Semigroup Forum Version of October 4, 2005
Springer-Verlag New York Inc.

æ

Semigroups of Binary Stochastic Matrices and Convergence of Nonhomogeneous Markov Chains

Mark Beintema and Gregory Budzban

Communicated by K. H. Hofmann

ABSTRACT. We give conditions which determine when a stochastic matrix can be written as a convex combination of binary stochastic matrices from a particular class of semigroups. Using the equivalence between convolution products of measures and products in the semigroup algebra of binary stochastic matrices, verifiable conditions based only on the entries in the individual transition matrices are given for the convergence of certain nonhomogeneous Markov chains.

1. Introduction

Markov chains provide one of the most useful models of stochastic behavior available. In a variety of settings, the behavior of a stochastic system seems to depend largely on the current state of the system. While the convergence theory for homogeneous Markov chains is well understood, the analogous theory for nonhomogeneous Markov chains, where the transition matrix changes over time, is far from complete even in the finite state case. In particular, most known results concentrate on conditions for weak and strong ergodicity [see Seneta (**11**) pg. 144 and pg. 151]. Convergence, in the absence of weak ergodicity, is rarely considered. What would be ideal would be to determine necessary and sufficient conditions for the convergence of nonhomogeneous Markov Chain that are based entirely on the entries of the individual transition matrices, and not on any conditions of the products of the matrices. Conditions based on the products are, in general, very difficult to verify.

One of the most important applications of nonhomogeneous Markov chains is the stochastic optimization technique known as simulated annealing. Most of the work on simulated annealing assumes the irreducibility of the chain or equivalently, the existence of only one ergodic component in the chain $(3, \mathbf{8})$.

In an attempt to speed up the convergence of the annealing algorithm one may wish to use a priori information to introduce both transient states and distinct ergodic components.

Under these conditions the literature is sparse. Gidas (**4**) is an exception and he has some interesting examples in the appendix of his article where, due to the reducibility of the chain, "phase transitions" occur. Yet even in these examples Gidas assume that the sequence of transition matrices (A_k) is such that $\lim_k A_k$ exists. This is natural in the context of simulated annealing and is induced by the temperature parameter. In this paper we wish to begin to consider the following possibilities.

(1) No assumptions concerning the irreducibility of chain are made. Both transient states and more than one ergodic component are possible.
(2) No assumptions concerning the existence of $\lim_k A_k$.
(3) No assumptions where each element of the sequence (A_k) belongs to a special class of matrices, such as the scrambling matrices or regular matrices of Seneta (Seneta, pg. 143).

The objects of study in this paper are completely simple semigroups of binary stochastic matrices. A completely simple semigroup is a union of isomorphic groups with no proper ideals. Given the well known correspondence between $n \times n$ binary stochastic matrices and functions on a set of n elements, a completely simple semigroup in this setting can be shown to be a union of isomorphic subgroups characterized by what we call compatible ranges and partitions of a set with n elements. We assume a knowledge of the basic principles of semigroup theory, most notably Green's equivalences and Rees' structure theory for completely simple semigroups. For basic definitions and results concerning semigroups, we recommend Howie [**7**]. More detailed information concerning the transformation semigroup T_X can be found in the classic text of Clifford and Preston [**10**]. For background concerning probability measures and the convergence of random walks on semigroups, the reader is referred to Mukherjea and Högnäs [**6**].

The following notational conventions are in force throughout the paper. If A is a matrix, then we write $[A]_{ij}$ to denote the (i,j)-th entry of A. When we write A_{ij} we mean the (i,j)-th block in some block decomposition of A. (This will be clear in context.) We write (A_n) to denote a sequence of matrices, and $A_{ij}(n)$ to indicate the appropriate block of the n-th matrix in the sequence.

2. The Semigroup T_X

Let X be a set with n elements, say $X = \{1, 2, \ldots, n\}$. The set of all maps from X into itself will be denoted T_X. Of course this is a semigroup under the operation of composition of mappings. We write $(i)f$ for the image of i under f. Thus the product fg indicates that g follows f. If $f \in T_X$, we write $R(f)$ for the range of f and set $\pi(f) = \{f^{-1}(j) \mid j \in R(f)\}$. Thus $\pi(f)$ is a partition of X. The rank of f is $\text{rank}(f) = |R(f)| = |\pi(f)|$. The following observation is immediate:

Proposition 2.1. *Let $f \in T_X$. Then f is idempotent if and only if f is the identity map on $R(f)$.* ■

The semigroup T_X is studied in connection with Markov chains in [**5**], where the following results are proved.

Theorem 2.2 (Högnäs). *Let $\mathcal{S} \subset T_X$, X a finite set.*
(1) The kernel K of S is the set of elements of minimal rank.
(2) If $f \in K$, then the L-class of f is $L_f = \{g \in K \mid R(f) = R(g)\}$.
(3) If $f \in K$, then the R-class of f is $R_f = \{g \in K \mid \pi(f) = \pi(g)\}$.
■

Let $\pi = \{P_1, \ldots, P_r\}$ be a partition of X, and let R be an r-subset of X. We say that π and R are *compatible* if $|R \cap P_k| = 1$ for each $k = 1, \ldots, r$. In the terminology of Clifford and Preston [**10**], R is a cross section of π. It is possible to construct an idempotent e with $\pi(e) = \pi$ and $R(e) = R$ precisely when π and R are compatible. That is, the intersection of an L-class L_f and an R-class R_g is a group if and only if $R(f)$ and $\pi(g)$ are compatible. We will designate such an H-class by $H(\pi, R)$. It is useful to think of the resulting group as a subgroup of the full group of permutations on r letters.

Proposition 2.3. *Let $f \in H(\pi_1, R_1)$, $g \in H(\pi_2, R_2)$. If π_2 and R_1 are compatible, then* $\operatorname{rank}(f) = \operatorname{rank}(fg) = \operatorname{rank}(g)$.

Proof. Assume π_2 and R_1 are compatible. Clearly $\operatorname{rank}(f) = \operatorname{rank}(g)$. Since $R(fg) \subset R(g)$, it suffices to show that $R(fg) = R(g)$. More specifically, it is enough to show that $g|_{R_1}$ is injective.

Let $R_1 = \{j_1, \dots, j_r\}$ and $R_2 = \{k_1, \dots, k_r\}$, so that $\pi_2 = \{g^{-1}(k_1), \dots, g^{-1}(k_r)\}$. Now for each $\ell = 1, \dots, r$, there is a uniquely determined i such that $j_i \in R_1 \cap g^{-1}(k_\ell)$. That is, given ℓ there is precisely one index i so that $(j_i)g = k_\ell$. We deduce that $g|_{R_1}$ is injective. ■

Each element $f \in T_X$ can be represented by an $n \times n$ binary matrix T_f as follows:

$$\Big[T_f\Big]_{ij} = \begin{cases} 1 & \text{if} \quad (i)f = j \\ 0 & \text{otherwise} \end{cases}$$

Such a matrix is called a *transformation matrix*. It is easy to see that $T_f T_g = T_{fg}$ and $\operatorname{rank} f = \operatorname{rank} T_f$ (in the usual matrix theoretic sense). Clearly every binary stochastic matrix can be thought of as a transformation matrix, and vice versa. Birkhoff's theorem (see e.g. [**1**]) asserts that every doubly stochastic matrix can be written as a convex combination of permutation matrices. In the next section we will generalize this result, giving necessary and sufficient conditions for certain stochastic matrices to be expressible as a convex combination of transformation matrices from a completely simple semigroup.

3. Binary Stochastic Matrices

Throughout this section we let $\pi = \{P_1, P_2, \dots, P_r\}$ be a fixed partition of X, with $|P_j| = p_j$ for each $j = 1, \dots, r$. By relabeling elements of X if necessary, we can assume that $P_1 = \{1, 2, \dots, p_1\}$, $P_2 = \{p_1 + 1, p_1 + 2, \dots, p_1 + p_2\}$, etc. Let $R_1, R_2, \dots, R_N$ be the collection of compatible ranges for π; a simple counting argument shows that $N = p_1 \cdot p_2 \cdot \dots \cdot p_r$. Now we let $\mathcal{S}_\pi$ be the semigroup generated by π and its compatible ranges; that is,

$$\mathcal{S}_\pi = \bigcup_{i=1}^{N} H(\pi, R_i),$$

where each $H(\pi, R_i)$ is the full group of permutations on r letters. Since all elements have the same fixed (therefore minimal) rank, $\mathcal{S}_\pi$ is completely simple by Theorem 2.2. Let $\mathcal{S}$ be the corresponding semigroup of binary stochastic matrices.

Let $\mathcal{P}$ be the set of all $n \times n$ matrices which are partitioned according to π. That is, the set of matrices of the form

$$A = \begin{bmatrix} A_{11} & A_{12} & . & . & . & A_{1r} \\ A_{21} & A_{22} & . & . & . & A_{2r} \\ & & & & & \\ & & & & & \\ A_{r1} & A_{r2} & . & . & . & A_{rr} \end{bmatrix}$$

where each A_{ij} is a $p_i \times p_j$ matrix. Note that $\mathcal{S} \subset \mathcal{P}$.

Let $\mathcal{M}_r(\mathbf{Z}_2)$ denote the set of $r \times r$ $(0,1)$-matrices, and define a map $\Phi : \mathcal{P} \to \mathcal{M}_r(\mathbf{Z}_2)$ as follows:

$$\Big[\Phi(A)\Big]_{ij} = \begin{cases} 1 & \text{if} \quad A_{ij} \neq 0 \\ 0 & \text{if} \quad A_{ij} = 0 \end{cases}$$

Let $A \in \mathcal{P}$. Submatrices of the form

$$A_{\bullet j} = \begin{bmatrix} A_{1j} \\ A_{2j} \\ \vdots \\ A_{rj} \end{bmatrix} \qquad \text{and} \qquad A_{i\bullet} = [\, A_{i1}\, A_{i2}\, \ldots\, A_{ir}\,]$$

are called *block columns* and *block rows* respectively. The term block line will refer to either a block column or row. We define the *block term rank* of A, denoted $\rho(A)$, to be the maximum number of nonzero blocks, no two on a line. Thus the block term rank of $A \in \mathcal{P}$ is the usual term rank of the matrix $\Phi(A)$. We immediately deduce the following analogue of König's Theorem [**1**]:

Proposition 3.1. *Let $A \in \mathcal{P}$. Then $\rho(A)$ is equal to the minimal number of block lines needed to cover all the nonzero blocks of A.* ■

Now each matrix $B \in \mathcal{S}$ can be represented by a pair $\langle \sigma, R_i \rangle$, where σ is a permutation and the corresponding permutation matrix $\Phi(B)$ determines the placement of nonzero blocks. Each nonzero block has exactly one nonzero column (a column of 1's). The range R_i then determines the placement of the nonzero columns within the blocks.

As an example, let $X = \{1, \ldots, 7\}$, $\pi = \Big\{\{1,2\}, \{3,4,5\}, \{6,7\}\Big\}$, $R = \{2,3,7\}$. Also let $\sigma = (23)$ and $\psi = (132)$ be permutations (written in cycle notation). Then

$$\langle \sigma, R \rangle = \begin{bmatrix} 0&1&0&0&0&0&0 \\ 0&1&0&0&0&0&0 \\ 0&0&0&0&0&0&1 \\ 0&0&0&0&0&0&1 \\ 0&0&0&0&0&0&1 \\ 0&0&1&0&0&0&0 \\ 0&0&1&0&0&0&0 \end{bmatrix} \quad \text{and} \quad \langle \psi, R \rangle = \begin{bmatrix} 0&0&0&0&0&0&1 \\ 0&0&0&0&0&0&1 \\ 0&1&0&0&0&0&0 \\ 0&1&0&0&0&0&0 \\ 0&1&0&0&0&0&0 \\ 0&0&1&0&0&0&0 \\ 0&0&1&0&0&0&0 \end{bmatrix}.$$

Definition . We say that $A \in \mathcal{P}$ is *block-column balanced* if for all $1 \leq k < \ell \leq r$, we have

$$\sum_{i=1}^{r} \frac{1}{p_i} |A_{ik}| = \sum_{i=1}^{r} \frac{1}{p_i} |A_{i\ell}|,$$

where $|A|$ is the sum of the entries in A.

Lemma 3.2. *Any linear combination of block-column balanced matrices from $\mathcal{P}$ is again block-column balanced.*

Proof. Assume $A^{(1)}, A^{(2)} \in \mathcal{P}$ are block-column balanced, fix an index j, and set

$$M_1 = \sum_{i=1}^{r} \frac{1}{p_i} |A_{ij}^{(1)}|, \quad M_2 = \sum_{i=1}^{r} \frac{1}{p_i} |A_{ij}^{(2)}|.$$

Let c_1, c_2 be scalars, and set $A = c_1 A^{(1)} + c_2 A^{(2)}$. For $1 \le k \le r$ we have

$$\begin{aligned}\sum_{i=1}^{r} \frac{1}{p_i}|A_{ik}| &= \sum_{i=1}^{r} \frac{1}{p_i}\left[c_1|A_{ik}^{(1)}| + c_2|A_{ik}^{(2)}|\right] \\ &= c_1 \sum_{i=1}^{r} \frac{1}{p_i}|A_{ik}^{(1)}| + c_2 \sum_{i=1}^{r} \frac{1}{p_i}|A_{ik}^{(2)}| \\ &= c_1 M_1 + c_2 M_2.\end{aligned}$$

Thus $A = c_1 A^{(1)} + c_2 A^{(2)}$ is block-column balanced. The result follows easily by induction. ■

Lemma 3.3. *Let $A \in \mathcal{P}$ be a nonzero matrix with constant row sums. If A is block-column balanced, then $\rho(A) = r$.*

Proof. Clearly $\rho(A) \le r$. Thus we can find e block rows and f block columns which cover all nonzero blocks with $e + f \le r$. Let I and J be the respective set of indices for rows and columns in the block cover of A, so $|I| = e, |J| = f$. Multiplying A by a nonzero scalar will not affect the block-column balance property, so we can assume that A is stochastic. Thus

$$\sum_{j=1}^{r} \frac{1}{p_i}|A_{ij}| = 1$$

for each $i = 1, \dots, r$. Thus we also have

$$\sum_{i \in I} \sum_{j=1}^{r} \frac{1}{p_i}|A_{ij}| = e \qquad \text{and} \qquad \sum_{i=1}^{r} \sum_{j=1}^{r} \frac{1}{p_i}|A_{ij}| = r. \tag{1}$$

Let j be fixed, and set $M = \sum_{i=1}^{r} \frac{1}{p_i}|A_{ij}|$. Summing over all block columns and changing the order of summation in (1) above, we get

$$r \cdot M = \sum_{j=1}^{r} \sum_{i=1}^{r} \frac{1}{p_i}|A_{ij}| = r,$$

and consequently $M = 1$. It follows that

$$\sum_{j \in J} \sum_{i=1}^{r} \frac{1}{p_i}|A_{ij}| = f. \tag{2}$$

Combining (1) and (2) and observing that the blocks with row indices from I and column indices from J must cover all nonzero entries, we get $e + f \ge r$. We deduce that $\rho(A) = r$. ■

We now present the main result of this section.

Theorem 3.4. *Let $A \in \mathcal{P}$ be a stochastic matrix. Then A is a convex combination of binary stochastic matrices from the completely simple semigroup $\mathcal{S}$ if and only if each A_{ij} has identical rows and A is block-column balanced.*

Proof. If $B \in \mathcal{S}$ and $1 \le k \le r$, then

$$\frac{1}{p_i}|B_{ik}| = \begin{cases} 1 & \text{if} \quad [\Phi(B)]_{ik} = 1 \\ 0 & \text{otherwise} \end{cases}.$$

I.e., $\frac{1}{p_i}|B_{ik}| = [\Phi(B)]_{ik}$. Since $\Phi(B)$ is a permutation matrix, we see that $\sum_{i=1}^{r} \frac{1}{p_i}|B_{ik}| = 1$ for all k. It follows that each $B \in \mathcal{S}$ is block-column balanced. By Lemma 3.2, each convex combination of binary stochastic matrices from $\mathcal{S}$ is block-column balanced. Moreover, each $B \in \mathcal{S}$ has identical rows within blocks, and hence a convex combination of matrices from $\mathcal{S}$ also has this property.

To prove the converse, assume that A is block-column balanced. We use induction on the number of nonzero *entries* of A. By Lemma 3.3, we can find r nonzero blocks, no two on a block line. Let σ_1 be the corresponding permutation. Within each nonzero block, choose the column with the smallest positive entry. Let R_1 be the corresponding range, and set $B^{(1)} = \langle \sigma_1, R_1 \rangle$. Let c_1 be the smallest of all the nonzero entries of A. Since each of A and $B^{(1)}$ are block column balanced, so is $A' = A - c_1 B^{(1)}$. Note A' has constant row sum $1 - c_1$, whence Lemma 3.3 yields $\rho(A') = r$. Moreover, A' has fewer nonzero entries than A, so we can iterate the process above until A is written as a convex combination of matrices from $\mathcal{S}$. ∎

We end this section with a result similar to Theorem 3.4. Let R be a fixed r-element subset of X. Let $\pi_1, \pi_2, \ldots, \pi_N$ be the collection of all compatible partitions, and set $\mathcal{S} = \mathcal{S}_R = \bigcup_{i=1}^{N} H(\pi_i, R)$. Once again, we are taking $H(\pi_i, R)$ to be the full group of permutations on r letters. Finally, given an $n \times n$ matrix A, we write $A[R]$ to denote the principal submatrix with row and column indices from R.

Proposition 3.5. *An $n \times n$ stochastic matrix A can be written as a convex combination of matrices from $\mathcal{S}$ if and only if $A[R]$ is doubly stochastic.*

Proof. It is easy to see that if $B \in \mathcal{S}$, then $B[R]$ is a permutation matrix. Thus if A is a convex combination of matrices from $\mathcal{S}$, $A[R]$ is a convex combination of permutation matrices, hence doubly stochastic.

Now assume A is an $n \times n$ stochastic matrix, with $A[R]$ doubly stochastic. Again we show that A can be written as a convex combination of elements of $\mathcal{S}$ by induction on the number of nonzero entries in A. Note that A must have all of its nonzero entries in those columns whose index lies in R.

Let c_1 be the smallest nonzero entry in $A[R]$, and c_2 be the smallest nonzero entry of A not in $A[R]$. Assume c_1, c_2 are the (k, ℓ) and (u, v) entries, respectively. It is an easy consequence of Birkhoff's theorem that we can find r nonzero entries in $A[R]$, such that no two on a line and one of which is the (k, ℓ) entry. Once this is accomplished, we choose one nonzero entry from each row i, $i \notin R$; this is done arbitrarily except for the stipulation that in row u, we choose the (u, v) entry. Let B be the $(0,1)$-matrix with ones in precisely the positions chosen above. Note that $B \in \mathcal{S}$. Let $c = \min\{c_1, c_2\}$.

Now $A' = A - cB$ is a nonnegative matrix with constant row sum $1 - c$, hence a scalar multiple of a stochastic matrix. Also $A'[R]$ is a scalar multiple of a doubly stochastic matrix. Finally note A' has fewer nonzero entries that A, so by the inductive hypothesis $\frac{1}{1-c}A'$ can be written as a convex combination of matrices from $\mathcal{S}$. It easily follows that A can be so written.

Remark . A is a convex combination of matrices from $\mathcal{S}_R$ if and only if A is block column balanced with respect to π for each π compatible with R.

4. Convergence Theorems

In this section, we use Theorem 3.4 to derive sufficient conditions for the convergence of nonhomogeneous products of stochastic matrices. We will rely on the equivalence of products of elements in a semigroup algebra and convolution products of measures on the semigroup.

Let $\mathcal{S} = \{s_1, s_2, ...s_k\}$ be a finite semigroup. Let $\mathcal{A}(\mathcal{S})$ be the real semigroup algebra generated by S. That is, $\mathcal{A}(\mathcal{S})$ is a set of formal sums with addition and multiplication defined in the standard way. An important subset of $\mathcal{A}(\mathcal{S})$ is defined by

$$\mathcal{A}_1(S) = \left\{ \sum_{i=1}^{k} p_i s_i \,\middle|\, p_i \geq 0,\ \sum_{i=1}^{k} p_i = 1 \right\}.$$

Notice that $\mathcal{A}_1(\mathcal{S})$ is closed under multiplication. Let $\mathrm{Prob}(\mathcal{S})$ be the set of all probability measures on $\mathcal{S}$. Clearly, there is a one-to-one correspondence between elements of $\mathrm{Prob}(\mathcal{S})$ and elements of $\mathcal{A}_1(\mathcal{S})$. If $\mu \in \mathrm{Prob}(\mathcal{S})$, then

$$a = \sum_{i=1}^{k} \mu(s_i) s_i \in \mathcal{A}_1(\mathcal{S}). \tag{3}$$

If $a \in \mathcal{A}_1(S)$ and $\mu \in \mathrm{Prob}(S)$ are related as in (3), we will write $a \sim \mu$. Now recall that if $\mu_1, \mu_2 \in \mathrm{Prob}(\mathcal{S})$, then the convolution product is defined by

$$\mu_1 * \mu_2(s) = \sum_{s_i s_j = s} \mu_1(s_i)\mu_2(s_j).$$

It is elementary to show that if $a_1 \sim \mu_1$ and $a_2 \sim \mu_2$, then $a_1 \cdot a_2 \sim \mu_1 * \mu_2$. Notice that since every $n \times n$ stochastic matrix can be written as a convex combination of binary stochastic matrices, and the set of $n \times n$ binary stochastic matrices forms a finite semigroup, the preceding discussion applies.

Now let (X_k) be a finite state nonhomogeneous Markov chain with transition matrices (A_k). Let

$$A_{k,n} = A_{k+1} \cdot A_{k+2} \cdot \ldots \cdot A_n.$$

If conditions could be found for the convergence of $A_{k,n}$ for all k, then the chain (X_k) would converge regardless of initial distribution, for every starting time $k \geq 0$.

The following theorem was proved in [**2**]. Here the convolution product $\mu_{k,n}$ is defined similarly to $A_{k,n}$.

Theorem 4.1. *Let S be a finite completely simple semigroup and $(\mu_n) \subset \mathrm{Prob}(S)$. Suppose that $\lim_{n\to\infty} \mu_n(Sx)$ exists for all $x \in S$, and that for some idempotent $e \in S$,*

$$\liminf_{n\to\infty} \mu_n(e) > 0.$$

Then $\mu_{k,n}$ converges weakly as $n \to \infty$ for all $k \geq 0$. ■

With the above discussion in mind, it is clear that if $A_i \sim \mu_i$ for all i and $\mu_{k,n}$ converges weakly, then $A_{k,n}$ converges in the standard topology for matrices.

For a given partition π of $\{1, 2, \ldots, n\}$, let $\mathcal{S}_\pi$ be defined as in Section 3. We identify $\mathcal{S}_\pi$ with the isomorphic semigroup $\mathcal{S}$ of binary stochastic matrices.

Note that since π is fixed, each $H(\pi, R_i)$ is in fact an L-class. Since $\mathcal{S}$ is completely simple,

$$H(\pi, R_i) = Ł_B = \mathcal{S} \cdot B,$$

where B is any element of $H(\pi, R_i)$, and $\mathcal{S} \cdot B = \{AB | A \in \mathcal{S}\}$. For convenience, we will label the elements of $H(\pi, R_i)$ distinctly as $B_k^{(i)}$, where $1 \le k \le r!$. The idempotent of $H(\pi, R_i)$ will be labeled as $B_1^{(i)}$.

Let (A_n) be a sequence of block-column balanced stochastic matrices such that each block $A_{ij}(n)$ has identical rows. Then by Theorem 3.4, we see that for each n, A_n has a decomposition in terms of binary stochastic matrices from a completely simple semigroup. Let

$$A_n = \sum_{j=1}^{N} \sum_{k=1}^{r!} p_k^j(n) B_k^{(j)}(n) \tag{4}$$

Then we have the following theorem.

Theorem 4.2. *Let (A_n) be a sequence of block-column balanced stochastic matrices such that each block $A_{ij}(n)$ has identical rows. Let the decomposition of A_n be as in (4) above. Suppose that for some idempotent $B_1^{(\ell)}$,*

$$\liminf_{n\to\infty} p_1^\ell(n) > 0$$

and for each j, $1 \le j \le N$,

$$\lim_{n\to\infty} \sum_{k=1}^{r!} p_k^j(n) \text{ exists.}$$

Then $A_{k,n}$ converges as $n \to \infty$ for all $k \ge 0$.

Proof. If we set $\mu_n\left(B_k^{(j)}(n)\right) = p_k^j(n)$, then $\mu_n \sim A_n$ for all n and the conditions of Theorem 4.1 are satisfied. Thus $\mu_{k,n}$ converges weakly for all k and this implies $A_{k,n}$ converges for all k. ∎

We will close this section with an example illustrating Theorem 4.2. Recall that a sequence of stochastic matrices (A_k) is weakly ergodic if the limit points of the products $A_{k,n}$ are matrices with identical rows. If, in addition to weak ergodicity, the products $A_{k,n}$ converge, then (A_k) is said to be strongly ergodic.

Example 4.3.

Let (A_k) be a sequence of stochastic matrices with

$$A_k = \begin{bmatrix} \frac{1}{8} + r_k & s_k & \frac{7}{8} - (r_k + s_k) \\ \frac{1}{8} + r_k & s_k & \frac{7}{8} - (r_k + s_k) \\ \frac{1}{8} - r_k & \frac{3}{4} - s_k & \frac{1}{8} + r_k + s_k \end{bmatrix}$$

where the only restriction on r_k, s_k are

$$0 \le r_k \le \frac{1}{8}$$

$$0 \le s_k \le \frac{3}{4}$$

Let

$$B_1 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad B_2 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$B_3 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \qquad B_4 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Now $S = \{B_1, B_2, B_3, B_4\}$ is a completely simple semigroup of matrices with

$$\pi = \{(1,2),(3)\}, \quad R_1 = \{1,3\}, \quad R_2 = \{2,3\}.$$
$$\text{and } H(\pi, R_1) = \{B_1, B_3\}, \quad H(\pi, R_2) = \{B_2, B_4\}.$$

Notice

$$A_k = \sum_{i=1}^{4} p_i^k B_i$$

with

$$p_1^k = \frac{1}{8} + r_k$$
$$p_2^k = s_k$$
$$p_3^k = \frac{1}{8} - r_k$$
$$p_4^k = \frac{3}{4} - s_k.$$

For all k, $p_1^k + p_3^k = \frac{1}{4}$, $p_2^k + p_4^k = \frac{3}{4}$. Also

$$\lim_k \inf(p_1^k) \geq \frac{1}{8} > 0$$

where $B_1 = B_1^2$ is idempotent. Thus by Theorem 4.2, $A_{k,n}$ converges for all k, as $n \to \infty$.

Now in particular let

$$r_k = \begin{cases} 0 & k = 2^n \\ \frac{1}{8} & \text{otherwise} \end{cases}$$

$$s_k = \begin{cases} \frac{1}{2} & k = 2^n \\ \frac{3}{4} & \text{otherwise} \end{cases}$$

Thus

$$A_k = \begin{cases} \begin{bmatrix} \frac{1}{8} & \frac{1}{2} & \frac{3}{8} \\ \frac{1}{8} & \frac{1}{2} & \frac{3}{8} \\ \frac{1}{8} & \frac{1}{4} & \frac{5}{8} \end{bmatrix} & k = 2^n \\ \begin{bmatrix} \frac{1}{4} & \frac{3}{4} & 0 \\ \frac{1}{4} & \frac{3}{4} & 0 \\ 0 & 0 & 1 \end{bmatrix} & \text{otherwise} \end{cases}$$

By Bernstein's condition (see **11**, pg. 141) since

$$\begin{aligned}&\sum_{k}^{\infty} \max_j \,[\min_i(a_{ij}^{k})]\\ &= \sum_{n=1}^{\infty} \max_j \,[\min(a_{ij}^{2^n})]\\ &= \infty,\end{aligned}$$

the sequence (A_k) is weakly ergodic. We have shown $A_{k,n}$ converges by Theorem 4.2, thus (A_k) is strongly ergodic.

Notice the elements of (A_k) have been chosen so that they:

(1) Do not all contain a single ergodic class,
(2) Are not all regular,
(3) Do not converge.

Thus, to the best of our knowledge, no known sufficient conditions for strong ergodicity would apply in this situation.

Concluding Remarks

The problem of generalizing Theorem 3.4 to handle semigroups with more than a single partition remains open. The situation becomes considerably more difficult, but we feel the notion of block-column balance will prove instrumental for the general case as well. The interplay between the structure of semigroups of binary stochastic matrices and the convergence of nonhomogeneous Markov chains should continue to be fruitful.

References

[1] R. Brualdi and H. J. Ryser, *Combinatorial Matrix Theory*, Cambridge University Press, Cambridge, 1991.

[2] G. Budzban and I. Ruzsa, *Some Results Concerning Convergence of Convolution Products of Probability Measures on Discrete Semigroups*, To appear in J. Theoretical Probability.

[3] D. Geman and S. Geman, *Stochastic relaxation, Gibbs distributions, and the Baysian restoration of images*, IEEE Transactions PAMI **6** (1984), 721–741.

[4] B. Gidas, *Nonstationary Markov Chains and Convergence of the Annealing Algorithm*, J. Stat. Phys. **39** (1985), 73–131.

[5] G. Högnäs, *Random Semigroup Acts on a Finite Set*, J. Austral. Math. Soc. **23 (Ser. A)** (1977), 481–498.

[6] G. Högnäs and A. Mukherjea, *Probability Measures on Semigroups: Convolution Products, Random Walks, and Random Matrices*, Plenum Press, New York, 1995.

[7] J. M. Howie, *An Introduction to Semigroup Theory*, Academic Press, London, 1976.

[8] C. Hwang and S. Sheu, *Singular Perturbed Markov Chains and Exact Behaviors of Simulated Annealing Processes*, Journal of Theoretical Probability **5** (1992), 223–249.

[9] D. Isaacson and R. Madsen, *Markov Chains: Theory and Applications*, Wiley, New York, 1976.

[10] A.H. Clifford and G.B. Preston, *The Algebraic Theory of Semigroups, Vol 1*, AMS, Providence, 1961.

[11] E. Seneta, *Non-Negative Matrices and Markov Chains*, Springer-Verlag, New York, 1980.

Department of Mathematics
Southern Illinois University
Carbondale, Illinois 62901–4408

email:gbudzban@ math.siu.edu