

19.7 Birkhoff's Theorem

Chapter 18 §08 met the doubly stochastic matrices, the nonnegative matrices whose rows and columns all add up to 1. It showed that every permutation matrix is an extreme point of the set Ω_n they form, stated Birkhoff's theorem that there are no other extreme points, proved that this is equivalent to "every doubly stochastic matrix is an average of permutation matrices", and left the proof to this chapter. This section supplies it. The proof is a count: a corner of Ω_n has so many zero entries that some row holds a single 1, and deleting that row and its column leaves a corner of Ω_{n-1} . We then bound how many permutation matrices an average needs, turn the theorem into a procedure for finding them, and use it to give the second proof of Schur's majorization theorem that Chapter 18 §08 sketched.

Throughout, the field is $\mathbb{R}$, and inequalities between matrices or vectors are entrywise, as in §01. Recall Definition 18.8.4: $\mathbf{M} \in M_n(\mathbb{R})$ is **doubly stochastic** if $m_{ij} \geq 0$ for all i, j , and every row and every column adds up to 1; in terms of the all-ones vector, $\mathbf{M}\mathbf{1} = \mathbf{1}$ and $\mathbf{1}^\top \mathbf{M} = \mathbf{1}^\top$. So $\mathbf{M}$ is stochastic in the book's column sense (Definition 9.11.1), and so is $\mathbf{M}^\top$.

19.7.1 What is left to prove

Here is where Chapter 18 §08 stopped. The set Ω_n is convex and compact. The $n!$ permutation matrices $\mathbf{P}_\sigma$, $\sigma \in S_n$ (Definition 3.4.2), lie in it, and each is an extreme point (Proposition 18.8.9). The two forms of Birkhoff's theorem, "the extreme points are the permutation matrices" and " Ω_n is the convex hull of the permutation matrices", were shown to be equivalent there, using Minkowski's theorem (Theorem 18.8.5) in one direction and Proposition 18.8.2 in the other. For $n = 2$, the Quick check after the statement found Ω_2 by hand: it is the segment from $\mathbf{I}$ to the swap matrix.

So one implication is missing: **every extreme point of Ω_n is a permutation matrix**. For $n \geq 3$ nothing as simple as the 2×2 computation is available, since Ω_3 already has four free parameters. What does the work instead is a result of Chapter 18 §09 that recognizes a corner of a polyhedron from its tight constraints.

19.7.2 The doubly stochastic matrices form a polyhedron

Recall Definition 18.9.2 and Theorem 18.9.2. A polyhedron is a set $\{\mathbf{x} \in \mathbb{R}^N : \mathbf{A}\mathbf{x} \leq \mathbf{b}\}$; its extreme points are called vertices; and a point $\mathbf{x}$ of it is a vertex exactly when the rows $\mathbf{a}_i^\top$ of the constraints that hold with equality at $\mathbf{x}$, the **active** rows, include N linearly independent ones. An equation is allowed, as a pair of inequalities $\mathbf{a}^\top \mathbf{x} \leq \beta$ and $-\mathbf{a}^\top \mathbf{x} \leq -\beta$; both are active at every point of the polyhedron.

To put Ω_n in this form, list the n^2 entries of a matrix $\mathbf{X}$ as one column, the vectorization $\text{vec } \mathbf{X} \in \mathbb{R}^{n^2}$ of Chapter 8. A linear expression $\sum_{i,j} b_{ij}x_{ij}$ in the entries is then the dot product of $\text{vec } \mathbf{X}$ with $\text{vec } \mathbf{B}$, and we may treat the coefficient matrix $\mathbf{B} \in M_n(\mathbb{R})$ itself as the row of the constraint. Three families of coefficient matrices appear. With $\mathbf{E}_{ij}$ the matrix unit (a 1 in position (i, j) , zeros elsewhere), let

$$\mathbf{R}_i = \mathbf{e}_i \mathbf{1}^\top, \quad \mathbf{C}_j = \mathbf{1} \mathbf{e}_j^\top \quad (1 \leq i, j \leq n),$$

the matrices whose i -th row, respectively j -th column, consists of ones and whose other entries are zero. The letter $\mathbf{C}_j$ is local to this section; elsewhere $\mathbf{C}(p)$ is a companion matrix and $\mathbf{C}_k(\lambda)$

a real Jordan block. Then $\sum_{k,l}(\mathbf{R}_i)_{kl}x_{kl}$ is the i -th row sum of $\mathbf{X}$, and $\mathbf{C}_j$ gives the j -th column sum. So $\Omega_n \subseteq M_n(\mathbb{R}) \cong \mathbb{R}^{n^2}$ is the polyhedron cut out by

- (i) the n^2 inequalities $-x_{ij} \leq 0$, with rows $-\mathbf{E}_{ij}$;
- (ii) the n row equations “row sum i equals 1”, with rows $\pm\mathbf{R}_i$;
- (iii) the n column equations “column sum j equals 1”, with rows $\pm\mathbf{C}_j$.

At a point $\mathbf{M} \in \Omega_n$, every row $\pm\mathbf{R}_i$ and $\pm\mathbf{C}_j$ is active, and $-\mathbf{E}_{ij}$ is active exactly when $m_{ij} = 0$. The whole argument turns on how much the $2n$ equations are worth. They are not independent: adding all row sums and adding all column sums both give the sum of all entries, that is,

$$\mathbf{R}_1 + \cdots + \mathbf{R}_n = \mathbf{J} = \mathbf{C}_1 + \cdots + \mathbf{C}_n,$$

with $\mathbf{J}$ the all-ones matrix. The next lemma says that this is the only relation.

Lemma 19.7.1. The Row and Column Sum Conditions

Let $n \geq 1$, and let

$$W_n = \{\mathbf{X} \in M_n(\mathbb{R}) : \mathbf{X}\mathbf{1} = \mathbf{0}, \mathbf{1}^\top \mathbf{X} = \mathbf{0}^\top\}$$

be the space of real matrices whose rows and columns all add up to 0. Then:

- (a) $\dim W_n = (n-1)^2$;
- (b) the $2n$ matrices $\mathbf{R}_1, \dots, \mathbf{R}_n, \mathbf{C}_1, \dots, \mathbf{C}_n$ span a subspace of $M_n(\mathbb{R})$ of dimension $2n-1$.

Idea. A matrix in W_n is determined by its top-left $(n-1) \times (n-1)$ block, which can be anything: the last column and the last row are then forced by the zero sums, and the corner comes out consistent. That gives (a). For (b), W_n is exactly the set of matrices killed by the $2n$ sum conditions, so rank-nullity converts one count into the other.

Proof.

- (a) If $n = 1$, $W_1 = \{\mathbf{0}\}$ and $(n-1)^2 = 0$. Let $n \geq 2$, and let $\varphi: W_n \rightarrow M_{n-1}(\mathbb{R})$ delete the last row and the last column; it is linear. It is injective: if $\mathbf{X} \in W_n$ has top-left block zero, then for $i < n$ the i -th row sum is $x_{in} = 0$, for $j < n$ the j -th column sum is $x_{nj} = 0$, and then the last row sum is $x_{nn} = 0$. So $\mathbf{X} = \mathbf{0}$. It is surjective: given $\mathbf{B} \in M_{n-1}(\mathbb{R})$, with row sums r_i , column sums c_j , and total $s = \sum_i r_i = \sum_j c_j$, let $\mathbf{X}$ have top-left block $\mathbf{B}$, last column $(-r_1, \dots, -r_{n-1}, s)$, and last row $(-c_1, \dots, -c_{n-1}, s)$. The i -th row sum for $i < n$ is $r_i - r_i = 0$, the j -th column sum for $j < n$ is $c_j - c_j = 0$, and the last row and last column both sum to $-s + s = 0$. So $\mathbf{X} \in W_n$ and $\varphi(\mathbf{X}) = \mathbf{B}$. Hence φ is an isomorphism and $\dim W_n = \dim M_{n-1}(\mathbb{R}) = (n-1)^2$.
- (b) Let $\mathbf{K}$ be the $2n \times n^2$ matrix whose rows are the transposed vectorizations $(\text{vec } \mathbf{R}_i)^\top$ and $(\text{vec } \mathbf{C}_j)^\top$, for $1 \leq i, j \leq n$. Then $\mathbf{K} \text{vec } \mathbf{X}$ lists the row sums and the column sums of $\mathbf{X}$, so the null space of $\mathbf{K}$ is $\text{vec}(W_n)$, of dimension $(n-1)^2$ by (a), as vec is an isomorphism. By Theorem 3.5.7, $\text{rank } \mathbf{K} = n^2 - (n-1)^2 = 2n-1$, and by Theorem 3.5.5 this is the dimension of the span of the rows of $\mathbf{K}$. Since vec is an isomorphism, the $2n$ matrices $\mathbf{R}_i, \mathbf{C}_j$ span a subspace of dimension $2n-1$, as claimed.

■

So of the $n^2 + 4n$ constraints defining Ω_n , the $4n$ that come from equations contribute only $2n - 1$ dimensions to the active rows, wherever we stand. The rest must come from zero entries.

Quick check. A vertex of Ω_3 needs 9 linearly independent active rows. How many zero entries must it therefore have, at least? Does the identity matrix $\mathbf{I}_3$ meet this requirement?

Solution. The row and column equations supply at most $2 \cdot 3 - 1 = 5$ independent rows, by Lemma 19.7.1 (b). Each zero entry adds one row $-\mathbf{E}_{ij}$, so at least $9 - 5 = 4$ zero entries are needed; that is, at most 5 nonzero entries. The identity has 6 zero entries and 3 nonzero ones, so it meets the requirement with room to spare.

19.7.3 Birkhoff's theorem

Now the main theorem. We state both forms, since both are used, but prove only the missing implication.

Theorem 19.7.2. Birkhoff's Theorem

Let $n \geq 1$. The extreme points of Ω_n are exactly the $n!$ permutation matrices of size n . Equivalently, every doubly stochastic matrix is a convex combination of permutation matrices:

$$\Omega_n = \text{conv}\{\mathbf{P}_\sigma : \sigma \in S_n\}.$$

Idea. Step roadmap. Let $\mathbf{M}$ be an extreme point of Ω_n and induct on n .

① **Count.** $\mathbf{M}$ is a vertex of the polyhedron above, so it has n^2 independent active rows. At most $2n - 1$ of them come from the equations, so at least $n^2 - (2n - 1)$ entries of $\mathbf{M}$ are zero, and at most $2n - 1$ are nonzero.

② **Pigeonhole.** Every row has a nonzero entry, since it adds up to 1. With n rows and fewer than $2n$ nonzero entries, some row has exactly one, and that entry is 1. Its column then has no room for anything else.

③ **Strip.** Delete that row and column. What is left is doubly stochastic of size $n - 1$, and, the key point, it is again **extreme**: any way of splitting it would split $\mathbf{M}$.

④ **Induct.** The smaller matrix is a permutation matrix, and putting the 1 back gives one of size n .

Proof. By Proposition 18.8.9, every permutation matrix is an extreme point of Ω_n . Distinct permutations give distinct matrices, since column j of $\mathbf{P}_\sigma$ is $\mathbf{e}_{\sigma(j)}$, so there are $n!$ of them. The equivalence of the two forms was proved in Chapter 18 §08, from Theorem 18.8.5 and Proposition 18.8.2. It remains to show that every extreme point of Ω_n is a permutation matrix. We prove this by induction on $n \geq 1$.

For $n = 1$, $\Omega_1 = \{(1)\}$, and $(1) = \mathbf{P}_{\text{id}}$.

Let $n \geq 2$, assume the statement for $n - 1$, and let $\mathbf{M}$ be an extreme point of Ω_n .

Step 1: at most $2n - 1$ nonzero entries. As shown above, Ω_n is a polyhedron in $M_n(\mathbb{R}) \cong \mathbb{R}^{n^2}$, and $\mathbf{M}$ is one of its vertices (Definition 18.9.2). By Theorem 18.9.2, the active rows at $\mathbf{M}$ include n^2 linearly independent ones. Let z be the number of zero entries of $\mathbf{M}$. The active rows are $\pm \mathbf{R}_i$, $\pm \mathbf{C}_j$, and $-\mathbf{E}_{ij}$ for the z positions with $m_{ij} = 0$. All of them lie in the subspace

$$\text{span}\{\mathbf{R}_1, \dots, \mathbf{R}_n, \mathbf{C}_1, \dots, \mathbf{C}_n\} + \text{span}\{\mathbf{E}_{ij} : m_{ij} = 0\},$$

whose dimension is at most $(2n - 1) + z$, by Lemma 19.7.1 (b) and the dimension formula for sums (Theorem 1.7.3). A subspace containing n^2 linearly independent vectors has dimension at least n^2 , by Theorem 1.6.3 (a) applied in that subspace. Hence $n^2 \leq 2n - 1 + z$, and the number of nonzero entries of $\mathbf{M}$ is $n^2 - z \leq 2n - 1$.

Step 2: a row with a single 1. Each row of $\mathbf{M}$ adds up to 1, so it has at least one nonzero entry. If every row had at least two, $\mathbf{M}$ would have at least $2n$ nonzero entries, contradicting Step 1. So some row i has exactly one nonzero entry, say m_{ij} , and the row sum gives $m_{ij} = 1$. Column j consists of nonnegative entries adding up to 1, one of which is $m_{ij} = 1$; so $m_{kj} = 0$ for every $k \neq i$.

Step 3: the reduced matrix is extreme. Let $\mathbf{M}' \in M_{n-1}(\mathbb{R})$ be $\mathbf{M}$ with row i and column j deleted. Its entries are nonnegative. Each of its rows is a row $k \neq i$ of $\mathbf{M}$ with the entry $m_{kj} = 0$ removed, so it still adds up to 1; likewise each of its columns is a column $l \neq j$ of $\mathbf{M}$ with $m_{il} = 0$ removed. So $\mathbf{M}' \in \Omega_{n-1}$. For $\mathbf{Y} \in M_{n-1}(\mathbb{R})$, let $E(\mathbf{Y}) \in M_n(\mathbb{R})$ be the matrix obtained by inserting a new row at position i and a new column at position j , with a 1 where they cross and zeros elsewhere along them. Then $E(\mathbf{M}') = \mathbf{M}$; $E(\mathbf{Y}) \in \Omega_n$ whenever $\mathbf{Y} \in \Omega_{n-1}$, since the inserted row and column add up to 1 and add 0 to every old row and column; $E(\mathbf{Y}) = E(\mathbf{Z})$ only if $\mathbf{Y} = \mathbf{Z}$; and $E(\frac{1}{2}(\mathbf{Y} + \mathbf{Z})) = \frac{1}{2}(E(\mathbf{Y}) + E(\mathbf{Z}))$, because the inserted entries are the same for $\mathbf{Y}$, $\mathbf{Z}$ and their midpoint. Now suppose $\mathbf{M}' = \frac{1}{2}(\mathbf{Y} + \mathbf{Z})$ with $\mathbf{Y}, \mathbf{Z} \in \Omega_{n-1}$. Then $\mathbf{M} = \frac{1}{2}(E(\mathbf{Y}) + E(\mathbf{Z}))$ with $E(\mathbf{Y}), E(\mathbf{Z}) \in \Omega_n$, and since $\mathbf{M}$ is extreme, $E(\mathbf{Y}) = E(\mathbf{Z})$, so $\mathbf{Y} = \mathbf{Z}$. By the midpoint test (Lemma 18.8.1), $\mathbf{M}'$ is an extreme point of Ω_{n-1} .

Step 4: induction. By the induction hypothesis, $\mathbf{M}'$ is a permutation matrix: it has exactly one 1 in each row and each column and zeros elsewhere. Then so does $\mathbf{M} = E(\mathbf{M}')$: row i and column j contain only the inserted 1, and every other row and column is one of $\mathbf{M}'$ with a 0 inserted. By the remark after Definition 3.4.2, $\mathbf{M}$ is a permutation matrix. This completes the induction and proves the theorem. ■

This pays Chapter 18 §08's promise. With Minkowski's theorem it describes Ω_n completely: a compact convex set with exactly $n!$ corners, each a permutation matrix.

Step 3 is not decoration. It is tempting to finish from Step 1 alone, by an induction on the number of nonzero entries, and that induction does not close.

Warning. Few nonzero entries do not by themselves make a permutation matrix. The matrix

$$\mathbf{N} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \in \Omega_3$$

has $5 = 2 \cdot 3 - 1$ nonzero entries, the most Step 1 allows, and it is not a permutation matrix; it is

the midpoint of $\mathbf{I}_3$ and the permutation matrix that swaps the last two coordinates. Deleting its first row and column leaves $\frac{1}{2}\mathbf{J}_2$, with 4 nonzero entries, more than the $2 \cdot 2 - 1 = 3$ that a vertex of Ω_2 may have. The count must be applied afresh at each size, and that needs the smaller matrix to be extreme, which is what Step 3 proves.

19.7.4 How many permutation matrices

Birkhoff's theorem writes every doubly stochastic matrix as an average of permutation matrices, but a priori with up to $n!$ terms. Carathéodory's theorem (Theorem 18.2.2) cuts this down to one more than the dimension, and the dimension is the one computed in Lemma 19.7.1.

Corollary 19.7.3. Birkhoff's Theorem with a Count

Let $n \geq 1$. Every $\mathbf{M} \in \Omega_n$ is a convex combination of at most $(n - 1)^2 + 1$ permutation matrices.

Idea. All of Ω_n lies in the coset of matrices with row and column sums 1, which is $\frac{1}{n}\mathbf{J} + W_n$, of dimension $(n - 1)^2$. Translate to W_n and apply Carathéodory's theorem there.

Proof. Let $\mathbf{M} \in \Omega_n$. By Theorem 19.7.2 and Theorem 18.1.3, $\mathbf{M} = \sum_{k=1}^m t_k \mathbf{P}_k$ for some permutation matrices $\mathbf{P}_k$ and weights $(t_1, \dots, t_m) \in \Delta_m$. Put $\mathbf{X}_0 = \frac{1}{n}\mathbf{J} \in \Omega_n$. Every doubly stochastic matrix $\mathbf{X}$ satisfies $(\mathbf{X} - \mathbf{X}_0)\mathbf{1} = \mathbf{1} - \mathbf{1} = \mathbf{0}$ and $\mathbf{1}^\top(\mathbf{X} - \mathbf{X}_0) = \mathbf{0}^\top$, so $\mathbf{X} - \mathbf{X}_0 \in W_n$. As the weights add up to 1,

$$\mathbf{M} - \mathbf{X}_0 = \sum_{k=1}^m t_k (\mathbf{P}_k - \mathbf{X}_0),$$

so $\mathbf{M} - \mathbf{X}_0$ lies in the convex hull of $S = \{\mathbf{P}_\sigma - \mathbf{X}_0 : \sigma \in S_n\}$, a subset of the real vector space W_n , which has dimension $(n - 1)^2$ by Lemma 19.7.1 (a). By Theorem 18.2.2, $\mathbf{M} - \mathbf{X}_0 = \sum_{k=1}^r s_k (\mathbf{P}'_k - \mathbf{X}_0)$ for some $r \leq (n - 1)^2 + 1$, permutation matrices $\mathbf{P}'_k$, and weights $(s_1, \dots, s_r) \in \Delta_r$. Adding $\mathbf{X}_0 = \sum_k s_k \mathbf{X}_0$ back gives $\mathbf{M} = \sum_{k=1}^r s_k \mathbf{P}'_k$, as claimed. ■

For $n = 2$ the bound is 2, and every point strictly inside the segment Ω_2 needs both ends. For $n = 3$, Exercise C1 shows that 5 can be needed. That $(n - 1)^2 + 1$ is best possible for every $n \geq 2$ is a known result which we state without proof. Compare the density matrices of Chapter 18 §08, where the spectral theorem did much better than Carathéodory's count. Here there is no such structure to exploit.

19.7.5 Finding the permutations

The proof of Birkhoff's theorem says that the permutation matrices suffice. It does not say how to find them. One consequence of the theorem does, because it only concerns which entries are zero.

Corollary 19.7.4. A Positive Diagonal

Let $\mathbf{M} \in \Omega_n$. Then there is a permutation $\sigma \in S_n$ with $m_{\sigma(j)j} > 0$ for **every** $j = 1, \dots, n$.

Proof. By Theorem 19.7.2, $\mathbf{M} = \sum_{\tau \in S_n} t_\tau \mathbf{P}_\tau$ with every $t_\tau \geq 0$ and $\sum_\tau t_\tau = 1$, so some $t_\sigma > 0$. The $(\sigma(j), j)$ entry of $\mathbf{P}_\sigma$ is 1, and every term of the sum has nonnegative entries. Hence $m_{\sigma(j)j} \geq t_\sigma > 0$ for every j . ■

In words, one can pick one positive entry in each column, all in different rows. This gives a procedure, often called **peeling**. Given $\mathbf{M} \in \Omega_n$, choose σ as in the corollary and let $t = \min_j m_{\sigma(j)j}$, so $0 < t \leq 1$.

- If $t = 1$, then each column j has its whole sum 1 at position $\sigma(j)$, and $\mathbf{M} = \mathbf{P}_\sigma$.
- If $t < 1$, then $\mathbf{M} - t\mathbf{P}_\sigma$ has nonnegative entries and all row and column sums $1 - t$, so $\mathbf{M}_1 = (\mathbf{M} - t\mathbf{P}_\sigma)/(1 - t) \in \Omega_n$ and $\mathbf{M} = t\mathbf{P}_\sigma + (1 - t)\mathbf{M}_1$. Every zero of $\mathbf{M}$ is a zero of $\mathbf{M}_1$, since $\mathbf{P}_\sigma$ has its 1's at positive entries of $\mathbf{M}$, and the entry where the minimum t was attained becomes a new zero.

Repeat with $\mathbf{M}_1$, producing $\mathbf{M}_2, \mathbf{M}_3, \dots$, and write $\mathbf{M}_0 = \mathbf{M}$. A doubly stochastic matrix has at least n nonzero entries, one per column, and each round removes at least one, so the procedure stops after at most $n^2 - n + 1$ rounds, with a permutation matrix. Unwinding the rounds writes $\mathbf{M}$ as a convex combination of the permutation matrices found. In fact peeling never needs more than the $(n - 1)^2 + 1$ of Corollary 19.7.3. Let U_k be the subspace of those $\mathbf{X} \in W_n$ with $x_{ij} = 0$ wherever $\mathbf{M}_k$ has a zero; since zeros of $\mathbf{M}_k$ stay zeros, $U_{k+1} \subseteq U_k$, and $\mathbf{M}_k - \mathbf{M}_{k+1}$ lies in U_k but not in U_{k+1} , being positive at the new zero. So each round with $t < 1$ lowers $\dim U_k$, which starts at most $\dim W_n = (n - 1)^2$ by Lemma 19.7.1 (a) and never drops below 0; there are at most $(n - 1)^2$ such rounds, and one final round. Still, peeling finds a decomposition, not necessarily the shortest one.

Example. Peeling a 4×4 matrix

Write

$$\mathbf{M} = \frac{1}{8} \begin{pmatrix} 4 & 2 & 2 & 0 \\ 1 & 3 & 0 & 4 \\ 3 & 0 & 4 & 1 \\ 0 & 3 & 2 & 3 \end{pmatrix}$$

as a convex combination of permutation matrices.

Solution. The entries are nonnegative, each row of $8\mathbf{M}$ adds up to 8, and so does each column: $4 + 1 + 3 + 0$, $2 + 3 + 0 + 3$, $2 + 0 + 4 + 2$, $0 + 4 + 1 + 3$. So $\mathbf{M} \in \Omega_4$. We peel $\mathbf{N} = 8\mathbf{M}$, whose row and column sums start at 8; subtracting w times a permutation matrix lowers them all by w . A permutation is recorded by the positions $(\sigma(1), 1), \dots, (\sigma(4), 4)$ of its 1's, one per column.

Round 1. The diagonal of $\mathbf{N}$ is 4, 3, 4, 3, all positive, so take $\sigma_1 = \text{id}$ with weight $w_1 = 3$.

Subtracting $3\mathbf{I}$ leaves

$$\begin{pmatrix} 1 & 2 & 2 & 0 \\ 1 & 0 & 0 & 4 \\ 3 & 0 & 1 & 1 \\ 0 & 3 & 2 & 0 \end{pmatrix} \quad (\text{sums } 5).$$

Round 2. Column 2 is now nonzero only in rows 1 and 4. Try row 4 there: the positions $(1, 1)$, $(4, 2)$, $(3, 3)$, $(2, 4)$ carry 1, 3, 1, 4, so take this σ_2 with weight $w_2 = 1$:

$$\begin{pmatrix} 0 & 2 & 2 & 0 \\ 1 & 0 & 0 & 3 \\ 3 & 0 & 0 & 1 \\ 0 & 2 & 2 & 0 \end{pmatrix} \quad (\text{sums } 4).$$

Round 3. The positions $(3, 1)$, $(1, 2)$, $(4, 3)$, $(2, 4)$ carry 3, 2, 2, 3; weight $w_3 = 2$:

$$\begin{pmatrix} 0 & 0 & 2 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 2 & 0 & 0 \end{pmatrix} \quad (\text{sums } 2).$$

Round 4. The positions $(2, 1)$, $(4, 2)$, $(1, 3)$, $(3, 4)$ carry 1, 2, 2, 1; weight $w_4 = 1$. What remains has a single 1 in positions $(3, 1)$, $(4, 2)$, $(1, 3)$, $(2, 4)$, a permutation matrix $\mathbf{P}_{\sigma_5}$, with weight $w_5 = 1$.

The weights $3 + 1 + 2 + 1 + 1 = 8$ account for the whole sum, and dividing by 8,

$$\mathbf{M} = \frac{3}{8} \mathbf{P}_{\sigma_1} + \frac{1}{8} \mathbf{P}_{\sigma_2} + \frac{2}{8} \mathbf{P}_{\sigma_3} + \frac{1}{8} \mathbf{P}_{\sigma_4} + \frac{1}{8} \mathbf{P}_{\sigma_5},$$

where $\sigma_1 = \text{id}$, and $\sigma_2, \sigma_3, \sigma_4, \sigma_5$ send 1, 2, 3, 4 to 1, 4, 3, 2, to 3, 1, 4, 2, to 2, 4, 1, 3, and to 3, 4, 1, 2 respectively. Five terms, within the bound $(4 - 1)^2 + 1 = 10$. To check, add up any entry: position $(1, 3)$ is covered by σ_4 and σ_5 , giving $\frac{1}{8} + \frac{1}{8} = \frac{2}{8}$, which is m_{13} .

Remark. The route can be run the other way. [Corollary 19.7.4](#) is a statement about the zero pattern of $\mathbf{M}$ alone, and it can be proved directly by combinatorics: by Hall's marriage theorem, or by the Frobenius–König theorem, which says that every $n \times n$ matrix in which each choice of one entry per column, in distinct rows, meets a zero has an $r \times s$ block of zeros with $r + s = n + 1$. With the corollary in hand first, peeling proves Birkhoff's theorem constructively. Both routes are standard; the one taken here uses only what Chapter 18 had already built.

19.7.6 Schur's theorem again

Chapter 17 §08 proved Schur's majorization theorem, Theorem 17.8.2: for a Hermitian $\mathbf{A}$, the diagonal vector $\mathbf{d}(\mathbf{A})$ is majorized by the eigenvalue vector $\boldsymbol{\lambda}(\mathbf{A})$. The proof there went through Ky Fan's maximum principle. Chapter 18 §08 noted that Birkhoff's theorem would give a second proof, and sketched it. Here it is in full. The first ingredient says that the diagonal is a doubly

stochastic image of the spectrum.

Lemma 19.7.5. The Diagonal Is a Doubly Stochastic Image of the Spectrum

Let $F = \mathbb{R}$ or $\mathbb{C}$, let $\mathbf{A} \in M_n(F)$ be Hermitian, and let $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^*$ with $\mathbf{U}$ unitary and $\mathbf{D} = \text{diag}(\lambda_1(\mathbf{A}), \dots, \lambda_n(\mathbf{A}))$. Then the matrix $\mathbf{S} = (|u_{ij}|^2)$ is doubly stochastic, and

$$\mathbf{d}(\mathbf{A}) = \mathbf{S}\boldsymbol{\lambda}(\mathbf{A}).$$

Such a $\mathbf{U}$ exists by Corollary 12.4.3, after reordering the columns of $\mathbf{U}$, since a Hermitian matrix is normal: the diagonal entries of the $\mathbf{D}$ it gives are the eigenvalues of $\mathbf{A}$ with multiplicity, because similar matrices have the same characteristic polynomial $p_{\mathbf{A}}$ (Theorem 7.10.3), and they are real (Theorem 12.2.1), so they can be put in decreasing order.

Proof. Write $\lambda_j = \lambda_j(\mathbf{A})$. By the definition of the matrix product, $a_{ii} = \sum_j u_{ij}\lambda_j\overline{u_{ij}} = \sum_j |u_{ij}|^2\lambda_j$, which is the i -th entry of $\mathbf{S}\boldsymbol{\lambda}(\mathbf{A})$. The entries of $\mathbf{S}$ are nonnegative. Row i of $\mathbf{S}$ adds up to $\sum_j |u_{ij}|^2 = (\mathbf{U}\mathbf{U}^*)_{ii} = 1$, and column j adds up to $\sum_i |u_{ij}|^2 = (\mathbf{U}^*\mathbf{U})_{jj} = 1$, since $\mathbf{U}$ is unitary. So $\mathbf{S} \in \Omega_n$. ■

A doubly stochastic matrix of the form $(|u_{ij}|^2)$ is called **unistochastic**. Not every doubly stochastic matrix is one (Exercise C3 asks for a witness), but the lemma needs only that $\mathbf{S} \in \Omega_n$. The second ingredient is Birkhoff's theorem, which makes $\mathbf{S}\boldsymbol{\lambda}$ an average of rearrangements of $\boldsymbol{\lambda}$.

Proposition 19.7.6. Doubly Stochastic Maps Even Out

Let $\mathbf{S} \in \Omega_n$ and $\mathbf{y} \in \mathbb{R}^n$. Then $\mathbf{S}\mathbf{y} \prec \mathbf{y}$.

Idea. A permutation matrix only rearranges $\mathbf{y}$, and rearrangements are invisible to majorization. Birkhoff's theorem writes $\mathbf{S}\mathbf{y}$ as an average of rearrangements, and the vectors majorized by $\mathbf{y}$ form a convex set.

Proof. By Theorem 19.7.2, $\mathbf{S} = \sum_{k=1}^m t_k \mathbf{P}_{\sigma_k}$ with $(t_1, \dots, t_m) \in \Delta_m$, so $\mathbf{S}\mathbf{y} = \sum_k t_k \mathbf{P}_{\sigma_k} \mathbf{y}$. Each $\mathbf{P}_{\sigma} \mathbf{y} = \sum_j y_j \mathbf{e}_{\sigma(j)}$ has entry y_j in position $\sigma(j)$, so it is a rearrangement of $\mathbf{y}$. As observed after Definition 17.8.1, majorization sees only which numbers occur and how often, so $\mathbf{y} \prec \mathbf{y}$ (Proposition 17.8.1 (a)) gives $\mathbf{P}_{\sigma} \mathbf{y} \prec \mathbf{y}$. By Example (The vectors majorized by a fixed vector) the set $\{\mathbf{x} : \mathbf{x} \prec \mathbf{y}\}$ is convex, and a convex set contains every convex combination of its points (Lemma 18.1.2). Hence $\mathbf{S}\mathbf{y} \prec \mathbf{y}$. ■

Second proof of Theorem 17.8.2. Let $\mathbf{A}$ be Hermitian. By Lemma 19.7.5, $\mathbf{d}(\mathbf{A}) = \mathbf{S}\boldsymbol{\lambda}(\mathbf{A})$ with $\mathbf{S} \in \Omega_n$, and by Proposition 19.7.6, $\mathbf{S}\boldsymbol{\lambda}(\mathbf{A}) \prec \boldsymbol{\lambda}(\mathbf{A})$. So $\mathbf{d}(\mathbf{A}) \prec \boldsymbol{\lambda}(\mathbf{A})$. $\square$

The first proof compared running totals using a variational principle; this one never compares running totals directly. It sees Schur's theorem as a statement about averaging: the diagonal

entries are weighted averages of the eigenvalues, with weights that form a doubly stochastic matrix, and averaging can only even out a vector.

Example. The averaging made visible

For

$$\mathbf{A} = \begin{pmatrix} 3 & 2 & 0 \\ 2 & 4 & 2 \\ 0 & 2 & 5 \end{pmatrix}$$

of Chapter 17 §08, whose eigenvalues are $\lambda(\mathbf{A}) = (7, 4, 1)$ with unit eigenvectors $\mathbf{u}_1 = \frac{1}{3}(1, 2, 2)$, $\mathbf{u}_2 = \frac{1}{3}(-2, -1, 2)$, $\mathbf{u}_3 = \frac{1}{3}(2, -2, 1)$ (Example (Coordinate vectors against eigenvectors)), compute $\mathbf{S}$, check $\mathbf{d}(\mathbf{A}) = \mathbf{S}\lambda(\mathbf{A})$, and write $\mathbf{d}(\mathbf{A})$ as an average of rearrangements of $\lambda(\mathbf{A})$.

Solution. Take $\mathbf{U}$ with columns $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$; it is orthogonal and $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^\top$. Squaring its entries,

$$\mathbf{U} = \frac{1}{3} \begin{pmatrix} 1 & -2 & 2 \\ 2 & -1 & -2 \\ 2 & 2 & 1 \end{pmatrix}, \quad \mathbf{S} = \frac{1}{9} \begin{pmatrix} 1 & 4 & 4 \\ 4 & 1 & 4 \\ 4 & 4 & 1 \end{pmatrix}.$$

Every row and column of $9\mathbf{S}$ adds up to 9. Then

$$\mathbf{S}\lambda(\mathbf{A}) = \frac{1}{9} \begin{pmatrix} 7 + 16 + 4 \\ 28 + 4 + 4 \\ 28 + 16 + 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix} = \mathbf{d}(\mathbf{A}).$$

To decompose $\mathbf{S}$, let $\mathbf{P}$ be the permutation matrix of $\sigma = (1\ 2\ 3)$, $\sigma(1) = 2$, $\sigma(2) = 3$, $\sigma(3) = 1$. Then $\mathbf{P} + \mathbf{P}^2 = \mathbf{J} - \mathbf{I}$, since $\mathbf{P}$ and $\mathbf{P}^2$ have their 1's in the off-diagonal positions and never in the same place, so

$$\mathbf{S} = \frac{1}{9}\mathbf{I} + \frac{4}{9}\mathbf{P} + \frac{4}{9}\mathbf{P}^2.$$

With $\mathbf{P}\lambda = (1, 7, 4)$ and $\mathbf{P}^2\lambda = (4, 1, 7)$, the cyclic shifts of $(7, 4, 1)$,

$$\mathbf{d}(\mathbf{A}) = \frac{1}{9}(7, 4, 1) + \frac{4}{9}(1, 7, 4) + \frac{4}{9}(4, 1, 7) = (3, 4, 5).$$

Each diagonal entry is an average of all three eigenvalues, with weight $\frac{1}{9}$ on one of them and $\frac{4}{9}$ on the other two. That is why the diagonal $(3, 4, 5)$ is so much flatter than $(7, 4, 1)$.

Together with Theorem 17.8.5, the lemma and the proposition already show that the diagonals of the Hermitian matrices with spectrum λ are exactly the vectors $\mathbf{S}\lambda$ with $\mathbf{S} \in \Omega_n$. The lemma puts every such diagonal among the $\mathbf{S}\lambda$; the proposition puts every $\mathbf{S}\lambda$ among the $\mathbf{x} \prec \lambda$; and Schur–Horn realizes every $\mathbf{x} \prec \lambda$ as a diagonal. So all three sets coincide, and in particular every $\mathbf{x} \prec \lambda$ is $\mathbf{S}\lambda$ for some $\mathbf{S} \in \Omega_n$. This converse of the proposition is part of the theorem of Hardy, Littlewood, and Pólya, which Chapter 21 takes up.

19.7.7 Exercises

19.7.7.1 A. Check your understanding

Exercise. A1

- (a) State Birkhoff's theorem in both forms.
- (b) Name the results of Chapter 18 that the proof of [Theorem 19.7.2](#) uses, and say what each contributes.
- (c) What upper bound does Step 1 of the proof give for the number of nonzero entries of a vertex of Ω_4 ? How many does a vertex actually have?
- (d) True or false: every matrix in Ω_n with at most $2n - 1$ nonzero entries is a permutation matrix. Justify your answer.
- (e) True or false: the product of two doubly stochastic matrices is doubly stochastic. Justify your answer.
- (f) True or false: every $\mathbf{M} \in \Omega_3$ is a convex combination of at most five permutation matrices. Justify your answer.

Solution.

- (a) The extreme points of Ω_n are exactly the $n!$ permutation matrices; equivalently, Ω_n is the convex hull of the permutation matrices, that is, every doubly stochastic matrix is a convex combination of them.
- (b) Theorem 18.9.2, which says that a vertex has n^2 independent active constraints and so forces many zero entries; Lemma 18.8.1, the midpoint test, used to show that the stripped matrix is again extreme; and Proposition 18.8.9 with the equivalence of the two forms, which rests on Theorem 18.8.5 and Proposition 18.8.2.
- (c) At most $2 \cdot 4 - 1 = 7$. By [Theorem 19.7.2](#) a vertex is a permutation matrix, with exactly 4 nonzero entries. The bound of Step 1 is far from sharp; it is only the first step of an induction.
- (d) False. The matrix $\mathbf{N}$ in the warning after [Theorem 19.7.2](#) lies in Ω_3 and has $5 = 2 \cdot 3 - 1$ nonzero entries, but it is not a permutation matrix.
- (e) True. If $\mathbf{A}, \mathbf{B} \in \Omega_n$, the entries of $\mathbf{AB}$ are sums of products of nonnegative numbers, and $\mathbf{AB}\mathbf{1} = \mathbf{A}\mathbf{1} = \mathbf{1}$, $\mathbf{1}^\top \mathbf{AB} = \mathbf{1}^\top \mathbf{B} = \mathbf{1}^\top$.
- (f) True, by [Corollary 19.7.3](#) with $n = 3$: $(3 - 1)^2 + 1 = 5$.

19.7.7.2 B. Practice**Exercise. B1: Which are corners?**

For each matrix, determine whether it lies in Ω_3 and, if so, whether it is an extreme point of Ω_3 . Justify your answers.

$$\mathbf{M}_1 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad \mathbf{M}_2 = \frac{1}{2} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix},$$

$$\mathbf{M}_3 = \frac{1}{2} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{M}_4 = \begin{pmatrix} 1 & 1 & -1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Solution. $\mathbf{M}_1$ has exactly one 1 in each row and each column and zeros elsewhere, so it is a permutation matrix, in Ω_3 and extreme by Proposition 18.8.9.

$\mathbf{M}_2$ has nonnegative entries and every row and column of $2\mathbf{M}_2$ adds up to 2, so $\mathbf{M}_2 \in \Omega_3$. It is not a permutation matrix, so by Theorem 19.7.2 it is not extreme. Explicitly, $\mathbf{M}_2 = \frac{1}{2}\mathbf{I} + \frac{1}{2}\mathbf{P}$ with $\mathbf{P}$ the permutation matrix with 1's at $(1, 2), (2, 3), (3, 1)$, and $\mathbf{I} \neq \mathbf{P}$.

$\mathbf{M}_3$ is not in Ω_3 : its rows add up to $1, 1, \frac{1}{2}$.

$\mathbf{M}_4$ is not in Ω_3 : its rows and columns add up to 1 (rows 1, 1, 1; columns 1, 1, 1), but the entry -1 is negative. It is in the coset $\frac{1}{3}\mathbf{J} + W_3$ of the proof of Corollary 19.7.3, which shows that the nonnegativity conditions matter.

Exercise. B2: Peeling a 3×3 matrix

Let

$$\mathbf{M} = \frac{1}{6} \begin{pmatrix} 3 & 3 & 0 \\ 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}.$$

Check that $\mathbf{M} \in \Omega_3$, and write $\mathbf{M}$ as a convex combination of permutation matrices by peeling. Hence confirm the bound of Corollary 19.7.3 for this $\mathbf{M}$.

Solution. The entries are nonnegative; the rows of $6\mathbf{M}$ add up to $3+3+0, 1+2+3, 2+1+3$, and the columns to $3+1+2, 3+2+1, 0+3+3$, all 6. So $\mathbf{M} \in \Omega_3$. Peel $\mathbf{N} = 6\mathbf{M}$.

Round 1. The diagonal 3, 2, 3 is positive; subtract $2\mathbf{I}$:

$$\begin{pmatrix} 1 & 3 & 0 \\ 1 & 0 & 3 \\ 2 & 1 & 1 \end{pmatrix} \quad (\text{sums } 4).$$

Round 2. The positions $(3, 1), (1, 2), (2, 3)$ carry 2, 3, 3; subtract $2\mathbf{P}_1$, where $\mathbf{P}_1$ has its 1's there:

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \quad (\text{sums } 2).$$

Round 3. The positions $(1, 1), (3, 2), (2, 3)$ carry 1, 1, 1; subtract $\mathbf{P}_2$, with 1's there. What remains has 1's at $(2, 1), (1, 2), (3, 3)$: a permutation matrix $\mathbf{P}_3$, with weight 1.

So $6\mathbf{M} = 2\mathbf{I} + 2\mathbf{P}_1 + \mathbf{P}_2 + \mathbf{P}_3$ and

$$\mathbf{M} = \frac{1}{3}\mathbf{I} + \frac{1}{3}\mathbf{P}_1 + \frac{1}{6}\mathbf{P}_2 + \frac{1}{6}\mathbf{P}_3,$$

with weights nonnegative and adding up to 1. Hence $\mathbf{M}$ uses four permutation matrices, within the bound $(3-1)^2 + 1 = 5$.

Exercise. B3: Schur's theorem by averaging

Let $\mathbf{A} = \begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix}$.

- (a) Find $\lambda(\mathbf{A})$ and an orthogonal $\mathbf{U}$ with $\mathbf{A} = \mathbf{U} \operatorname{diag}(\lambda(\mathbf{A})) \mathbf{U}^\top$.
- (b) Compute $\mathbf{S} = (u_{ij}^2)$ and check that $\mathbf{d}(\mathbf{A}) = \mathbf{S}\lambda(\mathbf{A})$.
- (c) Write $\mathbf{S}$ as a convex combination of permutation matrices, and hence $\mathbf{d}(\mathbf{A})$ as an average of rearrangements of $\lambda(\mathbf{A})$. Confirm $\mathbf{d}(\mathbf{A}) \prec \lambda(\mathbf{A})$ from Definition 17.8.1.

Solution.

- (a) $\mathbf{A}(2, 1) = (10, 5) = 5(2, 1)$ and $\mathbf{A}(1, -2) = (0, 0)$, so $\lambda(\mathbf{A}) = (5, 0)$. With the unit eigenvectors $\frac{1}{\sqrt{5}}(2, 1)$ and $\frac{1}{\sqrt{5}}(1, -2)$ as columns,

$$\mathbf{U} = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 & 1 \\ 1 & -2 \end{pmatrix},$$

which is orthogonal, and $\mathbf{A} = \mathbf{U} \operatorname{diag}(5, 0) \mathbf{U}^\top$ by Corollary 12.5.2.

- (b) $\mathbf{S} = \frac{1}{5} \begin{pmatrix} 4 & 1 \\ 1 & 4 \end{pmatrix}$, and $\mathbf{S}(5, 0) = (4, 1) = \mathbf{d}(\mathbf{A})$.

- (c) $\mathbf{S} = \frac{4}{5}\mathbf{I} + \frac{1}{5} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, so $\mathbf{d}(\mathbf{A}) = \frac{4}{5}(5, 0) + \frac{1}{5}(0, 5) = (4, 1)$. Directly: both vectors are decreasing, the totals are $5 = 5$, and at $k = 1, 4 \leq 5$. So $\mathbf{d}(\mathbf{A}) \prec \lambda(\mathbf{A})$.

19.7.7.3 C. Going deeper**Exercise. C1: Five permutation matrices can be needed**

Let $\mathbf{P}$ be the permutation matrix of the 3-cycle $1 \mapsto 2 \mapsto 3 \mapsto 1$, and let $\mathbf{T}_{12}, \mathbf{T}_{13}, \mathbf{T}_{23}$ be those of the transpositions $(1\ 2), (1\ 3), (2\ 3)$. The six permutation matrices of size 3 are $\mathbf{I}, \mathbf{P}, \mathbf{P}^2$ (the **even** ones) and $\mathbf{T}_{12}, \mathbf{T}_{13}, \mathbf{T}_{23}$ (the **odd** ones).

- (a) Show that each of the nine positions (i, j) holds a 1 in exactly one even and exactly one odd permutation matrix, and that each even and each odd permutation matrix share exactly one position of a 1.
- (b) Deduce that if $\sum_{\sigma} c_{\sigma} \mathbf{P}_{\sigma} = \mathbf{O}$ for real numbers c_{σ} , then $c_{\sigma} = c$ for the even σ and $c_{\sigma} = -c$ for the odd σ , for some $c \in \mathbb{R}$.
- (c) Let $\mathbf{M} = \frac{1}{15}(\mathbf{I} + 2\mathbf{P} + 3\mathbf{P}^2 + 4\mathbf{T}_{12} + 5\mathbf{T}_{13})$. Show that every way of writing $\mathbf{M}$ as a convex combination of permutation matrices uses at least five of them with positive weight.

Hint: in (a), a permutation of $\{1, 2, 3\}$ is determined by where it sends any two points.

Solution.

- (a) A permutation matrix $\mathbf{P}_{\sigma}$ has a 1 in position (i, j) exactly when $\sigma(j) = i$. The permutations of $\{1, 2, 3\}$ with $\sigma(j) = i$ are determined by where they send the other

two points, which must go to the other two values in one of two ways, so there are exactly two. If σ is one of them, the other is $\tau\sigma$, with τ the transposition exchanging the two remaining values, so their signs are opposite (Definition 7.3.2, Theorem 7.3.1): one is even and one is odd. For the second claim, let σ be even and τ odd. They cannot agree at all three points, since they differ; they cannot agree at exactly two, since two agreements force the third; and if they agreed nowhere, $\tau^{-1}\sigma$ would fix no point and so be a 3-cycle, which is even, while $\tau^{-1}\sigma$ is odd. So they agree at exactly one point j , that is, they share exactly one position $(\sigma(j), j)$.

- (b) Look at position (i, j) : by (a) exactly one even σ and one odd τ have a 1 there, so the equation gives $c_\sigma + c_\tau = 0$. By (a) again, every pair of an even σ and an odd τ arises in this way at their shared position, so $c_\sigma = -c_\tau$ for every such pair. Fixing one odd τ , all even c_σ equal $c = -c_\tau$, and then every odd $c_{\tau'} = -c$.
- (c) The given weights are $\frac{1}{15}(1, 2, 3)$ on $\mathbf{I}, \mathbf{P}, \mathbf{P}^2$ and $\frac{1}{15}(4, 5, 0)$ on $\mathbf{T}_{12}, \mathbf{T}_{13}, \mathbf{T}_{23}$; they are nonnegative with sum 1. If $\mathbf{M} = \sum_\sigma t_\sigma \mathbf{P}_\sigma$ is any other convex combination, the difference of the two weight lists satisfies (b), so the even weights are $\frac{1}{15}(1, 2, 3) + s$ and the odd ones $\frac{1}{15}(4, 5, 0) - s$ for some real s . Nonnegativity needs $s \geq -\frac{1}{15}$ (from $\mathbf{I}$) and $s \leq 0$ (from $\mathbf{T}_{23}$). For $-\frac{1}{15} < s < 0$, all six weights are positive. At $s = 0$ exactly one vanishes, that of $\mathbf{T}_{23}$; at $s = -\frac{1}{15}$ exactly one vanishes, that of $\mathbf{I}$, the others being $\frac{1}{15}(1, 2)$ and $\frac{1}{15}(5, 6, 1)$. So every representation has at least five positive weights, and the bound $(3 - 1)^2 + 1 = 5$ of Corollary 19.7.3 is attained.

Exercise. C2: Integer matrices with equal line sums

Let $k \geq 1$, and let $\mathbf{N} \in M_n(\mathbb{R})$ have nonnegative **integer** entries, with every row and every column adding up to k .

- (a) Prove that $\mathbf{N}$ is a sum of k permutation matrices, not necessarily distinct.

- (b) Carry this out for $\mathbf{N} = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & 2 \\ 1 & 1 & 1 \end{pmatrix}$.

Hint: apply Corollary 19.7.4 to $\frac{1}{k}\mathbf{N}$.

Solution.

- (a) By induction on k . The matrix $\frac{1}{k}\mathbf{N}$ is doubly stochastic, so by Corollary 19.7.4 there is σ with $n_{\sigma(j)j} > 0$ for every j . These are positive integers, so each is at least 1, and $\mathbf{N} - \mathbf{P}_\sigma$ has nonnegative integer entries with every row and column adding up to $k - 1$. If $k = 1$, then $\mathbf{N} - \mathbf{P}_\sigma$ has line sums 0 and nonnegative entries, so it is $\mathbf{O}$ and $\mathbf{N} = \mathbf{P}_\sigma$. If $k \geq 2$, the induction hypothesis writes $\mathbf{N} - \mathbf{P}_\sigma$ as a sum of $k - 1$ permutation matrices, and adding $\mathbf{P}_\sigma$ gives k .
- (b) Every row and column adds up to 3. The diagonal 2, 1, 1 is positive; subtracting $\mathbf{I}$ leaves

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 2 \\ 1 & 1 & 0 \end{pmatrix}.$$

The positions $(1, 1), (3, 2), (2, 3)$ are positive; subtracting that permutation matrix leaves the permutation matrix with 1's at $(3, 1), (1, 2), (2, 3)$. So $\mathbf{N}$ is $\mathbf{I}$ plus these

two permutation matrices.

Exercise. C3: The dimension, and a matrix that is not unistochastic

Let $n \geq 2$.

(a) Prove that if $\mathbf{X} \in W_n$ has every $|x_{ij}| \leq \frac{1}{n}$, then $\frac{1}{n}\mathbf{J} + \mathbf{X} \in \Omega_n$. Deduce that $\text{aff } \Omega_n = \frac{1}{n}\mathbf{J} + W_n$ and $\dim \Omega_n = (n-1)^2$.

(b) Show that $\mathbf{M} = \frac{1}{2} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$ is doubly stochastic but is not of the form $(|u_{ij}|^2)$ for any unitary $\mathbf{U} \in M_3(\mathbb{C})$.

Hint for (b): two columns of $\mathbf{U}$ must be orthogonal; look at which rows both can be nonzero in.

Solution.

(a) The entries of $\frac{1}{n}\mathbf{J} + \mathbf{X}$ are $\frac{1}{n} + x_{ij} \geq \frac{1}{n} - \frac{1}{n} = 0$, and its rows and columns add up to $1 + 0 = 1$, since $\mathbf{X} \in W_n$. So it lies in Ω_n . The proof of [Corollary 19.7.3](#) showed $\Omega_n \subseteq \frac{1}{n}\mathbf{J} + W_n$, a coset, which is an affine set containing Ω_n , so $\text{aff } \Omega_n \subseteq \frac{1}{n}\mathbf{J} + W_n$. Conversely, let $\mathbf{X} \in W_n$, $\mathbf{X} \neq \mathbf{O}$, and put $c = n \max_{i,j} |x_{ij}| > 0$. Then $\frac{1}{n}\mathbf{J}$ and $\frac{1}{n}\mathbf{J} + \frac{1}{c}\mathbf{X}$ lie in Ω_n , and the affine combination

$$\frac{1}{n}\mathbf{J} + \mathbf{X} = (1-c) \frac{1}{n}\mathbf{J} + c \left(\frac{1}{n}\mathbf{J} + \frac{1}{c}\mathbf{X} \right)$$

lies in $\text{aff } \Omega_n$. So $\text{aff } \Omega_n = \frac{1}{n}\mathbf{J} + W_n$, and $\dim \Omega_n = \dim W_n = (n-1)^2$ by [Lemma 19.7.1](#) (a).

(b) The entries are nonnegative and every row and column adds up to 1. Suppose $\mathbf{M} = (|u_{ij}|^2)$ with $\mathbf{U}$ unitary. The zero entries $m_{13} = m_{21} = m_{32} = 0$ force $u_{13} = u_{21} = u_{32} = 0$. Columns 1 and 2 of $\mathbf{U}$ are $(u_{11}, 0, u_{31})$ and $(u_{12}, u_{22}, 0)$, and they are orthogonal because $\mathbf{U}$ is unitary:

$$0 = u_{11}\overline{u_{12}} + 0 \cdot \overline{u_{22}} + u_{31} \cdot 0 = u_{11}\overline{u_{12}}.$$

But $|u_{11}|^2 = |u_{12}|^2 = \frac{1}{2}$, so $u_{11}\overline{u_{12}} \neq 0$, a contradiction. Hence $\mathbf{M}$ is not unistochastic.