UCLA
COMPUTATIONAL AND APPLIED MATHEMATICS

A Geomtric Approach to Ergodic Non-Homogeneous Markov Chains

Jianhong Shen

January 1999

CAM Report 99-4

Department of Mathematics
University of California, Los Angeles
Los Angeles, CA. 90095-1555

A GEOMETRIC APPROACH TO ERGODIC NON-HOMOGENEOUS MARKOV CHAINS

JIANHONG SHEN

ABSTRACT. Inspired by the recent work of Daubechies and Lagarias on a set of matrices with convergent infinite products, we establish a new geometric approach to the classical problem of (weakly) ergodic non-homogeneous Markov chains. The existing key inequalities (related to the Hajnal inequality) in the literature are unified in this geometric picture. A more general inequality is established. Important quantities introduced by various authors are easily interpreted. A quantitative connection is established between the classical work of Hajnal and the more recent one of Daubechies and Lagarias.

1. OVERVIEW

In this paper, we restrict ourselves to Markov chains on finite states.

Section 2 and 3 together serve as an introductory section, where the historical background and recent development are briefly reviewed, and our motivation and new results are introduced.

In Section 2, we review the literature of ergodic non-homogeneous Markov chains [1, 2, 4, 5, 8, 10]. The basic meaning of ergodicity is explained. Section 3 displays the efforts of several authors to characterize the conditions for ergodicity. Hajnal's inequality [5] is singled out for its significant role in this course. Our most general inequality is introduced. New concepts such as Σ-*Markov chains* and *LCP (left-convergent-products*, Daubechies and Lagarias [2]) are defined. We also claim that the projected jointed spectral radius of Daubechies and Lagarias can be characterized in a new way.

Section 4 shows the motivation of our geometric approach by working out the simple example of Cantorian (non-homogeneous) Markov chains. It

1991 *Mathematics Subject Classification.* 15A51, 60J10, 15A60.

Key words and phrases. Simplices, stochastic matrices, LCP, ergodicity, Hajnal inequality, Cantorian Markov chains, joint spectral radius.

allows us to see the many advantages of treating transition (stochastic) matrices as transforms of simplices. This example is inspired by the connection we find among the examples that Daubechies and Lagarias discussed in [2].

Section 5 is the mainland. In our geometric approach, a stochastic matrix is understood as a transform of simplices. The classical stochastic matrices such as *Markov matrices* and *scrambling matrices* are easily understood in this picture. Classical quantities that were used to characterize ergodicity find simple geometric interpretations. Remarkably, our approach links ergodicity to a much more familiar one in complete metric spaces. The essential meaning of Hajnal's inequality is understood with our most general result, whose proof is reduced to a simple relation between a simplex and its linearly contracted copy. We thereby generalize the results of Markov (see Seneta [8]), Hajnal [5], Dobrušin [3], and Paz and Reichaw [6].

In the last section, we define a digital character $\pi(\Sigma)$ for a set of stochastic matrices Σ, which is called the *Hajnal rate*. It leads to a necessary and sufficient condition for ergodicity. We also show that in the ergodic case, the Hajnal rate is exactly the *projected joint spectral radius* introduced by Daubechies and Lagarias in [2].

Most classical results were only about a *finite* set of stochastic matrices ([1, 2, 10]). The generalization to a compact set of stochastic matrices continues the author's work in [9].

2. Preliminaries

Given n distinct states $|1\rangle, |2\rangle, \cdots, |n\rangle$, a homogeneous Markov chain is characterized uniquely by its one-step transition matrix $P = (p_{ij})_{n\times n}$. Here p_{ij} is the probability for a "particle" in state $|i\rangle$ to transit to $|j\rangle$ directly. Since at each step, the particle must evolve to some new state,

$$p_{i1} + p_{i2} + \cdots + p_{in} = 1.$$

Such a non-negative matrix whose row sums are always 1 is called a *stochastic matrix*. To put it in a familiar way, this paper is about products of stochastic matrices and their convergence as the number of matrices increases to infinity.

A (homogeneous) Markov chain P is said to be *ergodic* if (in Hajnal's language [5])

"The particle forgets its past."

It can be made rigorously as follows. Assume that initially no state is preferable, i.e. at time step $t = 0$, the probability of finding the particle in any state $|i\rangle$ is $1/n$. Suppose that during a certain experiment, the particle is observed in some state $|j\rangle$ at some time step $t \gg 1$. Then the ergodicity condition means that almost nothing can be concluded about the initial state of this *particular* particle, i.e.

$$\text{Prob}(X_0 = |i\rangle \mid X_t = |j\rangle) \simeq \frac{1}{n},$$

for any state $|i\rangle$. Here event $X_s = |k\rangle$ means the particle is found in state $|k\rangle$ at time step s. Let p^t_{ij} denote the probability for a particle initially in state $|i\rangle$ to be in state $|j\rangle$ at time t, and p^0_i the probability for $X_0 = |i\rangle$. From Bayes' formula

$$\text{Prob}(X_0 = |i\rangle \mid X_t = |j\rangle) = \frac{p^t_{ij} p^0_i}{\sum_{k=1}^n p^t_{kj} p^0_k},$$

we conclude that

$$p^t_{ij} p^0_i \simeq p^t_{kj} p^0_k,$$

for any i and k. Since it is assumed that $p^0_i \equiv 1/n$ for all i, we must have

$$p^t_{ij} \simeq p^t_{kj}.$$

In the homogeneous case, the matrix $(p^t_{ij})_{n\times n}$ is in fact P^t. The ergodicity condition therefore demands that as t increases, each column of P^t tend to be a constant one (though this constant may change from time to time).

Throughout the paper, $\mathbf{J}$ always denotes the *column* vector $(1, 1, \cdots, 1)^T$. A p.d.f (probability density function) $\mathbf{p}$ is a nonnegative *row* vector $(p_1, p_2, \cdots, p_n)$ with a unit sum. From above, a rigorous definition of ergodicity is

Definition 2.1 (Ergodicity. Hajnal [5]). A (homogeneous) Markov chain whose transition matrix is P is said to be ergodic if for any $\epsilon > 0$, there exists M, such that for any $m > M$, one can find a p.d.f. $\mathbf{p}$,

$$\|P^m - \mathbf{J}\cdot\mathbf{p}\| \le \epsilon. \tag{1}$$

Since all matrix norms are equivalent, it is unnecessary to specify the type of norm we are using.

In the homogeneous case, Condition (1) is equivalent to:

$$P^\infty = \lim_{m\to\infty} P^m \quad \text{exists and} \quad \text{rank}(P^\infty) = 1. \tag{2}$$

That (2) implies (1) is trivial. From (1), for any $n \geq 0$ and $m > M$,

$$\begin{aligned} \|P^{m+n} - P^m\| &= \|(P^{m+n} - \mathbf{J}\cdot\mathbf{p}) + (\mathbf{J}\cdot\mathbf{p} - P^m)\| \\ &\leq \|P^n(P^m - \mathbf{J}\cdot\mathbf{p})\| + \|\mathbf{J}\cdot\mathbf{p} - P^m\| \\ &\leq C\|P^m - \mathbf{J}\cdot\mathbf{p}\| \leq C\epsilon. \end{aligned}$$

(The constant C depends merely on the norm we are using since the set of stochastic matrices is bounded.) Therefore the sequence P^m does converge to some P^∞, which must have rank 1. This is the condition (2).

The picture changes as one considers the non-homogeneous case. A Markov chain is said to be *non-homogeneous* if the one-step transition matrix $P(t)$ changes with time step t. In applications, such a model is usually more successful due to the many changing factors that control the transition process. For a non-homogeneous Markov chain, the associated *transition set* Σ is defined to be

$$\Sigma = \{P \mid \text{at some time } \ t, \ P(t) = P\},$$

which is a set of stochastic matrices. In this paper, we shall consider frequently the opposite direction: given a set of stochastic matrices Σ, we are interested in all Σ-Markov chains. A Σ-Markov chain is one whose transition matrices are all taken from Σ.

Given a non-homogeneous Markov chain $(P(1), P(2), \cdots)$, define

$$P^{i:j} = P(i+1)P(i+2)\cdots P(j).$$

The Markov chain is said to be *ergodic* if for any given $i \geq 0$, as $j \to \infty$,

$$\text{each column of } P^{i:j} \text{ gets to be a constant column.} \tag{3}$$

(Though the constant may change with j.) Accurate statement like (1) can be put down similarly. However, unlike in the homogeneous case, generally one should not expect that $\lim_{j\to\infty} P^{i:j}$ exists.

Definition 2.2 (Sub-Markov chain). Given two Markov chains whose transition matrices are $(P(1), P(2), \cdots)$ and $(Q(1), Q(2), \cdots)$, respectively, we say the latter is a sub-Markov chain of the former if there exists a sequence $0 \leq m_0 < m_1 < m_2 < \cdots$, such that

$$Q(k) = P^{m_{k-1}:m_k}, \qquad k = 1, 2, \cdots.$$

Lemma 2.1. *A Markov chain is ergodic if and only if one of its sub-Markov chains is ergodic.*

Proof. It suffices to prove the direction from the right to the left. The other one is trivial. Suppose that the sub-Markov chain $(Q(1) = P^{m_0:m_1}, Q(2) = P^{m_1:m_2}, \cdots)$ is ergodic. We need show that the statement (3) is true. Suppose

$$m_{k-1} < i \leq m_k \leq m_{K-1} \leq j < m_K,$$

for some k and K. With i being fixed, as $j \to \infty$, so is K. Hence for any $\epsilon > 0$, there exists J, such that for any $j > J$, there is a p.d.f. $\mathbf{p}$, such that

$$\|Q^{k:K} - \mathbf{J} \cdot \mathbf{p}\| \leq \epsilon.$$

Now that

$$P^{i:j} = P^{i:m_k} \cdot Q^{k:K} \cdot P^{m_K:j},$$

setting $\mathbf{q} = \mathbf{p} \cdot P^{m_K:j}$ (which depends on j), we have

$$\begin{aligned}
\|P^{i:j} - \mathbf{J} \cdot \mathbf{q}\| &\leq \|P^{i:m_k}(Q^{k:K} \cdot P^{m_K:j} - \mathbf{J} \cdot \mathbf{q})\| \\
&\leq C_1 \|Q^{k:K} \cdot P^{m_K:j} - \mathbf{J} \cdot \mathbf{q}\| \\
&= C_1 \|(Q^{k:K} - \mathbf{J} \cdot \mathbf{p}) P^{m_K:j}\| \\
&\leq C_2 \|Q^{k:K} - \mathbf{J} \cdot \mathbf{p}\| \leq C_2 \epsilon.
\end{aligned}$$

Here C_1 and C_2 only depend on the norm. This completes the proof. □

Several authors have studied the conditions for ergodicity and the asymptotic behavior of $P^{i:j}$, which will be seen in details as we proceed.

3. Hajnal's inequality and the joint spectral radius

To a large extent, this paper has been inspired by the recent work of Daubechies and Lagarias [2] on the convergence of infinite matrix products, where non-homogeneous Markov chains served as one typical example for their general results. The geometric approach was discovered after our studying the relation between non-homogeneous Markov chains and the second class of their examples — the iterated function systems (see the next section). In this section, we review the existing results and outline some of our major contributions.

3.1. Hajnal's inequality. A stochastic matrix P is called a *Markov matrix* (Seneta [8]) if

$$P \text{ contains a positive column.} \tag{4}$$

By definition, if a non-homogeneous Markov chain $\{P(t) \mid t = 1, 2, \cdots\}$ is ergodic, then for each fixed i, as j gets large enough, $P^{i;j}$ must be a Markov matrix, since it is close to a rank one stochastic matrix. Based on his memorylessness picture of ergodicity mentioned above, Hajnal [5] introduced the powerful concept – *scrambling matrix.*

Definition 3.1 (Scrambling). Let P be the one-step transition matrix of a given Markov chain. P is said to be a *scrambling matrix* if for any two distinct states $|i\rangle$ and $|j\rangle$, there always exists a state $|k\rangle$, such that both one-step transitions are possible: $|i\rangle \to |k\rangle$, and $|j\rangle \to |k\rangle$; or equivalently, p_{ik} and p_{jk} are both positive.

In another word, for each pair of rows in P, there is a column such that the intersections are both positive. For any two vectors

$$\mathbf{a} = (a_1, \cdots, a_n), \qquad \mathbf{b} = (b_1, \cdots, b_n),$$

we define a new vector

$$\mathbf{a} \wedge \mathbf{b} = (\min(a_1, b_1), \cdots, \min(a_n, b_n)).$$

Denote the i-th row of P by $\mathbf{p}_i$, a p.d.f. To "measure" the scramblingness, Hajnal [5] defined

$$\lambda(P) = \min_{i,j} \|\mathbf{p}_i \wedge \mathbf{p}_j\|_1.$$

Here,

$$\|\mathbf{a}\|_q = (|a_1|^q + \cdots + |a_n|^q)^{\frac{1}{q}}.$$

It is easy to see from the definition that

(i) P is scrambling if and only if $\lambda(P) > 0$.
(ii) $\lambda(P) = 1$ if and only if P has rank 1. Especially, if $\lambda(P) = 1$, P is a Markov matrix.

Following these notations, the key inequality in the theory of ergodic Markov chains — Hajnal's inequality [5] can be stated simply. Define the *Hajnal diameter* of a stochastic matrix P as

$$\text{diam}_\infty(P) = \max_{i,j} \|\mathbf{p}_i - \mathbf{p}_j\|_\infty.$$

Classical Result 3.1 (Hajnal's inequality). *If P and Q are both stochastic matrices, then*

$$\text{diam}_\infty(PQ) \le (1 - \lambda(P))\text{diam}_\infty(Q).$$

To see its fundamental role in the literature, check some of the major classical results whose proofs were closely based on this inequality: Hajnal [5] provided a necessary and sufficient condition for ergodicity; Wolfowitz [10] connected ergodicity to the condition of SIA (see also Section 3); and finally, Anthonisse and Tijms [1] obtained the exponential convergence rate for infinite products of certain stochastic matrices.

Note that Hajnal and other authors before us did not use the notations we use here, nor adopt the point of view we have here that important quantities emerging unconsciously in the literature are in fact closely related geometric objects such as norms and diameters. From this geometric standpoint, not only that we can understand better Hajnal's inequality and its generalizations by other authors (Dobrušin [3]; Paz and Reichaw [6]), but also we are able to establish its most general form (Section 5).

Theorem 3.1 (Generalized Hajnal's Inequality). *Let* $\|\cdot\|$ *be a vector norm in* $\mathbb{R}^n$. *For any stochastic matrix* P *with row vectors* $\mathbf{p}_1, \cdots, \mathbf{p}_n$, *define*

$$\operatorname{diam}(P) = \max_{i,j} \|\mathbf{p}_i - \mathbf{p}_j\|.$$

Then for any two stochastic matrices P *and* Q,

$$\operatorname{diam}(PQ) \leq (1 - \lambda(P)) \operatorname{diam}(Q).$$

3.2. SIA set.

Definition 3.2 (SIA). A stochastic matrix P is said to be a SIA matrix (stochastic, irreducible and aperiodic; See Wolfowitz [10]) if Condition (2) is satisfied. A set of stochastic matrices Σ is called a SIA set if any *finite* product in Σ is a SIA matrix.

We have seen that a homogeneous Markov chain is ergodic if and only if the associated one-step transition matrix is a SIA matrix. Wolfowitz [10] generalized this result to the non-homogeneous case.

Classical Result 3.2. *Suppose a finite set* Σ *of stochastic matrices is a SIA set. Then, any* Σ*-Markov chain is ergodic.*

Wolfowitz in fact showed that all Σ-Markov chains are *uniformly* ergodic. The major tool of his proof is again the Hajnal inequality.

3.3. **The projected joint spectral radius.** Recently, motivated by the matrix aspect of the compactly supported wavelets, Daubechies and Lagarias [2] studied an interesting class of matrix sets, which we shall call

LCP (left-convergent-product) sets to fit our context. Their work added a new chapter to the theory of ergodic non-homogeneous Markov chains.

Definition 3.3 (LCP set). A set Σ of matrices (unnecessary to be stochastic) is said to be an *LCP set* if for any infinite sequence $(A_1, A_2, \cdots) \in \Sigma^\infty$,

$$A_\infty = \lim_{m\to\infty} A_m A_{m-1} \cdots A_1 \qquad \text{exists.}$$

(Pay attention to the direction of infiniteness.)

An LCP set is called a *V-LCP* (or vanishing LCP) set if the above limit is always a zero matrix.

An LCP set is called a *U-LCP* (or uniform LCP) set if the above limit converges uniformly for all sequences (see also Shen [9]).

A classical result of Daubechies-Lagarias [2] shows

Classical Result 3.3. *If Σ is a $U-LCP$ set, then*

(i) The eigenvalue $\lambda = 1$ is geometrically simple for any matrix in Σ (i.e. no 1-Jordan block of size greater than 1).

(ii) All matrices in Σ share a common 1-eigenspace $\mathbf{E}_1$.

(iii) Let Q denote any (skew) projection along $\mathbf{E}_1$ (i.e. $Q^2 = Q$ and the kernel of Q is exactly $\mathbf{E}_1$). Then

$$Q\Sigma Q^T = \{QAQ^T \mid A \in \Sigma\}$$

is a V-LCP set.

Besides, a finite set Σ of stochastic matrices has the property that any Σ-Markov chain is ergodic if and only if it is a rank 1 U-LCP set.

The dimension of $\mathbf{E}_1$ is called the *rank* of the U-LCP set.

Suppose Σ is a set of stochastic matrices. To analyze convergence, both Hajnal [5] and Daubechies-Lagarias [2] introduced two characteristic quantities: the *Hajnal diameter* $\mathrm{diam}_\infty(\cdot)$ we defined earlier and the *projected joint spectral radius.* The latter concept is a more general one than the former. The Hajnal diameter, which may only be meaningful for stochastic matrices, however, appears to be intuitive and more intrinsic for studying Markov chains.

In the last section, we shall establish a quantitative relation between their work.

Theorem 3.2 (Projected joint spectral radius v.s. Hajnal's rate). *Suppose Σ is a finite set of stochastic matrices so that all Σ-Markov chains are ergodic. Let $\mathbf{E}_1$ be the common 1-eigenspace of Σ and Q any skew projection* *along* $\mathbf{E}_1$. *Then*

$$\rho(Q\Sigma Q^T) = \pi(\Sigma).$$

In the theorem, ρ is the classical joint spectral radius (Daubechies-Lagarias [2]; Rota-Strang [7]. Also see Section 6). $\pi(\Sigma)$ is called the *Hajnal rate* and is defined by

$$\operatorname{diam}(\Sigma_m) = \sup_{B\in\Sigma_m} \operatorname{diam}(B)$$
$$\pi(\Sigma) = \limsup_{m\to\infty} \operatorname{diam}(\Sigma_m)^{1/m}.$$

Here $\Sigma_m = \{A_1A_2\cdots A_m \mid A_i \in \Sigma\}$.

As a result,

Corollary 3.1. *Let Σ be a finite set of stochastic matrices. Then all Σ-Markov chains are ergodic if and only if $\pi(\Sigma) < 1$.*

This concludes the introduction part of the paper.

4. An example: Cantorian Markov chains

What has motivated our geometric approach is this simple example.

Define

$$P_0 = \begin{bmatrix} 1 & 0 \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix}, \qquad P_1 = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ 0 & 1 \end{bmatrix},$$

and $\Sigma = \{P_0, P_1\}$. For any real number $x \in (0,1]$, define a Markov chain, which we call the *Cantorian x-Markov chain.* Let $x = 0.t_1t_2\cdots$ be the unique *infinite* dyadic expansion, i.e.

$$x = \frac{t_1}{2} + \frac{t_2}{4} + \cdots, \qquad t_i \in \{0,1\}.$$

The x-Markov chain is one such that at each time step n, the one-step transition matrix is P_{t_n}. Define the associated Cantorian number

$$c_x = \frac{2t_1}{3} + \frac{2t_2}{9} + \frac{2t_3}{27} + \cdots.$$

Then c_x obviously belongs to the ordinary Cantor set.

Proposition 4.1. *Define*

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

Then

$$P_x = \lim_{m\to\infty} P_{t_m} P_{t_{m-1}} \cdots P_{t_1}.$$

Moreover, the convergence is uniform with respect to all $x \in (0, 1]$.

This means that Σ is a U-LCP. Hence any Cantorian x-Markov chain is ergodic. We now interpret Proposition 4.1 in a geometric way, rather than from a probabilistic approach.

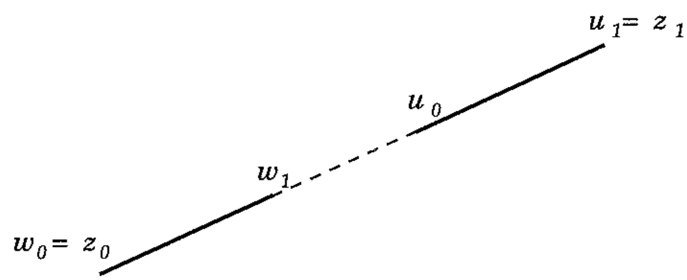

FIGURE 1. Cantorian Markov chains

In the complex plane (or 2-D real linear space), any pair of distinct points z_0, z_1 defines a line segment, denoted by $[z_0 z_1]$. Both P_0 and P_1 can be seen as a mapping of line segments (see Figure 1):

$$P_0 : [z_0 z_1] \to [w_0 w_1], \qquad P_1 : [z_0 z_1] \to [u_0 u_1],$$

with

$$\begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = P_0 \begin{pmatrix} z_0 \\ z_1 \end{pmatrix} = \begin{pmatrix} z_0 \\ \frac{2}{3} z_0 + \frac{1}{3} z_1 \end{pmatrix},$$

$$\begin{pmatrix} u_0 \\ u_1 \end{pmatrix} = P_1 \begin{pmatrix} z_0 \\ z_1 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} z_0 + \frac{2}{3} z_1 \\ z_1 \end{pmatrix}.$$

Now that

$$\|w_1 - w_0\| = \|u_1 - u_0\| = \frac{1}{3}\|z_1 - z_0\|, \tag{5}$$

the two maps are both *contracting*! Iteration of Eq. (5) implies that if

$$\begin{pmatrix} v_0 \\ v_1 \end{pmatrix} = P_{t_m} \cdots P_{t_1} \begin{pmatrix} z_0 \\ z_1 \end{pmatrix}, \qquad t_i \in \{0, 1\},$$

then

$$\|v_0 - v_1\| \le \frac{1}{3^m}\|z_0 - z_1\|. \tag{6}$$

If we apply the above analysis to the two special line segments [01] and [10], then the x-Markov chain is easily seen to be ergodic.

On the other hand, given a line segment $[z_0 z_1]$, if we iteratively apply $P_{t_1}, P_{t_2}, \cdots$, the geometric picture clearly shows that we are essentially constructing one branch of the classical Cantor dust on $[z_0 z_1]$! Take $z_0 = 0$ and $z_1 = 1$. Suppose

$$\begin{pmatrix} v_0^m \\ v_1^m \end{pmatrix} = P_{t_m} \cdots P_{t_1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Then $[v_0^m v_1^m]$ is one of the 2^m segments at the m-th step of constructing the classical Cantor dust on the unit interval. Hence $v_0^m, v_1^m \to c_x$ as $m \to \infty$ if $x = 0.t_1 t_2 \cdots$. Similarly, if $z_0 = 1$ and $z_1 = 0$, then v_0^m and v_1^m converge to $1 - c_x$.

This simple special example contains all the necessary geometric ingredients of studying ergodicity. First, the "contraction rate" $\frac{1}{3}$ plays a crucial role in the convergence process. Secondly, the convergence is pretty much linked to a familiar one in metric spaces. Let Ω be a complete metric space. Suppose

$$K_1 \supset K_2 \supset \cdots$$

is a chain of nonempty compact sets such that the diameters tend to zero. Then there must exist a unique point $\mathbf{v} \in \Omega$, such that

$$\bigcap_{m=1}^{\infty} K_m = \{\mathbf{v}\}.$$

This geometric picture is developed for more general ergodic non-homogeneous Markov chains in the next section. The number $\frac{1}{3}$ shall be replaced by new appropriate quantities (i.e. $\mu(\Sigma)$ and $\lambda(\Sigma)$ in the next section). At this point, please notice the connection among the unique point $\mathbf{v}$ in this abstract setting, the rank one of the U-LCP Σ, and the rank one condition of ergodicity.

One major advantage of the geometric picture is that most analytical quantities and relations find their simple explanations. Before us, those quantities were introduced to the literature mostly from the probabilistic intuition or analytical point of view. The geometric correspondence established here unifies the theory of ergodic Markov chains.

5. Stochastic matrices as transforms of simplices

5.1. **Simplices in $\mathbb{R}^n$.** For any ordered n vectors (vertices) $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ in $\mathbb{R}^n$, the associated n-simplex is their convex hull, denoted by $[\mathbf{v}_1\mathbf{v}_2\cdots\mathbf{v}_n]$. That is,

$$[\mathbf{v}_1\cdots\mathbf{v}_n] = \{\sum_{i=1}^n a_i\mathbf{v}_i \mid a_i \geq 0, \sum_{i=1}^n a_i = 1\}.$$

From example, for $n = 2$, $[\mathbf{v}_1\mathbf{v}_2]$ is a line segment; and for $n = 3$, $[\mathbf{v}_1\mathbf{v}_2\mathbf{v}_3]$ is a triangle. The simplex is called *non-degenerate* if the n vertices are linearly independent. A non-degenerate n-simplex is $n-1$-dimensional.

Let

$$\mathbf{e}_i = [0, \cdots, 0, 1, 0, \cdots, 0],$$

for $i = 1, \cdots, n$, be the canonical basis of $\mathbb{R}^n$. The associated simplex $[\mathbf{e}_1\cdots\mathbf{e}_n]$ is called the *canonical simplex.*

For a given stochastic matrix $P = (p_{ij})_{n\times n}$, we define a transform of simplices

$$[\mathbf{v}_1\cdots\mathbf{v}_n] \to [\mathbf{v}_1^P\cdots\mathbf{v}_n^P],$$

by requiring

$$\mathbf{v}_k^P = p_{k1}\mathbf{v}_1 + p_{k2}\mathbf{v}_2 + \cdots + p_{kn}\mathbf{v}_n.$$

It is obvious that

Proposition 5.1.

$$[\mathbf{v}_1^P\cdots\mathbf{v}_n^P] \subset [\mathbf{v}_1\cdots\mathbf{v}_n].$$

A special case is

$$[\mathbf{e}_1^P\cdots\mathbf{e}_2^P] = [\mathbf{p}_1\cdots\mathbf{p}_n],$$

where $\mathbf{p}_i$ is the i-th row vector of P. For simplicity, we denote this simplex by $[P]$.

Proposition 5.2. *There exists a one to one correspondence between the set of (n by n) stochastic matrices and the set of all (ordered) n-simplices contained in the canonical simplex.*

The concepts such as Markov matrices and scrambling matrices (Section 3) from probability considerations have a "visualizable" geometric correspondence.

Proposition 5.3 (Markov matrices). *A stochastic matrix P is a Markov matrix if and only if for any non-degenerate simplex $[\mathbf{v}_1 \cdots \mathbf{v}_n]$, there exists a vertex $\mathbf{v}_k$, such that*

$$[\mathbf{v}_1^P \cdots \mathbf{v}_n^P] \bigcap [\mathbf{v}_1 \cdots \mathbf{v}_{k-1} \widehat{\mathbf{v}}_k \mathbf{v}_{k+1} \cdots \mathbf{v}_n] = \emptyset.$$

Here, by convention, $\widehat{\bullet}$ means having the element $\bullet$ dropped. (See Figure 2.)

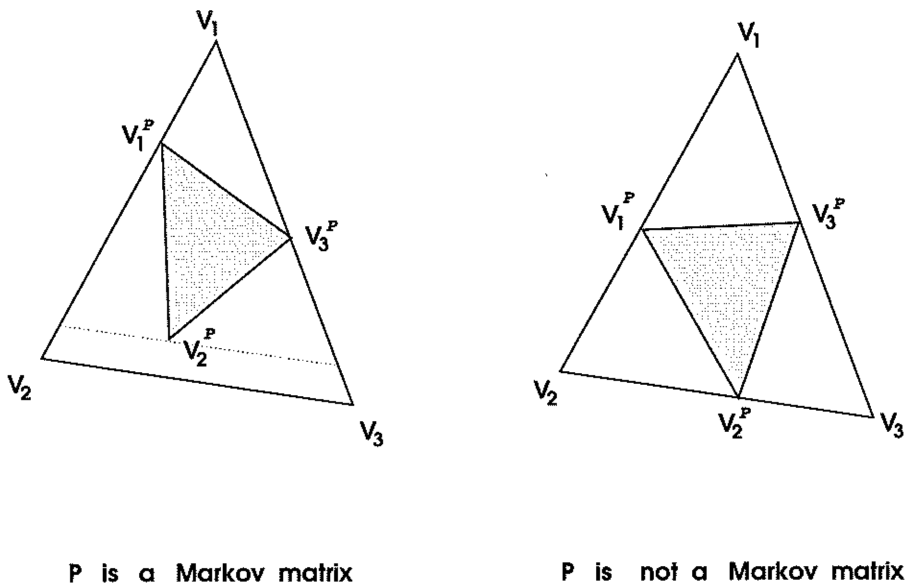

FIGURE 2. Markov matrices.

Similarly,

Proposition 5.4 (Scrambling matrices). *A stochastic matrix P is scrambling if and only if for any non-degenerate n-simplex $[\mathbf{v}_1 \cdots \mathbf{v}_n]$, any i and j, there exists k, such that*

$$[\mathbf{v}_i^P \mathbf{v}_j^P] \bigcap [\mathbf{v}_1 \cdots \mathbf{v}_{k-1} \widehat{\mathbf{v}}_k \mathbf{v}_{k+1} \cdots \mathbf{v}_n] = \emptyset.$$

(See Figure 3.)

5.2. **Geometric approach to Markov matrices.** We illustrate our geometric approach by first proving the following classical result mostly due to Markov.

Theorem 5.1. *If Σ is a finite set of Markov matrices, then any Σ-Markov chain is ergodic.*

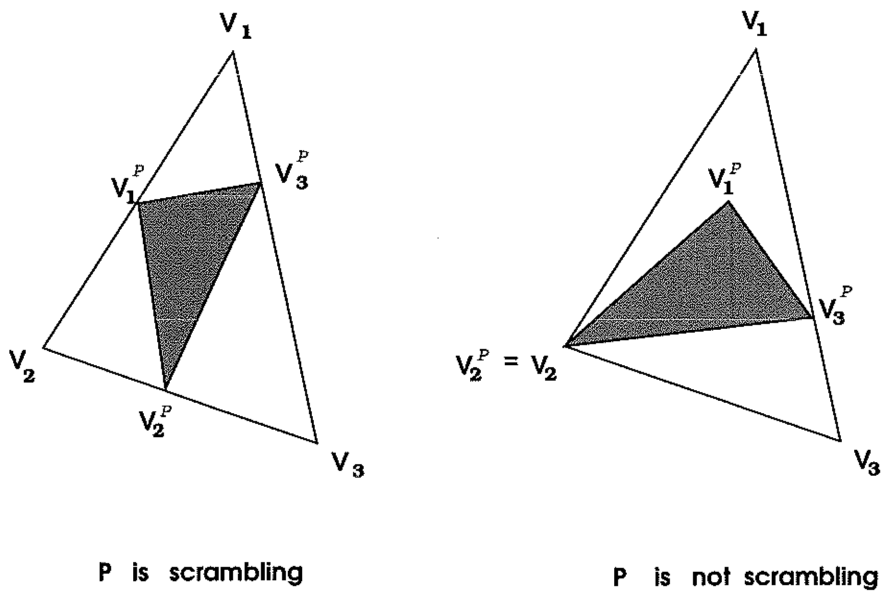

FIGURE 3. Scrambling matrices.

Proof. (Referring to Figure 4.) Fix a non-degenerate simplex $[\mathbf{v}_1 \cdots \mathbf{v}_n]$. For any $P \in \Sigma$, there exists k, such that

$$[\mathbf{v}_1^P \cdots \mathbf{v}_n^P] \bigcap [\mathbf{v}_1 \cdots \mathbf{v}_{k-1} \widehat{\mathbf{v}}_k \mathbf{v}_{k+1} \cdots \mathbf{v}_n] = \emptyset.$$

Define

$$\mu_*(P) = \min_i p_{ik},$$

and

$$\begin{cases} \mathbf{w}_i = (1-\mu_*(P))\mathbf{v}_i + \mu_*(P)\mathbf{v}_k, i \neq k. \\ \mathbf{w}_k = \mathbf{v}_k. \end{cases}$$

It is easy to see that

$$[\mathbf{v}_1^P \cdots \mathbf{v}_n^P] \subset [\mathbf{w}_1 \cdots \mathbf{w}_n].$$

Notice that the new $\mathbf{w}$-simplex is simply a uniform linear contraction of the original one with $\mathbf{v}_k$ as the fixed point, and the contraction rate $1-\mu_*(P)$. Hence, under any linear norm,

$$\begin{aligned} \operatorname{diam}([\mathbf{v}_1^P \cdots \mathbf{v}_n^P]) &\leq \operatorname{diam}([\mathbf{w}_1 \cdots \mathbf{w}_n]) \\ &\leq (1-\mu_*(P))\operatorname{diam}([\mathbf{v}_1 \cdots \mathbf{v}_n]) \\ &\leq (1-\mu_*)\operatorname{diam}([\mathbf{v}_1 \cdots \mathbf{v}_n]), \end{aligned}$$

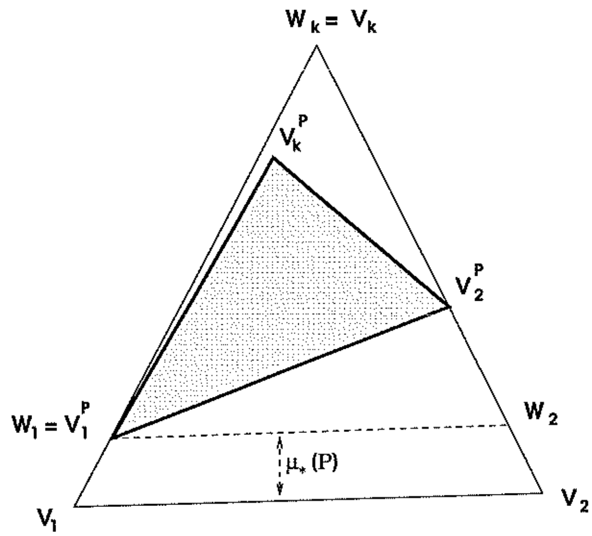

FIGURE 4. Proof of Theorem 5.1.

where,

$$\mu_* = \min_{P \in \Sigma} \mu_*(P) > 0.$$

For any infinite sequence $(P_1, P_2, \cdots) \in \Sigma^\infty$, Let $\mathbf{v}_i^{P_1 \cdots P_m}$ denote the i-th vertex of the simplex obtained from $[\mathbf{v}_1 \cdots \mathbf{v}_m]$ by successively applying the transforms $P_1, \cdots, P_m$. Then the above inequality implies

$$\operatorname{diam}([\mathbf{v}_1^{P_1 \cdots P_m} \cdots \mathbf{v}_n^{P_1 \cdots P_m}]) \leq (1 - \mu_*)^m \operatorname{diam}([\mathbf{v}_1 \cdots \mathbf{v}_n]) \to 0,$$

as $m \to \infty$. On the other hand, the sequence of simplices

$$\{ [\mathbf{v}_1^{P_1 \cdots P_m} \cdots \mathbf{v}_n^{P_1 \cdots P_m}] \}_{m=1}^\infty$$

is a non-increasing chain of compact sets (by Proposition 5.1). Therefore, there exists a unique point, say $\mathbf{v}$, such that

$$\{\mathbf{v}\} = \bigcap_m [\mathbf{v}_1^{P_1 \cdots P_m} \cdots \mathbf{v}_n^{P_1 \cdots P_m}].$$

Especially, if we take the canonical simplex, and denote $\mathbf{v}$ by $\mathbf{p}$, then

$$\lim_{m \to \infty} P_m \cdots P_1 = \lim_{m \to \infty} \begin{bmatrix} \mathbf{e}_1^{P_1 \cdots P_m} \\ \vdots \\ \mathbf{e}_n^{P_1 \cdots P_m} \end{bmatrix} = \begin{bmatrix} \mathbf{p} \\ \vdots \\ \mathbf{p} \end{bmatrix} = \mathbf{J} \cdot \mathbf{p}.$$

Since the convergence is controlled by the factor $1 - \mu_* < 1$, the set Σ is a U-LCP. Hence any Σ-Markov chain is ergodic. This completes the proof. □

Remark 5.1.

(i) The above contraction rate $1-\mu_*$ can be improved as follows. Define

$$\mu(P) = \|\mathbf{p}_1 \wedge \mathbf{p}_2 \wedge \cdots \wedge \mathbf{p}_n\|_\infty,$$
$$\mu(\Sigma) = \max_{P\in\Sigma} \mu(P).$$

With μ_* being replaced by μ, the above proof still works. Therefore,

$$\mathrm{diam}([\mathbf{v}_1^P \cdots \mathbf{v}_n^P]) \le (1-\mu(\Sigma))\mathrm{diam}([\mathbf{v}_1 \cdots \mathbf{v}_n]),$$

for any $P \in \Sigma$ and simplex $[\mathbf{v}_1 \cdots \mathbf{v}_n]$. It is not difficult to see that this intrinsic number $1-\mu(\Sigma)$ gives the tight converging rate in certain cases.

(ii) If we take L_1 norm in $\mathbb{R}^n$ and apply the above analysis to the canonical simplex, then

$$\begin{aligned}\mathrm{diam}([P_m \cdots P_1]) &\le \prod_{i=1}^m (1-\mu(P_i))\mathrm{diam}([\mathbf{e}_1 \cdots \mathbf{e}_n]) \\ &= 2\prod_{i=1}^m (1-\mu(P_i)),\end{aligned}$$

since under L_1 norm, the diameter of the canonical simplex is 2. This inequality is called the "Markov Theorem" (see Seneta [8]). Hence the Markov Theorem is a special case of ours.

(iii) The combination of the above analysis and Lemma 2.1 leads to a result mostly due to Markov.

Classical Result 5.1 (Seneta [8]). *A non-homogeneous Markov chain whose transition matrices are $P(1), P(2), \cdots$ is ergodic (See Section 2) if and only if there exists sequence of index $0 \le m_0 < m_1 < \cdots$, such that*

$$\mu(P^{m_0:m_1}) + \mu(P^{m_1:m_2}) + \cdots = \infty.$$

Example — The Sierpińskian Markov Chains.

Take any three 3 by 3 stochastic matrices

$$P_1 = \begin{bmatrix} a & * & * \\ b & * & * \\ c & * & * \end{bmatrix}, \quad \begin{bmatrix} * & d & * \\ * & e & * \\ * & f & * \end{bmatrix}, \quad \begin{bmatrix} * & * & g \\ * & * & h \\ * & * & i \end{bmatrix},$$

such that

$$\min(a, b, \cdots, h, i) \geq \frac{1}{2}.$$

In addition, suppose all matrices are non-singular. Define $\Sigma = \{P_1, P_2, P_3\}$. We call any Σ-Markov chain a *Sierpińskian Markov chain.* From the above theorem, any Sierpińskian Markov chain is ergodic. This example generalizes the Cantorian Markov chain. The corresponding geometric picture is very close to the construction of the classical Sierpiński gasket (see Yamaguti, et al. [11]).

5.3. **Geometric approach to scrambling matrices.** Since the diameter of a simplex is in fact a quantity determined by the pairs of its vertices, it is therefore plausible from the geometric point of view, to consider a "pairwise" version of Markov matrices — scrambling matrices (see Section 3).

For scrambling matrices, the quantity $\mu(P)$ defined above is too "coarse." The better one is the following $\lambda(P)$, first introduced by Hajnal [5].

Theorem 5.2 (Generalized Hajnal's inequality). *For a stochastic matrix P, define*

$$\lambda(P) = \min_{i,j} \|\mathbf{p}_i \wedge \mathbf{p}_j\|_1.$$

Then, under any norm of $\mathbb{R}^n$,

$$\operatorname{diam}([\mathbf{v}_1^P \cdots \mathbf{v}_n^P]) \leq (1 - \lambda(P))\operatorname{diam}([\mathbf{v}_1 \cdots \mathbf{v}_n]).$$

Proof. (Referring to Figure 5.) For any *given* pair of $i \neq j$, suppose

$$\mathbf{p}_i \wedge \mathbf{p}_j = (\delta_1, \delta_2, \cdots, \delta_n).$$

From the given simplex $[\mathbf{v}_1 \cdots \mathbf{v}_n]$, construct

$$\mathbf{w}_k = (1 - \sum_{l \neq k} \delta_l)\mathbf{v}_k + \sum_{l \neq k} \delta_l \mathbf{v}_l, \tag{7}$$

for $k = 1, 2, \cdots, n$. Then it is easy to see

$$\mathbf{v}_i^P, \mathbf{v}_j^P \in [\mathbf{w}_1 \cdots \mathbf{w}_n].$$

Hence,

$$\|\mathbf{v}_i^P - \mathbf{v}_j^P\| \leq \operatorname{diam}([\mathbf{w}_1 \cdots \mathbf{w}_n]).$$

If we show that

$$\operatorname{diam}([\mathbf{w}_1 \cdots \mathbf{w}_n]) = (1 - \sum_{k=1}^{n} \delta_k)\operatorname{diam}([\mathbf{v}_1 \cdots \mathbf{v}_n]), \tag{8}$$

then the proof is complete since i and j are arbitrary and

$$\lambda(P) \leq \sum_{k=1}^{n} \delta_k = \|\mathbf{p}_i \wedge \mathbf{p}_j\|_1.$$

Therefore, it suffices to show that (8) is true. This is the following lemma. □

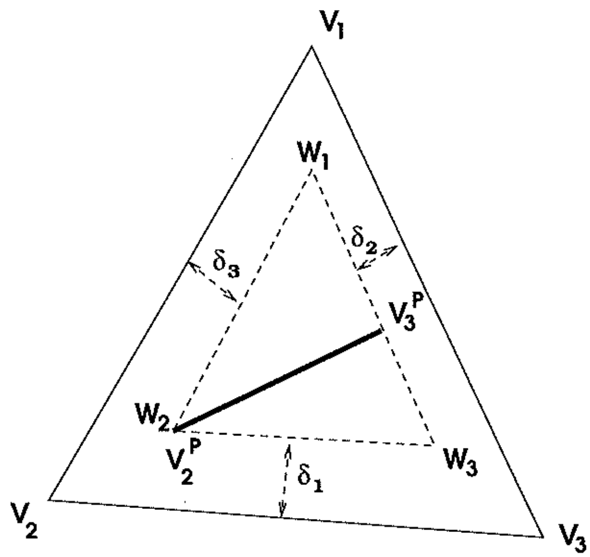

FIGURE 5. Proof of Theorem 5.2 ($i = 2, j = 3$).

Lemma 5.1. *Suppose* $\mathbf{w}_1, \cdots, \mathbf{w}_n$ *are constructed from a given set of vectors* $\mathbf{v}_1, \cdots, \mathbf{v}_n$ *by Eq.* (7) *for some* $\delta_k \geq 0$ *and* $\sum_{k=1}^{n} \delta_k \leq 1$. *Then* (8) *is true.*

Proof. It is simply because

$$\mathbf{w}_i - \mathbf{w}_j = (1 - \delta_1 - \cdots - \delta_n)(\mathbf{v}_i - \mathbf{v}_j),$$

for any i and j. □

Remark 5.2. The proof of Theorem 5.2 and this last equation make two points very clear. First, the generalized Hajnal inequality is tight in some special cases (for example, $\mathbf{v}_k^P = \mathbf{w}_k$, for $k = 1, \cdots, n$). Secondly, the quantity $\lambda(P)$ defined by Hajnal is better to be understood from the last equation. It is essentially a concept belonging to linear spaces, "orthogonal" to the norm we use. Historically, $\lambda(P)$ was frequently mixed with the special norm defining diam(P) (See Seneta [8], for example).

Corollary 5.1. *For any stochastic matrix P and linear norm in $\mathbb{R}^n$,*

$$\operatorname{diam}([P]) \le (1-\lambda(P)) \operatorname{diam}([\mathbf{e}_1 \cdots \mathbf{e}_n]).$$

Proof. Apply the theorem to the canonical simplex. □

If we take the L_∞ norm

$$\|(a,b,\cdots,d)\|_\infty = \max\{|a|,|b|,\cdots,|d|\},$$

then the classical Hajnal inequality is obtained.

Corollary 5.2. *Let $R=(r_{ij})=PQ=(p_{ij})(q_{ij})$ be a product of stochastic matrices. Then*

$$\max_k \max_{i,j} |r_{ik}-r_{jk}| \le (1-\lambda(P)) \max_k \max_{i,j} |q_{ik}-q_{jk}|.$$

If, instead, we now take the L_1 norm

$$\|(a,b,\cdots,d)\|_1 = |a|+|b|+\cdots+|d|,$$

then the diameter of the canonical simplex becomes 2. This leads to the generalized Hajnal inequality by Dobrušin [3] and Paz-Reichaw [6]:

Corollary 5.3. *Suppose $P,Q,\cdots,R$ are all stochastic matrices. Then*

$$\frac{1}{2}\operatorname{diam}([PQ\cdots R]) \le (1-\lambda(R))(1-\lambda(Q))\cdots(1-\lambda(P)),$$

or equivalently,

$$1-\lambda(PQ\cdots R) \le (1-\lambda(R))(1-\lambda(Q))\cdots(1-\lambda(P)).$$

Therefore, Theorem 5.2 is the most general form of the Hajnal inequality. The role of the number $\lambda(P)$ becomes very clear in it. For the first time, $\lambda(P)$ is separated from the special norm one uses to define the diameter.

To show more applications of Theorem 5.2, we generalize some classical results.

Proposition 5.5 (Being scramblingness means ergodicity). *Let Σ be a compact set of scrambling stochastic matrices. Then any Σ-Markov chain is ergodic.*

Proof. For any $P \in \Sigma$, $\lambda(P) > 0$ since P is scrambling. Obviously $\lambda(P)$ is a continuous function of stochastic matrices (as a compact subset in the matrix space). Since Σ is compact, we conclude that

$$\lambda(\Sigma) := \inf_{P\in\Sigma} \lambda(P) > 0.$$

Hence,

$$\operatorname{diam}([P_1 \cdots P_m]) \leq (1 - \lambda(\Sigma))^m \operatorname{diam}([\mathbf{e}_1 \cdots \mathbf{e}_n]),$$

for any $P_1, \cdots, P_m \in \Sigma$. This implies that any Σ-Markov chain is ergodic (in fact, uniformly ergodic in the obvious sense).

□

Proposition 5.6 (Being SIA means ergodicity). *Let Σ be a compact SIA set (see Section 3). Then any Σ-Markov chain is ergodic.*

Proof. Wolfowitz [10] showed that there exists a finite M such that

$$\Sigma_M = \{P_1 \cdots P_M \mid P_i \in \Sigma\},$$

is a scrambling set. It is easy to see that Σ_M is compact because of the compactness assumption on Σ. Therefore any Σ_M-Markov chain is ergodic by the preceding proposition. This implies that any Σ-Markov chain has an ergodic sub-Markov chain. The proof is complete thanking to Lemma 2.1.

□

Example — 3-State Markov chains.

We say a stochastic matrix P is *fuzzy* if there exists no *deterministic* transition between any two states. That is, the following is impossible: for some i and j,

$$p_{ij} = \operatorname{Prob}(|i\rangle \to |j\rangle) = 1.$$

Proposition 5.7. *Let Σ be a* compact *set of* 3 *by* 3 *fuzzy stochastic matrices. Then any Σ-Markov chain is ergodic.*

Proof. It is easy to show that a fuzzy 3 by 3 stochastic matrix must be scrambling. □

Remark 5.3. In studying the convergence of infinite products of matrices, it is always beneficial working on a compact set of matrices. For general results outside the scope of stochastic matrices, see the recent work by the author [9]. Here is one example showing that the compactness condition in the above discussion is essential. For $n = 1, 2, \cdots$, define

$$P_n = \begin{bmatrix} 1 - \frac{2}{n} & \frac{1}{n} & \frac{1}{n} \\ \frac{1}{n} & 1 - \frac{2}{n} & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & 1 - \frac{2}{n} \end{bmatrix} = (1 - \frac{3}{n}) I_3 + \frac{1}{n} \mathbf{J} \cdot \mathbf{J}^T.$$

Then all P_n's are scrambling. Let Σ be the collection of all P_n. Notice that it is not compact since the unique cluster "point" I_3 does not belong to it. Using the product formula for any two 3 by 3 stochastic matrices of the form $aI_3 + b\mathbf{J} \cdot \mathbf{J}^T$ and $cI_3 + d\mathbf{J} \cdot \mathbf{J}^T$:

$$(aI_3 + b\mathbf{J} \cdot \mathbf{J}^T)(cI_3 + d\mathbf{J} \cdot \mathbf{J}^T) = acI_3 + \frac{1-ac}{3}\mathbf{J} \cdot \mathbf{J}^T,$$

one can easily verify that a Σ-Markov chain defined by $(P_1, P_2, P_4, \cdots)$ is not ergodic.

6. The Hajnal rate $\pi(\Sigma)$ and joint spectral radius

Given a set of stochastic matrices Σ, define

$$\Sigma_m = \{P_1 \cdots P_m \mid P_i \in \Sigma\}$$

for all $m = 1, 2, \cdots$. Under any given norm of $\mathbb{R}^n$, define

$$\operatorname{diam}(\Sigma_m) = \sup_{Q \in \Sigma_m} \operatorname{diam}([Q]).$$

Then we define the *Hajnal rate* $\pi(\Sigma)$ to be

$$\pi(\Sigma) = \limsup_{m \to \infty} \operatorname{diam}(\Sigma)^{1/m}.$$

Suppose that Σ is compact and any Σ-Markov chain is ergodic. Then Σ must be a SIA set. Wolfowitz's lemma (see also the proof of Proposition 5.6) asserts that there exists M, such that any matrix in Σ_M is scrambling. Set

$$\lambda_0 = \inf_{Q \in \Sigma_M} \lambda(Q).$$

Then $\lambda_0 > 0$ since Σ_M is a compact scrambling set.

For any integer m, there exist unique k and $r < M$ such that $m = kM + r$. Then for any $Q = P_1 \cdots P_m \in \Sigma_m$,

$$\begin{aligned}\operatorname{diam}([Q]) &\le (1 - \lambda(P_1 \cdots P_M)) \cdots (1 - \lambda(P_{(k-1)M+1} \cdots P_{kM}))\operatorname{diam}([P_{kM+1} \cdots P_m]) \\ &\le (1 - \lambda_0)^k \operatorname{diam}([\mathbf{e}_1 \cdots \mathbf{e}_n]),\end{aligned}$$

from which, it is easily deduced that

$$\pi(\Sigma) \le (1 - \lambda_0)^{\frac{1}{M}} < 1.$$

Conversely, for any set of stochastic matrices Σ, if $\pi(\Sigma) < 1$, then it is obviously ergodic. Therefore, we obtain

Theorem 6.1. *Let Σ be a compact set of stochastic matrices. Then all Σ-Markoc chains are ergodic if and only if its Hajnal rate $\pi(\Sigma) < 1$.*

An immediate application is the generalization of Daubechies-Lagarias' result [2] to a compact set.

Proposition 6.1. *Let Σ be a compact set of stochastic matrices. Then all Σ-Markov chains are ergodic if and only if Σ is a U-LCP of rank 1 (see Section 3).*

Proof. The sufficiency is trivial. Now suppose all Σ-Markov chains are ergodic. Then $\pi(\Sigma) < 1$. Take any $d \in (\pi(\Sigma), 1)$. Then there exists M, such that for all $m > M$, and $Q_m = P_m \cdots P_1 \in \Sigma_m$,

$$\operatorname{diam}([P_m \cdots P_1]) \leq d^m.$$

Let $\mathbf{p}$ be the first row of Q_m. Then

$$\|Q_m - \mathbf{J} \cdot \mathbf{p}\| \leq C_0 d^m,$$

where the constant C_0 only depends on the matrix norm we use. Hence, for any $k \geq 0$,

$$\begin{aligned} \|P_{m+k} \cdots P_1 - P_m \cdots P_1\| &= \|P_{m+k} \cdots P_{m+1}(Q_m - \mathbf{J} \cdot \mathbf{p}) - (Q_m - \mathbf{J} \cdot \mathbf{p})\| \\ &\leq C_1 \|Q_m - \mathbf{J} \cdot \mathbf{p}\| + \|Q_m - \mathbf{J} \cdot \mathbf{p}\| \\ &\leq C_2 d^m, \end{aligned}$$

where, all C_i only depends on the matrix norm. This means that Σ is a U-LCP. That it has rank 1 is obvious. □

On the other hand, if Σ is a rank 1 U-LCP, it has a common 1-eigenspace $\mathbf{E}_1(\Sigma)$ (see Section 3). Let $\mathbf{V}$ be any linear complement of $\mathbf{E}_1(\Sigma)$ in $\mathbb{R}^n$,

$$\mathbf{E}_1(\Sigma) \oplus \mathbf{V} = \mathbb{R}^n.$$

Denote by Q the (skew) projection onto $\mathbf{V}$ along $\mathbf{E}_1(\Sigma)$, then Daubechies-Lagarias showed that

$$\rho_0(\Sigma) := \rho(Q \Sigma Q^T) < 1.$$

The definition of joint spectral radius $\rho(\bullet)$ of a set of matrices is similar to that of $\pi(\bullet)$, whose first introduction was by Rota and Strang [7]. It was Daubechies and Lagarias who found their significance in characterizing convergence of matrix products. For a set Λ of matrices (of the same size), it is defined by

$$\|\Lambda\| = \sup_{A \in \Lambda} \|A\|,$$

$$\rho(\Lambda) = \limsup_{m \to \infty} \|\Lambda_m\|^{\frac{1}{m}}.$$

A few comments are in order.

(i) Both π and ρ are independent of the special (vector or matrix) norm we use, since all norms are equivalent in finite dimensional Banach space.

(ii) It seems to the author that the Hajnal rate $\pi(\Sigma)$ is "more" intrinsic than the projected joint spectral radius $\rho_0(\Sigma) = \rho(Q\Sigma Q^T)$, since the latter involves a choice of $\mathbf{V}$. However, the joint spectral radius is a more general concept.

Theorem 6.2. *Suppose Σ is a compact set of stochastic matrices such that any Σ-Markov chain is ergodic. Then*

$$\pi(\Sigma) = \rho_0(\Sigma).$$

Proof. First, we show $\pi(\Sigma) \leq \rho_0(\Sigma)$. In the case of stochastic matrices, the common 1-eigenspace $\mathbf{E}_1(\Sigma)$ is spaned by $\mathbf{J}$. Let $\mathbf{V}$ be any complement of $\mathbf{E}_1(\Sigma)$ in $\mathbb{R}^n$. Take a basis $\mathbf{v}_1, \cdots, \mathbf{v}_n$ so that $\mathbf{v}_1 = \mathbf{J}$, and the remaining $n-1$ vectors span $\mathbf{V}$. Define a matrix

$$A = (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n).$$

Then for any $P \in \Sigma$,

$$A^{-1}PA = \begin{pmatrix} 1 & \alpha_P \\ 0 & \hat{P} \end{pmatrix},$$

where $\hat{P} = QPQ^T$ and Q is the projection onto $\mathbf{V}$ along $\mathbf{E}_1(\Sigma)$. Take any $d : \rho_0(\Sigma) < d < 1$. There exists M, so that for any $m > M$,

$$\|\hat{P}_1 \cdots \hat{P}_m\| \leq d^m, \qquad P_1, \cdots, P_m \in \Sigma.$$

Assume

$$A^{-1}(P_1 \cdots P_m)A = \begin{pmatrix} 1 & \alpha_{P_1 \cdots P_m} \\ 0 & \widehat{P_1 \cdots P_m} \end{pmatrix}.$$

Then $\widehat{P_1 \cdots P_m} = \hat{P}_1 \cdots \hat{P}_m$, and

$$\|A^{-1}(P_1 \cdots P_m)A - \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \cdot (1, \alpha_{P_1 \cdots P_m})\| \leq \| \begin{pmatrix} 0 & 0 \\ 0 & \widehat{P_1 \cdots P_m} \end{pmatrix} \| \quad \leq Cd^m,$$

Here C (and all the following C_i's) only depends on the norm we use and the choice of A. Therefore,

$$\|P_1 \cdots P_m - A \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \cdot (1, \alpha_{P_1 \cdots P_m}) A^{-1} \| \leq C_1 d^m,$$

$$\|P_1 \cdots P_m - \mathbf{J} \cdot \mathbf{q}\| \leq C_1 d^m,$$

where $\mathbf{q} = (1, \alpha_{P_1 \cdots P_M}) A^{-1}$. It says that all row vectors of $P_1 \cdots P_m$ lie inside the $C_1 d^m$–neighborhood of $\mathbf{q}$. Hence,

$$\mathrm{diam}([P_1 \cdots P_M]) \leq C_2 d^m.$$

This implies that $\pi(\Sigma) \leq d$. Since $d > \rho_0(\Sigma)$ is arbitrary, we complet the proof of the first part.

Now we show $\rho_0(\Sigma) \leq \pi(\Sigma)$. This is done by almost converting the above steps. Take any $d : \pi(\Sigma) < d < 1$. Then there exists M, so that for all $m > M$,

$$\mathrm{diam}([P_1 \cdots P_m]) \leq d^m.$$

Especially, by taking $\mathbf{p}$ to be the first row of $P_1 \cdots P_m$,

$$\|P_1 \cdots P_m - \mathbf{J} \cdot \mathbf{p}\| \leq C d^m.$$

Let A be defined as above. Then

$$\|A^{-1} P_1 \cdots P_m A - A^{-1} \mathbf{J} \cdot \mathbf{p} A\| \leq C_1 d^m,$$

or,

$$\| \begin{pmatrix} 1 & \alpha_{P_1 \cdots P_m} \\ 0 & \widehat{P_1 \cdots P_m} \end{pmatrix} - \begin{pmatrix} 1 & \beta \\ 0 & 0 \end{pmatrix} \| \leq C_2 d^m.$$

This implies that

$$\|\hat{P}_1 \cdots \hat{P}_m\| = \|\widehat{P_1 \cdots P_m}\| \leq C_3 d^m.$$

Therefore, $\rho_0(\Sigma) \leq d$. The proof is complete since $d > \pi(\Sigma)$ is arbitary.

□

References

[1] J. M. Anthonisse and H. Tijms. Exponential convergence of products of stochastic matrices. *J. Math. Anal. Appl.*, 59:360–364, 1977.

[2] I. Daubechies and J. C. Lagarias. Sets of matrices all infinite products of which converge. *Linear Alg. Appl.*, 161:227–263, 1992.

[3] R. L. Dobrušin. Central limit theorem for non-stationary Markov chains, I, II. *Theor. Probability Appl.*, 1:65–80, 329–383, 1956.

[4] J. Hajnal. The ergodic properties of nonhomogeneous finite Markov chains. *Proc. Camb. Phil. Soc.*, 52:67–77, 1956.

[5] J. Hajnal. Weak ergodicity in nonhomogeneous Markov chains. *Proc. Camb. Phil. Soc.*, 54:233–246, 1958.

[6] A. Paz and M. Reichaw. Ergodic theorems for sequences of infinite stochastic stochastic matrices. *Proc. Camb. Phil. Soc.*, 63:777–784, 1967.

[7] G.-C. Rota and G. Strang. A note on the joint spectral radius. *Indag. Math.*, 22:379–381, 1960.

[8] E. Seneta. On the historical development of the theory of finite inhomogeneous Markov chains. *Proc. Camb. Phil. Soc.*, 74:507–513, 1973.

[9] J. Shen. Compactification of a set of matrices with convergent infinite products. CAM Report 98-46, Dept. of Math., UCLA. Submitted, 1998.

[10] J. Wolfowitz. Products of indecomposable, aperiodic, stochastic matrices. *Proc. Amer. Math. Soc.*, 14:733–737, 1963.

[11] M. Yamaguti, M. Hata, and J. Kigami. *Mathematics of Fractals*, volume 167 of *Translations of Mathematical Monographs*. Amer. Math. Soc., 1997.

Department of Mathematics, 7354 Mathematical Sciences Building, UCLA, Los Angeles, CA 90095-1555. *jhshen@math.ucla.edu.*