

## ON THE CONVERGENCE OF DOUBLY STOCHASTIC MARKOV CHAINS

LUDOVICK BOUTHAT, NICOLAS DOYON, JAVAD MASHREGHI,  
AND FRÉDÉRIC MORNEAU-GUÉRIN

**Abstract:** We characterize the asymptotic behavior of time-homogeneous doubly stochastic Markov chains. Our investigation revolves around understanding the dynamics of products of doubly stochastic matrices, which in turn allows us to fully characterize three distinct behaviors: cyclicity, convergence towards a special equilibrium matrix, and divergence. Notably, we introduce a novel and comprehensive sufficient condition for the convergence of an infinite product of doubly stochastic matrices.

**2020 Mathematics Subject Classification:** 15B51, 60J10.

**Key words:** doubly stochastic matrices, Markov chains.

### 1. Introduction

A discrete-time Markov chain is a sequence  $X_0, X_1, X_2, \dots$ , or  $(X_n)_{n \geq 0}$ , of random variables, the so called *states*, such that the probability of moving to a future state is independent of the previous states and only depends on the current state; this “memoryless” property is referred to as the Markov property. The possible values of  $X_i$ s form an at most countable set  $S$  of elements  $s_j$ , called the state space of the chain.

Markov chains are characterized by their transition matrices at each discrete time step  $k$ , where the product of these matrices describes the probability of transitioning from one state to another within a given number of time steps. Specifically, if  $P_k$  is the transition matrix of the Markov chain at time  $k$ , the entry  $(i, j)$  in the matrix  $P_1 P_2 \cdots P_m$  represents the probability that  $X_m = s_j$ , given that  $X_0 = s_i$ . In cases where the state space is finite, the transition matrices are stochastic matrices, i.e., square matrices with non-negative real entries and with each row summing to 1. A Markov chain is said to be time-homogeneous, or simply homogeneous, when there exists a stochastic matrix  $P$  such that  $P_k = P$  for all  $k \geq 1$ . Otherwise, it is said to be non-homogeneous. For detailed information on Markov chains and stochastic processes, we refer to the excellent book by Levin, Peres, and Wilmer [19]. For an accessible yet detailed discussion of finite Markov chains (including a study of some of the issues that will be discussed in this article but in a slightly different setting), we refer the reader to [16].

Building on this setup, the central object becomes the infinite product of transition matrices, whose convergence has been studied extensively. Indeed, in 1963 Wolfowitz established sufficient conditions for the convergence of products of indecomposable and aperiodic stochastic matrices [31]. His results ensured that, under these conditions, the limiting matrix has identical rows, corresponding to a unique stationary distribution. Building on this, in 1976 Hajnal extended the analysis to products of general non-negative matrices, proving that under wide conditions, the product of a large number of square non-negative matrices, independent of or-

der, is close to a positive matrix of rank 1 [11]. A year later, Chatterjee and Seneta refined these ideas in the context of repeated averaging, proving that under suitable connectivity and aperiodicity assumptions, successive products of stochastic matrices converge to a rank-one stochastic matrix, thereby formalizing the emergence of consensus in distributed systems [6].

In some applications, we require the transition matrices of Markov chains to be *doubly stochastic*. For example, such matrices appear in the study of gossip algorithms by Boyd et al. [4] and in the analysis of non-selective quantum measurements by Vourdas [30].

Recall that an  $n \times n$  matrix is doubly stochastic if all of its entries are non-negative real numbers and if each row and column sums to 1. Birkhoff's theorem provides a geometric characterization of doubly stochastic matrices, asserting that every such matrix can be written as a convex combination of  $n \times n$  permutation matrices [13, Theorem 8.7.2]. The associated Birkhoff–von Neumann algorithm offers a constructive procedure for obtaining this decomposition. In the present paper, we shall denote by  $\Omega_n$  the set of all  $n \times n$  doubly stochastic matrices. Note that  $\Omega_n$  is closed under classical matrix multiplication. Therefore, given a doubly stochastic Markov chain (i.e., a Markov chain with transition matrices  $D_k \in \Omega_n$  for all  $k$ ), any matrix given by a product of the form  $D_1 D_2 \cdots D_m$  is again doubly stochastic.

Doubly stochastic Markov chains received considerable interest in the latter half of the 20th century. However, since the literature on doubly stochastic Markov chains is very varied (both in terms of the language used and the specialized jargon employed), it has proved difficult to compile a comprehensive list of known results and to fully trace their origins. Nevertheless, we will endeavor to provide a substantial, if incomplete, overview in the following.

In 1965, V. M. Maksimov first turned his attention to homogeneous doubly stochastic Markov chains. He considered the convergence of  $D^m$  as  $m \rightarrow \infty$  for  $D \in \Omega_n$ , and obtained some results relating to the convergence of such limits which depends on the Birkhoff decomposition of  $D$  [20] (see [1] for more detail on the Birkhoff decomposition). In view of this line of research, it does not seem unreasonable to assume that Maksimov as well as some of his contemporaries were aware of the following key result, even though the earliest written mention of it that we have been able to trace is in a 1979 book by Fritz, Huppert, and Willems [8, p. 30]: given  $D \in \Omega_n$  with 1 as a simple eigenvalue and such that  $\lim_{m \rightarrow \infty} D^m$  exists, then

$$(1.1) \quad \lim_{m \rightarrow \infty} D^m = \frac{1}{n} \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix}_{n \times n} =: J_n.$$

In 1970, Maksimov delved deeper into this issue and considered the case of non-homogeneous doubly stochastic Markov chains [21]. In this context, it is the associated sequence of transition matrices  $(D_k)_{k \geq 1}$  that is considered. A sequence of such matrices is said to be *convergent* if, for any  $k \geq 1$ , the matrix product  $D_k D_{k+1} \cdots D_m = B_{k,m}$ , converges as  $m \rightarrow \infty$  to some  $B_k$ . If all these limit matrices  $B_k$  coincide (that is, if  $B_k$  is independent of  $k$ ), then we say that the chain  $(D_k)_{k \geq 1}$  is *properly convergent*.

Using this notation, Maksimov provided necessary and sufficient conditions ensuring convergence and proper convergence of the infinite product of a given sequence  $(D_k)_{k \geq 1} \subseteq \Omega_n$ . However, these general conditions relied heavily on the structure of the zero entries in the doubly stochastic matrices. Thus, these conditions prove most useful when the  $D_k$ s are connected by some regularity law. For the general case, Maksimov was able to provide convenient sufficient conditions for (proper)

convergence. In particular, he proved that if  $d_{ii}^{(k)} \geq \varepsilon > 0$  hold uniformly (i.e., for all  $i$  and  $k \geq 1$ ), and if  $\sum_{k=1}^{\infty} d_{ij}^{(k)} + d_{ji}^{(k)} = \infty$  for all  $i$  and  $j$ , then  $(D_k)_{k \geq 1}$  converges properly.

In 1980, Schwarz took interest in the non-homogeneous problem and tried to improve upon the sufficient condition given by Maksimov [28]. Therein, he showed among other things that if a doubly stochastic Markov chain  $(D_k)_{k \geq 1}$  satisfies the condition  $\sum_{k=1}^{\infty} \nu(D_k) = \infty$ , where  $\nu(D) := \min_{i,j} d_{ij}$ , then the infinite product  $D_1 D_2 D_3 \cdots$  converges (in the sense that  $L = \lim_{i \rightarrow \infty} D_1 \cdots D_i$  exists) to the uniform matrix  $J_n$ .

Then, in 1987, Du provided further insight by showing that if  $D$  is an irreducible  $n \times n$  doubly stochastic matrix and  $\text{tr}(D) > 0$ , then the limit  $\lim_{m \rightarrow \infty} D^m$  exists and is equal to  $J_n$  [7]. Fast forwarding a few years, in 2001, Hwang and Pyo addressed a similar problem by characterizing the doubly stochastic matrices  $D$  whose powers  $D^m$ ,  $m \geq 1$ , eventually stop, i.e.,  $D^m = D^p$ ,  $m \geq p$ , for some positive integer  $p$  [15]. Somewhat more recently, in 2009, Rénier determined the limits of the powers of locally finite infinite doubly stochastic matrices [26]. Finally, we note that in 2010, Hunter studied the stationary distribution, the generalized inverses of Markovian kernels, the mean and variance of the first passage times, and the expected times to mixing of Markov chains with doubly stochastic transition matrices, which are some of the more important properties of Markov chains [14].

Recent results on the geometry of the set of doubly stochastic matrices gave some insight which allowed us to improve upon some of the known sufficient conditions mentioned above [2, 3]. It is the main purpose of this paper to present this new result.

Additionally, we have identified what appears to be two inaccuracies relayed recently in the literature. Indeed, in [14, Theorem 1] it is stated that for any  $n \times n$  doubly stochastic matrix  $P$ ,  $\lim_{k \rightarrow \infty} P^k = J_n$ . However, the identity matrix serves as a direct counterexample to this statement. Examining the argument used in the paper, the Perron–Frobenius theorem is used, so it may be safely assumed that one needs to make the additional assumption that  $D$  is irreducible. Nonetheless, even then it is not that simple. Indeed, in [26, Theorem 5.1] it is stated that if  $A \in \Omega_n$  is irreducible, and not a permutation matrix, then  $\lim_{k \rightarrow \infty} A^k = J_n$ . Upon closer inspection, one may note that this result also appears to require refinement. For example, taking  $A = \begin{bmatrix} 0 & J_2 \\ J_2 & 0 \end{bmatrix}$ , we encounter a counterexample where  $A$  is irreducible but exhibits periodic behavior. Specifically, it is easy to show that  $A^k = A$  for every odd  $k > 0$  and that  $A^k = \begin{bmatrix} J_2 & 0 \\ 0 & J_2 \end{bmatrix}$  for every even  $k > 0$ . Given these subtleties, we also aim to clarify and provide more insight on the sequence  $(D^m)_{m \geq 1}$ , where  $D \in \Omega_n$ , as we characterize the limiting behavior of this sequence as  $m \rightarrow \infty$ .

We begin our study in Section 2 by establishing some preliminary facts and results about doubly stochastic matrices. Then, in Section 3, we focus on the following question: when does the infinite product  $D_1 D_2 D_3 \cdots$  converge to the uniform matrix  $J_n$ ? Therein, we provide a new sufficient condition which improves a result of Schwarz [28]. We then proceed in Section 4 to characterize the behavior of any homogeneous doubly stochastic Markov chain; that is, we characterize the behavior of the sequence  $(D^m)_{m \geq 1}$  as  $m \rightarrow \infty$  for an arbitrary doubly stochastic matrix  $D$ . In particular, we show that  $D^m$  converges when  $m \rightarrow \infty$  if and only if  $D$  is permutation-similar to a direct sum of primitive doubly stochastic matrices and we characterize the doubly stochastic matrices which describe a cycle of order  $p$ .

Let us now introduce some notations that shall be used throughout the text. Given any pair of matrices  $A$  and  $B$  of size  $m \times n$  and  $p \times q$  respectively, the *direct sum* of  $A$  and  $B$  is the block matrix of size  $(m + p) \times (n + q)$  defined by:

$$A \oplus B := \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}.$$

Likewise, the *Kronecker product* of  $A$  and  $B$  is defined as the  $pm \times qn$  block matrix

$$A \otimes B = \begin{bmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \dots & a_{mn}B \end{bmatrix}.$$

The family of all complex  $n \times n$  matrices is denoted by  $M_n$ . The spectral radius of a matrix  $A$  is  $\rho(A) := \max |\lambda|$ , where the maximum is taken over all eigenvalues of  $A$ . We denote the number of eigenvalues of modulus  $\rho(A)$ , multiplicity counted, by  $h(A)$ , and if there is no confusion, simply by  $h$ .

## 2. Preliminary

A square matrix  $A$  is *reducible* if it can be transformed to a block lower-triangular form by simultaneous row and column permutations, i.e., if there exists a permutation matrix  $P$  such that

$$P^\top AP = \begin{bmatrix} A_1 & 0 \\ A_3 & A_2 \end{bmatrix},$$

where  $A_1$  and  $A_2$  are square matrices. A square matrix which is not reducible is said to be *irreducible*. One of the most important results on irreducible non-negative matrices is the celebrated Perron–Frobenius theorem [9, 25]. The theorem comprises numerous statements (for instance, it says that the principal eigenvector is unique up to scaling and has all entries of the same sign). We present here only those relevant to the present discussion. For a comprehensive treatment of the Perron–Frobenius theorem, see [13, Chapter 8].

**Perron–Frobenius theorem.** *Let  $A \in M_n$ ,  $n \geq 2$ , be irreducible with non-negative entries. Then the following hold.*

- (i) *There exists a number  $\mu > 0$ , called the Perron–Frobenius eigenvalue of  $A$ , such that  $\mu$  is a simple eigenvalue of  $A$  and any other eigenvalue  $\lambda$  of  $A$  verifies  $|\lambda| \leq \mu$ . Thus  $\rho(A) = \mu$ .*
- (ii) *The Perron–Frobenius eigenvalue of  $A$  fulfils the row-sum inequalities*

$$(2.1) \quad \min_{1 \leq i \leq n} \sum_{j=1}^n a_{ij} \leq \mu \leq \max_{1 \leq i \leq n} \sum_{j=1}^n a_{ij}.$$

- (iii) *The matrix  $A$  is similar to  $e^{2\pi i/h} A$  and the spectrum of  $A$  is invariant under multiplication by  $e^{2\pi i/h}$ .*
- (iv) *If  $h > 1$ , there exists a permutation matrix  $P$  such that*

$$(2.2) \quad P^\top AP = \begin{bmatrix} 0 & A_{1,2} & 0 & \dots & 0 \\ 0 & 0 & A_{2,3} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & A_{h-1,h} \\ A_{h,1} & 0 & 0 & \dots & 0 \end{bmatrix},$$

*where the non-zero blocks are square matrices.*

- (v) *If  $h = 1$ , then the limit*

$$\lim_{m \rightarrow \infty} \frac{A^m}{\mu^m}$$

*exists. This limit is called the Perron projection of  $A$ .*

In the case of doubly stochastic matrices, Perfect and Mirsky showed that reducible doubly stochastic matrices assume a very particular form as a direct sum of smaller irreducible doubly stochastic matrices [24]. To be more precise, if  $D \in \Omega_n$  is reducible, then there exist an  $n \times n$  permutation matrix  $P$  and an integer  $r \geq 2$  such that

$$(2.3) \quad P^\top DP = D_1 \oplus D_2 \oplus \cdots \oplus D_r,$$

where the  $D_j$ s are irreducible doubly stochastic matrices. Hence, upon applying the Perron–Frobenius theorem to each summand, a fundamental result emerges: the eigenvalues of any doubly stochastic matrix  $D$  lie within the closed unit disk  $\overline{\mathbb{D}}$  and  $\rho(D) = 1$ . In fact, for irreducible doubly stochastic matrices, parts (i) and (ii) of the Perron–Frobenius theorem guarantee that  $\rho(D) = 1$ . For reducible matrices, it follows from (2.3) that

$$\rho(D) = \rho(P^\top DP) = \rho(D_1 \oplus D_2 \oplus \cdots \oplus D_r) = \max_{1 \leq k \leq r} \rho(D_k) = 1.$$

Therefore, for a doubly stochastic matrix  $D$ , the quantity  $h$  represents the number of unimodular eigenvalues of  $D$ .

When  $D \in \Omega_n$  is irreducible, (2.2) reveals that  $D$  is permutation-similar to a matrix with a special form, reminiscent of the  $h \times h$  *circular shift matrix*

$$K_h := \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \end{bmatrix},$$

with square matrices instead of 1s. In the particular case where  $D$  is not only irreducible, but also doubly stochastic, then Marcus, Minc, and Moyls showed that  $h$  must divide  $n$  and that each non-zero square matrix has the same size (namely  $n' := n/h$ ), which is not necessarily the case when  $D$  is not doubly stochastic [23]. More specifically, the authors proved that there exists a permutation matrix  $P$  such that

$$(2.4) \quad P^\top DP = \begin{bmatrix} 0 & D_1 & 0 & \cdots & 0 \\ 0 & 0 & D_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & D_{h-1} \\ D_h & 0 & 0 & \cdots & 0 \end{bmatrix},$$

where  $D_j \in \Omega_{n'}$ ,  $1 \leq j \leq h$ , and  $n' = n/h$ . Note that we can also write

$$(2.5) \quad P^\top DP = (D_1 \oplus \cdots \oplus D_h)(K_h \otimes I_{n'}),$$

where  $K_h \otimes I_{n'}$  is the Kronecker product of  $K_h$  and  $I_{n'}$ , and the ordering of factors and summands on the right-hand side is important.

If  $e$  denotes the uniform vector  $(1, 1, \dots, 1)^\top$  and  $D$  is any doubly stochastic matrix, then  $De = e$ . This is a direct manifestation of the fact that 1 is always an eigenvalue of any doubly stochastic matrix. Furthermore, if  $D$  is irreducible, then 1 is the Perron–Frobenius eigenvalue of  $D$ , and statement (i) of the Perron–Frobenius theorem ensures that this eigenvalue is simple. Notably, this fact serves as a characterizing property of irreducibility in doubly stochastic matrices. To see this, suppose that  $D \in \Omega_n$  is reducible. In this case, (2.3) guarantees the existence of a permutation matrix  $P$  such that  $P^\top DP = D_1 \oplus \cdots \oplus D_r$ , where  $r \geq 2$  and the  $D_j$ s are doubly stochastic matrices. Given that the eigenvalues of  $D$  encompass all the eigenvalues of  $D_1, D_2, \dots, D_r$ ,

and as each of the  $D_j$ s possesses 1 as an eigenvalue, it follows that the geometric multiplicity of 1 as an eigenvalue of  $D$  is strictly greater than 1. It thus follows that

$$(2.6) \quad D \in \Omega_n \text{ is irreducible} \iff 1 \text{ is a simple eigenvalue of } D.$$

Note that for an irreducible matrix  $D$ , it is possible that  $h(D) > 1$ . For example, for the irreducible matrix

$$D := \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

we have  $h(D) = 2$ .

A matrix  $A \in M_n$  is said to be *primitive* if it is non-negative, irreducible, and has a *unique* (simple) eigenvalue of maximal modulus (i.e.,  $h(A) = 1$ ). In the case of doubly stochastic matrices, this definition can be simplified. Indeed, if a doubly stochastic matrix  $D$  is not primitive, then either it is reducible (in which case the geometric multiplicity of 1 as an eigenvalue is strictly greater than 1), or else  $h(D) > 1$ . In both cases, we end up with  $h(D) > 1$ . Therefore,

$$(2.7) \quad D \in \Omega_n \text{ is primitive} \iff h(D) = 1.$$

Lastly, we need the following lemma, partially due to Marcus, who established the equivalence of (i) and (iv) in [22, Lemma 2].

**Lemma 2.8.** *Let  $D \in \Omega_n$ . Then the following are equivalent.*

- (i)  $D = J_n$ .
- (ii) 0 is an eigenvalue of geometric multiplicity  $n - 1$  for  $D$ .
- (iii) 0 is a singular value of geometric multiplicity  $n - 1$  for  $D$ .
- (iv)  $\text{rank}(D) = 1$ .

*Proof:* (i)  $\implies$  (ii) It suffices to note that 0 is an eigenvalue of  $J_n$  associated with the  $n - 1$  linearly independent eigenvectors

$$\vec{v} := (1, \omega_n^k, \omega_n^{2k}, \dots, \omega_n^{(n-1)k})^\top, \quad 1 \leq k \leq n - 1,$$

where  $\omega_n := e^{-2\pi i/n}$ . The other eigenvalue is 1, which necessarily must have multiplicity 1.

(ii)  $\implies$  (iii) Suppose that 0 is an eigenvalue of geometric multiplicity  $n - 1$  for  $D$  and observe that if  $Dv = 0$ , then  $D^\top Dv = 0$ . Thus 0 is a singular value of  $D$  associated with the vector  $v$ . It follows that  $D$  has 0 as a singular value of geometric multiplicity  $n - 1$ . But, since  $D^\top D$  is a doubly stochastic matrix, and since 1 is always an eigenvalue of any doubly stochastic matrix, we find that  $D$  has exactly one non-zero singular value.

(iii)  $\implies$  (iv) From the assumption, it immediately follows that  $\text{rank}(D) = 1$ .

(iv)  $\implies$  (i) The rows of  $D$  span a subspace of dimension 1. Hence, they must be equal since if two rows were distinct, the sum of the elements of these two rows would be different, which is impossible since  $D$  is (doubly) stochastic. Hence,

$$D = \begin{bmatrix} d_1 & d_2 & \dots & d_n \\ \vdots & \vdots & \ddots & \vdots \\ d_1 & d_2 & \dots & d_n \end{bmatrix}.$$

Since the elements of each column of  $D$  sum to 1, we also find that  $nd_j = 1$  for each  $1 \leq j \leq n$ . Therefore,  $d_j = 1/n$  for all  $1 \leq j \leq n$  and it follows that  $D = J_n$ .  $\square$

### 3. Non-homogeneous Markov chains

In this section, we study the case of non-homogeneous doubly stochastic Markov chains. If  $(D_\ell)_{\ell \geq 1}$  is a given sequence of  $n \times n$  doubly stochastic matrices, we seek to determine sufficient conditions to ensure that

$$(3.1) \quad \lim_{m \rightarrow \infty} D_1 D_2 \cdots D_m = J_n.$$

This question was initially studied by Maksimov [20, 21] and Schwarz [28]. In particular, the latter author showed that if

$$(3.2) \quad \sum_{\ell=1}^{\infty} v(D_\ell) = +\infty,$$

where

$$(3.3) \quad v(D) := \min_{i,j} d_{ij},$$

then (3.1) holds. In what follows, we present a more general result to ensure (3.1).

If we label the eigenvalues of  $D \in \Omega_n$  as  $1 = \lambda_1, \lambda_2, \dots, \lambda_n \in \mathbb{C}$ , the eigenvalues  $\lambda_i$  for  $i \geq 2$  are referred to as the *non-stochastic* eigenvalues of  $D$ . In precise terms, these are the eigenvalues of  $D$  whose associated eigenvectors are not scalar multiples of the uniform vector  $e$ .

**Lemma 3.4.** *Let  $D \in \Omega_n$  with eigenvalues  $1 = \lambda_1, \lambda_2, \dots, \lambda_n \in \mathbb{C}$ , and let  $\alpha \in \mathbb{C}$  be a complex number. Then the eigenvalues of  $D - \alpha J_n$  are  $1 - \alpha, \lambda_2, \dots, \lambda_n$ .*

*Proof:* To establish the claim, first observe that

- (i) the matrices  $D$  and  $J_n$  commute;
- (ii) the eigenvalue  $\lambda_1 = 1$  of  $D$  is associated with the all-ones eigenvector  $e$ ;
- (iii) and the eigenvalues of  $J_n$  are 1 with geometric multiplicity 1 (associated with the eigenvector  $e$ ), and 0 with geometric multiplicity  $n - 1$ .

It follows from Theorem 2.3.3 in [13] that there is a unitary matrix  $U$  whose first column is equal to  $\frac{1}{\sqrt{n}}e$  such that

$$U^* D U = \begin{bmatrix} 1 & 0 & \dots & 0 \\ * & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \dots & \lambda_n \end{bmatrix} \quad \text{and} \quad U^* J_n U = \begin{bmatrix} 1 & 0 & \dots & 0 \\ * & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \dots & 0 \end{bmatrix}.$$

Thus,

$$U^*(D - \alpha J_n)U = \begin{bmatrix} 1 & 0 & \dots & 0 \\ * & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \dots & \lambda_n \end{bmatrix} - \begin{bmatrix} \alpha & 0 & \dots & 0 \\ * & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \dots & 0 \end{bmatrix} = \begin{bmatrix} 1 - \alpha & 0 & \dots & 0 \\ * & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \dots & \lambda_n \end{bmatrix}.$$

The conclusion follows from this decomposition. □

We now introduce the main result and subsequently explore its scope and how it relates to previously established findings.

**Theorem 3.5.** *Let  $(D_\ell)_{\ell \geq 1}$  be a sequence of doubly stochastic matrices in  $\Omega_n$ . If*

$$\sum_{\ell=1}^{\infty} (1 - \sigma_2(D_\ell)) = +\infty,$$

where  $\sigma_2(D_\ell)$  denotes the second largest singular value of  $D_\ell$ , then

$$\lim_{m \rightarrow \infty} D_1 D_2 \cdots D_m = J_n.$$

*Proof:* The peculiar identity

$$D_1 D_2 \cdots D_m - J_n = (D_1 - J_n)(D_2 - J_n) \cdots (D_m - J_n)$$

is proved by induction. By the submultiplicative property of the spectral norm  $\|\cdot\|_2$ , we have

$$(3.6) \quad \|D_1 \cdots D_m - J_n\|_2 \leq \|D_1 - J_n\|_2 \cdots \|D_m - J_n\|_2.$$

Moreover, Birkhoff’s theorem and the permutation invariance of both  $J$  and the spectral norm ensures that

$$(3.7) \quad \begin{aligned} \|D - J_n\|_2 &= \left\| \sum_i \alpha_i (P_i - J_n) \right\|_2 \\ &\leq \sum_i \alpha_i \|P_i - J_n\|_2 \\ &= \sum_i \alpha_i \|I_n - J_n\|_2 = 1, \end{aligned}$$

for all  $D \in \Omega_n$ . Here, we used the easily verified identity  $\|I_n - J_n\|_2 = 1$ .

Recall that if a sequence  $(a_\ell)_{\ell \geq 1}$  of real numbers satisfy  $0 \leq a_\ell \leq 1$  for all  $\ell$ , then  $\sum_{\ell=1}^{\infty} (1 - a_\ell) = +\infty$  implies that  $\prod_{\ell=1}^{\infty} a_\ell = 0$ . Indeed, since  $0 \leq x \leq e^{x-1}$  for all  $0 \leq x \leq 1$ , it follows that

$$0 \leq \prod_{k=1}^m a_k \leq \prod_{k=1}^m e^{a_k - 1} = e^{-\sum_{k=1}^m (1 - a_k)}.$$

Since  $e^{-x}$  is continuous, letting  $m \rightarrow \infty$  yields  $\prod_{k=1}^{\infty} a_k = 0$ . Hence, by (3.7),

$$\sum_{\ell=1}^{\infty} (1 - \|D_\ell - J_n\|_2) = +\infty \implies \prod_{\ell=1}^{\infty} \|D_\ell - J_n\|_2 = 0.$$

Then, by (3.6), the condition on the left-hand side implies

$$\lim_{m \rightarrow \infty} D_1 \cdots D_m = J_n.$$

To finish, we find yet another interpretation for  $\|D_\ell - J_n\|_2$ . It is well known that  $\|D_\ell - J_n\|_2$  equals the greatest singular value of  $D_\ell - J_n$ . But,

$$(D_\ell - J_n)^*(D_\ell - J_n) = (D_\ell^\top - J_n)(D_\ell - J_n) = D_\ell^\top D_\ell - J_n.$$

Thus, the largest eigenvalue of  $(D_\ell - J_n)^*(D_\ell - J_n)$  coincides with the largest eigenvalue of  $D_\ell^\top D_\ell - J_n$ , which, by Lemma 3.4, is the largest non-stochastic eigenvalue of  $D_\ell^\top D_\ell$ , which in turn is the second largest eigenvalue of  $D_\ell^\top D_\ell$ . Hence,

$$\|D_\ell - J_n\|_2 = \sigma_2(D_\ell). \quad \square$$

Let us show that Theorem 3.5 implies Schwarz's result. To establish this fact, we need to make a short digression and perform a delicate calculation. Given a real number  $\alpha$ , we have

$$\|D - \alpha J_n\|_2 = \sigma_{\max}(D - \alpha J_n) = \rho^{1/2}(D^\top D - \alpha(2 - \alpha)J_n),$$

since  $(D - \alpha J_n)^\top(D - \alpha J_n) = D^\top D - \alpha(2 - \alpha)J_n$ . Lemma 3.4 then ensures that the eigenvalues of the matrix  $D^\top D - \alpha(2 - \alpha)J_n$  are  $1 - \alpha(2 - \alpha) = (1 - \alpha)^2$ , as well as the  $n - 1$  non-stochastic eigenvalues of  $D^\top D$  (which are the same as the  $n - 1$  non-stochastic eigenvalues of  $D - \alpha J_n$ ). Consequently,

$$\|D - \alpha J_n\|_2 = \max\{|1 - \alpha|, \|D - J_n\|_2\},$$

from which it also follows that

$$(3.8) \quad \|D - J_n\|_2 = \min_{\alpha} \|D - \alpha J_n\|_2 \leq \|D - nv(D)J_n\|_2,$$

where we recall  $v(D)$ , which was defined in (3.3).

On the one hand, for  $D \neq J_n$ , the matrix  $\frac{1}{1-nv(D)}(D - nv(D)J_n)$  is doubly stochastic, since each row/column sums to 1 and each coefficient is greater than  $\frac{1}{1-nv(D)}(v(D) - nv(D)\frac{1}{n}) = 0$ . This implies that the spectral norm of this latter matrix is 1, since the spectral norm of every doubly stochastic matrix is 1. Thus

$$\|D - nv(D)J_n\|_2 = (1 - nv(D)) \left\| \frac{D - nv(D)J_n}{1 - nv(D)} \right\|_2 = 1 - nv(D).$$

As a result, by (3.8), we deduce  $\|D - J_n\|_2 \leq 1 - nv(D)$ . Thus,

$$(3.9) \quad nv(D) \leq 1 - \|D - J_n\|_2 = 1 - \sigma_2(D)$$

for every doubly stochastic matrix  $D \neq J_n$ .

On the other hand, one can check that (3.9) is trivially verified if  $D = J_n$ . Hence this relation holds for any  $D \in \Omega_n$  and we conclude that

$$\sum_{\ell=1}^{\infty} v(D_\ell) = +\infty \implies \sum_{\ell=1}^{\infty} (1 - \sigma_2(D_\ell)) = +\infty.$$

Consequently, whenever the infinite product of a sequence  $D_1, D_2, \dots \in \Omega_n$  verifies Schwarz's sufficient condition (3.2) for convergence, it also verifies the convergence criterion stated in Theorem 3.5.

In order to show that the latter criterion is not simply a reformulation of (3.2), but rather a proper generalization, consider the doubly stochastic matrix

$$D_\ell = \begin{bmatrix} 1 - \varepsilon_\ell - \varepsilon_\ell^2 & \varepsilon_\ell & \varepsilon_\ell^2 \\ \varepsilon_\ell & 1 - 2\varepsilon_\ell & \varepsilon_\ell \\ \varepsilon_\ell^2 & \varepsilon_\ell & 1 - \varepsilon_\ell - \varepsilon_\ell^2 \end{bmatrix},$$

where  $0 \leq \varepsilon_\ell \leq \frac{1}{2}$  goes to 0 as  $\ell \rightarrow \infty$ . Clearly,  $v(D_\ell) = \varepsilon_\ell^2$ . Moreover, an easy computation reveals that the eigenvalues of  $D$  are 1,  $1 - 3\varepsilon_\ell$ , and  $1 - \varepsilon_\ell - 2\varepsilon_\ell^2$ . Since the matrix is positive semi-definite, they are also equal to the singular values. Hence, if  $\varepsilon_\ell = \frac{1}{4\ell}$ , then Schwarz's criteria yield

$$\sum_{\ell=1}^{\infty} v(D_\ell) = \frac{1}{16} \sum_{\ell=1}^{\infty} \frac{1}{\ell^2} < \infty,$$

while Theorem 3.5 gives

$$\sum_{\ell=1}^{\infty} (1 - \sigma_2(D_\ell)) = \frac{1}{4} \sum_{\ell=1}^{\infty} \left( \frac{1}{\ell} + \frac{1}{2\ell^2} \right) = \infty.$$

Hence,  $D_1 D_2 D_3 \cdots = J_n$ , and the above discussion establishes the more general scope of Theorem 3.5.

However, grouping the matrices together by adjacent pairs and looking at the smallest coefficient of the resulting matrices might allow one to apply Schwarz’s criteria, while being easier to verify than Theorem 3.5. We show that this is not possible, even if we allow the matrices to be grouped in blocks of large sizes (i.e., to look at the smallest entry of  $D_1 D_2 \cdots D_k$ , then  $D_{k+1} D_{k+2} \cdots D_{2k}$ , where  $k$  is any fixed integer). Indeed, note that the matrices  $D_\ell$  can be diagonalized as

$$D_\ell = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -2 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \alpha_\ell & 0 \\ 0 & 0 & \beta_\ell \end{bmatrix} \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{6} & -\frac{1}{3} & \frac{1}{6} \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix},$$

where  $\alpha_\ell := 1 - 3\varepsilon_\ell$  and  $\beta_\ell := 1 - \varepsilon_\ell - 2\varepsilon_\ell^2$ . Hence, if  $A_{\ell,k} := \alpha_\ell \cdots \alpha_{\ell+k-1}$  and  $B_{\ell,k} := \beta_\ell \cdots \beta_{\ell+k-1}$ , then

$$\begin{aligned} D_\ell \cdots D_{\ell+k} &= \begin{bmatrix} 1 & 1 & -1 \\ 1 & -2 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & A_{\ell,k} & 0 \\ 0 & 0 & B_{\ell,k} \end{bmatrix} \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{6} & -\frac{1}{3} & \frac{1}{6} \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix} \\ &= J_3 + \frac{1}{6} \begin{bmatrix} A_{\ell,k} + 3B_{\ell,k} & -2A_{\ell,k} & A_{\ell,k} - 3B_{\ell,k} \\ -2A_{\ell,k} & 4A_{\ell,k} & -2A_{\ell,k} \\ A_{\ell,k} - 3B_{\ell,k} & -2A_{\ell,k} & A_{\ell,k} + 3B_{\ell,k} \end{bmatrix}. \end{aligned}$$

Since  $A_{\ell,k}, B_{\ell,k} \geq 0$ , the smallest entry of  $D_\ell \cdots D_{\ell+k}$  is realized either by  $\frac{1}{3} - 2A_{\ell,k}$  or by  $\frac{1}{3} + A_{\ell,k} - 3B_{\ell,k}$ . But since  $\alpha_\ell \leq \beta_\ell$  for all  $\ell$ , we have  $A_{\ell,k} \leq B_{\ell,k}$ , which implies that  $A_{\ell,k} - 3B_{\ell,k} \leq A_{\ell,k} - 3A_{\ell,k} = -2A_{\ell,k}$ . Hence, the smallest coefficient of  $D_\ell \cdots D_{\ell+k}$  is attained by  $\frac{1}{3} + \frac{1}{6}A_{\ell,k} - \frac{1}{2}B_{\ell,k}$ . It thus follows that

$$\sum_{n=1}^{\infty} v(D_{(n-1)k+1} \cdots D_{nk}) = \sum_{n=1}^{\infty} \left( \frac{1}{3} + \frac{1}{6}A_{(n-1)k+1,k} - \frac{1}{2}B_{(n-1)k+1,k} \right).$$

But now, observe that Stirling’s approximation yields

$$\begin{aligned} A_{(n-1)k+1,k} &= \alpha_{(n-1)k+1} \cdots \alpha_{nk+1} = (1 - 3\varepsilon_{(n-1)k+1}) \cdots (1 - 3\varepsilon_{nk+1}) \\ &= \left( 1 - \frac{3}{4((n-1)k+1)} \right) \cdots \left( 1 - \frac{3}{4(nk+1)} \right) \\ &= \frac{(n-1)k + \frac{1}{4}}{(n-1)k+1} \cdots \frac{nk + \frac{1}{4}}{nk+1} = \frac{\Gamma(nk + \frac{5}{4})\Gamma((n-1)k+1)}{\Gamma(nk+2)\Gamma((n-1)k + \frac{1}{4})} \\ &\asymp \frac{(nk + \frac{1}{4})^{\frac{3}{4}} \cdot ((n-1)k)^{\frac{1}{2}}}{(nk+1)^{\frac{3}{2}} ((n-1)k - \frac{3}{4})^{-\frac{1}{4}}} \\ &\asymp 1 - \frac{3(k+2)}{4kn} - \frac{3k^2 - 30k - 57}{32k^2n^2} + O(n^{-3}). \end{aligned}$$

Similarly,

$$B_{(n-1)k+1,k} \asymp 1 - \frac{k+2}{4kn} - \frac{3k^2+2k-5}{32k^2n^2} + O(n^{-3}).$$

Therefore, for all positive integer values of  $k$ , we finally get

$$\begin{aligned} \sum_{n=1}^{\infty} v(D_{(n-1)k+1} \cdots D_{nk}) &= \sum_{n=1}^{\infty} \left( \frac{1}{3} + \frac{1}{6} A_{(n-1)k+1,k} - \frac{1}{2} B_{(n-1)k+1,k} \right) \\ &\asymp \sum_{n=1}^{\infty} \left( \frac{1}{3} + \frac{1}{6} \left( 1 - \frac{3(k+2)}{4kn} - \frac{3k^2-30k-57}{32k^2n^2} + O(n^{-3}) \right) \right) \\ &\quad - \frac{1}{2} \left( 1 - \frac{k+2}{4kn} - \frac{3k^2+2k-5}{32k^2n^2} + O(n^{-3}) \right) \\ &= \sum_{n=1}^{\infty} \left( \frac{k^2+6k+7}{32k^2n^2} + O(n^{-3}) \right) \\ &< \infty. \end{aligned}$$

Hence, there exist matrices for which the condition of Theorem 3.5 is stronger than Schwarz's criteria, but for which the latter does not work no matter how many matrices are grouped together.

Moreover, the condition provided in Theorem 3.5 is also convenient if the doubly stochastic Markov chain is not predetermined. Indeed, let

$$D = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix}.$$

Then it is easily verified that a direct application of Schwarz's criteria does not allow one to conclude that  $D^m \rightarrow J_3$  as  $m \rightarrow \infty$ , while Theorem 3.5 does. However, applying Schwarz's result to the matrix  $D^2$  does allow one to show that the infinite product converges to  $J_3$  in less computation than Theorem 3.5 would require. Now, suppose that the matrix  $D_k$  is the matrix  $D$  with a probability of  $1/7$ , and any of the  $3 \times 3$  permutation matrices with an equal probability of  $1/7$ . Although probabilistic considerations have not appeared so far – and will play only a modest role here – they will take on a more significant role in later sections. It is nevertheless natural to introduce this simple stochastic variant now. In this random setting, however, there is no straightforward way to group the matrices  $D_k$  together to apply Schwarz's result. However, it is clear that  $1 - \sigma_2(D_k)$  is equal to 0 with a probability of  $6/7$  and  $1/2$  with a probability of  $1/7$ . Hence, an application of Theorem 3.5 allows one to conclude that the infinite product  $D_1 D_2 \cdots D_m$  converges to  $J_3$  almost surely as  $m \rightarrow \infty$ .

During the reviewing process of this document, the papers [10, 11] were brought to our attention. Among the contributions of these publications, [11, Theorem 3] was particularly relevant to the present research. Indeed, for a sequence  $(P_k)_{k \geq 1}$  of stochastic matrices, define  $U_{r,k} = [u_{ij}^{(r,k)}]$  to be the stochastic matrix

$$(3.10) \quad U_{r,k} := P_{r+1} P_{r+2} \cdots P_{r+k}.$$

The sequence  $(P_k)_{k \geq 1}$  is said to be *weakly ergodic* if for all  $i, j, s = 1, 2, \dots, n$  and all  $r \geq 0$  we have

$$u_{i,s}^{(r,k)} - u_{j,s}^{(r,k)} \xrightarrow{k \rightarrow \infty} 0,$$

that is, if all rows of  $U_{r,k}$  tend to become the same as  $k \rightarrow \infty$ . Moreover, call any scalar function  $\mu(\cdot)$  that is continuous on the set of  $n \times n$  stochastic matrices  $P$  and satisfies  $0 \leq \mu(P) \leq 1$  a *coefficient of ergodicity*. This coefficient is said to be *proper* if  $\mu(P) = 1$ , or, equivalently, if all rows of  $P$  are identical.

Hajnal proved in [11, Theorem 3] that if  $1 - m(\cdot)$  and  $\mu(\cdot)$  are two proper coefficients of ergodicity satisfying

$$(3.11) \quad m(P_1 P_2 \cdots P_k) \leq \prod_{i=1}^k (1 - \mu(P_i)),$$

for any stochastic matrices  $P_i$  ( $i = 1, \dots, k$ ) and any  $k \geq 1$ , then  $(P_k)_{k \geq 1}$  is weakly ergodic if and only if there exists a strictly increasing subsequence  $(i_j)_{j \geq 1}$  of the positive integers satisfying

$$\sum_{j=1}^{\infty} \mu(U_{i_j, i_{j+1} - i_j}) = \infty.$$

Now, let us make the following three observations:

- (a)  $J_n$  is the only doubly stochastic matrix for which every row is identical.
- (b)  $\mu(D) := 1 - \|D - J_n\|_2$  is a proper coefficient of ergodicity on doubly stochastic matrices.
- (c) If  $m(D) = 1 - \mu(D) = \|D - J_n\|_2$ , then (3.11) is satisfied (see (3.6)).

As consequences, we find that (1) a sequence of doubly stochastic  $(D_k)_{k \geq 1}$  is weakly ergodic if and only if for all  $r \geq 0$ ,  $\lim_{k \rightarrow \infty} D_{r+1} \cdots D_{r+k} = J_n$ ; and (2) a coefficient of ergodicity  $\mu(\cdot)$  is proper if and only if  $\mu(D) = 1 \iff D = J_n$ . Moreover, we find by defining  $m(D) = 1 - \mu(D) = \|D - J_n\|_2$  that Hajnal's results reduce, after similar observations to those made in the proof of Theorem 3.5, to the equivalence

$$(3.12) \quad \lim_{k \rightarrow \infty} D_{r+1} D_{r+2} \cdots D_{r+k} = J_n, \quad \forall r \geq 0,$$



$$(3.13) \quad \exists \text{ increasing sequence } (i_j) \text{ such that } \sum_{j=1}^{\infty} (1 - \sigma_2(D_{i_{j+1}} \cdots D_{i_j})) = \infty.$$

Moreover, it is clear (simply choose  $i_j = j$  for all  $j \geq 1$ ) that the sufficient condition of Theorem 3.5 implies (3.13), and that (3.12) implies that  $D_1 D_2 \cdots = J_n$ . Hence, Theorem 3.5 can be derived from Hajnal's 1958 result.

As a final remark to this section, observe that inequality (3.9) is of independent interest, as it gives us an easy-to-compute upper bound on the size of the second largest eigenvalue of a symmetric doubly stochastic matrix. Such an estimate is useful in various contexts. For instance, Boyd et al. showed in [4] that the averaging time of any gossip algorithm depends on the second largest eigenvalue of a symmetric doubly stochastic matrix: the smaller this eigenvalue, the faster the averaging algorithm. These algorithms, somehow based on how an epidemic spreads, aim to distribute a certain amount of information to all members of a group in a decentralized manner (i.e., with no master node aggregating information or distributing tasks). Such algorithms have multiple applications in communication [17], in cybersecurity [18], and in big data [5].

#### 4. Homogeneous Markov chains

In this section, we consider the behavior of the sequence  $D^m$  as  $m \rightarrow \infty$ , where  $D$  is a doubly stochastic matrix. As a matter of fact, in all generality, one can use the Jordan canonical form of  $A \in \mathbb{C}^{n \times n}$  to compute the powers of  $A$  and the limit of  $A^m$  as  $m \rightarrow \infty$ . Using this method, Hensel showed that  $\lim_{m \rightarrow \infty} A^m$  exists if and only if there exists a nonsingular matrix  $P$  such that

$$P^{-1}AP = \begin{bmatrix} I & 0 \\ 0 & K \end{bmatrix},$$

where  $\rho(K) < 1$  [12]. In this case,

$$\lim_{m \rightarrow \infty} A^m = P \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} P^{-1}$$

and in particular,  $\lim_{m \rightarrow \infty} A^m = 0$  if and only if  $\rho(A) < 1$ . However, in many cases, this condition is difficult to verify, and the result does not provide significant insights into the behavior of the sequence  $(A^m)_{m \geq 1}$  when  $\lim_{m \rightarrow \infty} A^m$  does not exist. Therefore, using properties of  $\Omega_n$  and tools from matrix theory, we aim to provide a more detailed characterization of the behavior of the sequence  $(D^m)_{m \geq 1}$  as  $m \rightarrow \infty$  for any doubly stochastic matrix  $D$ . In particular, we shall see that one of three phenomena can occur:

- (a) The sequence forms a cycle, i.e.,  $D^{p+1} = D$  for some integer  $p > 1$ .
- (b) The sequence converges to a doubly stochastic matrix of the form

$$P^\top (J_{n_1} \oplus \cdots \oplus J_{n_r}) P,$$

where  $P$  is a permutation matrix.

- (c) The sequence does not converge and is not cyclic. However, the sequence follows a pattern which can be loosely described as *converging to a cycle*.

To describe more precisely the above cases, we say that a matrix  $A$  is *cyclic of order  $p$*  if there is a positive integer  $p$  satisfying  $A^{p+1} = A$  and  $A^{m+1} \neq A$  for all  $1 \leq m < p$ . Moreover, in the language of matrix theory, the matrix  $A$  is called *convergent* if  $\lim_{m \rightarrow \infty} A^m = 0$  and *semi-convergent* whenever  $\lim_{m \rightarrow \infty} A^m$  exists. Here, we are only concerned with semi-convergence since no doubly stochastic matrix is convergent. Furthermore, in our setting, the concepts of proper convergence and convergence (in the sense of Markov chains), defined in the introduction, coincide with that of semi-convergence. Hence, we shall adopt the classical terminology of semi-convergence mentioned above to avoid any potential confusion.

The work on the convergence problem of doubly stochastic matrices mainly began around 1965. Since then, several results have been proved. However, in recent years, there seem to be misconceptions around some of these results, and there are a handful of incorrect theorems in the literature. We aim to provide a complete characterization of the doubly stochastic matrices  $D$  satisfying  $\lim_{k \rightarrow \infty} D^k = J_n$  in Theorem 4.10 and Theorem 4.12.

As the first step, the crucial property (2.3) allows us, without any loss of generality, to assume that the matrix  $D \in \Omega_n$  is irreducible. Indeed, for any permutation matrix  $P$  and any doubly stochastic matrix  $D$ , the behavior of the matrix  $D^m$  when  $m \rightarrow \infty$  is entirely determined by that of  $(P^\top DP)^m = P^\top D^m P$ . Moreover, if  $P^\top DP = D_1 \oplus D_2 \oplus \cdots \oplus D_r$ , then

$$(P^\top DP)^m = D_1^m \oplus D_2^m \oplus \cdots \oplus D_r^m.$$

Thus, it is enough to study the behavior of  $D_j^m$  for the irreducible doubly stochastic matrices  $D_j$ ,  $1 \leq j \leq r$ , to characterize the behavior of the sequence  $(D^m)_{m \geq 1}$ .

Hence, using the same argument with (2.4), we observe that it suffices to first prove our results for irreducible doubly stochastic matrices of the form

$$(4.1) \quad D = \begin{bmatrix} 0 & D_1 & 0 & \dots & 0 \\ 0 & 0 & D_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & D_{h-1} \\ D_h & 0 & 0 & \dots & 0 \end{bmatrix} = \left( \bigoplus_{i=1}^h D_i \right) (K_h \otimes I_{n/h}),$$

where  $h = h(D)$  and  $D_i \in \Omega_{n/h}$ ,  $1 \leq i \leq h$ .

This section will thus be divided into three different subsections. We begin by identifying the general behavior of the powers of irreducible doubly stochastic matrices. As a corollary, we are then able to characterize the irreducible doubly stochastic matrices which (1) exhibit a cycle of order  $p$ , (2) semi-converge, and (3) diverge. Moreover, we also determine the precise way in which the sequence  $(D^m)_{m \geq 1}$  diverges. Finally, we combine the results of these two subsections to characterize the behavior of  $D^m$  for a general doubly stochastic matrix  $D$ .

**4.1. Irreducible doubly stochastic matrices.** The following lemma allows us to easily compute the powers of the matrix  $D = (D_1 \oplus \dots \oplus D_h)(K_h \otimes I_{n'})$  encountered in (2.5). For the sake of notational simplicity, let us denote  $[m]$  as the number  $m \bmod h$ . We observe that the  $m$ th power of the matrix  $D$  has the structure of the  $m$ th power of the circular shift matrix  $K_h$ , where the non-zero coefficients are in the position  $(i, [i + m])$ ,  $1 \leq i \leq h$ , except that instead of having 1s in these entries, we have more complex matrices.

**Lemma 4.2.** *Let*

$$D = (D_1 \oplus \dots \oplus D_h)(K_h \otimes I_{n'}) \in M_n(\mathbb{C}),$$

where  $h, n' \in \mathbb{N}$ ,  $n = hn'$ , and  $D_j \in M_{n'}(\mathbb{C})$  for every  $1 \leq j \leq h$ . Then

$$D^m = \left( \bigoplus_{i=1}^h \prod_{j=1}^m D_{[i+j-1]} \right) (K_h^m \otimes I_{n'}).$$

*Remark.* Since we are in a non-commutative setting, the ordering in products is important. In the above formula, the product should be interpreted as the left multiplication, i.e.,

$$\prod_{j=1}^m A_j := A_1 A_2 \dots A_m,$$

where  $A_j$  are arbitrary  $n \times n$  matrices.

*Proof:* The proof is by induction. For  $m = 1$ , the result is trivial. Now suppose that the statement holds true for  $m - 1$ , and write  $B_i = \prod_{j=1}^{m-1} D_{[i+j-1]}$ . Then

$$D^m = D^{m-1} D = \left( \bigoplus_{i=1}^h \prod_{j=1}^{m-1} D_{[i+j-1]} \right) (K_h^{m-1} \otimes I_{n'}) D$$

$$= \begin{bmatrix} 0 & \dots & 0 & B_1 & \dots & 0 \\ 0 & \dots & 0 & 0 & \ddots & 0 \\ \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ B_{h+1-[m-1]} & \dots & 0 & 0 & \dots & 0 \\ 0 & \ddots & 0 & 0 & \dots & 0 \\ 0 & \dots & B_h & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} 0 & D_1 & 0 & \dots & 0 \\ 0 & 0 & D_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & D_{h-1} \\ D_h & 0 & 0 & \dots & 0 \end{bmatrix},$$

where each  $B_j$  lies in the position  $(j, [j + m - 1])$ ,  $1 \leq j \leq h$ . The block multiplication of these matrices is everywhere zero, except when we multiply  $B_j$  by  $D_{[j+m-1]}$ . Hence, we obtain

$$\begin{aligned}
 D^m &= \begin{bmatrix} 0 & \dots & 0 & B_1 D_{[m]} & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & B_{h-[m]} D_{[h-1]} \\ B_{h+1-[m]} D_{[h]} & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & B_h D_{[m-1]} & 0 & \dots & 0 \end{bmatrix} \\
 &= \left( \bigoplus_{i=1}^h \prod_{j=1}^m D_{[i+j-1]} \right) (K_h^m \otimes I_{n'}) \quad \square
 \end{aligned}$$

In Lemma 3.4, we see that for any doubly stochastic matrix  $D$  with eigenvalues  $1, \lambda_2, \dots, \lambda_n$ , where 1 is the Perron root of  $D$  associated with the all-ones eigenvector, the eigenvalues of  $D - J_n$  are precisely  $0, \lambda_2, \dots, \lambda_n$ . Consequently, we have  $h(D) = 1$  if and only if  $\rho(D - J_n) < 1$ , which itself is equivalent to the fact that  $\lim_{m \rightarrow \infty} (D - J_n)^m = 0$ . Moreover, we also saw that  $(D - J_n)^m = D^m - J_n$ . Therefore, it follows directly that

$$(4.3) \quad \lim_{m \rightarrow \infty} D^m = J_n \iff h(D) = 1.$$

In his 1965 paper, Maksimov provided a more general result by showing that, under certain conditions, there exists a number  $p$  such that  $\lim_{m \rightarrow \infty} D^{mp+r}$  exists, for any  $1 \leq r \leq p$ , and he precisely determined the value of this limit. However, his result relied on a Birkhoff decomposition of  $D$  and on subgroups and quotient groups [1]. This approach introduced challenging conditions to verify, and, moreover, the eventual limit is difficult to compute. The following result provides a clear picture of the period  $p$  as well as an explicit formula for the limit. Recall the decomposition (2.4), which is used below.

**Theorem 4.4.** *Let  $D \in \Omega_n$  be an irreducible doubly stochastic matrix and let  $P$  be a permutation matrix such that*

$$P^\top D P = \left( \bigoplus_{i=1}^h D_i \right) (K_h \otimes I_{n'}),$$

where  $h := h(D)$  and  $D_i \in \Omega_{n'}$ ,  $1 \leq i \leq h$ , and  $n' = n/h$ . Then, for every  $1 \leq r \leq h$ ,

$$D^{mh+r} \xrightarrow{m \rightarrow \infty} P(K_h^r \otimes J_{n'})P^\top.$$

*Proof:* According to Lemma 4.2, we have

$$P^\top D^{mh+r} P = \left( \bigoplus_{i=1}^h \prod_{j=1}^{mh+r} D_{[i+j-1]} \right) (K_h^{mh+r} \otimes I_{n'}).$$

Notice that  $K_h^{mh+r} = (K_h^h)^m K_h^r = I_h^m K_h^r = K_h^r$ . Hence, we have

$$\begin{aligned} \bigoplus_{i=1}^h \prod_{j=1}^{mh+r} D_{[i+j-1]} &= \bigoplus_{i=1}^h \left( \prod_{k=0}^{m-1} \prod_{j=kh+1}^{(k+1)h} D_{[i+j-1]} \right) \left( \prod_{j=mh+1}^{mh+r} D_{[i+j-1]} \right) \\ &= \bigoplus_{i=1}^h \left( \prod_{k=0}^{m-1} \prod_{j=1}^h D_{[i+j+kh-1]} \right) \left( \prod_{j=1}^r D_{[i+j+mh-1]} \right) \\ &= \bigoplus_{i=1}^h \left( \prod_{k=0}^{m-1} \prod_{j=1}^h D_{[i+j-1]} \right) \left( \prod_{j=1}^r D_{[i+j-1]} \right) \\ &= \bigoplus_{i=1}^h \left( \prod_{j=1}^h D_{[i+j-1]} \right)^m \left( \prod_{j=1}^r D_{[i+j-1]} \right). \end{aligned}$$

For simplicity, let us write

$$(4.5) \quad D_{i,r} := D_{[i]} D_{[i+1]} \cdots D_{[i+r-1]}, \quad r \geq 1.$$

Using this notation, it follows that

$$P^\top D^{mh+r} P = \left( \bigoplus_{i=1}^h D_{i,h}^m D_{i,r} \right) (K_h^r \otimes I_{n'}).$$

Now, observe that

$$\bigoplus_{i=1}^h D_{i,h} = \bigoplus_{i=1}^h \prod_{j=1}^h D_{[i+j-1]} = P^\top D^h P.$$

Hence, the number of eigenvalues of unit modulus of  $\bigoplus_{i=1}^h D_{i,h}$  and  $D^h$  is the same, and since the number of unimodular eigenvalues of  $D^k$  is independent of  $k \geq 1$ , we have

$$h = h(D) = h(D^h) = h(P^\top D^h P) = h(\bigoplus_{i=1}^h D_{i,h}) = \sum_{i=1}^h h(D_{i,h}) \geq h,$$

where the inequality follows from the fact that each  $D_{i,h}$  is doubly stochastic. Therefore, the inequality must be saturated, i.e.,

$$h(D_{i,h}) = 1, \quad 1 \leq i \leq h.$$

Hence, it follows from (4.3) and from the fact that  $J_n A = J_n$  for every  $A \in \Omega_n$  that

$$\begin{aligned} \lim_{m \rightarrow \infty} P^\top D^{mh+r} P &= \lim_{m \rightarrow \infty} \left( \bigoplus_{i=1}^h D_{i,h}^m D_{i,r} \right) (K_h^r \otimes I_{n'}) \\ &= \left( \bigoplus_{i=1}^h J_{n'} D_{i,r} \right) (K_h^r \otimes I_{n'}) \\ &= \left( \bigoplus_{i=1}^h J_{n'} \right) (K_h^r \otimes I_{n'}) \\ &= K_h^r \otimes J_{n'}. \end{aligned}$$

□

Theorem 4.4 tells us that the structure of the matrix  $P^\top D^m P$  follows a cyclic pattern, while each individual non-zero submatrix semi-converges to the matrix  $J_{n'}$ . However, we can say even more. Indeed, we can show that the semi-convergence of each individual element is monotone. Therefore, while  $P^\top D^m P$  does not itself converge in general, each submatrix gets closer to  $J_{n'}$  in a monotonic way.

**Corollary 4.6.** *Let  $D \in \Omega_n$  be an irreducible doubly stochastic matrix and let  $P$  be a permutation matrix such that*

$$P^\top D P = \left( \bigoplus_{i=1}^h D_i \right) (K_h \otimes I_{n'}),$$

where  $h := h(D)$  and  $D_i \in \Omega_{n'}$ ,  $1 \leq i \leq h$ , and  $n' = n/h$ . Then

$$P^\top D^m P = \left( \bigoplus_{i=1}^h D_{i,m} \right) (K_h^m \otimes I_{n'})$$

and

$$\|D_{i,m+1} - J_{n'}\|_2 \leq \|D_{i,m} - J_{n'}\|_2, \quad 1 \leq i \leq h,$$

where  $\|\cdot\|_2$  denotes the spectral norm.

*Proof:* The first assertion is a restatement of Lemma 4.2. Then observe that

$$D_{i,m+1} = \prod_{j=1}^{m+1} D_{[i+j-1]} = \left( \prod_{j=1}^m D_{[i+j-1]} \right) D_{[i+m]} = D_{i,m} D_{[i+m]}.$$

Since  $J_n D = J_n$  for any  $D \in \Omega_n$ , we have

$$D_{i,m+1} - J_{n'} = D_{i,m} D_{[i+m]} - J_{n'} = (D_{i,m} - J_{n'}) D_{[i+m]}.$$

Additionally, as  $D^\top D \in \Omega_n$  and the spectral radius of any doubly stochastic matrix is equal to 1, it directly follows that  $\|D\|_2 = \sqrt{\rho(D^\top D)} = 1$  for any doubly stochastic matrix  $D$ . Therefore,

$$\begin{aligned} \|D_{i,m+1} - J_{n'}\|_2 &= \|(D_{i,m} - J_{n'}) D_{[i+m]}\|_2 \\ &\leq \|D_{i,m} - J_{n'}\|_2 \|D_{[i+m]}\|_2 \\ &= \|D_{i,m} - J_{n'}\|_2, \end{aligned}$$

which concludes the proof. □

**4.2. Cyclic and primitive doubly stochastic matrices.** In the following theorem, we characterize the irreducible doubly stochastic matrices which describe a cycle of order  $p$ .

**Theorem 4.7.** *Let  $D$  be an  $n \times n$  irreducible doubly stochastic matrix. Then  $D$  is cyclic if and only if there exists a permutation matrix  $P$  such that*

$$(4.8) \quad P^\top D P = K_h \otimes J_{n'},$$

where  $h := h(D)$  and  $n' = n/h$ . Moreover, if  $D$  is cyclic of order  $p$ , then  $p-1$  divides  $h$ .

*Proof:* Suppose that  $D^p = D$ . By (2.4), there is a permutation matrix  $P$  such that

$$P^\top DP = \left( \bigoplus_{i=1}^h D_i \right) (K_h \otimes I_{n'}),$$

where  $h = h(D)$ ,  $D_i \in \Omega_{n'}$ ,  $1 \leq i \leq h$ , and  $n' = n/h$ . Let  $\lambda_1, \lambda_2, \dots, \lambda_n$  denote the eigenvalues of  $D$  and let  $U$  be the unitary matrix in a Schur triangularization of  $D$  (see [13, Theorem 2.3.1]), i.e., a unitary matrix which satisfies

$$U^*DU = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ * & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \dots & \lambda_n \end{bmatrix}.$$

Then, since  $D = D^p$ , we have

$$\begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ * & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \dots & \lambda_n \end{bmatrix} = U^*DU = U^*D^pU = (U^*DU)^p = \begin{bmatrix} \lambda_1^p & 0 & \dots & 0 \\ * & \lambda_2^p & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \dots & \lambda_n^p \end{bmatrix}.$$

Consequently,  $\lambda_k = \lambda_k^p$  for all  $k \geq 1$ ; indeed, if a matrix satisfies a polynomial equation, then each of its eigenvalues must be a root of that polynomial. Thus, the eigenvalues of the matrix  $D$  are either of unit modulus, or they are equal to 0. However, by the definition of  $h$ , we know that  $D$  has exactly  $h$  eigenvalues of unit modulus. Furthermore, the Perron–Frobenius theorem guarantees that the spectrum of  $D$  is invariant under multiplication by  $e^{\frac{2\pi i}{h}}$ . Therefore, the unimodular eigenvalues of  $D$  are precisely  $e^{\frac{2\pi ki}{h}}$ ,  $1 \leq k \leq h$ , along with  $n - h$  other eigenvalues, all equal to 0.

In particular, it follows that the eigenvalues of  $D^h$  are 1s (with geometric multiplicity  $h$ ) and 0s (with geometric multiplicity  $n - h$ ). However, by Lemma 4.2,

$$(4.9) \quad \begin{aligned} P^\top D^h P &= \begin{bmatrix} D_1 \cdots D_h & 0 & \dots & 0 \\ 0 & D_2 \cdots D_h D_1 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & D_h D_1 \cdots D_{h-1} \end{bmatrix} \\ &=: \begin{bmatrix} D_{1,h} & 0 & \dots & 0 \\ 0 & D_{2,h} & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & D_{h,h} \end{bmatrix}. \end{aligned}$$

Recall the definition of  $D_{i,r}$  from (4.5). Since all the matrices in the main diagonal are  $n' \times n'$  doubly stochastic matrices, they must all have at least one eigenvalue equal to 1. Thus, since  $D$  has precisely  $h$  unimodular eigenvalues, each doubly stochastic matrix in the main diagonal must contain a unique unimodular eigenvalue (equal to 1) while every other eigenvalue is equal to 0. Hence, it follows from Lemma 2.8 that  $D_{k,h} = J_{n'}$  for every  $1 \leq k \leq h$ .

Lastly, because of the particular form of  $D$ , given by (4.8), the order  $p$  must be of the form  $p = 1 + mh$ , for some integer  $m \geq 0$ , since otherwise we would find using Lemma 4.2 that the non-zero block matrices of  $P^\top DP$  and  $P^\top D^p P$  would not coincide, which is a contradiction.

Therefore, on the one hand, we have

$$P^\top DP = \begin{bmatrix} 0 & D_1 & 0 & \dots & 0 \\ 0 & 0 & D_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & D_{h-1} \\ D_h & 0 & 0 & \dots & 0 \end{bmatrix}.$$

On the other hand,

$$\begin{aligned} P^\top DP &= P^\top D^p P = (P^\top DP)^{1+mh} = P^\top DP (P^\top D^h P)^m \\ &= P^\top DP (J_{n'} \oplus \dots \oplus J_{n'})^m = P^\top DP (J_{n'} \oplus \dots \oplus J_{n'}) \\ &= \begin{bmatrix} 0 & J_{n'} & 0 & \dots & 0 \\ 0 & 0 & J_{n'} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & J_{n'} \\ J_{n'} & 0 & 0 & \dots & 0 \end{bmatrix} = K_h \otimes J_{n'}. \end{aligned}$$

The above argument is reversible, and the reverse is even simpler. □

We know from (4.3) that if  $D \in \Omega_n$ , then  $\lim_{m \rightarrow \infty} D^m = J_n$  if and only if  $h(D) = 1$ . As an easy corollary of Theorem 4.4, we shall complete the picture by showing that  $\lim_{m \rightarrow \infty} D^m$  exists if and only if 1 is the only eigenvalue of unit modulus, i.e.,  $h(D) = 1$ . It is worth mentioning that Fritz, Huppert, and Willems provided a sufficient condition which was very close to being necessary [8] (see equation (1.1)). But our result given here is, on the one hand, the consequence of a much more general result and, on the other hand, necessary to fully characterize the limiting behavior of the sequence  $(D^m)_{m \geq 1}$ .

**Theorem 4.10.** *Let  $D \in \Omega_n$  be irreducible. Then  $\lim_{m \rightarrow \infty} D^m$  exists if and only if  $D$  is primitive. Moreover, if so, we necessarily have  $\lim_{m \rightarrow \infty} D^m = J_n$ .*

*Proof:* We have the representation equation (2.5). Moreover, by Theorem 4.4, for every  $1 \leq r \leq h$ , we have

$$D^{mh+r} \xrightarrow{m \rightarrow \infty} P(K_h^r \otimes J_{n'})P^\top.$$

Therefore, since  $h \geq 1$ , a necessary condition for the limit of the powers of  $D$  to exist is that  $h = 1$  (i.e., that  $D$  is primitive), otherwise different values of  $r$  would yield distinct subsequences with unequal limits. Moreover, if  $h = 1$ , then  $r = 1$ , and Theorem 4.4 once again gives

$$\lim_{m \rightarrow \infty} D^m = \lim_{m \rightarrow \infty} D^{m+1} = P(K_1 \otimes J_n)P^\top = J_n. \quad \square$$

**4.3. The general case.** In this last part we consider a general doubly stochastic matrix  $D$ , which is not necessarily irreducible. In this case, the behavior of  $D^m$  as  $m \rightarrow \infty$  depends on the asymptotic behavior of its irreducible submatrices when it is expressed as the direct sum  $P^\top DP = D_1 \oplus \dots \oplus D_r$ . Therefore, it is possible for some components to converge while other components may diverge or be cyclic. We use the previously established results to characterize the general structure of doubly stochastic matrices for which the powers converge or exhibit a cycle. Moreover, we also show that the powers of a uniformly chosen doubly stochastic matrix will *almost always* converge to the uniform matrix  $J_n$ . We begin by characterizing the doubly stochastic matrices which describe a cycle of order  $p$ .

**Theorem 4.11.** *Let  $D \in \Omega_n$ . Then  $D$  is cyclic of order  $p$  if and only if there exist a permutation matrix  $P$  and positive integers  $r, h_i, k_i, 1 \leq i \leq r$ , satisfying  $p = 1 + \text{LCM}(h_1, h_2, \dots, h_r)$  and  $h_1k_1 + h_2k_2 + \dots + h_rk_r = n$  such that*

$$P^\top DP = \bigoplus_{j=1}^r (K_{h_j} \otimes J_{k_j}).$$

*In that case,  $D$  has  $r$  eigenvalues equal to 1 and  $h(D) = h_1 + h_2 + \dots + h_r$ .*

*Proof:* According to the representation equation (2.3), there is a permutation matrix  $P$  such that

$$P^\top DP = D_1 \oplus D_2 \oplus \dots \oplus D_r,$$

where each  $D_i \in \Omega_{n_i}$  is irreducible with  $n_1 + \dots + n_r = n$ . Now,  $D$  is cyclic of order  $p$  if and only if  $P^\top DP$  is cyclic of order  $p$ . The latter happens if and only if  $D_i^p = D_i$  for all  $1 \leq i \leq r$  and, moreover, for each integer  $m$  with  $1 < m < p$ , there exists at least one index  $1 \leq k \leq r$  such that  $D_k^m \neq D_k$ . In particular, every submatrix  $D_i$  must be cyclic of order  $p_i | p$ . Hence, Theorem 4.7 guarantees that for each  $D_i$ , there exists a permutation matrix  $P$  such that

$$P_i^\top D_i P_i = K_{p_i} \otimes J_{n_i/p_i}.$$

Therefore, writing  $h_i := p_i$  and  $k_i := n_i/p_i$ , we find that there exists a permutation matrix  $Q$  such that

$$Q^\top DQ = \bigoplus_{j=1}^r (K_{h_j} \otimes J_{k_j}).$$

Moreover, each submatrix is cyclic of order  $h_j$  and thus the smallest number  $p$  such that  $D^p = D$  is  $p = \text{LCM}(h_1, h_2, \dots, h_r)$ . □

We now focus on characterizing the doubly stochastic matrices for which the powers semi-converge to some doubly stochastic matrix. As in the previous case, we do not assume that the matrix is irreducible. Recall (2.3), which is used below.

**Theorem 4.12.** *Let  $D \in \Omega_n$ , and let  $h = h(D)$ . Let  $P$  be a permutation matrix such that*

$$P^\top DP = D_1 \oplus D_2 \oplus \dots \oplus D_r,$$

*where each  $D_i \in \Omega_{n_i}$  is irreducible. Then  $D^m$  semi-converges as  $m \rightarrow \infty$  if and only if  $D$  has precisely  $h$  eigenvalues equal to 1. Moreover, in this case,  $r = h$  and  $\lim_{m \rightarrow \infty} D^m = P(J_{n_1} \oplus J_{n_2} \oplus \dots \oplus J_{n_r})P^\top$ .*

*Proof:* The proof is a direct consequence of (2.3) and Theorem 4.10. □

### 5. A probabilistic result

When randomly selecting a doubly stochastic matrix of order  $n$ , what level of certainty can we establish regarding the convergence of its powers? In this regard and with respect to the Lebesgue measure, we can show that the powers of *almost all* doubly stochastic matrices will converge to the matrix  $J_n$ . To verify this claim, recall that if the coefficients of an  $n \times n$  matrix  $A$  are non-negative, then  $A$  is primitive if and only if the coefficients of  $A^m$  are positive for some  $m \geq 1$  [13, Theorem 8.5.2]. Hence, if the coefficients of  $A \in \Omega_n$  are positive, then  $A$  is primitive and Theorem 4.10 ensures that  $A^m \rightarrow J_n$  as  $m \rightarrow \infty$ . Consequently, it is enough to convince ourselves that *almost all* doubly stochastic matrices are positive. Intuitively, it suffices to observe that the set  $\Omega_n^+$  of all-component positive doubly stochastic matrices is precisely the

interior of the polytope  $\Omega_n$  (relative to the standard topology on  $\mathbb{R}^{n^2}$ ). Hence, since  $\Omega_n^+$  is non-empty, we should expect a uniformly chosen random doubly stochastic matrix to be in  $\Omega_n^+$ .

**Theorem 5.1.** *For almost all doubly stochastic matrices  $D$ ,  $D^m$  converges to the matrix  $J_n$  as  $m \rightarrow \infty$ .*

*Proof:* As discussed before the theorem, it suffices to show that almost all doubly stochastic matrices are in  $\Omega_n^+$ . We define  $\mathcal{E}_{n-1}$ ,  $n \geq 2$ , the subset of all  $(n-1) \times (n-1)$  matrices, as follows.

- (a) The matrices in  $\mathcal{E}_{n-1}$  are non-negative (componentwise).
- (b) The row sums and column sums never exceed 1.
- (c) The sum of all entries is larger than or equal to  $n - 2$ .

In other words,  $\mathcal{E}_{n-1}$  is the  $(n - 1)^2$ -dimensional convex polytope in  $\mathbb{R}^{(n-1) \times (n-1)}$  consisting of the  $(n - 1) \times (n - 1)$  matrices  $A = (a_{i,j})$  in the intersection of the closed half-planes

$$(5.2) \quad a_{i,j} \geq 0 \quad (1 \leq i, j \leq n - 1),$$

$$(5.3) \quad \sum_{j=1}^{n-1} a_{i,j} \leq 1 \quad (1 \leq i \leq n - 1),$$

$$(5.4) \quad \sum_{i=1}^{n-1} a_{i,j} \leq 1 \quad (1 \leq j \leq n - 1),$$

$$(5.5) \quad \sum_{i,j=1}^{n-1} a_{i,j} \geq n - 2.$$

It is easily verified that the mapping  $\Phi: \Omega_n \rightarrow \mathcal{E}_{n-1}$ , given by

$$\begin{bmatrix} a_{1,1} & \cdots & a_{1,n-1} & a_{1,n} \\ \vdots & \ddots & \vdots & \vdots \\ a_{n-1,1} & \cdots & a_{n-1,n-1} & a_{n-1,n} \\ a_{n,1} & \cdots & a_{n,n-1} & a_{n,n} \end{bmatrix} \mapsto \begin{bmatrix} a_{1,1} & \cdots & a_{1,n-1} \\ \vdots & \ddots & \vdots \\ a_{n-1,1} & \cdots & a_{n-1,n-1} \end{bmatrix},$$

along with its inverse  $\Phi^{-1}: \mathcal{E}_{n-1} \rightarrow \Omega_n$ ,

$$\begin{bmatrix} a_{1,1} & \cdots & a_{1,n-1} \\ \vdots & \ddots & \vdots \\ a_{n-1,1} & \cdots & a_{n-1,n-1} \end{bmatrix} \mapsto \begin{bmatrix} a_{1,1} & \cdots & a_{1,n-1} & 1 - \sum a_{1,j} \\ \vdots & \ddots & \vdots & \vdots \\ a_{n-1,1} & \cdots & a_{n-1,n-1} & 1 - \sum a_{n-1,j} \\ 1 - \sum a_{i,1} & \cdots & 1 - \sum a_{i,n-1} & 2 - n + \sum a_{i,j} \end{bmatrix},$$

is one-to-one.

Let  $\Omega_n^0$  denote the set of  $n \times n$  doubly stochastic matrices with at least one null coefficient. By the same token, we consider  $\Omega_n^0(i, j)$ , the set of  $n \times n$  doubly stochastic matrices with the coefficient in position  $(i, j)$  being zero. Then put

$$\mathcal{E}_{n-1}^0 := \Phi(\Omega_n^0), \quad \text{and} \quad \mathcal{E}_{n-1}^0(i, j) := \Phi(\Omega_n^0(i, j)).$$

Now the following relations hold.

- (a) If  $1 \leq i, j \leq n - 1$ , then  $\mathcal{E}_{n-1}^0(i, j)$  corresponds to the polytope  $\mathcal{E}_{n-1}^0$  with an equality in (5.2) corresponding to coefficients  $(i, j)$ .
- (b) If  $j = n$  and  $i < n$ , then  $\mathcal{E}_{n-1}^0(i, j)$  corresponds to the polytope  $\mathcal{E}_{n-1}^0$  with an equality in (5.3) corresponding to the coefficient  $i$ .
- (c) If  $i = n$  and  $j < n$ , then  $\mathcal{E}_{n-1}^0(i, j)$  corresponds to the polytope  $\mathcal{E}_{n-1}^0$  with an equality in (5.4) corresponding to the coefficient  $j$ .
- (d) If  $i = j = n$ , then  $\mathcal{E}_{n-1}^0(i, j)$  corresponds to the polytope  $\mathcal{E}_{n-1}^0$  with an equality in (5.5).

In each case, we lose a degree of freedom in the parameters of  $\mathcal{E}_{n-1}^0$ . Consequently, we have  $\dim(\mathcal{E}_{n-1}^0(i, j)) = (n-1)^2 - 1 < (n-1)^2$ . Therefore, if  $m$  denotes the Lebesgue measure on  $\mathbb{R}^{(n-1)^2}$ , we trivially have  $m(\mathcal{E}_{n-1}^0(i, j)) = 0$  for all  $1 \leq i, j \leq n - 1$ . It thus follows that

$$m(\mathcal{E}_{n-1}^0) = m\left(\bigcup_{i,j} \mathcal{E}_{n-1}^0(i, j)\right) \leq \sum_{i,j} m(\mathcal{E}_{n-1}^0(i, j)) = 0.$$

Moreover, observe that

$$\left[\frac{n-2}{(n-1)^2}, \frac{1}{n}\right]^{(n-1)^2} \subseteq \mathcal{E}_{n-1},$$

since for any element  $A$  in the former set,  $\Phi^{-1}(A) \in \Omega_n$ . Consequently,

$$m(\mathcal{E}_{n-1}) \geq m\left(\left[\frac{n-2}{(n-1)^2}, \frac{1}{n}\right]^{(n-1)^2}\right) = (n(n-1)^2)^{-(n-1)^2} > 0.$$

Hence, almost all elements of  $\mathcal{E}_{n-1}$  belong to  $\mathcal{E}_{n-1} \setminus \mathcal{E}_{n-1}^0$ , the range of  $\Phi$  of the positive doubly stochastic matrices. Since  $\Phi$  satisfies the conditions of Theorem 31 in [27], the conclusion follows. □

*Remark.* An  $n \times n$  doubly stochastic matrix is completely determined by its  $(n-1) \times (n-1)$  upper left entries. There is thus a bijection between the set of  $n \times n$  doubly stochastic matrices and the set  $\mathcal{V}$  of vectors of  $v \in (\mathbb{R}^+)^{(n-1)^2}$  satisfying the conditions

- (a)  $\sum_{j=1}^{n-1} v_{m(n-1)+j} \leq 1, \quad \forall m$  such that  $0 \leq m \leq n-2$ ,
- (b)  $\sum_{j=0}^{n-2} v_{j(n-1)+m} \leq 1, \quad \forall m$  such that  $1 \leq m \leq n-1$ .

This bijection between doubly stochastic matrices and the convex set  $\mathcal{V}$  makes it possible to sample matrices using the Lebesgue measure in  $\mathbb{R}^{(n-1)^2}$ .

From a practical point of view, random doubly stochastic matrices are usually generated by the Sinkhorn process, which consists of taking a random non-negative matrix and iteratively normalizing the rows and columns. This process always converges linearly to a doubly stochastic matrix, as long as the initial non-negative matrix does not have *too many zeros* (the formal condition is that the matrix has total support; for more information, we refer to [29]). The important observation here is that the zero coefficients in the doubly stochastic matrix obtained in the limit are precisely in the same positions as the initial non-negative matrix. Consequently, as long as the initial random non-negative matrices are obtained in such a way that almost

all of them are positive, then almost all doubly stochastic matrices resulting from the Sinkhorn process will be positive and thus will converge to the matrix  $J_n$ .

**Acknowledgements.** We sincerely thank the two anonymous referees for their insightful comments, which have significantly improved the quality of our article. We are also grateful to them for drawing our attention to the works of Hajnal [10, 11] – classics, to be sure, but presented in a different context and therefore not previously on our radar – as well as to the related contributions of Wolfowitz [31] and Chatterjee [6]. Their careful reading and thoughtful suggestions have been especially valuable.

## References

- [1] G. BIRKHOFF, Tres observaciones sobre el algebra lineal, *Univ. Nac. Tucumán Rev. Ser. A* **5** (1946), 147–151.
- [2] L. BOUTHAT, J. MASHREGHI, AND F. MORNEAU-GUÉRIN, On the geometry of the Birkhoff polytope I: the operator  $\ell_n^p$ -norms, *Acta Sci. Math. (Szeged)* **91(1-2)** (2025), 227–245. DOI: 10.1007/s44146-024-00152-8.
- [3] L. BOUTHAT, J. MASHREGHI, AND F. MORNEAU-GUÉRIN, On the geometry of the Birkhoff polytope II: the Schatten  $p$ -norms, *Acta Sci. Math. (Szeged)* **91(3-4)** (2025), 401–419. DOI: 10.1007/s44146-024-00153-7.
- [4] S. BOYD, A. GHOSH, B. PRABHAKAR, AND D. SHAH, Randomized gossip algorithms, *IEEE Trans. Inform. Theory* **52(6)** (2006), 2508–2530. DOI: 10.1109/TIT.2006.874516.
- [5] X. CAO, S. GAO, AND L. CHEN, Gossip-based load balance strategy in big data systems with hierarchical processors, *Wireless Pers. Commun.* **98** (2018), 157–172. DOI: 10.1007/s11277-017-4861-4.
- [6] S. CHATTERJEE AND E. SENETA, Towards consensus: some convergence theorems on repeated averaging, *J. Appl. Probability* **14(1)** (1977), 89–97. DOI: 10.2307/3213262.
- [7] HONG KE DU, The limit of powers of doubly stochastic matrices (Chinese), *Math. Practice Theory* **3** (1987), 76–77.
- [8] F.-J. FRITZ, B. HUPPERT, AND W. WILLEMS, *Stochastische Matrizen*, Hochschultext [University Textbooks], Springer-Verlag, Berlin-New York, 1979. DOI: 10.1007/978-3-642-67131-9.
- [9] G. FROBENIUS, Über Matrizen aus nicht negativen Elementen, *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften* (1912), 456–477.
- [10] J. HAJNAL, Weak ergodicity in non-homogeneous Markov chains, *Proc. Cambridge Philos. Soc.* **54(2)** (1958), 233–246. DOI: 10.1017/S0305004100033399.
- [11] J. HAJNAL, On products of non-negative matrices, *Math. Proc. Cambridge Philos. Soc.* **79(3)** (1976), 521–530. DOI: 10.1017/S030500410005252X.
- [12] K. HENSEL, Über Potenzreihen von Matrizen, *J. Reine Angew. Math.* **1926(155)** (1926), 107–110. DOI: 10.1515/crll.1926.155.107.
- [13] R. A. HORN AND C. R. JOHNSON, *Matrix Analysis*, Second edition, Cambridge University Press, Cambridge, 2013. DOI: 10.1017/CB09781139020411.
- [14] J. J. HUNTER, Some stochastic properties of “semi-magic” and “magic” Markov chains, *Linear Algebra Appl.* **433(5)** (2010), 893–907. DOI: 10.1016/j.laa.2010.04.021.
- [15] S.-G. HWANG AND S.-S. PYO, Doubly stochastic matrices whose powers eventually stop, *Linear Algebra Appl.* **330(1-3)** (2001), 25–30. DOI: 10.1016/S0024-3795(01)00260-9.
- [16] J. G. KEMENY AND J. L. SNELL, *Finite Markov Chains*, Reprinting of the 1960 original, Undergrad. Texts Math., Springer-Verlag, New York-Heidelberg, 1976.
- [17] A. KOLOSKOVA, S. STICH, AND M. JAGGI, Decentralized stochastic optimization and gossip algorithms with compressed communication, in: *Proceedings of the 36th International Conference on Machine Learning*, PMLR **97**, 2019, pp. 3478–3487.
- [18] R. LAZZERETTI, S. HORN, P. BRACA, AND P. WILLETT, Secure multi-party consensus gossip algorithms, in: *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 7406–7410. DOI: 10.1109/ICASSP.2014.6855039.
- [19] D. A. LEVIN, Y. PERES, AND E. L. WILMER, *Markov Chains and Mixing Times*, American Mathematical Society, Providence, RI, 2009. DOI: 10.1090/mbk/058.
- [20] V. M. MAKSIMOV, On the relation between limit theorems on finite groups and an ergodic theorem for Markov chains with doubly-stochastic transition matrix (Russian), *Teor. Veroyatnost. i Primenen.* **10(3)** (1965), 544–547. DOI: 10.1137/1110060.

- [21] V. M. MAKSIMOV, Convergence of non-homogeneous bistochastic Markov chains, (Russian), *Teor. Veroyatnost. i Primenen.* **15** (1970), 622–636. English translation in: *Theor. Probability Appl.* **15** (1970), 604–618. DOI: 10.1137/1115070.
- [22] M. MARCUS, On subdeterminants of doubly stochastic matrices, *Illinois J. Math.* **1(4)** (1957), 583–590. DOI: 10.1215/IJM/1255380681.
- [23] M. MARCUS, H. MINC, AND B. MOYLS, Some results on non-negative matrices, *J. Res. Nat. Bur. Standards Sect. B* **65B(3)** (1961), 205–209. DOI: 10.6028/jres.065b.019.
- [24] H. PERFECT AND L. MIRSKY, Spectral properties of doubly-stochastic matrices, *Monatsh. Math.* **69** (1965), 35–57. DOI: 10.1007/BF01313442.
- [25] O. PERRON, Zur Theorie der Matrizes, *Math. Ann.* **64(2)** (1907), 248–263. DOI: 10.1007/BF01449896.
- [26] S. RÉNIER, Characterization, parity, and power sequences of locally finite doubly stochastic matrices, *Discrete Math.* **309(23–24)** (2009), 6563–6571. DOI: 10.1016/j.disc.2009.07.005.
- [27] C. A. ROGERS, *Hausdorff Measures*, Cambridge University Press, London-New York, 1970. DOI: 10.1017/S001309150000969X.
- [28] Š. SCHWARZ, Infinite products of doubly stochastic matrices (Russian), *Acta Math. Univ. Comenian.* **39** (1980), 131–150.
- [29] R. D. SINKHORN AND P. J. KNOPP, Concerning nonnegative matrices and doubly stochastic matrices, *Pacific J. Math.* **21(2)** (1967), 343–348. DOI: 10.2140/PJM.1967.21.343.
- [30] A. VOUDAS, Markov chains with doubly stochastic transition matrices and application to a sequence of non-selective quantum measurements, *Phys. A* **593** (2022), Paper no. 126911, 17 pp. DOI: 10.1016/j.physa.2022.126911.
- [31] J. WOLFOVITZ, Products of indecomposable, aperiodic, stochastic matrices, *Proc. Amer. Math. Soc.* **14** (1963), 733–737. DOI: 10.2307/2034984.

Ludovick Bouthat, Nicolas Doyon, Javad Mashreghi

Département de mathématiques et de statistique, Université Laval, 2325 Rue de l'Université, Québec, G1V 0A6, QC, Canada

*E-mail address:* ludovick.bouthat.1@ulaval.ca

ORCID: 0000-0002-2094-3028

*E-mail address:* nicolas.doyon@mat.ulaval.ca

ORCID: 0009-0006-5095-4007

*E-mail address:* javad.mashreghi@mat.ulaval.ca

ORCID: 0000-0002-7969-9576

Frédéric Morneau-Guérin

Département Éducation, Université TÉLUQ, 455 Rue du Parvis, Québec, G1K 9H6, QC, Canada

*E-mail address:* frederic.morneau-guerin@teluq.ca

ORCID: 0000-0001-7610-4648

Received on November 8, 2024.

Accepted on February 23, 2026.