment problem, Contributions to the Theory of Games Vol. 2, 1953. Annals of Mathematics Studies No. 28, Princeton University Press, 5–12.

[20] V. A. Yemelichev, M. M. Kovalev, and M. K. Kravtsov, Polytopes, graphs, and optimisation, Cambridge University Press, 1984.

[21] D. Zeilberger, Proof of the alternating sign matrix conjecture, 1996. Electron. J. Combin. **3**, no. 2.

[22] G. M. Ziegler, Lectures on Polytopes, Springer-Verlag, New York, 1995.