Alternating sign matrices, extensions and related cones

2 authors:

Some of the authors of this publication are also working on these related projects:

Project   Combinatorial matrix theory; alternating sign matrices View project

Alternating sign matrices, extensions and related cones

Richard A. Brualdi[*]
Geir Dahl[†]

December 1, 2016

Abstract

An *alternating sign matrix*, or ASM, is a $(0, \pm1)$-matrix where the nonzero entries in each row and column alternate in sign, and where each row and column sum is 1. We study the convex cone generated by ASMs of order n, called the ASM cone, as well as several related cones and polytopes. Some decomposition results are shown, and we find a minimal Hilbert basis of the ASM cone. The notion of (±1)-doubly stochastic matrices and a generalization of ASMs are introduced and various properties are shown. For instance, we give a new short proof of the linear characterization of the ASM polytope, in fact for a more general polytope. Finally, we investigate faces of the ASM polytope, in particular edges associated with permutation matrices.

Key words. Alternating sign matrix, doubly stochastic matrix, decomposition, cones.
AMS subject classifications. 05B20, 15B48.

1 Introduction

An *alternating sign matrix* ([20]), or ASM-matrix, is a $(0, \pm1)$-matrix where the nonzero entries in each row and column alternate in sign, and where each

---
[*]Department of Mathematics, University of Wisconsin, Madison, WI 53706, USA. `brualdi@math.wisc.edu`
[†]Department of Mathematics, University of Oslo, Norway. `geird@math.uio.no`

row and column sum is 1. Let $\mathcal{P}_n$ be the set of $n \times n$ permutation matrices and let $\mathcal{A}_n$ be the set of $n \times n$ alternating sign matrices. The convex hull of $\mathcal{P}_n$ is the polytope Ω_n of doubly stochastic matrices with $\dim \Omega_n = (n-1)^2$ whose set of extreme points is $\mathcal{P}_n$ ([4, 21]). Let Λ_n be the convex hull of $\mathcal{A}_n$. The next result was shown in [3, 20], but we give a short argument for this fact.

Theorem 1 *The set of extreme points of Λ_n is $\mathcal{A}_n$.*

Proof.   Let $x \in \{0, \pm 1\}^n$ be an alternating sign vector (a row or column in an ASM) which is a convex (positive) combination of other alternating sign vectors $x^{(k)}$s. In any position where x has a 1, each $x^{(k)}$ also has a 1, and the same holds for -1. Moreover, there must be a position j where x has 0 and some $x^{(k)}$ has 1, and therefore some $x^{(p)}$ has a -1. Choosing this position j closest possible to a position where x, and therefore $x^{(k)}$ and $x^{(p)}$, has a 1, we get a contradiction to the alternating property. This argument implies that no ASM can be a convex combination of ASMs different from it.   □

It is also easy to see that $\dim \Lambda_n = (n-1)^2$. Note that we have $\mathcal{P}_n \subseteq \mathcal{A}_n$, and $\Omega_n \subseteq \Lambda_n$. It is an elementary fact that the maximum number of linearly independent permutation matrices in $\mathcal{P}_n$ equals $(n-1)^2 + 1$.

ASMs were defined by Mills, Robbins, and Ramsey who conjectured [16] that the number of $n \times n$ ASMs equals

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

This formula was proved in 1996 by Zeilberger [22], and independently by Kuperberg [14] who showed a bijection between ASMs and configurations of statistical physics known as square ice. An exposition of the fascinating origins of ASMs and a history of this conjecture including its connection with other combinatorial objects can be found in the book [9]. Spencer [19] showed that the above formula is asymptotic to

$$\left(\frac{3\sqrt{3}}{4} \right)^{n^2}.$$

Let X be a set in a vector space. We let $\mathrm{Cone}(X)$ denote the convex cone consisting of all nonnegative linear combinations of elements in X. A convex

2

cone is called *rational* if it is generated by a set of rational vectors. X is called a *Hilbert basis* provided every integer vector in $\mathrm{Cone}(X)$ is expressible as a nonnegative *integer* linear combination of the vectors in X. It is proved in [13] that every pointed[1] rational cone is generated by a *unique* minimal Hilbert basis. We refer to this unique minimal Hilbert basis of a pointed rational cone $\mathrm{Cone}(X)$ as the *H-basis* of $\mathrm{Cone}(X)$. It is a classical result of Carathéodory that in a rational cone $\mathrm{Cone}(X)$ of dimension d generated by a finite set $X = \{x_1, x_2, \ldots, x_k\}$ of rational vectors, every vector in the cone can be expressed as a nonnegative linear combination of d vectors in X. The integer analogue of Carathéodory's theorem asks: What is the smallest integer $h(d)$ such that every integer vector in a pointed rational cone C of dimension d can be expressed as a nonnegative integer linear combination of $h(d)$ vectors in the H-basis of C? It is known [11, 10] that

$$\lfloor (7/6)d \rfloor \le h(d) \le 2d - 1. \tag{1}$$

Let $\mathrm{Cone}(\mathcal{P}_n)$ and $\mathrm{Cone}(\mathcal{A}_n)$ be, respectively, the cones generated by $\mathcal{P}_n$ and $\mathcal{A}_n$. Then $\mathrm{Cone}(\mathcal{P}_n) \subseteq \mathrm{Cone}(\mathcal{A}_n)$, and both these cones have dimension $(n-1)^2 + 1$. Since they are generated by rational matrices, they are *rational cones*. Moreover, both cones are pointed since, if A is an ASM, then $-A$ is not.

The paper is organized as follows. In Section 2 we introduce some cones containing $\mathrm{Cone}(\mathcal{A}_n)$, and establish a so-called signed decomposition of an ASM in terms of permutation matrices. Moreover, some properties of the pattern of matrices in the ASM polytope Λ_n are found. The set of $n \times n$ matrices having all line sums 1 and a related polytope are studied in Section 3, while Section 4 characterizes the sum of signed permutation matrices, and shows related results. We introduce, in Section 5, a notion generalizing ASMs based on a majorization concept. In particular, the notion of r-ASM is studied. We obtain a polyhedral result characterizing the convex hull of such r-ASMs, and obtain a decomposition result for r-ASMs as sums of ASMs. From this we determine the H-basis of $\mathrm{Cone}(\mathcal{A}_n)$. The final section focuses on faces of the ASM polytope, in particular on edges of that polytope containing permutation matrices. Characterizations of such edges are found. Finally, we introduce and prove some results for a certain combinatorial distance of ASMs to the set of permutation matrices.

---

[1] A cone is pointed provided it contains no line.

Notation: Let $M_{m,n}$ (resp. M_n) denote the vector space of all real $m \times n$ (resp. $n \times n$) matrices. The $n \times n$ all ones matrix is denoted by J_n (or just J). A vector $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ is called *nonnegative* if $x_i \geq 0$ ($i \leq n$). The *support* of a matrix $A = [a_{ij}]$, denoted by supp(A), is the set of positions (i,j) with $a_{ij} \neq 0$. Let the all ones vector (of suitable dimension) be denoted by e. For a set S (in a vector space) Span(S) (resp. Cone(S)) denotes the subspace spanned by S (resp. the (convex) cone spanned by S). Similarly, Conv(S) denotes the convex hull of S.

2 The ASM cone and related cones and polytopes

In this section, after discussing a classical decomposition involving $n \times n$ permutation matrices and the cone they generate, we derive an analogous decomposition result for $n \times n$ ASMs, and discuss the cone and polytope they generate.

The following classical decomposition result ([6, 7]) will be used later.

Theorem 2 *Let A be an $n \times n$ nonnegative integral matrix with equal row and column sums, say equal to k. Then A may be decomposed as the sum of k permutation matrices.*

For a matrix $A = [a_{ij}] \in M_n$, let $R_i(A) = \sum_{j=1}^{n} a_{ij}$ and $S_j(A) = \sum_{i=1}^{n} a_{ij}$ be the ith row sum and jth column sum, respectively. Define

$$\mathcal{L}_n = \{A \in M_n : R_1(A) = \cdots = R_n(A) = S_1(A) = \cdots = S_n(A)\}$$

and

$$\mathcal{K}_n = \{A \in \mathcal{L}_n : A \geq O\}.$$

$\mathcal{L}_n$ consists of the $n \times n$ matrices with all row and column sums equal, and $\mathcal{K}_n$ is the subset of nonnegative such matrices.

Lemma 3 (*i*) $\mathcal{L}_n$ *is a subspace of M_n and*

$$\mathcal{L}_n = \text{Span}(\mathcal{P}_n) = \text{Cone}(\mathcal{P}_n \cup (-\mathcal{P}_n)) = \text{Cone}(\mathcal{P}_n \cup \{-J_n\}).$$

Each integral matrix in $\mathcal{L}_n$ may be written as an integral nonnegative linear combination of matrices in $\mathcal{P}_n \cup (-\mathcal{P}_n)$ and in $\mathcal{P}_n \cup \{-J_n\}$.

(ii) $\mathcal{K}_n$ *is a pointed rational cone and*

$$\mathcal{K}_n = \mathrm{Cone}(\mathcal{P}_n).$$

Each integral matrix in $\mathcal{K}_n$ may be written as an integral nonnegative linear combination of matrices in $\mathcal{P}_n$.

Proof. Statement (ii) follows from Theorem 2 (which again is shown in [6]). $\mathcal{L}_n$ is clearly a subspace and it contains each permutation matrix and $-J_n$, so $\mathrm{Cone}(\mathcal{P}_n \cup (-\mathcal{P}_n)) \subseteq \mathcal{L}_n$. Let $A \in \mathcal{L}_n$. Then $A - \alpha J_n \geq O$ where α is the minimum of all entries in A and 0. By (ii), $A - \alpha J_n$ is a nonnegative linear combination of permutation matrices. Thus $A \in \mathrm{Cone}(\mathcal{P}_n \cup \{-J_n\})$. Also observe that J_n may be decomposed as a sum of n permutation matrices (by Theorem 2). Then property (i) follows from these comments. $\square$

Example 1 The following shows a matrix A in $\mathcal{L}_3$ and a decomposition of A:

$$A = \begin{bmatrix} -2 & -1 & -1 \\ 0 & -2 & -2 \\ -2 & -1 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} + (-2) \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

$\square$

Theorem 4 $\mathcal{P}_n$ *is the H-basis of* $\mathrm{Cone}(\mathcal{P}_n)$.

Proof. By Lemma 3 every integral matrix in $\mathrm{Cone}(\mathcal{P}_n)$ may be written as an integral nonnegative linear combination of matrices in $\mathcal{P}_n$. Therefore, $\mathcal{P}_n$ is a Hilbert basis of $\mathrm{Cone}(\mathcal{P}_n)$. Moreover, $\mathcal{P}_n$ is a *minimal* Hilbert basis of this cone, since no permutation matrix can be expressed as a nonnegative linear combination of the permutation matrices different from it. Actually, such a linear combination of at least two matrices will have row sum at least 2 in the first row. As $\mathrm{Cone}(\mathcal{P}_n)$ is pointed, the minimal Hilbert basis is unique, and it must therefore be $\mathcal{P}_n$. $\square$

A natural question is to find to the H-basis of $\mathrm{Cone}(\mathcal{A}_n)$. We study this question in Section 5.

We now turn to ASMs and prove that any ASM may be "decomposed" in terms of permutation matrices.

Corollary 5 *Let $A \in \mathcal{A}_n$. Then there exist distinct permutation matrices $P_i \in \mathcal{P}_n$ and $\lambda_i \in \{-1, 1\}$ $(i \leq N)$ for some N such that*

$$A = \lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_N P_N. \tag{2}$$

Here N may be chosen so that $N \leq 2n + 1$.

Proof. The first part follows from Lemma 3, but to find N, we repeat the construction. Let $B = A + J_n$ where J_n is the $n \times n$ all ones matrix. Then B is a nonnegative integral matrix with all row and column sums being $n + 1$. By Theorem 2, B may be written as a sum of $n+1$ permutation matrices, say $B = \sum_{i=1}^{n+1} P_i$. Moreover, applying Theorem 2 to J_n, there are n permutation matrices P_i $(i = n + 2, \ldots, 2n + 1)$ such that $J_n = \sum_{i=n+2}^{2n+1} P_i$. This gives

$$A = B - J_n = \sum_{i=1}^{n+1} P_i - \sum_{i=n+2}^{2n+1} P_i$$

as desired. □

Example 2 Let

$$A + J_4 = \begin{bmatrix} & 1 & & \\ 1 & -1 & 1 & \\ & 1 & -1 & 1 \\ & & 1 & \end{bmatrix} + J_4 = \begin{bmatrix} 1 & 2 & 1 & 1 \\ 2 & 0 & 2 & 1 \\ 1 & 2 & 0 & 2 \\ 1 & 1 & 2 & 1 \end{bmatrix}$$

$$= 2P_1 + P_2 + P_3 + P_4$$

$$= 2\begin{bmatrix} & 1 & & \\ 1 & & & \\ & & & 1 \\ & & 1 & \end{bmatrix} + \begin{bmatrix} & & & 1 \\ & & 1 & \\ & 1 & & \\ 1 & & & \end{bmatrix} + \begin{bmatrix} & & 1 & \\ & & & 1 \\ 1 & & & \\ & 1 & & \end{bmatrix} + \begin{bmatrix} 1 & & & \\ & & 1 & \\ & 1 & & \\ & & & 1 \end{bmatrix}.$$

Now $J_4 = P_1 + P_2 + P_3 + I$ so that

$$A = P_1 + P_4 - I.$$

□

We call the representation of A in this theorem a *signed decomposition*. Note that a matrix having a signed decomposition might not be an ASM. In fact, Corollary 5 and its proof are valid for any $(0, \pm 1)$-matrix in which each row and column sums equals 1. From our results so far we have

$$\mathcal{K}_n = \mathrm{Cone}(\mathcal{P}_n) \subseteq \mathrm{Cone}(\mathcal{A}_n) \subseteq \mathrm{Cone}(\mathcal{P}_n \cup \{-J_n\}) = \mathcal{L}_n.$$

The exact value of N in Corollary 5 may be smaller than $2n+1$, as Example 2 shows. A smaller value of N, for a given matrix A, may be obtained as follows. Let $A \in \mathcal{A}_n$. Let $k(A)$ be the maximum number of (-1)'s in a row or column of A. By the alternating property of an ASM, $k(A) + 1$ is the maximum number of $+1$'s in a row or column of A. Then $k(A)$ equals the smallest number k of permutation matrices $P_1, P_2, \ldots, P_k$ such that

$$\sum_{i=1}^{k(A)} P_i + A \geq O. \tag{3}$$

This follows from the fact (see [7]) that a nonnegative integral matrix with maximum line sum k, may be decomposed into a sum of k subpermutation matrices.

Corollary 6 *For any given ASM A, the smallest N in a signed decomposition of A is $N = 2k(A) + 1$, i.e., the maximum number of nonzeros in a row or column of A.*

Proof. We use the same technique as in the proof of Corollary 5. Now, let $B = \sum_{i=1}^{k(A)} P_i + A \geq O$, and argue as above. Note that B has all row and columns sums equal to $k(A) + 1$, so B may be decomposed as a sum of $k(A) + 1$ permutation matrices. That gives the desired signed decomposition with $N = 2k(A) + 1$ permutation matrices. Since A is an ASM, N also equals the maximum number of nonzeros in a row or column of A.

Moreover, in A there is a row or column with $N = 2k(A) + 1$ nonzeros. Since a permutation matrix, or its negative, contains exactly one nonzero in any row or column, there must be at least N signed permutation matrices in a decomposition of A. $\qquad\square$

In Example 2 we have such a decomposition with $k_A = 1$ (only one permutation matrix is needed to cover the single -1).

We now turn to the *ASM polytope* Λ_n defined to be the convex hull of the $n \times n$ ASMs. The following result was shown in [3, 20].

7

Theorem 7 ([3, 20]) Λ_n *equals the set of* $n \times n$ *matrices* $A = [a_{ij}]$ *with all row and column sums being* 1 *and satisfying*

$$\sum_{j=1}^{q} a_{ij} \geq 0, \quad \sum_{j=q+1}^{n} a_{ij} \geq 0 \quad (1 \leq q \leq n, \ 1 \leq i \leq n),$$
$$\sum_{i=1}^{p} a_{ij} \geq 0, \quad \sum_{i=p+1}^{n} a_{ij} \geq 0 \quad (1 \leq p \leq n, \ 1 \leq j \leq n). \tag{4}$$

This implies the following result for the ASM cone.

Corollary 8 $\text{Cone}(\mathcal{A}_n)$ *equals the set of* $n \times n$ *matrices* $A = [a_{ij}]$ *satisfying* $A \in \mathcal{L}_n$ *and the linear inequalities* (4).

Proof. Each ASM satisfies the linear inequalities in (4), due to the alternating structure in each line, and so does every nonnegative linear combination of these.

Conversely, let $A = [a_{ij}] \in \mathcal{L}_n$ be a nonzero matrix that satisfies (4). Let α be the common row and column sum in A. If $\alpha = 0$, then the inequalities (and induction) shows that $A = O$. So assume $\alpha > 0$. Define $B = (1/\alpha)A$. Then all row and column sums of B are 1, and it is easy to see that B satisfies the inequalities defining the ASM polytope (4), so, by Theorem 7, B is a convex combination of ASMs. But then $A = \alpha B$ is nonnegative combination of ASMs. □

We now consider the patterns of matrices in the ASM polytope Λ_n. Patterns of ASMs were also considered in [8]. A matrix in Λ_n will be called a *fractional ASM*. (The term "fractional" here is used as in the area of polyhedral combinatorics, e.g., fractional matching.)

Let $A = [a_{ij}]$ be a fractional ASM. Let the pattern of A be the $n \times n$ $(0,1)$-matrix $P(A) = [p_{ij}]$ where $p_{ij} = 0$ if $a_{ij} = 0$ and $p_{ij} = 1$ if $a_{ij} \neq 0$.

Note that the polytope Ω_n of $n \times n$ doubly stochastic matrices is contained in Λ_n. We want to find properties of patterns of fractional ASMs. Note that a fractional ASM may contain an $r \times s$ zero submatrix with $r + s = n + 1$; this is not true for a doubly stochastic matrix (due to König's minmax theorem). For example, with $n = 3$, the ASM

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

has a 2×2 zero submatrix with $2 + 2 = 4$.

Theorem 9 *Let A be an $n \times n$ fractional ASM. Then A does not contain an $r \times s$ zero submatrix with consecutive rows or consecutive columns where $r + s = n + 1$.*

Proof. Suppose to the contrary that A contains an $r \times s$ zero submatrix with $r + s = n + 1$ and, say, consecutive rows. Then the $r \times n$ submatrix A' determined by these r rows $k+1, k+2, \ldots, k+r$ have a total sum of r. The matrix A' has s columns of zeros. The other $n - s = r - 1$ columns of A' each have sum between -1 and 1. This is because for each such column, the sum in rows $k+1, k+2, \ldots, k+r$ is between 0 and 1, since the sum in rows 1 to k is at least zero, the sum in rows $k + r + 1, \ldots, n$ is at least zero, and the total sum in the column is 1. It follows that A' has s columns summing to zero and $n - s = r - 1$ columns summing to at most 1. Thus the total sum of the columns of A' is at most $r - 1$, a contradiction since the total sum of the rows of A' equals r. □

For instance, when $n = 3$, the condition in Theorem 9 excludes zero rows ($r = 1$, $s = 3$), zero columns ($r = 3$, $s = 1$) and zero 2×2 submatrices with consecutive rows or columns.

Consider the pattern

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

Then P satisfies the condition in Theorem 9, but it is not the pattern of any fractional ASM (because the pattern of A is less than or equal to P, and line sums being 1 implies that $A = P$, but then a column sum is 2). This shows that the converse of the theorem does not hold, so the patterns of fractional ASMs satisfy further requirements.

Theorem 10 *Let A be a fractional ASM. Then if A has an $r \times s$ zero submatrix A_1 whose complementary $(n - r) \times (n - s)$ submatrix A_2 is also a zero matrix, then $r + s = n$.*

Proof. After permutations,

$$A = \begin{bmatrix} O_{r,s} & A_{1,2} \\ A_{2,1} & O_{n-r,n-s} \end{bmatrix}$$

where $A_{1,2}$ is $r \times (n-s)$ and A_{21} is $(n-r) \times s$. Since all row and column sums of A equal 1, we have $r = n - s$. $\qquad\square$

Note that Theorem 10 holds provided only that A is a real matrix with all row and column sums equal to 1. Theorem 9 uses the alternating property in that consecutive elements in a row or column always sum to a number between -1 and 1.

3   Signed doubly stochastic matrices

Motivated by our discussion of the polytope generated by $n \times n$ ASMs, we investigate in this section, the affine set of doubly stochastic matrices with the restriction of nonnegativity of entries removed.

Let $\mathcal{P}_n^{\pm}$ be the set of $n \times n$ *signed permutation matrices* (so permutation matrices where some 1's may be replaced with -1's). Then every $n \times n$ ASM is an integral nonnegative linear combination of $n \times n$ signed permutation matrices. This follows from Corollary 5 since we can just use permutation matrices and their negatives (a negative of a permutation matrix is a signed permutation matrix). For instance

$$
\begin{bmatrix} & 1 & & \\ 1 & -1 & 1 & \\ & 1 & -1 & 1 \\ & & 1 & \end{bmatrix}
=
\begin{bmatrix} & & & 1 \\ & & 1 & \\ & 1 & & \\ 1 & & & \end{bmatrix}
+
\begin{bmatrix} & & & -1 \\ & -1 & & \\ & & -1 & \\ -1 & & & \end{bmatrix}
+
\begin{bmatrix} & 1 & & \\ 1 & & & \\ & & & 1 \\ & & 1 & \end{bmatrix}.
$$

In a signed decomposition of an ASM one uses permutation matrices and their negatives, and Corollary 6 gives the minimum number required in such a decomposition. A natural question is therefore if fewer matrices are needed when one allows signed permutation matrices, and this is answered in the negative in the following corollary.

Corollary 11 *For any given ASM A, the smallest number of signed permutation matrices with sum A is $N = 2k(A) + 1$.*

Proof.   Corollary 6 shows that there exists a decomposion of A with $N = 2k(A) + 1$ signed permutation matrices (actually permutation matrices or their negatives). Moreover, in A there is a row or column with $N = 2k(A)+1$ nonzeros. Since a signed permutation matrix contains exactly one nonzero

in any row or column, there must be at least N signed permutation matrices in a decomposition of A. □

We define an $n \times n$ real matrix to be $(\pm)$-*doubly stochastic* provided all row and column sums equal 1, and we let $\Omega_n^{\pm}$ denote the set of all such matrices. $\Omega_n^{\pm}$ is an affine set and

$$\Omega_n^{\pm} = \{A \in \mathcal{L}_n : A = [a_{ij}], \ \sum_{j=1}^{n} a_{1j} = 1\}.$$

Using that a real matrix may be written as a difference between two non-negative matrices, we shall find properties of $\Omega_n^{\pm}$ by investigating the closely related set

$$\widehat{\Omega}_n^{\pm} = \{(A_1, A_2) \in M_n \times M_n : A_1, A_2 \geq O, \ A_1 - A_2 \in \Omega_n^{\pm}\}$$

where the patterns of A_1 and A_2 are disjoint (no common 1's). If $a \in \mathbb{R}$, define $a^+ = \max\{a, 0\}$ and $a^- = \max\{-a, 0\}$. So, $a = a^+ - a^-$. For a matrix $A = [a_{ij}]$, define $A^+ = [a_{ij}^+]$ and $A^- = [a_{ij}^-]$ and, therefore, $A = A^+ - A^-$. If $A \in \Omega_n^{\pm}$, then $(A^+, A^-) \in \widehat{\Omega}_n^{\pm}$. It is clear that $\widehat{\Omega}_n^{\pm}$ is a pointed polyhedron which is unbounded.

Let D_n denote the directed graph with vertices $u_1, u_2, \ldots, u_n$ and $v_1, v_2, \ldots, v_n$ and directed arcs (u_i, v_j) and (its backward arc) (v_j, u_i) for $1 \leq i, j \leq n$; this graph is very similar to the notion of complete flow grid in [20], see Section 6. Let $I = \{u_1, u_2, \ldots, u_n\}$, $J = \{v_1, v_2, \ldots, v_n\}$ and $V = I \cup J$. We let $K_{n,n}$ be the corresponding complete bipartite graph, obtained by replacing the arcs (u_i, v_j) and (v_j, u_i) by an undirected edge $u_i v_j$ $(1 \leq i, j \leq n)$. Let T be a spanning tree in $K_{n,n}$, and let $E(T)$ denote its edge set. Consider an edge $e = u_i v_j \in E(T)$. If e is removed from T, the subgraph $T \setminus \{e\}$ consists of two disjoint trees (subtrees of T), and we let V_e^I (resp. V_e^J) be the vertex set of the subtree containing u_i (resp. v_j). Define a matrix $A = A(T) = [a_{ij}]$ by

$$a_{ij} = |J \cap V_e^J| - |I \cap V_e^J|$$

for each (i, j) with $u_i v_j \in E(T)$, and $a_{ij} = 0$ otherwise. (We also have $a_{ij} = |I \cap V_e^I| - |J \cap V_e^I|$.) Note that if the edge $u_i v_j$ is a leaf, so either V_e^I or V_e^J consists of a single vertex, then $a_{ij} = 1$. We say that $e = u_i v_j \in E(T)$ is *balanced* (for T) if $|J \cap V_e^J| = |I \cap V_e^J|$, so $a_{ij} = 0$.

11

Theorem 12 *Let T be a spanning tree in $K_{n,n}$, and let $A = A(T)$. Then (A^+, A^-) is an extreme point of $\widehat{\Omega}_n^{\pm}$. Moreover, any extreme point of $\widehat{\Omega}_n^{\pm}$ corresponds to some spanning tree in this way.*

Proof. Let (B, C) be an extreme point of $\widehat{\Omega}_n^{\pm}$ and define $A = B - C$. Let $A = [a_{ij}]$, $B = [b_{ij}]$ and $C = [c_{ij}]$. Then

$$
\begin{aligned}
1 &= \sum_{j=1}^{n} a_{ij} = \sum_{j=1}^{n} b_{ij} - \sum_{j=1}^{n} c_{ij} \quad (i \le n), \\
1 &= \sum_{i=1}^{n} a_{ij} = \sum_{i=1}^{n} b_{ij} - \sum_{i=1}^{n} c_{ij} \quad (j \le n).
\end{aligned}
\tag{5}
$$

These equations show that (B, C) is a nonnegative flow in the digraph D_n where b_{ij} is the flow along the arc (u_i, v_j) and c_{ij} is the flow along the backward arc (v_j, u_i). The divergence (net flow out) in vertex u_i is 1, and the divergence in vertex v_j is -1; so vertices in I are supply vertices and those in J are demand vertices. From network flow theory (see [1]) we know that extreme points of the set of nonnegative flows with given divergence, correspond to spanning trees in D_n, such that the flow is zero in all arcs outside the spanning tree. Now, a (directed) spanning tree in D_n contains at most one of the two arcs (u_i, v_j) and (v_j, u_i) (otherwise there is a cycle). This means that spanning trees in D_n correspond to (undirected) spanning trees in $K_{n,n}$. So, let T be a spanning tree in $K_{n,n}$. It only remains to verify that the matrix $A = A(T)$, as defined above, is such that (A^+, A^-) is the tree solution corresponding to the directed spanning tree associated with T. But, this tree solution is unique as the coefficient matrix associated with this tree is an invertible matrix of order $2n - 1$. This matrix is obtained from the coefficient matrix of size $2n \times 2n^2$ in (5) by deleting one equation (which is redundant) and only using the columns corresponding to the selected $2n - 1$ arcs in the tree. Therefore it is sufficient to verify that $A = A(T)$ with $B = A^+$ and $C = A^-$ satisfies (5). Let $i \le n$. Then

$$
\begin{aligned}
\sum_{j=1}^{n} a_{ij} &= \sum_{j:u_i v_j \in E(T)} a_{ij} \\
&= \sum_{j:u_i v_j \in E(T)} (|J \cap V_e^J| - |I \cap V_e^J|) \\
&= \sum_{j:u_i v_j \in E(T)} |J \cap V_e^J| - \sum_{j:v_i v_j \in E(T)} |I \cap V_e^J| \\
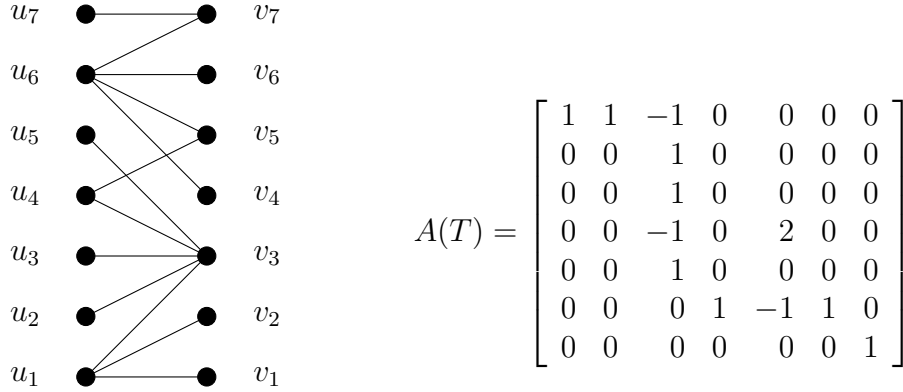&= n - (n-1) = 1
\end{aligned}
$$

12

Figure 1: A spanning tree T and the matrix $A(T)$.

$$A(T) = \begin{bmatrix} 1 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

A similar calculation shows that $\sum_{i=1}^{n} a_{ij} = 1$. This completes the proof. □

Example 3 Let $n = 7$. Figure 1 shows a spanning tree T and the associated matrix $A(T)$. The edge $u_6 v_7$ is balanced, so $a_{67} = 0$. □

   Remark: The set of $n \times n$ $(\pm)$-doubly stochastic matrices with entries between -1 and 1, is a bounded polyhedron, and therefore a polytope, say $\mathcal{K}_n^{\pm 1}$. All extreme points of $\mathcal{K}_n^{\pm 1}$ are integral, so the extreme points are all $(0, \pm 1)$-matrices with line sums 1. The integrality follows from the fact that the vertex-edge incidence matrix of a bipartite graph is totally unimodular (see [18]). Thus Theorem 12 implies that any $(\pm)$-doubly stochastic matrix may be written as a convex combination of matrices $A(T)$ for certain spanning trees T in $K_{n,n}$.
   We now study the patterns of $(\pm)$-doubly stochastic matrices. Let $A = [a_{ij}]$ be an $n \times n$ $(0, 1)$-matrix and assume that A has term rank n. Without loss of generality we may assume that $I_n \leq A$ and that

$$A = \begin{bmatrix} A_1 & A_{1,2} & \cdots & A_{1,k} \\ O & A_2 & \cdots & A_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & A_k \end{bmatrix}$$

where $A_1, A_2, \ldots, A_k$ are fully indecomposable matrices. Define a graph $G = G(A)$ with vertex set $\{1, 2, \ldots, k\}$ with an edge between i and j if and only

13

if $A_{ij} \neq O$. This graph is not connected if and only if A is a nontrivial direct sum, in which case we have two smaller matrices to consider in place of A. So we may assume that G is connected. Thus if P and Q are permutation matrices such that

$$PAQ = \begin{bmatrix} X_1 & X_{1,2} \\ O & X_2 \end{bmatrix} \tag{6}$$

where X_1 and X_2 are square non-vacuous matrices, then $X_{1,2}$ by the connectivity assumption cannot be a zero matrix. If whenever (6) holds, then $X_{1,2}$ contains at least two 1's, and we say that A is 2-*connected*. In fact, this is equivalent to the graph G being 2-connected, that is, connected and no bridge (an edge whose removal disconnects the graph), equivalently, G is connected and every edge is in a cycle.

Since each A_i is fully indecomposable, there exists a (nonnegative) doubly stochastic matrix S_i whose pattern is A_i.

Theorem 13 *Assume that A is 2-connected. Then there exists a $(\pm)$-doubly stochastic matrix with nonzero pattern A which is nonnegative in the positions of all of the diagonal blocks A_i.*

Proof. For each i, let S_i be a (nonnegative) doubly stochastic matrix whose pattern is A_i. Thus $\oplus_{i=1}^k S_i$ is a (nonnegative) doubly stochastic matrix. Since A_i is fully indecomposable, the bipartite graph of S_i has the property that there is a chain of positive entries between any pair of its vertices (alternate vertical and horizontal steps).

If some A_{ij} with $i \neq j$ contains two 1's in the same column, in its rows p and q, then using the chain between p and q and alternating ϵ (as small as needed) and $-\epsilon$, we can obtain a $(\pm)$-doubly stochastic matrix where we keep the pattern of S_i and retain its nonnegativity, and also get nonzeros in the positions of these two 1's. We do a similar construction if the two 1's are in the same row. Finally if there are two 1's in A_{ij} but in different rows and columns, we use a chain in A_i and a chain in A_j and again get nonzeros in the positions of the 1's, retaining the double stochasticity and the nonnegativity of the diagonal blocks (by appropriate choice of ϵ's).

Now we have only deal with at most one entry in each of the A_{ij}. But, with the same procedure, using the fact that each edge of G belongs to a cycle and the chains in each of the bipartite graphs of the S_i, we get a $(\pm)$-doubly stochastic matrix which is nonnegative in the positions of A_i and has the pattern of A. $\qquad\square$

4 Signed permutation matrices and associated polytope

In this section we continue our investigations of the previous section by imposing a boundedness condition which reduces our affine set to a polytope.

Let $\Pi^{\pm 1}$ be the convex hull of the $n \times n$ signed permutation matrices (i.e., of the so-called hyperoctahedral group). Recall that a square $(0, 1)$-matrix is a *subpermutation matrix* provided it has at most one 1 in each row and column, and that a square nonnegative matrix is *doubly substochastic* (DSS) provided all row and column sums are at most 1. For a real matrix $A = [a_{ij}]$, let $|A| = [|a_{ij}|]$. Then A is *generalized doubly substochastic* (GDSS) provided $|A|$ is doubly substochastic. The following theorem is known ([17]), but for completeness we give a proof here.

Theorem 14 $\Pi^{\pm 1}$ *is the set of $n \times n$ GDSS matrices.*

Proof. First an $n \times n$ nonnegative matrix A is DSS if and only if it is in the convex hull of the subpermutation matrices (of ranks $0, 1, 2, \ldots, n$): One way is trivial. Now assume that A is DSS. Then

$$B = \begin{bmatrix} A & D_1 \\ D_2 & A^T \end{bmatrix},$$

where D_1 (resp. D_2) is the diagonal matrix of 1 minus the row sums (resp. column sums) of A, is doubly stochastic. So B is a convex combination of permutation matrices. Therefore A is a convex combination of subpermutation matrices.

Now let $A = [a_{ij}]$ be a GDSS matrix so that $|A|$ is a convex combination of subpermutation matrices. A subpermutation matrix is an average of two signed permutation matrices (complete to a permutation matrix using 1's and complete to a signed permutation matrix using -1's). So $|A|$ is a convex combination of a set X of signed permutation matrices. For each entry a_{ij} of A that is negative, change the (i, j)-entry of each $P \in X$ to its negative (so 0 stays as 0). Then A becomes a convex combination of signed permutation matrices. □

Note that for real variables $x_1, x_2, \ldots, x_n$ the inequality $\sum_{j=1}^{n} |x_j| \leq 1$ is equivalent to the system of inequalities $\sum_{j=1}^{n} s_j x_j \leq 1$ for all $s_j \in \{-1, 1\}$ ($j \leq n$). Using this in combination with Theorem 14 one obtains a characterization of $\Pi^{\pm 1}$ in terms of linear inequalities.

The cone generated by the $n \times n$ signed permutation matrices is all of M_n. To see this, let $A = [a_{ij}]$ be any $n \times n$ real matrix with $r = \max\{|a_{ij}| : 1 \le i, j \le n\}$. Then $(1/(rn))A$ is GDSS, so it is a convex combination of signed permutation matrices. Thus A is a nonnegative linear combination of signed permutation matrices, that is, the cone generated by the $n \times n$ signed permutation matrices is M_n.

Suppose we now only consider nonnegative integral matrices. For instance

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Then $(1/3)A$ is DSS and is a convex combination of six signed permutation matrices, in fact, their average (decompose $J_{2,3}$ into three subpermutation matrices Q_1, Q_2, Q_3 of rank 2, and extend each and their negatives to 3×3 signed permutation matrices using 1's in row 1). This gives $2A$ as a nonnegative integral linear combination of six signed permutation matrices. But A cannot be expressed as a nonnegative integral linear combination of signed permutation matrices; see the next theorem.

Theorem 15 *An $n \times n$ integral matrix A is the sum of signed permutation matrices (that is, a nonnegative integral linear combination of signed permutation matrices) if and only if all row and column sums are even or all are odd.*

Proof. Since adding a signed permutation matrix either increases or decreases a row sum (resp. column sum) by 1, if A is the sum of signed permutation matrices then all row and column sums have the same parity.

Now assume that all row and column sums of A have the same parity. By adding a multiple of J_n if necessary (since $-J_n$ is the sum of signed permutation matrices all of whose nonzeros equal -1), we can assume that A is nonnegative. Also by adding I_n if necessary (since $-I_n$ is a signed permutation matrix), we can also assume that all row and column sums of A are even.

Now let the row sum and column sum vectors of A be the vectors $R = (r_1, r_2, \ldots, r_n)$ and $S = (s_1, s_2, \ldots, s_n)$ of positive even integers. Let

$$p = \max\{r_1, r_2, \ldots, r_n, s_1, s_2, \ldots, s_n\},$$

16

an even integer. Define

$$R' = (p - r_1, p - r_2, \ldots, p - r_n) \text{ and } S = (p - s_1, p - s_2, \ldots, p - s_n).$$

Then R' and S' are nonnegative, even integer vectors with the same sum of components. Let

$$R'' = (1/2)R' \text{ and } S'' = (1/2)S'.$$

There exists a nonnegative integral matrix $B = [b_{ij}]$ with row sum vector R'' and column sum vector S''. The matrix $C = A + 2B = [c_{ij}]$ is a nonnegative integral matrix with row and column sum vector equal to $(p, p, \ldots, p)$, and hence C is the sum of p permutation matrices,

$$C = \sum_{l=1}^{p} Q_l.$$

Now C is obtained from A by increasing each entry a_{ij} by a nonnegative even integer. Of the c_{ij} permutation matrices Q_l that contain a 1 in position (i, j) change b_{ij} of those 1's to -1's. We do this independently for each position (i, j). The result is that each Q_l has become a signed permutation matrix Q_l' and

$$A = \sum_{l=1}^{p} Q_l'.$$

$\square$

Example 4 Let

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}.$$

Then A is a sum of signed permutation matrices

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

but A is not the sum of $\pm$(permutation matrices). $\square$

If A is an $n \times n$ integral matrix with all row and column sums even or all odd, then the proof of Theorem 2 provides an algorithm to write A as a nonnegative linear combination of signed permutation matrices:

17

As in the proof construct the matrix B (easy algorithm) to obtain the matrix $C = A + 2B$ with constant p row and column sums. Then use Birkhoff's algorithm (as described in [5]) to express C as a nonnegative linear combination of permutation matrices:

$$C = \sum_{i=1}^{t} Q_i = \sum_{j=1}^{k} d_j P_j \quad \text{(integral } d_j > 0\text{)},$$

where $k \leq (n-1)^2 + 1$. Now proceed as in the proof of the theorem, changing some of the 1's in the Q_i to -1's. This expresses A as a sum of signed permutation matrices and so as a positive integral linear combination of signed permutation matrices.

5 Sum majorization and r-ASMs

We now introduce a generalization of ASMs which relates to the notion of majorization. Based on this we will prove a decomposition result involving sums of ASMs, and further use this to determine the H-basis of the ASM cone.

Let $B = [b_{ij}]$ be an $n \times n$ nonnegative matrix. We say that an $n \times n$ matrix $A = [a_{ij}]$ is *sum-majorized* by B if

$$\begin{aligned}
0 \leq \textstyle\sum_{j'=1}^{j} a_{ij'} \leq b_{ij} & \quad (1 \leq i, j \leq n), \\
0 \leq \textstyle\sum_{i'=1}^{i} a_{i'j} \leq b_{ij} & \quad (1 \leq i, j \leq n), \\
\textstyle\sum_{j=1}^{n} a_{ij} = b_{in} & \quad (1 \leq i \leq n), \\
\textstyle\sum_{i=1}^{n} a_{ij} = b_{nj} & \quad (1 \leq j \leq n).
\end{aligned} \tag{7}$$

In particular, if $B = J$ (the all ones matrix), we see that an integral matrix A is sum-majorized by J if and only if A is an ASM. Moreover, by Theorem 7 [3, 20]) a real matrix is sum-majorized by J if and only if it is in Λ_n.

Another interesting special case is when $B = rJ$ for a positive integer r. We call an integral matrix A an r-*ASM* provided that A is sum-majorized by rJ. Actually, this notion was studied in [3] under another name, namely higher spin ASMs. It follows that the convex hull of the r-ASMs is the convex hull of the set of matrices obtained from ASMs by multiplying by r. To see this, observe that if A is in the convex hull of r-ASMs, then $(1/r)A$ is in the convex hull of ASMs and so is a convex combination of ASMs; thus

multiplying by r we see that A is a convex combination of r times ASMs. So the extreme points are nothing other than r times the ASMs. For instance, the 2-ASM

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

is not an extreme point of the convex hull of 2-ASMs, since A is the convex combination

$$A = (1/2) \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} + (1/2) \begin{bmatrix} 0 & 2 & 0 \\ 0 & 0 & 2 \\ 2 & 0 & 0 \end{bmatrix}.$$

We now return to a matrix B as above and shall study the convex hull of all integral matrices A that are sum-majorized by B.

Let $D = (V, E)$ denote the directed graph with vertices v_{ij} $(1 \leq i, j \leq n)$ and arcs $(v_{ij}, v_{i+1,j})$ and $(v_{ij}, v_{i,j+1})$ for all i, j (where indices are defined). Thus, the vertices correspond to the positions of an $n \times n$ matrix, and arcs from a position go to the neighbor below or to the right.

Theorem 16 *Let $B = [b_{ij}]$ be an $n \times n$ nonnegative matrix. The convex hull of all integral matrices that are sum-majorized by B equals the set of real matrices $A = [a_{ij}]$ satisfying the linear system in (7).*

Proof. Let $\mathcal{P}^* \subseteq M_n$ be the polyhedron consisting of all *real* matrices $A = [a_{ij}]$ satisfying the linear system in (7). For $A = [a_{ij}]$ define $\Sigma(A)$ as the matrix $S = [s_{ij}]$ where

$$s_{ij} = \sum_{1 \leq i' \leq i,\, 1 \leq j' \leq j} a_{i'j'} \quad (1 \leq i, j \leq n).$$

Then the map $T : A \to \Sigma(A)$ is an isomorphism, and therefore $\mathcal{P}^*$ and its image

$$\Sigma(\mathcal{P}^*) = \{\Sigma(A) : A \in \mathcal{P}^*\}$$

are isomorphic. Let $A = [a_{ij}]$ and $S = [s_{ij}] = T(A)$. Then $A = T^{-1}(S)$ is given by

$$a_{ij} = s_{ij} + s_{i-1,j-1} - s_{i-1,j} - s_{i,j-1} \quad (1 \leq i, j \leq n)$$

19

where we define $s_{0j} = 0$ $(1 \leq j \leq n)$ and $s_{i0} = 0$ $(1 \leq i \leq n)$. Moreover

$$
\begin{aligned}
\sum_{i'=1}^{i} a_{i'j} &= s_{ij} - s_{i,j-1} \quad (1 \leq i, j \leq n), \\
\sum_{j'=1}^{j} a_{ij'} &= s_{ij} - s_{i-1,j} \quad (1 \leq i, j \leq n), \\
\sum_{1 \leq i' \leq i, 1 \leq j \leq n} a_{i'j} &= s_{in} \quad (1 \leq i \leq n), \\
\sum_{1 \leq i \leq n, 1 \leq j' \leq j} a_{ij'} &= s_{nj} \quad (1 \leq j \leq n).
\end{aligned}
\tag{8}
$$

Note that this isomorphism T and its inverse T^{-1} preserve integrality, that is, an integral matrix is mapped by T and T^{-1} into an integral matrix.

Now, we claim that $\Sigma(\mathcal{P}^*)$ is the set of matrices $S = [s_{ij}]$ satisfying

$$
\begin{aligned}
0 \leq s_{ij} - s_{i,j-1} &\leq b_{ij} \quad (1 \leq i, j \leq n), \\
0 \leq s_{ij} - s_{i-1,j} &\leq b_{ij} \quad (1 \leq i, j \leq n), \\
s_{in} &= \sum_{i'=1}^{i} b_{i'n} \quad (1 \leq i \leq n), \\
s_{nj} &= \sum_{j'=1}^{j} b_{nj'} \quad (1 \leq j \leq n).
\end{aligned}
\tag{9}
$$

If fact, if A satisfies (7), then, due to (8), $S = \Sigma(A)$ satisfies (9). Conversely, assume $S = [s_{ij}]$ satisfies (9) and let $A = T^{-1}(S)$. Then $A = [a_{ij}]$ satisfies $T(A) = S$, so due to (8), A satisfies (7), as claimed.

The coefficient matrix in (9) is totally unimodular. To see this, note that the first two constraints define the arc-vertex incidence matrix of the directed graph D introduced above. Moreover, all the constants (right hand sides) in the system are integers, as B is integral. A standard result from polyhedral theory (see [18]) then implies that $\Sigma(\mathcal{P}^*)$ is an integral polyhedron, so all extreme points are integral. From the properties of the isomorphism, $\mathcal{P}^*$ is integral, and this shows the theorem. □

The proof above shows a connection between sum-majorization and network flow potentials in the digraph D. Note that Theorem 7, originally shown in [3, 20], is obtained as a corollary of Theorem 16; just let $B = J$. Thus, we have a short proof via total unimodularity of the linear characterization of the ASM polytope. As the proof shows, one can obtain an even more general integrality result, where the given matrix B is replaced by a pair (B, C) of matrices, one for partial row sum constraints and one for partial column sum constraints. The partial sum matrix $\Sigma(A)$ used in the proof of Theorem 16 is called a corner-sum matrix in [3] and it is there used to investigate properties of r-ASMs, which they called higher spin ASMs.

We now show that the polyhedral cone spanned by r-ASMs has the integer decomposition property. This result is an ASM analogue to Theorem 1 for permutation matrices.

Theorem 17 *Let A be a square matrix and r a positive integer. Then A is an r-ASM if and only if there exist ASMs $A_1, A_2, \ldots, A_r$ such that*

$$A = A_1 + A_2 + \cdots + A_r.$$

Proof. We use the construction (and notation) given in the proof of Theorem 16, with the matrix $B = rJ$.

Assume A is an r-ASM. Let $S = [s_{ij}] = \Sigma(A) = T(A)$. Then S is an integral matrix and satisfies (9). Note that S is nonnegative, as S is nondecreasing in every row and column, and $s_{0j} = 0$ ($1 \leq j \leq n$), $s_{i0} = 0$ ($1 \leq i \leq n$). Define $\mathcal{Q}_r$ as the set of nonnegative matrices satisfying (9). Note that

$$\mathcal{Q}_r = r\mathcal{Q}_1 = \{rC : C \in \mathcal{Q}_1\}.$$

$\mathcal{Q}_r$ is a polyhedron defined by a totally unimodular coeffient matrix, as explained in the proof of Theorem 16. So, S is an integral matrix in $\mathcal{Q}_r$. We may now apply a general result of Baum and Trotter, see Theorem 19.4 in [18], saying that if x is an integral vector satisfying $x \geq O$, $Mx \leq rb$, where M is a totally unimodular matrix, b is a nonnegative integral vector and r a positive integer, then there are integral vectors $x_1, x_2, \ldots, x_r \in \{x \geq O : Mx \leq b\}$ with $x = x_1 + x_2 + \cdots + x_r$. So, in our situation, this theorem shows the existence of r integral nonnegative matrices $S_1, S_2, \ldots, S_r \in \mathcal{Q}_1$ such that

$$S = S_1 + S_2 + \cdots + S_r.$$

Define $A_i = T^{-1}(S_i)$ ($i \leq r$). Since $S_i \in \mathcal{Q}_1$, by Theorem 16, A_i is an ASM ($i \leq r$). And, finally,

$$A = T^{-1}(S) = T^{-1}(\sum_{i=1}^{r} S_i) = \sum_{i=1}^{r} T^{-1}(S_i) = \sum_{i=1}^{r} A_i$$

which is the desired decomposition of A.

The converse statement, that the sum of r ASMs is an r-ASM, follows directly from the definition. $\square$

We now return to the question of a H-basis for the ASM cone.

Theorem 18 $\mathcal{A}_n$ *is the H-basis of* $\mathrm{Cone}(\mathcal{A}_n)$.

Proof. Let A be an integral matrix in $\mathrm{Cone}(\mathcal{A}_n)$. Then A must be an r-ASM, for some positive integer r. In fact, let $A = \sum_{j=1}^{N} \lambda_j A_j$ where the λ_j's are nonnegative and each A_j is an ASM. Now, each A_j satisfies the linear system (7) for $B = J$, and then $A = \sum_{j=1}^{N} \lambda_j A_j$ satisfies (7) for $B = rJ$ where $r = \sum_{j=1}^{N} \lambda_j$. This number r must be an integer, as the common line sum of A is $\sum_{j=1}^{N} \lambda_j = r$ and A is integral. So A is an r-ASM. By Theorem 17, A may be written as the sum of r ASMs. Thus, every integral matrix in $\mathrm{Cone}(\mathcal{A}_n)$ may be written as an integral nonnegative linear combination of matrices in $\mathcal{A}_n$. Therefore, $\mathcal{A}_n$ is a Hilbert basis of $\mathrm{Cone}(\mathcal{A}_n)$.

Moreover, $\mathcal{A}_n$ is a *minimal* Hilbert basis of this cone, since no ASM can be expressed as a integer nonnegative linear combination of the ASMs different from it. Actually, such a linear combination of at least two matrices will have row sum at least 2 in the first row. As $\mathrm{Cone}(\mathcal{A}_n)$ is pointed, the minimal Hilbert basis is unique, and it must therefore be $\mathcal{A}_n$. $\square$

Since $\mathcal{A}_n$ is the H-basis of $\mathrm{Cone}(\mathcal{A}_n)$ and the dimension of Λ_n is $(n-1)^2$, it follows from (1) that every integer matrix in $\mathrm{Cone}(\mathcal{A}_n)$ can be written as a positive integer linear combination of at most $2(n-1)^2 - 1$ ASMs.

6 Faces of the ASM polytope

In [20] one establishes a bijection between the set $\mathcal{A}_n$ of $n \times n$ ASMs and the set $\mathcal{F}_n$ of so-called *simple flow grids* of order n. Moreover, faces of the ASM polytope Λ_n may be described using this notion. In this section we discuss this and study edges of Λ_n and a certain distance concept.

The *complete flow grid* C_n is a digraph with vertices

$$\{(i,j) : 0 \le i \le n+1, 0 \le j \le n+1\} \setminus \{(0,0), (0,n+1), (n+1,0), (n+1,n+1)\}.$$

So there are $n^2 + 4n$ vertices, of which n^2 are *internal* (those with $1 \le i, j \le n$, in a square formation) and $4n$ are *boundary vertices*. The edges of C_n consist of all edges in both directions between internal vertices (that are adjacent in the grid, horizontally or vertically) and an edge from a vertex on the border of the square of internal vertices to the obvious boundary vertex.

Given $A \in \mathcal{A}_n$, the mentioned correspondence is: The vertex (position) corresponding to each 1 of A is a *source*, and so has an edge from it to each

of the neighboring vertices. The vertex (position) corresponding to each -1 of A is a *sink*, and so has an edge to it from each of the neighboring vertices. This determines the simple flow grid corresponding to A in that, to construct the remaining edges connecting neighboring vertices one continues the edges from each source until one either hits a sink or a boundary vertex. Note that, in a simple flow grid, there are four edges incident to each internal vertex: either four edges directed inward (when that entry is -1), four edges directed outward (when that entry is 1), or two horizontal edges pointing in the same direction and two vertical edges pointing in the same direction (when that entry is 0). In particular, both the inward degree and the outward degree of a vertex are even numbers and their sum is 4.

For example, if

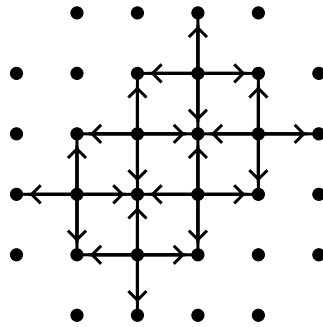$$A = \begin{bmatrix} & & 1 & \\ & 1 & -1 & 1 \\ 1 & -1 & 1 & \\ & 1 & & \end{bmatrix},$$

Figure 2: The simple flow grid of A (some missing arrows).

then we initially get the graph in Figure 2. Now continue arrows until one hits a boundary vertex. In this example, the -1's are surrounded by 1's, so there are no arrows into the sinks that have to be continued "backwards".

Note that in case the ASM A is actually a permutation matrix, then the corresponding simple flow grid is a union of n *directed* 4-*path stars* (four directed paths emanating from a vertex of various lengths) centered at each vertex corresponding to a 1, with no edges in common.

An $n \times n$ *elementary flow grid* is a subgraph of C_n whose edge set is the union of the edge sets of simple flow grids (so they are determined by a subset

$\mathcal{X} \subseteq \mathcal{A}_n$ of $n \times n$ ASMs). It is shown in [20] that the face lattice of $\mathcal{A}_n$ is isomorphic to the lattice of all $n \times n$ elementary flow grids; in particular, faces of the $n \times n$ ASM-polytope Λ_n are in one-to-one correspondence with $n \times n$ elementary flow grids. Moreover, it is possible to determine the dimension of a face of Λ_n (see [20]); it is in terms of *doubly directed regions* (here doubly directed means an edge in both directions between a pair of vertices) of an elementary flow grid. For a set $\mathcal{X}$ of $n \times n$ ASMs let $G(\mathcal{X})$ be the graph obtained from the elementary flow grid associated with $\mathcal{X}$ by replacing a pair of bidirected edges by an undirected edge, and removing the remaining edges. $G(\mathcal{X})$ is clearly planar, and we now prove the following connectivity result.

Theorem 19 *Let $\mathcal{X}$ be a set of at least two $n \times n$ ASMs. Then each connected component of $G(\mathcal{X})$ is 2-connected, or equivalently, every edge is contained in a cycle. Therefore, $G(\mathcal{X})$ determines connected plane regions bounded by closed curves (cycles).*

Proof. Consider a doubly directed edge. We can assume without loss of generality that this doubly directed edge is horizontal joining vertices (p, k) and $(p, k+1)$. Thus there is a simple flow grid, an ASM $A = [a_{ij}]$ in $\mathcal{X}$, with an edge $(p, k) \to (p, k+1)$ and another simple flow grid, an ASM $B = [b_{ij}]$ in $\mathcal{X}$ with an edge $(p, k+1) \to (p, k)$. Then it follows that $a_{pk} = 0$ or 1 and $b_{pk} = 0$ or -1. There are four possible combinations and each such combination yields another doubly directed edge at (p, k):

(i) $a_{pk} = 0$ and $b_{pk} = 0$: a doubly directed edge joining vertices $(p, k-1)$ and (p, k).

(ii) $a_{pk} = 1$ and $b_{pk} = 1$: a doubly directed edge joining vertices $(p, k-1)$ and (p, k).

(iii) $a_{pk} = 0$ and $b_{pk} = -1$: a doubly directed edge joining vertices (p, k) and either $(p-1, k)$ or $(p+1, k)$.

(iv) $a_{pk} = 1$ and $b_{pk} = 0$: a doubly directed edge joining vertices (p, k) and either $(p-1, k)$ or $(p+1, k)$.

Thus there exists another doubly directed edge at (p, k). □

Let $\mathcal{X}$ be a set of at least two $n \times n$ ASMs. Then the dimension of the face of Λ_n that corresponds to the elementary flow grid of $\mathcal{X}$, is the number of

plane-connected regions determined by the cycles of $G(\mathcal{X})$, but not counting the unbounded region outside the edges. More precisely, each such region is bounded by the edges of a cycle C, and any two points in the region may be connected by a curve which does not cross any of the edges or vertices of C. In particular, the edges of Λ_n (one-dimensional faces) correspond to elementary flow grids with exactly one cycle of doubly directed edges.

Now, consider two distinct permutation matrices $P = [p_{ij}]$ and $Q = [q_{ij}]$ of order n. Let $P \oplus_2 Q$ be the mod 2 sum of P and Q; this matrix has a 1 in position (i, j) if and only if exactly one of p_{ij} and q_{ij} is 1 $(i, j \leq n)$. Each row and column of $P \oplus_2 Q$ is either the zero vector or it contains two ones. These ones form disjoint cycles in the bipartite graph $K_{n,n}$, and we also call these *cycles* in $P \oplus_2 Q$. In fact, the cycles in $P \oplus_2 Q$ correspond to the permutation cycles of P relative to Q, that is, the permutation cycles of PQ^{-1} of length more than 1. Assume that there exist $i_1 < i < i_2$ and $j_1 < j < j_2$ such that $P \oplus_2 Q$ contains a 1 in all the positions (i, j_1), (i, j_2), (i_1, j), (i_2, j). Then, $p_{ij} = q_{ij} = 0$, and we say that (i, j) is a *crossing* of $P \oplus_2 Q$. With this terminology we obtain the following result on certain edges of Λ_n.

Theorem 20 *Let $P, Q \in \mathcal{P}_n$. Then P and Q form an edge of the $n \times n$ ASM polytope Λ_n if and only if $P \oplus_2 Q$ has a unique cycle and no crossing.*

Proof. Consider the elementary flow grid corresponding to $\mathcal{X} = \{P, Q\}$. For every line (row or column) where the 1 in P and the 1 in Q are in different positions, i.e., $P \oplus_2 Q$ contains two ones, there are doubly directed edges on the path between these two positions in the elementary flow grid. Then $G(\mathcal{X})$ contains vertices corresponding to all positions between two 1s in the same line of $P \oplus_2 Q$, and edges of paths between such pairs of 1s. It is clear that each vertex of $G(\mathcal{X})$ has degree 2 or 4.

Assume that P and Q form an edge of Λ_n. Then $G(\mathcal{X})$ contains exactly one cycle, and this implies that $P \oplus_2 Q$ has at most one cycle. Moreover, if $P \oplus_2 Q$ has a crossing, then the graph $G(\mathcal{X})$ has at least one vertex of degree 4. Then $G(\mathcal{X})$ has at least two cycles, and so does the flow grid; a contradiction. So $P \oplus_2 Q$ has no crossing.

Conversely, assume that $P \oplus_2 Q$ has a unique cycle and no crossing. Then $G(\mathcal{X})$ is a cycle (plus isolated vertices) so it determines only one bounded plane-connected region and therefore P and Q form an edge of Λ_n. $\qquad\square$

Note that this theorem shows that there is a stronger condition for two permutation matrices to form an edge of the ASM polytope Λ_n than what

25

is the case for the doubly stochastic polytope Ω_n. Recall that in Ω_n one-dimensional faces correspond to pairs of permutation matrices P, Q such that PQ^{-1} is a permutation matrix with exactly one cycle of length > 1. We refer to [6] for a treatment of faces, and edges, of Ω_n and to [12] for edges of an interesting subpolytope.

Example 5 Let

$$P = \begin{bmatrix} & 1 & & \\ & & 1 & \\ 1 & & & \\ & & & 1 \end{bmatrix} \quad \text{and} \quad Q = \begin{bmatrix} & & & 1 \\ 1 & & & \\ & & 1 & \\ & & 1 & \end{bmatrix}.$$

Here $P \oplus_2 Q$ has a unique cycle. The union of their corresponding simple flow grids is an elementary flow grid whose graph $G(\{P, Q\})$ has two cycles, see Figure 3. Moreover there is a crossing in $(2, 2)$, so P and Q do not form
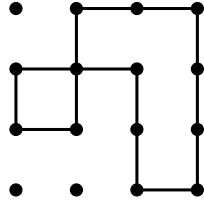

Figure 3: $G(\{P, Q\})$.

an edge of Λ_n. Actually, the face corresponding to the elementary flow grid has dimension 2. For example, the ASM

$$\begin{bmatrix} & 1 & & \\ 1 & -1 & 1 & \\ & 1 & & \\ & & & 1 \end{bmatrix}$$

is an extreme point of this two-dimensional face. Note, however, that P and Q form an edge of the polytope Ω_4. □

If P and Q are two permutation matrices of order n, then the graph $G(\{P, Q\})$ is obtained by drawing horizontal and vertical lines between 1's of P and Q in the same row or column. In Example 5 the 1's of P and Q correspond to the corners of the diagram presented.

From the discussion above of the results of [20] we have the following characterization of edges of Λ_n.

26

Theorem 21 ([20]) *Let $A, B \in \mathcal{A}_n$. Then A and B form an edge of the $n \times n$ ASM polytope Λ_n if and only if $G(\{A, B\})$ consists of a unique cycle (and isolated vertices).*

Example 6 Let $n = 6$. Let A_1 be the identity matrix, and A_2 be the permutation matrix obtained from A_1 by interchanging row 2 and 5. Then the graph $G(\{A_1, A_2\})$ is shown in Figure 4. There is only one bounded connected region, so the dimension of the face is 1, and therefore A_1 and A_2 form an edge of Λ_6.
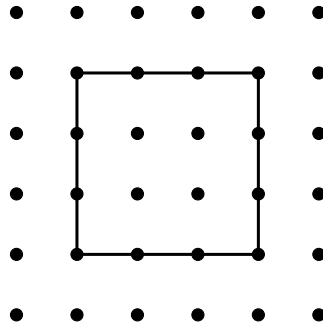

Figure 4: The graph $G(\{A_1, A_2\})$.

.

Next, let A_3 be the permutation matrix obtained from A_2 by interchanging row 3 and 4. The graph $G(\{A_1, A_2, A_3\})$ is shown in Figure 5. There are two bounded connected regions, so the dimension of the face is 2.   □

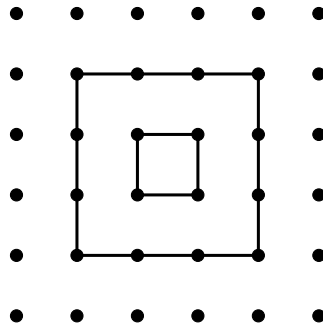Figure 5: The graph $G(\{A_1, A_2, A_3\})$.

Example 7 Let

$$
A = \begin{bmatrix} & & 1 & \\ & 1 & -1 & 1 \\ 1 & -1 & 1 & \\ & 1 & & \end{bmatrix} \quad \text{and} \quad P_1 = \begin{bmatrix} & & & 1 \\ 1 & & & \\ & & 1 & \\ & 1 & & \end{bmatrix}.
$$

The graph $G(\{A, P_1\})$ is shown in Figure 6.

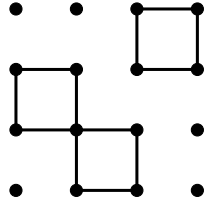

Figure 6: The graph $G(\{A, P_1\})$.

There are three bounded connected regions so by Theorem 21, A and P_1 do not form an edge of Λ_n. On the other hand, if

$$
P_2 = \begin{bmatrix} & & 1 & \\ & 1 & & \\ 1 & & & \\ & & & 1 \end{bmatrix},
$$

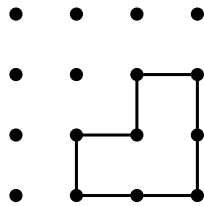then $G(\{A, P_2\})$ is the graph shown in Figure 7.

Figure 7: The graph $G(\{A, P_2\})$.

Here there is only one bounded connected region, so A and P_2 form an edge of Λ_n. □

Let P and Q be $n \times n$ permutation matrices. Let a be the number of cycles of length > 1 of P relative to Q (the number of permutation cycles

of PQ^{-1}). The dimension of the smallest face $\langle P, Q\rangle_{\Omega_n}$ of Ω_n containing P and Q equals a, and the face is a hypercube of dimension a whose vertices correspond to the subsets of the cycles (2 choices for each cycle) so 2^a vertices.

Now consider the smallest face $\langle P, Q\rangle_{\Lambda_n}$ of Λ_n containing P and Q. For $i \leq n$, let $j_1(i) \leq j_2(i)$ denote the columns in which P and Q (in some order) contain a 1 in row i. Similarly, for $j \leq n$, let $i_1(j) \leq i_2(j)$ denote the rows in which P and Q contain a 1 in column j. Define Z_1 as the set of (i, j) where $i \leq n$ and either $j < j_1(i)$ or $j > j_2(i)$. Similarly, let Z_2 be the set of (i, j) where $j \leq n$ and either $i < i_1(j)$ or $i > i_2(j)$. Finally, define

$$Z_{P,Q} = Z_1 \cup Z_2 \ \text{ and } \ E_{P,Q} = \{1, 2, \ldots, n\}^2 \setminus Z_{P,Q},$$

and let $X_{P,Q}$ be the $(0, 1)$-matrix with support $E_{P,Q}$. For instance, from P and Q in Example 5, we get

$$X_{P,Q} = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

Let $C_{P,Q}$ denote the set of positions (i, j) such that (i, j) is a crossing of $P \oplus_2 Q$.

Proposition 22 *Let P and Q be $n \times n$ permutation matrices. Then*

$$E_{P,Q} = \mathrm{supp}(P) \cup \mathrm{supp}(Q) \cup C_{P,Q}.$$

Proof. From the definition of $E_{P,Q}$ one sees that $(i, j) \in E_{P,Q}$ if and only if

$$i_1(j) \leq i \leq i_2(j) \ \text{ and } \ j_1(i) \leq j \leq j_2(i).$$

If all these four inequalities are strict, (i, j) is a crossing of $P \oplus_2 Q$, so $(i, j) \in C_{P,Q}$. Otherwise, when at least one inequality holds with equality, either P or Q has a 1 in position (i, j). This gives the result. $\qquad\square$

In general, for a polyhedron H, defined by a system of linear inequalities, one calls an inequality in the system an *implicit equality* if each point in H satisfies the inequality with equality ([18]).

Theorem 23 *Let P and Q be $n \times n$ permutation matrices. The face $\langle P, Q \rangle_{\Lambda_n}$ of Λ_n consists of all real $n \times n$ matrices $A = [a_{ij}]$ satisfying*

$$\sum_{j=j_1(i)}^{q} a_{ij} \geq 0, \quad \sum_{j=q+1}^{j_2(i)} a_{ij} \geq 0 \quad (j_1(i) \leq q \leq j_2(i), \ 1 \leq i \leq n),$$

$$\sum_{i=i_1(j)}^{p} a_{ij} \geq 0, \quad \sum_{i=p+1}^{i_2(j)} a_{ij} \geq 0 \quad (i_1(j) \leq p \leq i_2(j), \ 1 \leq j \leq n),$$

$$\sum_{j=j_1(i)}^{j_2(i)} a_{ij} = 1 \quad (1 \leq i \leq n), \tag{10}$$

$$\sum_{i=i_1(j)}^{i_2(j)} a_{ij} = 1 \quad (1 \leq j \leq n),$$

$$a_{ij} = 0 \quad ((i,j) \in Z_{P,Q}).$$

The only implicit equalities in the system (10) are the equations in the three last lines of (10), and the dimension of $\langle P, Q \rangle_{\Lambda_n}$ is

$$(n-1)^2 - |Z_{P,Q}| + t - 1$$

where t is the number of cycles of P relative to Q (that is, of PQ^{-1}). In particular, an ASM lies in the face $\langle P, Q \rangle_{\Lambda_n}$ if and only if all its nonzeros are in the set $E_{P,Q}$, and such matrices are the vertices of $\langle P, Q \rangle_{\Lambda_n}$.

Proof. Let $P = [p_{ij}]$ and $Q = [q_{ij}]$. $\langle P, Q \rangle_{\Lambda_n}$ is a face of Λ_n, and, by Theorem 7, Λ_n is characterized by the system (4) of linear inequalities, plus the linear equations saying that each row and column sum is 1. By polyhedral theory (see chapter 8 of [18]), any face of a polyhedron H is obtained by replacing some of the linear inequalities defining H by corresponding equalities. In our setting, we obtain the desired face by finding those inequalities in (4) that hold with equality for both P and Q. In fact, the inequalities in (4) that hold with equality for both P and Q hold with equality for all matrices in the face $\langle P, Q \rangle_{\Lambda_n}$, for otherwise we could find a smaller face containing P and Q (by putting such an inequality to equality). Clearly, an inequality which holds only for one of P and Q can not be put to equality.

Let $i \leq n$. For each $j < j_1(i)$,

$$\sum_{j'=1}^{j} p_{ij'} = 0 \quad \text{and} \quad \sum_{j'=1}^{j} q_{ij'} = 0.$$

Therefore, for each $i \leq n$, both P and Q satisfy the equations $a_{ij} = 0$ for $j < j_1(i)$, and, by similar arguments, $a_{ij} = 0$ for $j > j_2(i)$. Similarly, for each $j \leq n$, both P and Q satisfy $a_{ij} = 0$ for $i < i_1(j)$ or $i > i_2(j)$. This implies

that P and Q satisfy (10), so each matrix A in the face $\langle P, Q \rangle_{\Lambda_n}$ must also satisfy (10).

Finally, consider the remaining inequalities in (4), i.e., those that are different from those discussed in the previous paragraph. First, let $i \leq n$, and consider

$$(*) \quad \sum_{j=1}^{q} a_{ij} \geq 0$$

for some q with $j_1(i) \leq q \leq j_2(i)$. If $q = j_2(i)$, this inequality is redundant, as, by the discussion above, every matrix in $\langle P, Q \rangle_{\Lambda_n}$ satisfies $\sum_{j=1}^{q} a_{ij} = 1$. If $q < j_2(i)$, then exactly one of P and Q satisfies inequality $(*)$ with equality because, say, $p_{ij_1(i)} = 1$ and $q_{ij_2(i)} = 1$, then $\sum_{j=1}^{q} p_{ij} = 1$ while $\sum_{j=1}^{q} q_{ij} = 0$. So, the face $\langle P, Q \rangle_{\Lambda_n}$ is not contained in the hyperplane given by $\sum_{j=1}^{q} a_{ij} \geq 0$. Similar arguments also show that $\langle P, Q \rangle_{\Lambda_n}$ is not contained in the hyperplane given by any of the other remaining inequalities in (4). This shows that the only implicit equalities in the system (10) are those stated in the theorem. Now, the dimension of a polyhedron equals the dimension of the underlying space minus the rank of the coefficient matrix of the implicit equalities.

To compute this rank, note that one can assume that P and Q have no common 1's since they add nothing. One can regard the coefficient matrix as the vertex-edge incidence matrix of a graph with $2n$ vertices (corresponding to the equations summing to 1) and $|E_{P,Q}|$ edges (corresponding to the entries of $E_{P,Q}$. Then the rank is $2n - c$ where c is the number of connected components of the graph. The number of connected components is the number t of cycles of P relative to Q (that is, of PQ^{-1}). Thus the dimension of the face $\langle P, Q \rangle_{\Lambda_n}$ is

$$|E_{P,Q}| - (2n - c) = n^2 - |Z_{P,Q}| - 2n + t = (n-1)^2 - |Z_{P,Q}| + (t-1).$$

By the previous paragraph, the ASMs in the face $\langle P, Q \rangle_{\Lambda_n}$ are precisely the ASMs that satisfy (10), or equivalently, the ASMs with all nonzeros in the positions in $E_{P,Q}$. All ASMs are extreme points of Λ_n, so the last statement is due to Minkowski's theorem; a bounded polyhedron is the convex hull of its vertices (extreme points). $\square$

In Example 5, as mentioned above, $|Z_{P,Q}| = 7$, so the dimension of the corresponding face is $3^2 - 7 = 2$ as we saw before.

Using the notation above, we say that a position $(i, j) \notin Z_{P,Q}$ is *fixed* if $i_1(j) = i_2(j)$ or $j_1(i) = j_2(i)$; otherwise (i, j) is called *non-fixed*. If position

(i, j) is fixed, then each matrix $A = [a_{ij}] \in \langle P, Q \rangle_{\Lambda_n}$ satisfies $a_{ij} = 1$; this follows from Theorem 23. Concerning the non-fixed positions, we have the following result.

Corollary 24 *Let P and Q be $n \times n$ permutation matrices. For each non-fixed position (i, j) there exists an ASM in the face $\langle P, Q \rangle_{\Lambda_n}$ with a nonzero in position (i, j).*

Proof. If not, each ASM $A = [a_{ij}]$ in the face would satisfy $a_{ij} = 0$ and this contradicts the statement concering the implicit equalities in Theorem 23. □
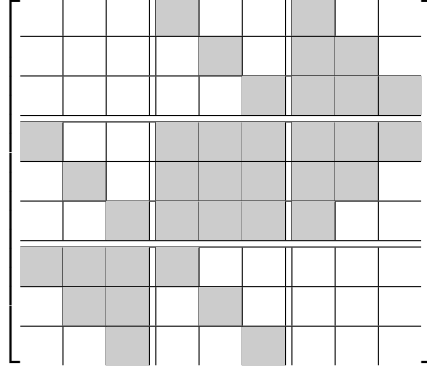
Example 8 Consider two 9×9 permutation matrices with 1's specified by a's and b's, respectively:

$$
\begin{bmatrix}
 & & & a & & & b & & \\
 & & & & a & & b & & \\
 & & & & & a & & b & \\
b & & & & & & & a & \\
 & b & & & & & a & & \\
 & & b & & & a & & & \\
a & & & b & & & & & \\
 & a & & & b & & & & \\
 & & a & & & b & & &
\end{bmatrix}.
$$

The shaded areas below give the crossings:

$$
\begin{bmatrix}
 & & & a & & & b & & \\
 & & & & a & & \blacksquare & b & \\
 & & & & & a & \blacksquare & \blacksquare & b \\
b & & & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & a \\
 & b & & \blacksquare & \blacksquare & \blacksquare & \blacksquare & a & \\
 & & b & \blacksquare & \blacksquare & \blacksquare & a & & \\
a & \blacksquare & \blacksquare & b & & & & & \\
 & a & \blacksquare & & b & & & & \\
 & & a & & & b & & &
\end{bmatrix}.
$$

Thus, the set $E_{P,Q}$ consists of the shaded positions in the following matrix, and any ASM in the face $\langle P, Q \rangle_{\Lambda_n}$ has its nonzeros only in such positions.

$\Box$

Let P and Q be distinct permutation matrices of order n. Due to Theorem 23, any permutation matrix R all of whose 1's are in $E_{P,Q}$ belongs to the face $\langle P, Q \rangle_{\Lambda_n}$ from which it follows that the number of permutation matrices in this face equals the permanent of $E_{P,Q}$.

In Example 6 the three permutation matrices have no pairwise crossings. This implies that the face of Λ_6 they determine not only has dimension 2 but the only ASMs in the face are permutation matrices. This argument holds in general, so we have the following result.

Corollary 25 *If permutation matrices $P_1, P_2, \ldots, P_r$ have no pairwise crossings, then the face they determine in Λ_n is the same as the face they determine in Ω_n.*

We now consider the final topic of this paper. For two $n \times n$ ASMs, let $d(A, B)$ be the *distance* between A and B in the 1-*skeleton* of Λ_n (the graph with vertex set $\mathcal{A}_n$ and an edge between A_1 and A_2 whenever these two ASMs form an edge of Λ_n). Thus $d(A, B)$ equals the minimum number of edges of a path connecting A and B in the 1-skeleton of Λ_n. Also define the distance to the set of permutation matrices as follows

$$d_{\mathcal{P}_n}(A) = \min\{d(A, P) : P \in \mathcal{P}_n\} \quad (A \in \mathcal{A}_n).$$

So, $d_{\mathcal{P}_n}(A) = 0$ if and only if A is a permutation matrix. For the ASM A in Example 7 we have $d_{\mathcal{P}_n}(A) = 1$ as $d(A, P_2) = 1$.

Theorem 26 *Let A be an $n \times n$ ASM, and let $\sigma_{neg}(A)$ be the number of (-1)'s in A. Then*

$$d_{\mathcal{P}_n}(A) \leq \sigma_{neg}(A).$$

Proof. If $\sigma_{neg}(A) = 0$, then $A = [a_{ij}]$ is a permutation matrix and the result is clear, so assume $\sigma_{neg}(A) \geq 1$. We shall construct an ASM B which is adjacent to A, so $d(A, B) = 1$, and satisfies $\sigma_{neg}(B) = \sigma_{neg}(A) - 1$.

Choose a position (i_1, j_1) with $a_{i_1 j_1} = -1$ and such that there is no $(i, j) \neq (i_1, j_1)$ with $i \geq i_1$, $j \geq j_1$, and $a_{ij} = -1$. Then there is a unique $i_2 > i_1$ with $a_{i_2 j_1} = 1$. Moreover, $a_{i_2 j} = 0$ for each $j > j_1$ (by the ASM property and since no such entry can be -1). Similarly, there is a unique $j_2 > j_1$ with $a_{i_1 j_2} = 1$. Then $a_{ij_2} = 0$ for each $i > i_1$. In particular, $a_{i_2 j_2} = 0$.

Let $B = [b_{ij}]$ be obtained from A by defining

$$b_{i_1 j_1} = b_{i_2 j_1} = b_{i_1 j_2} = 0, \ b_{i_2 j_2} = 1,$$

and $b_{ij} = a_{ij}$ for all other (i, j). Then B is an ASM. In fact, in row i_1 we removed the last two nonzeros, so this row is still alternating and has sum 1. In row i_2 we just moved the last nonzero, a 1, from column j_1 to column $j_2 > j_1$, so this row is still alternating and has sum 1. Similar arguments are valid for columns j_1 and j_2. No other rows or columns are changed, so this shows that B is an ASM. We also have $\sigma_{neg}(B) = \sigma_{neg}(A) - 1$. Moreover, the graph $G(A, B)$ contains a single cycle, corresponding to the rectangle given by the four vertices (i_1, j_1), (i_2, j_1), (i_1, j_2), (i_2, j_2). Therefore, by Theorem 21, A and B are adjacent.

Clearly, we may repeat this process, now applied to B, and eventually, after $\sigma_{neg}(A)$ steps we reach a permutation matrix. It follows that $d_{\mathcal{P}_n}(A) \leq \sigma_{neg}(A)$. $\qquad\square$

We next improve the bound in Theorem 26. For $1 \leq i, j \leq n$ define $I_{ij} = \{(i', j') : i \leq i' \leq n, j \leq j' \leq n\} \setminus \{(i, j)\}$, which is the rectangle of position below and to the right of (i, j), except (i, j) itself. Let $A = [a_{ij}]$ be an $n \times n$ $(0, \pm 1)$-matrix A. Define

$$W(A) = \{(i, j) : a_{ij} = -1, a_{i'j'} \geq 0 \ ((i', j') \in I_{ij})\}.$$

Let A^+ be the matrix obtained from A by changing the entry in each position $(i, j) \in W(A)$ from -1 to 0. Then $\sigma_{neg}(A^+) = \sigma_{neg}(A) - |W(A)|$. We construct a sequence of matrices, starting with A and using this operation until we have a nonnegative matrix:

$$A_0 = A, \ A_1 = A_0^+, A_2 = A_1^+, \ldots, A_k = A_{k-1}^+ \geq O$$

This number k is uniquely determined by A, and we define $\kappa(A) = k$. If A is nonnegative, we let $\kappa(A) = 0$.

Theorem 27 *Let A be an $n \times n$ ASM. Then*

$$d_{\mathcal{P}_n}(A) \leq \kappa(A).$$

Proof. We use the same construction as in the proof of Theorem 26, with the following modifications. Let B be obtained from A as follows: for each $(i, j) \in W(A)$ replace the submatrix corresponding to I_{ij} by the zero matrix, and, finally, let $b_{nn} = 1$. Then B is an ASM, and $d(A, B) = 1$, because $G(A, B)$ contains the single cycle corresponding to the boundary of $\cup_{(i,j) \in W(A)} I_{ij}$.

Then we repeat the process for smaller matrix obtained from B by deleting the last row and column. This gives the desired inequality. □

As the referee pointed out the construction used in the proof of Theorems 26 and 27 is similar to the construction of the pseudo-key of an ASM and the key of an ASM as given in [2] and [15].

We remark that completely similar results to Theorem 27 may be obtained by symmetry. In Theorem 27 we considered modifications using sets I_{ij} defined by lower-right positions. But a similar construction works for lower-left, upper-right and upper-left index sets, or even for a combination of these methods.

Example 9 Let $n \geq 3$ and let A_n be the $n \times n$ ASM with (-1) in positions $(2, n-1)$, $(3, n-2)$, ..., $(n-1, 2)$, and 1's in the adjacent positions, while there are zeros elsewhere. For instance, the matrix A in Example 7 equals A_4.

Then we see that $W(A_n)$ contains all the positions of the (-1)'s in A, so $\kappa(A) = 1$. So, by Theorem 27, $d_{\mathcal{P}_n}(A_n) = 1$. □

References

[1] R.K. Ahuja, T.L. Magnanti, J.B. Orlin, Network Flows: Theory, Algorithms, and Applications, Prentice-Hall, Upper Saddle River, NJ, USA, 1993.

[2] J.C. Aval, Keys and alternating sign matrices, Sem. Lothar. Combin. 59 (2007), Art. B59f, pp13.

[3] R.E. Behrend, V.A. Knight, Higher spin alternating sign matrices, Electron. J. Combin. 14 (2007) #1.

[4] G.D. Birkhoff, Tres observaciones sobre el algebra lineal, Univ. Nac. Tucumán Rev. Ser. A, (1946) 147–151.

[5] R.A. Brualdi, Notes on the Birkhoff algorithm for doubly stochastic matrices, Canad. Math. Bull., 25 (1982), 191–199.

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

[7] R.A. Brualdi, H.J. Ryser, Combinatorial Matrix Theory, Cambridge University Press, Cambridge, 1991.

[8] R.A. Brualdi, K.P. Kiernan, S.A. Meyer, M.S. Schroeder, Patterns of Alternation Sign Matrices, Linear Alg. Appl. 438 (2013) 3967–3990.

[9] D. Bressoud, Proofs and Confirmations: The Story of the Alternating Sign Matrix Conjecture, Math. Association of America, Cambridge University Press, 1999.

[10] W. Bruns, J. Gubeladze, M. Henk, A. Martin, R. Weismantel, A counterexample to an integer analogue of Carathéodory's theorem, J. Reine Angew Math. 510 (1999) 179–185.

[11] W. Cook, J. Fonlupt, A. Schrijver, An integer analogue of Carathéodory's theorem, J. Combin. Theory, Ser. B 40 (1986) 63–70.

[12] G. Dahl, Tridiagonal doubly stochastic matrices, Linear Algebra Appl. 390 (2004) 197–208.

[13] R.G. Jeroslow, Some basis theorems for integral monoids, Math. Oper. Res. 3 (1978) 145–154.

[14] G. Kuperberg, Another proof of the alternating-sign matrix conjecture, Internat. Math. Res. Notices (1996), 139–150.

[15] A. Lascoux, Chern and Yang through Ice, Preprint, 2002.

[16] W.H. Mills, D.P. Robbins, H. Rumsey, Alternating-sign matrices and descending plane partitions, J. Combin. Theory Ser. A, 34 (1983) 340–359.

[17] M.A. Hadian Nadoshan, A. Armandnejad, B-majorization and its linear preservers, Linear Algebra Appl. 478 (2015) 218–227.

[18] A. Schrijver, Theory of Linear and Integer Programming, Wiley-Interscience, Chichester, 1986.

[19] Joel Spencer, private communication.

[20] J. Striker, The alternating sign matrix polytope, Electron. J. Combin. 16 #R41 (2009).

[21] J. von Neumann, A certain zero-sum two-person game equivalent to an optimal assignment problem, in: Contributions to the Theory of Games, II (H.W. Kuhn and A.W. Tucker, eds), Ann. Math. Studies 28 (1953) 5–12, Princeton University Press, Princeton, N.J.

[22] D. Zeilberger, Proof of the alternating sign matrix conjecture, Electron. J. Combin. 3 (1996), no. 2, Research Paper 13, 84 pages.