A GEOMETRIC APPROACH TO THE DECOMPOSITION OF FINITE STOCHASTIC AUTOMATA

by

I-Ngo Chen, B. S. E., M. Sc.

Submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

Department of Electrical Engineering
Faculty of Pure and Applied Science
University of Ottawa
Ottawa, Ontario

June, 1968

ABSTRACT

This thesis is concerned with the analysis of finite stochastic automata. A geometric concept is introduced which regards a transition matrix as a mapping of the state-probability-distributions. The range of the mapping is a polyhedral convex set determined by points which are row vectors of the matrix. Interesting properties of such points in a convex polyhedron are investigated. Serial combinations of finite stochastic automata are studied and methods of decomposing them serially are given. Parallel combinations of finite stochastic automata are also studied. Sufficient and necessary conditions that a finite stochastic automaton admits parallel decomposition are formulated and proved. Based on the same geometric concept, implication of definite stochastic automata is studied and decision procedures are given.

ACKNOWLEDGEMENTS

The author wishes to express his gratitude to Professor C. L. Sheng for his guidance. Without his advice and encouragement, this work could not have been done.

The author is also grateful to Professors G. S. Glinski and S. G. S. Shiva for their criticism, encouragement, and many stimulating discussions during the course of this research. Appreciation is due to Dr. W. A. Davis, Dr. Shanker Singh, and Mr. W. K. Chung for their advice. The assistance of Mr. C. F. Chung and Mr. S. Y. Lee are acknowledged, especially for proof reading of the final manuscript. The author is indebted to Madame L. LeBlanc, who typed the final manuscript, to Mrs. D. Kocis-Thibert, who typed several parts of the drafts, to Mr. Rene LeHenaff, for his technical assistance, and to NRC, for financial support.

CONTENTS

PAGE

CONTENTS (Continued) PAGE

LIST OF SYMBOLS

σ_i : input symbol

Σ : set of σ_i

z_i : output symbol

Z : set of z_i

$\mathcal{C}$: combinational network

M : mapping

X_i : input variable

f_i : output variable

$\mathcal{M}_1$: Moore model sequential machine

$\mathcal{O}$: output function

$\mathcal{M}_2$: Mealy model sequential machine

D : unit delay

$\mathcal{M}_3$: finite automaton

s_i : state

S : set of states

s_o : initial state

F : set of final states

w_i, u_i : states probability distribution

W : set of w_i

U : subset of W

w_o : initial states probability distribution

Ω : finite stochastic automaton

A_{σ_i} : transition matrix when σ_i is applied

$a_{jk}(\sigma_i)$: entry of A_{σ_i}

x : tape of σ_i

Ω : finite stochastic automaton

η^F : column vector with component either 1 or 0

$\mathcal{M}_p$: Page's probabilistic sequential machine

λ : cut-point

$\mathcal{R}_E$: reduction relation

t : time

ν : time

$\mathcal{S}^n$: space

p_i : point

α_i : coefficient

G : convex polyhedral set

d_i : determining point

V : set of vertices

Λ_V : polyhedron determined by V

v : vector $v_i = (\beta_{i1}, \beta_{i2}, \ldots)$

β_{jk} : coefficient

$\oplus$: binary operator

O	:	null vector
Γ	:	group
Γ_j	:	subset of Γ containing j non-zero components
r	:	real integer
$\#(\Gamma)$	:	No. of elements of Γ
θ	:	imaginary point
N	:	point set
M_N	:	mapping
N_i	:	$N_i = M_N(\Gamma_i)$
p_{jk}	:	image of p in the kth j-dimensional face of Λ_V
J, K	:	subset of V
p_J	:	image of p in Λ_J
L	:	set of index i for $K_i \in V$
H	:	hyperplane
e_i	:	unit vector
E_n	:	polyhedral convex set determined by $e_1, \ldots. e_n$
R_A	:	range of the mapping M_A
ρ_i	:	element of Σ_2
$\mathcal{A}, \mathcal{B}$	:	column vector whose component are matrices
π, δ	:	uniform partition over V
$\bar{L}$	:	line segment
T	:	matrix
Q	:	matrix
Q'	:	matrix whose row vectors are all linearly independent

CHAPTER I

INTRODUCTION

The theory of finite automata has been well developed in the past decade. Much work has been done in generalizing finite automata to include the nondeterministic case. The purpose of pursuing such a generalization is not just for mathematical interest. As systems are going more and more complicated, the probability of failure of a system due to noise in transmission, disturbance in the system, imperfect maintenance, and other factors becomes of great concern. Actual systems are thus probabilistic and the study of probabilistic systems will have practical application in estimating and improving the reliability of a system. Udagawa and Inagaki [44, 45] have shown how the reliability of a sequential circuit can be accurately estimated and improved through the study of probabilistic systems. Others [3, 4, 5, 11] have shown how probabilistic systems can be used as mathematical models for the synthesis of learning systems. Stochastic automata with reinforcement scheme have been used by Varshavskii and Vorontsova, Fu, McMurtry, and McLaren [14, 15, 46] for the design of optimal and adaptive controllers.

The first paper dealing with all aspects of probabilistic systems was written by Rabin [39]. In his paper, Rabin generalized the concept of finite automata developed by Kleene, Rabin, and Scott [27, 40]. A probabilistic automaton (p. a.) is defined as a system with a finite set of states and a transition probability table. The system may receive input from an input alphabet.

If the system is in a certain state when a certain input is applied, the system can go to any one of the states of the system with a certain probability. By introducing a threshold value, called cut-point, for determining the acceptance or rejection by the system to any tape of input symbols, Rabin has proved that deterministic automata are included in probabilistic automata. Several interesting properties about probabilistic automata such as the number of states and the problem of stability have been investigated. The work of Rabin has been well extended by Paz [36]. Another approach was made by Carlyle [6, 7, 8] in generalizing finite automata in the sense of Moore [31]. A Carlyle's stochastic sequential machine is essentially characterized by a family of matrices each one of which indicates the state to state transition probabilities for a certain input sequence and a certain output sequence. Equivalence among machines was thus defined and reduced forms for stochastic sequential machines can be obtained. Another attempt in generalizing deterministic automata was made by Bacon [1] in the problem of decomposition. Decomposition theorems similar to those found by Hartmanis [24] in the deterministic case have been found and proved. Page [32] studied several equivalence relations among sequential machines and therewith obtained the behavioral equivalence between deterministic machines and probabilistic machines.

In the literature, the definition of a probabilistic system varies from one paper to another. The essential feature common to all different definitions is the probabilistic nature of the state transition which can be denoted by a state transition matrix. This is the sole object of our investigation. Thus in Chapter II, we shall define a stochastic automaton as a probabilistic automaton with final states undefined or a stochastic sequential machine with output undefined. It is seen that a stochastic automaton so defined possesses the common factors to all probabilistic machines of different models and can thus be transformed to any one of them. Starting from deterministic computation channels, the concept of finite stochastic automata is introduced through discussions on their relationships with related subjects.

Chapter III then introduces the geometricconcept which we shall use as our tool in attacking the problem of analysis of stochastic automata. The transition matrix of a stochastic automaton is regarded as a mapping of the state-probability-distribution. The range of the mapping is a polyhedral convex set, determined by points that are row vectors of the transition matrix. As the transition matrix is stochastic, all points in the range of the mapping are contained in a convex polyhedron. In order to reveal certain characteristics of a transition matrix, certain properties of points in the range of the mapping are investigated. Definitions of parallelism and projectivity are introduced. Dimension of the range of a mapping is discussed. The whole chapter thus provides the mathematical background for the rest of our investigations.

Based on the geometric concept discussed in Chapter III, Chapter IV discusses the problems of serial combination and serial decomposition of finite stochastic automata. Methods for decomposing a transition matrix into two matrices in series are given. Theorems assurring the validity of the methods are proved. However, as serial decomposition of finite stochastic automata has little practical application, more weight is put on the problem of serial combinations.

Chapter V deals with the problems of parallel combination and parallel decomposition of finite stochastic automata. Geometry meanings to the points in the range of the mapping of the combined matrix are discussed in detail. The necessary and sufficient conditions that a finite stochastic automaton admits a parallel decomposition are formulated and proved. Methods of testing and decomposing, together with an illustrating example, are given.

A special class of finite stochastic automata called definite stochastic automata is discussed in Chapter VI. Similar to deterministic automata, definite stochastic automata are defined as finite stochastic automata which have a certain fixed state probability distribution after a definite number of input symbols are applied and irrespective of the initial distribution. The geometric implication of the transition matrix of this kind of automata is discussed. Decision procedures and illustrating example are given.

The last Chapter of the thesis is a brief conclusion and remarks on the whole work.

CHAPTER II

FINITE STOCHASTIC AUTOMATA

2.1 Motivation

Any system which receives signals from an external source and which gives out signals is analytically an information transformer where the incoming information is carried by the input symbols to the system and the outgoing information is carried by the output symbols from the system. If the system is a communication channel, information is to be retained. If the system is a computation channel, information is usually to be reduced. In the later case, the transformation of information is governed by a transition function which can be explicitly expressed by a transition table. If the system has memory, part or all of the information received is stored in the system. The information to be transformed is now not only that from the external source but also that stored in the system. In case the system is noisy or the transformation of information is not deterministic but rather governed by certain fixed probability while the external input is still deterministic, the information stored in the system can not be surw, i.e. at any time instant, what we are sure of is not a certain state symbol but a certain states probability distribution. Such is the case of a stochastic automaton. In the following, we are going to describe the stochastic automata through discussion of their relations with related subjects.

2.2 Combinational networks, automata, and sequential machines.

In all the following discussions, all systems are assumed to be discrete. Starting from the simplest case, consider a computation channel which is noiseless and memory-free. We denote an input symbol to the system as σ_i, the set of all input symbols as Σ, while an output symbol of the system is denoted as z_j, and the set of all output symbols as Z. The input symbol σ_i is in fact a vector. The number of components is equal to the number of input variables. Thus σ_i is a certain ordered valuation of the input variables. Similarly, z_j is a certain ordered valuation of the output variables. The system transforms the information carried by σ_i to the information carried by z_j. Or we may say that the system is a mapping which maps an element σ_i in Σ onto an element z_j in Z. Such a system can be realized by a combinational network. In this sense, we define a combinational network as:

<u>Definition 2.1:</u>

A combinational network is a discrete system

$$\mathcal{C} = < \Sigma, M, Z >$$

where

Σ : a finite set of input symbols

Z : a finite set of output symbols

M : a mapping

$$\Sigma \xrightarrow{M} Z.$$

Such a system is diagramtically shown in Fig. 2.1.

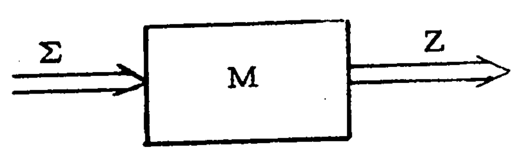

Fig. 2.1 Diagramatic representation of a combinational network.

Explicitly, $\mathcal{C}$ can be expressed by an output function table. For illustration, suppose the information input to a combinational network $\mathcal{C}_1$ is carried by binary variables X_1, X_2, and X_3 ; and the output information by the binary variables f_1 and f_2. Let the values of each binary variable be 0 and 1. Then

$$\Sigma = \{\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6, \sigma_7, \sigma_8\}$$

where

$$\sigma_1 = \boxed{0 \quad 0 \quad 0}$$
$$\sigma_2 = \boxed{0 \quad 0 \quad 1}$$
$$\sigma_3 = \boxed{0 \quad 1 \quad 0}$$
$$\sigma_4 = \boxed{0 \quad 1 \quad 1}$$
$$\sigma_5 = \boxed{1 \quad 0 \quad 0}$$
$$\sigma_6 = \boxed{1 \quad 0 \quad 1}$$
$$\sigma_7 = \boxed{1 \quad 1 \quad 0}$$
$$\sigma_8 = \boxed{1 \quad 1 \quad 1}$$

The first component of σ_i is the value of X_1, the second, that of X_2, and the third, that of X_3. Similarly,

$$Z = \{z_1, z_2, z_3, z_4\}$$

where

$$z_1 = \boxed{0 \quad 0}$$
$$z_2 = \boxed{0 \quad 1}$$
$$z_3 = \boxed{1 \quad 0}$$
$$z_4 = \boxed{1 \quad 1}\ .$$

The first and second components of z_i are the values of f_1 and f_2, respectively. If the mapping M is such that

$$M(\sigma_1) = M(\sigma_3) = M(\sigma_4) = z_1$$

$$M(\sigma_2) = M(\sigma_5) = z_2$$

$$M(\sigma_8) = z_3$$

$$M(\sigma_6) = M(\sigma_7) = z_4 ,$$

then the output function table will be that shown in Table 2.1.

Table 2.1 Output function table of C_1

	X_1	X_2	X_3	f_1	f_2	
σ_1	0	0	0	0	0	z_1
σ_2	0	0	1	0	1	z_2
σ_3	0	1	0	0	0	z_1
σ_4	0	1	1	0	0	z_1
σ_5	1	0	0	0	1	z_2
σ_6	1	0	1	1	1	z_4
σ_7	1	1	0	1	1	z_4
σ_8	1	1	1	1	0	z_3

If the system has memory, then the information to be transformed will be those of the external source together with those stored in the system. Let the information stored in the system be carried by the symbol s_i, called state. The set of all such symbols is S. If all of the information transformed by the system is stored in the system, then the mapping M will be

$$(\Sigma \times S) \xrightarrow{M} S.$$

Such is the case of a Moore model sequential machine.

Definition 2.2:

A Moore model sequential machine [31] is a system $\mathcal{M}_1$

$$\mathcal{M}_1 = < \Sigma , S, M, \mathcal{O}, Z >$$

where Σ , Z are as in Definition 2.1

S : a finite set of internal states

M : a mapping

$$(\Sigma \times S) \xrightarrow{M} S;$$

$$M(\sigma_i, s_j) = s_k \quad , \qquad \sigma_i \in \Sigma \quad ; \quad s_j , s_k \in S$$

$\mathcal{O}$: a mapping

$$S \xrightarrow{\mathcal{O}} Z ;$$

$$\mathcal{O}(s_i) = z_j \quad , \qquad s_i \in S \quad , \quad z_j \in Z .$$

A diagramatic expression of a Moore model sequential machine is shown in Fig. 2.2, where D is the unit delay.

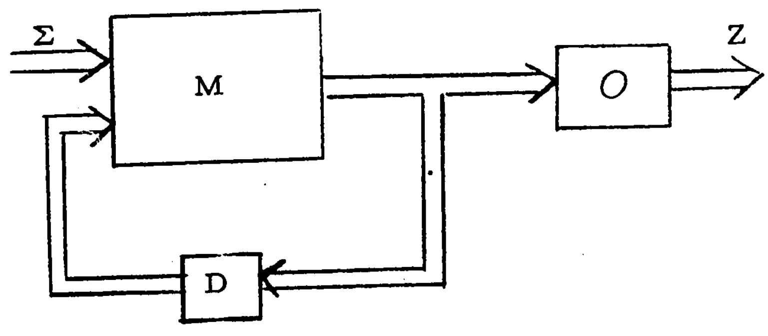

Fig. 2.2 Diagramatic representation of a Moore model sequential machine.

If neither all the information transformed by the system are stored in the system nor all of them are used as output, then we have a sequential machine of the Mealy type.

Definition 2.3:

A Mealy model sequential machine [29] is a system M_2

$$M_2 = < \Sigma , S, M, O , Z >$$

where Σ, S, M, and Z are as in Definition 2.2

$$(\Sigma \times S) \xrightarrow{O} Z.$$

A Mealy model sequential machine is shown in Fig. 2.3.

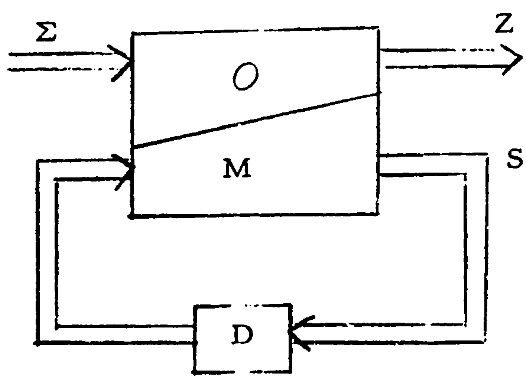

Fig. 2.3 Diagramatic representation of a Mealy model sequential machine.

Definition 2.4: (Rabin and Scott)

A finite automaton is a system M_3

$$M_3 = < \Sigma , S, M, s_o , F >$$

where Σ, S, and M are as in Definition 2.2

s_o : initial state

F : subset of S, called final states.

Thus except the emphasis of initial state, a finite automaton M_3 is a special Moore model sequential machine having a Z which contains only two elements; one associates with all states in F, and the other, with all states not in F.

From the above definitions, we see that the common part in all systems with memory is the triple $< \Sigma, S, M >$. This is the essential and most interesting part of a discrete system. As machines of different models can be transformed from one to another easily, we may assume, without lost of generality, that the output of a system with feedback depends on the internal state of the system.

2.3 Deterministic machines and stochastic automata

All the systems described in the previous section are said to be deterministic because at any time instant t, the probability that an input symbol σ_i is applied to the system, the probability that the system is in state s_j, and the probability that the transition $M(\sigma_i, s_j) = s_k$ takes place are all assumed to be either 1 or 0. If all probabilities are not restricted to be 1 or 0, then the output will be probabilistic.

Consider the case that the applying of the input σ_i is deterministic, but the system is probabilistic, i.e. the probability that the transition $M(\sigma_i, s_j) = s_k$ takes place is neither 1 nor 0, but is a certain value between them. In such a case, given an input symbol σ_i while it is in state s_j, the occurrence of next state is governed by a fixed probability distribution. Therefore, after a time of unit delay, the occurrence of next state is not deterministic, and what we can be sure of is a probability distribution over all the states of the system. Denote such a state-probability-distribution as w_i, and the set of all w_i' s as W. Then we may define a finite stochastic automaton as:

Definition 2.5:

A finite stochastic automaton is a system Ω

$$\Omega = < \Sigma , S, w_o, W, M >$$

where Σ , S are as in Definition 2.2

w_o : initial state-probability-distribution

W : set of all state-probability-distributions

M : a mapping

$$(\Sigma \times W) \xrightarrow{M} W.$$

The mapping M in the above definition is characterized by a set of matrices $\{A_{\sigma_i}\}$, $\sigma_i \in \Sigma$, where every row of the matrix A_{σ_i} is a stochastic row vector, i.e., every component is non-negative and all components sum up to unity. The entry $a_{jk}(\sigma_i)$ of a matrix A_{σ_i} is the probability that when the system is at state s_j and an external input symbol σ_i is applied, the next state will be s_k . If at time t, the state-probability-distribution is w_j, and an input symbol σ_i is applied, then the next state-probability-distribution will be w_k, where

$$w_k = M(\sigma_i , w_j) = w_j \cdot A_{\sigma_i} .$$

Let M_{σ_i} be a mapping characterized by the matrix A_{σ_i} . Then M is a set of mappings

$$M = \{ M_{\sigma_i} / \sigma_i \in \Sigma \} .$$

Let any length of sequence of the input symbols be called a <u>tape</u>. Let

$$x = \sigma_1 \sigma_2 \sigma_3 \ldots . \sigma_n \tag{2.1}$$

be such a tape. For a finite stochastic automaton Ω defined in Definition 2.5, starting from an initial state-probability-distribution w_o, the state-probability-distribution after a tape x in Eq. (2.1) is applied will be

$$w_o \cdot A_{\sigma_1} \cdot A_{\sigma_2} \ldots\ldots\ldots\ldots\ldots\ldots . A_{\sigma_n} = w_o \cdot A_x .$$

The finite stochastic automaton so defined possesses the essential feature of a probabilistic system. Any model of a probabilistic machine can be regarded as a finite stochastic automaton with certain specific features. For instance, a probabilistic combinational network is a finite stochastic automaton without memory. In this case, the set S of Definition 2.5 is the output of the probabilistic network and

$$\Sigma \xrightarrow{M} W .$$

<u>Definition 2.6:</u>

A finite stochastic automaton without memory is a finite stochastic automaton Ω where the mapping M is

$$\Sigma \xrightarrow{M} W.$$

For specific purpose, a deterministic output Z together with specific output function $\mathcal{O}$ may be added to a finite stochastic automaton. As in the deterministic case, Z may depend on the input and the state-probability-distribution or on the latter only, i.e.

$$(\Sigma \times W) \xrightarrow{\mathcal{O}} Z$$

or

$$W \xrightarrow{\mathcal{O}} Z.$$

Let F be a subset of S.

$$F = (s_{i_1}, \; s_{i_2}, \; \ldots\ldots\ldots\ldots\ldots, \; s_{i_r}).$$

Let η^F be a column vector with components η_i^F, where

$$\eta_i^F = \begin{cases} 1 & \text{if } i \in (i_1, \; i_2, \ldots\ldots\ldots, i_r) \\ 0 & \text{otherwise.} \end{cases}$$

Define U, a subset of W, as

$$U = \{ w_i \;/\; w_i \in W, \quad w_i \cdot \eta^F > \lambda \}$$

where λ is a real number, $0 \leq \lambda < 1$.

Let the initial state-probability-distribution w_o be assumed to be concentrated in one state s_o, and let $Z = \{1, 0\}$.

Define the output function as:

$$\mathcal{O}(w_i) = \begin{cases} 1 & \text{if } w_i \in U \\ 0 & \text{otherwise} \end{cases}$$

Then we have a probabilistic automaton with cut-point λ as defined by Rabin [39] and Paz [36]. In this case, a tape x is said to be accepted by the probabilistic automaton if and only if $\mathcal{O}(w_o \cdot A_x) = 1$.

In other case, let the output function $\mathcal{O}$ be defined as

$$\mathcal{O}(s_i) = z_i$$

where z_i is a real number and is called the output from state s_i.

Then we have a probabilistic sequential machine $\mathcal{M}_p$ as defined by Page [32] .

Evidently, machines which are deterministic, are a sub-class of machines, which are probabilistic. As pointed out by Rabin [39], every set of tapes definable by a deterministic automaton is definable by some probabilistic automata. However, the coverse is not true. Rabin and Paz both have given examples [39,36] showing that there exist events which are definable by probabilistic automata but not definable by deterministic automata.

The behavioral equivalences between probabilistic machines and deterministic machines has been shown by Page [32]. According to Page, the expected value of output for an input tape x of a probabilistic sequential machine $\mathcal{M}_p$ is defined as

$$E_{\mathcal{M}_p}(x) = w_o \cdot A_x \cdot Z \quad , \qquad \text{for } x \text{ in } \Sigma^*$$

where Z is a column vector whose ith component is z_i.

and $\Sigma^* = \bigcup_{n=o}^{\infty} \Sigma^n$

while Σ^o denotes the empty tape. The reduction relation $\mathcal{R}_E$ is thus defined as

$$x_i \mathcal{R}_E \, x_j \qquad \text{iff} \qquad E_{\mathcal{M}_p}(x_i x_k) = E_{\mathcal{M}_p}(x_j x_k)$$

$$\text{for all} \quad x_k \in \Sigma^* \quad \text{and}$$

$$\text{for all} \quad w_o \in W.$$

Defintion 2.6 (Page)

$rp_{\mathcal{M}_p}(x)$ is the response of $\mathcal{M}_p$ to input tape x. If $\mathcal{M}_p$ is deterministic, $rp_{\mathcal{M}_p}(x)$ is the state of $\mathcal{M}_p$ after an input of x. If $\mathcal{M}_p$ is stochastic, $rp_{\mathcal{M}_p}(x)$ is a random variable taking on values which are states with distribution $w_o \cdot A_x$.

Theorem 2.1 (Page):

The reduction relation $\mathcal{R}_E$ defined by a probabilistic sequential machine $\mathcal{M}_p$ has finite rank if and only if there exists a finite deterministic machine $\mathcal{M}'_p$ with a deterministic output $O_{\mathcal{M}'_p}$ such that $O_{\mathcal{M}'_p}(rp_{\mathcal{M}'_p}(x)) = E_{\mathcal{M}_p}(x), \quad \forall \; x \in \Sigma^*$.

The deterministic machine $\mathcal{M}'_p$ stated in Theorem 2.1 is a finite sequential machine with deterministic state transition while with each state of $\mathcal{M}'_p$, there is an expected value of output which is equal to that of one equivalence class of $\mathcal{R}_E$. This concept of expectation equivalence can be generalized to N-moment equivalence where an N-moment reduction relation $\mathcal{R}_N$ is defined. The equivalent deterministic machine in this case will have deterministic state transition with each state connected to a random device which gives the same expectation and N-1 moments as those of one congruence class of $\mathcal{R}_N$ of the probabilistic machine.

2.4 Stochastic automata and Markov chains.

A stochastic automaton defined by Definition 2.5 can be analyzed with the aid of the theory of Markov chains. For as can be seen, a stochastic automaton Ω with Σ containing only a single element is nothing but a Markov chain. If Σ is not a single element set and ~~that~~ the transition matrix is different for different input, then Ω is a non-homogeneous Markov chain. Udagawa and Inagaki [44] have shown that a stochastic automaton Ω with output set Z can be expressed as a simple Markov chain if the following assumptions are made.

1. the input is a simple Markov chain;
2. the output has no effect on the Markov process of the input;
3. the output at time $t = \nu$ depends only on the input and state at time $t = \nu$.

4. the state at time $t = \nu + 1$ depends only on the input, state, and output at time $t = \nu$.

Let the input symbol, the state symbol, and the output symbol combined together be called a total state symbol. A stochastic automaton with output Z is thus a transformation of the set of all total states to the set itself. This transformation is characterized by a Markov chain. By assigning an additional state s_w, called erroneous state, to the set S of a finite stochastic automaton Ω, the Markov chain obtained above is changed to a Markov chain with an absorbing state. Using such a chain, Udagawa and Inagaki have shown that the reliability of a sequential machine can be thus accurately evaluated.

CHAPTER III

PROPERTIES OF POINTS IN A SIMPLEX

The purpose of this chapter is to provide a mathematical background for the investigation in all the following chapters. As mentioned in the previous chapter, any probabilistic system can be regarded as an information transformer. The transformation of information is characterized by a transition matrix which is stochastic and can be regarded as a polyhedral convex set determined by points which are row vectors of the matrix. The transition matrix is thus characterized by such points. Investigation of the properties of such points would thus reveal the properties of the transition matrix, and in turn, those of a finite stochastic automaton.

3.1 Simplex

In this section, for reference purposes, we are going to introduce some terminology and definitions, although most of them are well-known.

A vector in an n-dimensional affine space $\mathcal{S}^n$ is also a point in the same space. Two points p_1 and p_2 determine a line $p_1 p_2$. Any point p on the line $p_1 p_2$ can be expressed as

$$p = \alpha_1 p_1 + \alpha_2 p_2 , \quad \alpha_1 + \alpha_2 = 1.$$

A convex set is a set containing points such that if points p_1 and p_2 are in the set, then the whole segment $\overline{p_1 p_2}$ is in the set G. An extreme point of a polyhedral convex set G is the intersection of (n-1) bounding hyperplanes of G where (n-1) is the dimension of G. Let $d_1, d_2, \ldots d_k$ be points of G. If any point p in G can be written

$$p = \alpha_1 d_1 + \ldots\ldots + \alpha_k d_k, \quad \sum_{i=1}^{k} \alpha_i = 1, \ \alpha_i \geq 0 \text{ for all } i, \qquad (3.1)$$

Then we call d_i's, the determining points of G. And we denote G as $G = [d_1, d_2, \ldots, d_k]$. If all d_i's are extreme points of G and $k = n$ where (n-1) is the dimension of G, then the expression of p in Eq. (3.1) is unique. Clearly, for a convex set G any set of determining points contains the set of all extreme points, but the point set determined by both of them are the same G. A well known theorem about a convex set G is worth mentioning here.

Theorem 3.1: (Kemeny, Mirkil, Snell and Thompson) [26]

Let G be a bounded polyhedral convex set. Then

(1) Every point in G can be written as a convex combination of extreme points of G.

(2) Every point p that is a convex combination of the extreme points of G belongs to G.

Theorem 3.1 is still valid if we replace the term "extreme points" in the theorem with "determining points". The relationship between extreme points and determining points is that the extreme points are not necessary linearly independent. Let $d_1, \ldots\ldots, d_n$ be points in S^n. They are called linearly independent if the vectors $d_2 - d_1$, $d_3 - d_1, \ldots\ldots, d_n - d_1$ are linearly independent. Let $d_1, \ldots\ldots, d_n$ be linearly independent. The convex set determined by these points is called a simplex of vertices $d_1, \ldots\ldots, d_n$. Every point in the simplex can be expressed as a convex combination of the vertices of the simplex. Conversely, every point that is a convex combination of the vertices of the simplex belongs to the simplex. A face of a simplex G is a simplex determined by vertices which are a subset of the vertices of G. A simplicial complex is a collection K of simplexes such that a). every face of any simplex belonging to K is in K : and

b). if G_1, G_2 are simplexes, $G_1 \in K$, then $G_1 \cap G_2$ is a common face of G_1 and G_2.

If K is finite, then it is called a finite simplicial complex. The union of point sets of all simplexes of a finite simplicial complex is called the polyhedron of K .

3.2 Points in a simplex

Let V be a collection of vertices where the number of vertices is n and each vertex is a point in an n-dimensional space. Assume that these n points are not contained in any (n-2)-dimensional hyperplane. Denote the simplex determined by these n vertices as Λ_V. A point p in Λ_V can thus be uniquely expressed as a convex combination of these n points. Let $d_1, d_2, \ldots, d_n$ be these n points, then

$$p = \alpha_1 d_1 + \alpha_2 d_2 + \ldots + \alpha_n d_n, \quad \sum_i^n \alpha_i = 1, \quad \alpha_i \geq 0.$$

Written in another way, we have

$$p = (\alpha_1, \alpha_2, \ldots, \alpha_n). \tag{3.2}$$

Let v be a vector such that its ith component is either 0 or α_i of Eq. (3.2), for i = 1, ..., n. Let Γ be a set of such vectors.

Let $v_i = (\beta_{i1}, \beta_{i2}, \ldots, \beta_{in})$

$v_j = (\beta_{j1}, \beta_{j2}, \ldots, \beta_{jn})$

be any two elements of Γ. Define a binary operation "$\oplus$" as :

$$v_i \oplus v_j = (\beta_{i1} \oplus \beta_{j1}, \ \beta_{i2} \oplus \beta_{j2}, \ldots\ldots, \ \beta_{in} \oplus \beta_{jn})$$

where $\beta_{ik} \oplus \beta_{jk} = \begin{cases} \beta_{ik} & \text{if} \quad \beta_{jk} = 0 \\ \beta_{jk} & \text{if} \quad \beta_{ik} = 0 \\ 0 & \text{otherwise} \end{cases}$

then Γ is a group with operator "$\oplus$", and with the element

$$O = (0, 0, \ldots\ldots, 0)$$

as the identity. The generating elements of the group Γ are those elements which takes only one component of p as its only non-zero component.

Let Γ_j, $j = 0, 1, 2, \ldots\ldots, n$

be the subset of Γ containing elements which have j non-zero components, then

$$\Gamma_i \cap \Gamma_j = \phi, \text{ for } i \neq j$$

$$\bigcup_j \Gamma_j = \Gamma$$

and all Γ_j's with even subscript form a subgroup of Γ. Thus for any point p in Λ_V, there associates a group Γ. The number of elements of Γ depends on the number of non-zero components of p. If p has r non-zero components, then

$$\#(\Gamma) = 2^r ,$$

where $\#(\Gamma$ denotes the number of elements of Γ.

since every element of Γ is a vector, each one except the O element can be normalized to have unity sum. This normalization in fact is a mapping M_N which maps Γ into a point set N in Λ_V, by defining $M_N(O) = \theta$

where θ is not existing in Λ_V and will be called an imaginary point. Every element of N is called an <u>image</u> of p.

Let $M_N(\Gamma_i) = N_i$, $i = 0, 1, 2, \ldots\ldots, n$.

We call N_i the set of all images of p in all (i - 1) - dimensional faces of Λ_V . Clearly

$$N_o = \{\theta\}$$

$$N_r = \{p\}$$

if p has r non-zero components.

Thus for any point p in Λ_V , there corresponds an image (may be θ) of p in every j-dimensional faces of Λ_V , for $j = 0, 1, 2, \ldots\ldots, n - 2$. Similarly for a point set G contained in Λ_V there corresponds an image set of G in every j-dimensional faces of Λ_V , for $j = 0, 1, 2, \ldots \ldots, n - 2$.

Denote the image of a point p in the kth j-dimensional face of Λ_V as p_{jk} . Two points p and q are <u>identical</u> if and only if

$$p_{jk} = q_{jk} \qquad \text{for all } j \text{ and } k \ .$$

If p and q are not identical, then they are <u>distinct</u> .

<u>3.3 Projectivity and parallelism</u>

<u>Definition 3.1:</u>

A line in a simplex Λ_V is said to be "projective" to a j-dimensional face of Λ_V if each point in the line segment within Λ_V has either the same image or θ in the face.

Note that the relation "projective" so defined is not commutative.

In the sequel, assume that J and K are subsets of V,

i.e. $V \supseteq J$

$V \supseteq K$

and that J has j points and K has k points. Denote the face of Λ_V determined by points of J as Λ_J. Denote the image of a point p in Λ_J as p_J, and that of another point q as q_J. Then we have

Lemma 3.1:

Let Λ_V , Λ_J, θ p, q, p_J and q_J be as defined above, then the line determined by p and q is projective Λ_J if

$$p_J = q_J$$

or $\quad p_J = \theta$

or $\quad q_J = \theta$.

This is evidently seen by Definition 3.1 and Eq. (3.2).

Definition 3.2:

A hyperplane in a simplex Λ_V is said to be projective to a j-dimensional face of Λ_V if every line in the hyperplane is projective to the face.

Lemma 3.2:

Let a point in Λ_J be p_J, and let the simplex formed by p_J and all points of $\bar{J}$ be denoted as $\Lambda_{(\bar{J} \cup p_J)}$, where $\bar{J}$ is the complement of J in V, then $\Lambda_{(\bar{J} \cup p_J)}$ is projective to Λ_J.

Proof:

Every point in $\Lambda_{(\overline{J} \cup p_J)}$ can be expressed as a convex combination of points in $(\overline{J} \cup p_J)$. Thus for any point in $\Lambda_{(\overline{J} \cup p_J)}$, if the coefficient of p_J is 0, then the image of this point in Λ_J is p_J. Therefore by Definitions 3.1 and 3.2 $\Lambda_{(\overline{J} \cup p_J)}$ is projective to Λ_J.

Q.E.D.

Corollary 3.1:

If $\quad J \cap K = \phi$

or $\quad J \cap K = \{d_\ell\}$ where d_ℓ is an element of V,

then Λ_J and Λ_K are projective to each other.

A point p in Λ_V can be expressed as a convex combination of its images in the faces of Λ_V. If

$$K_i \subseteq V, \text{ for } i \in L \qquad \text{where } \bigcup_{i \in L} K_i = V$$

and $K_i \cap K_j = \phi$, for $i \neq j$, $i, j \in L$

Let p_{K_i} be the image of p in Λ_{K_i},

then $$p = \sum_{i \in L} a_i \, p_{K_i}, \qquad 0 \leq a_i \leq 1, \qquad \sum_{i \in L} a_i = 1,$$

and this expression is unique.

Lemma 3.3:

If a point p in Λ_V has an image p_J in Λ_J, then p is contained in $\Lambda_{(\overline{J} \cup p_J)}$.

Proof:

p can be expressed as a convex combination of p_J and $p_{\bar{J}}$, where $p_{\bar{J}}$ is the image of p in $\Lambda_{\bar{J}}$. But $p_{\bar{J}}$ can be expressed as a convex combination of elements of $\bar{J}$. Thus p can be expressed as a convex combination of elements $(\bar{J} \cup p_J)$. By Theorem 3.1, p is in $\Lambda_{(\bar{J} \cup p_J)}$. Q.E.D.

Lemma 3.4:

If a hyperplane H in Λ_V is projective to Λ_J and to Λ_K, then H is projective to $\Lambda_{(J \cap K)}$ if $J \cap K \neq \phi$.

Proof:

This is obvious since if H is projective to Λ_J, H will be projective to all faces of Λ_J.

Lemma 3.5:

If a hyperplane H is Λ_V is projective to Λ_J and to Λ_K, and the images of H in Λ_J and Λ_K are not θ, then H is projective to $\Lambda_{(J \cup K)}$, if $J \cap K \neq \phi$ and the image of H in $\Lambda_{(J \cap K)}$ is not θ.

Proof:

Let the image of H in $\Lambda_{(J \cup K)}$ be $H_{(J \cup K)}$, then $H_{(j \cup K)}$ is projective to Λ_J and to Λ_K. Every point of $H_{(J \cup K)}$ can be expressed as a convex combination of the vertices of $(J \cup K)$. Since Λ_V is determined by n points and those n points are assumed before not to be contained in an $(n-2)$-dimensional hyperplane, the expression of points of $H_{(J \cup K)}$ in terms of vertices of $(J \cup K)$ is unique. Because $H_{(J \cup K)}$ is projective to Λ_J, every point of

$H_{(J \cup K)}$ will have image θ, or a certain point, say p_J, in Λ_J. Similarly, every point of $H_{(J \cup K)}$ will have image θ, or a certain p_K in Λ_K. By assumption, $H_{(J \cup K)}$ is not θ, p_{J_k}, the image of p_J in $\Lambda_{(J \cap K)}$, is not θ. Similarly, p_{K_j}, the image of p_K in $\Lambda_{(J \cap K)}$, is not θ. Since $H_{(J \cup K)}$ is projective to Λ_J and to Λ_K, by Lemma 3.4, $H_{(J \cup K)}$ is projective to $\Lambda_{(J \cap K)}$. Thus $p_{J_k} = p_{K_j}$.

Now for those points in $H_{(J \cup K)}$ which have image θ in Λ_J (Λ_K), their images in $\Lambda_K(\Lambda_J)$ must be θ also, for otherwise let q be any of these points, if q has image q_K (q_J) $\neq$ θ in $\Lambda_K(\Lambda_J)$, since q has image θ in Λ_J (Λ_K), q_K (q_J) will have image θ in $\Lambda_{(J \cap K)}$, thus q_K (q_J) is distinct to p_K (p_J) which has image p_{K_j} (p_{J_k}) in $\Lambda_{(J \cap K)}$. This is contradictory to the assumption that H is projective to Λ_K (Λ_J).

Thus those pcints in $H_{(J \cup K)}$ which have image θ in Λ_J have image θ in Λ_K and those which have image p_J in Λ_J have image p_K in Λ_K. The first set of points contains only θ while the second set of points contains only one element because all points in this set have the same images p_J and p_K in Λ_J and Λ_K respectively which implies that they have the same images in every faces of $\Lambda_{(J \cup K)}$ and are thus identical. Since $H_{(J \cup K)}$ contains only one point, H is projective to $\Lambda_{(J \cup K)}$ by Definitions 3.1 and 3.2.

Q. E. D.

<u>Definition 3.3:</u>

If $\quad p$ in Λ_V,

$$p = \alpha p_J + \sum_i \beta_i p_{K_i}$$

$$J \cap \bigcup_i K_i = \emptyset$$

then α is called the weight of p with respect to Λ_J.

Definition 3.4:

A hyperplane H in Λ_V is parallel to Λ_J if all its points have the same weight with respect to Λ_J.

Following this definition, some properties about points in Λ_V are immediately obtained which are given as Lemmas 3.6, 3.7, and 3.8 without proof.

Lemma 3.6:

Let Λ_J, Λ_K as defined before, then Λ_J and Λ_K are parallel to each other if and only if $J \cap K = \phi$.

As a matter of fact, Λ_J and Λ_K in Lemma 3.6 are parallel and projective to each other at the same time.

Lemma 3.7 :

If a hyperplane H in Λ_V is parallel to Λ_J and to Λ_K, then H is parallel to $\Lambda_{(J \cup K)}$ if $J \cap K = \phi$.

Lemma 3.8 :

If a hyperplane H in Λ_V is parallel to Λ_J and to Λ_K, then H is parallel to $\Lambda_{(J \cup K)}$ if $J \cap K \neq \phi$ and H is parallel to $\Lambda_{(J \cap K)}$.

3.4 Dimension of a hyperplane in a simplex

In this section, we are going to formulate some results about the dimension of a hyperplane which possesses the properties of parallelism or projectiveness.

Lemma 3.9 :

If a hyperplane H in Λ_V is projective to Λ_J, then the dimension of J is at most (n-j).

Proof:

Suppose a point p in H has image p_J in Λ_J, then by Definitions 3.1 and 3.2 and Lemma 3.3, H is contained in $\Lambda_{(\overline{J} \cup p_J)}$ and has dimension at most (n-j). Q.E.D.

Lemma 3.10:

Let a hyperplane H in Λ_V be projective to Λ_J and to Λ_K.

(1) If $J \cap K = \phi$

then the dimension of H is at most (n-j - k + 1)

(2) If $J \cap K \neq \phi$, but $H_{(J \cap K)} \neq \{\theta\}$,
then the dimension of H is at most $[n - \#(J \cup K)]$,
where $\#(J \cup K)$ denotes the number of elements of $(J \cup K)$.

Proof:

The truth of Part (1) is due to Lemma 3.9, and that of Part (2) is due to Lemmas 3.5 and 3.9.

Corollary 3.2:

If a hyperplane H in Λ_V is projective to Λ_J and to $\Lambda_{\overline{J}}$, then the dimension of H is at most one.

Lemma 3.11:

If a hyperplane H in Λ_V is parallel to Λ_J, then the dimension of H is at most (n-2).

Proof:

Since every point of H has the same weight with respect to Λ_J, there exists a linear equation among the components of any point of H other than the equation that all components of a point in Λ_V sum up to unity. This new equation indeed indicates a hyperplane in Λ_V with dimension (n-2). H being contained in this hyperplane so its dimension is at most equal to (n-2). Q.E.D.

Lemma 3.12 :

If a hyperplane H in Λ_V is parallel to Λ_J and to Λ_K, where $J \neq K$, $J \cup K \neq V$, then the dimension of H is at most (n-3).

Proof:

Since $J \neq K$ and H is parallel to Λ_J and to Λ_K. by Lemma 3.11, H must be contained in the intersection of two (n-2) -dimensional hyperplanes are distinct. Therefore the dimension of H is at most (n-3).

Proof:

Since $J \neq K$ and H is parallel to Λ_J and to Λ_K, by Lemma 3.11 H must be contained in the intersection of two (n-2) -dimensional hyperplanes. And since $J \cup K \neq V$, these two (n-2) -dimensional hyperplanes are distinct. Therefore the dimension of H is at most (n-3).

Q.E.D.

Corollary 3.3:

Lemma 3.12 if $J \cap K \neq \phi$ and H is also parallel to $\Lambda_{(J \cap K)}$, then the dimension of H is at most (n-4).

3.5 Geometric interpretation of a transition matrix.

For a finite stochastic automaton Ω defined in Defintion 2.5, corresponding to any input symbol σ_i, there is a state transition matrix A_{σ_i}. If the number of states of Ω is n, i.e.

$\#(S) = n$, A_{σ_i} is as shown in Eq. (3.3)

$$
A_{\sigma_i} = \begin{array}{c|cccc|}
 & s'_1 & s'_2 & \cdots & s'_n \\
\hline
s_1 & a_{11} & a_{12} & \cdots & a_{1n} \\
s_2 & a_{21} & a_{22} & \cdots & a_{2n} \\
\vdots & \vdots & \vdots & & \vdots \\
s_n & a_{n1} & a_{n2} & \cdots & a_{nn} \\
\hline
\end{array}
\quad
\begin{array}{l}
0 \leq a_{ij} \leq 1, \text{ for all } i, j \\
\sum_{j=1}^{n} a_{ij} = 1 \text{ for all } i
\end{array}
\qquad (3.3)
$$

Let $w_i = (w_{i1}, w_{i2}, \ldots\ldots\ldots\ldots, w_{in})$

$w_j = (w_{j1}, w_{j2}, \ldots\ldots\ldots\ldots, w_{jn})$

be row vectors with n-components where w_{ik} is the probability that at the time σ_i is applied, the automaton Ω is in state s_k, and $w_{j\ell}$ is the probability that after σ_i is applied and the state transition is completed, the automaton Ω will be in state s_ℓ. Clearly, w_i is the present state-probability-distribution and w_j is the next state-probability-distribution. If an input symbol σ_i is applied to Ω when the state-probability-distribution is w_i, then

$$w_j = w_i \cdot A_{\sigma_i} .$$

A_{σ_i} can thus be regarded as a mapping M_{σ_i} where

$$M_{\sigma_i}(w_i) = w_j .$$

The domain of the mapping M_{σ_i} is a polyhedral convex set E_n; the extreme points of E_n are unit vectors $e_1, e_2, \ldots\ldots, e_n$. (An ith unit vector e_i is a row vector where the ith component is 1 and all other components 0.) Any w_i in E_n may be written as a convex combination of the extreme points of E_n namely,

$$w_i = w_{i1}e_1 + w_{i2}e_2 + \ldots\ldots\ldots + w_{in}e_n, \quad \sum_{k=1}^{n} w_{ik} = 1,$$

$$w_{ik} \geq 0 \quad \text{for } k=1,2,\ldots\ldots,n.$$

Let the range of the mapping M_{σ_i} be denoted as R_{σ_i}.

<u>Theorem 3.2:</u>

For a mapping M_{σ_i} characterized by the stochastic matrix A_{σ_i} as in Eq. (3.3), the range R_{σ_i} of the mapping M_{σ_i} is a polyhedral convex set, and if the domain of M_{σ_i} is E_n, then $E_n \supseteq R_{\sigma_i}$.

Proof:

From Eq. (3.3), any point w_j in R_{σ_i} is

$$w_j = w_i \cdot A_{\sigma_i} = \left(\sum_{k=1}^{n} w_{ik}\, a_{k1}, \sum_{k=1}^{n} w_{ik}a_{k2}, \ldots\ldots, \sum_{k=1}^{n} w_{ik}a_{kn}\right)$$

$$= w_{i1}\,(a_{11}, a_{12}, \ldots\ldots, a_{1n}) + w_{i2}(q_{21}, a_{22}, \ldots\ldots, a_{2n}) + \ldots$$

$$\ldots\ldots + w_{in}\,(a_{n1}, a_{n2}, \ldots\ldots, a_{nn})$$

where

$$w_i = (w_{i1}, w_{i2}, \ldots\ldots, w_{in})$$

is a point in E_n

Let $\alpha_k = w_{ik}$ for $k = 1, 2, \ldots\ldots, n$;

and let

$$a_k = (a_{k1}, a_{k2}, \ldots\ldots, a_{kn}) \qquad \text{for } k=1,2,\ldots\ldots,n; \qquad (3.4)$$

then $w_j = \alpha_1 a_1 + \alpha_2 a_2 + \ldots\ldots\ldots + \alpha_n a_n$.

Since $\sum_{k=1}^{n} a_k = \sum_{k=1}^{n} w_{ik} = 1$

and $a_k \geq 0$ for $k=1,2,\ldots\ldots, n$;

R_{σ_i} is thus a polyhedral convex set with determining points a_1, a_2, ... , and a_n.

From Eq. (3.4)

$$a_k = (a_{k1}, a_{k2}, \ldots\ldots, a_{kn})$$
$$= a_{k1}e_1 + a_{k2}e_2 + \ldots\ldots + a_{kn}e_n .$$

By Theorem 3.1, a_k belongs to E_n. Similarly, every determining point of R_{σ_i} belongs to E_n. Therefore $E_n \supseteq R_{\sigma_i}$. Q. E. D.

The following example will illustrate the relation between the domain and the range of a square matrix of order 3.

<u>Example 3.1:</u>

For a mapping M_{σ_i} characterized by the transition matrix A_{σ_i} as shown in Eq. (3.5), the domain E_3 is the triangle determined by the unit vectors e_1, e_2, and e_3; and the range of the mapping M_{σ_i} is the triangle determined by three determining points a_1, a_2, and a_3 where

$$e_1 = (1,0,0)$$
$$e_2 = (0,1,0)$$
$$e_3 = (0,0,1)$$
$$a_1 = (0.6, 0.3, 0.1)$$
$$a_2 = (0.2, 0.5, 0.3)$$
$$a_3 = (0.3, 0.3, 0.4)$$

$$A_{\sigma_i} = \begin{array}{c|ccc|} & s'_1 & s'_2 & s'_3 \\ \hline s_1 & 0.6 & 0.3 & 0.1 \\ s_2 & 0.2 & 0.5 & 0.3 \\ s_3 & 0.3 & 0.3 & 0.4 \\ \hline \end{array} \tag{3.5}$$

The fact that $E \supset A_{\sigma_i}$ is evidently seen and can be shown as in Fig.3.1.

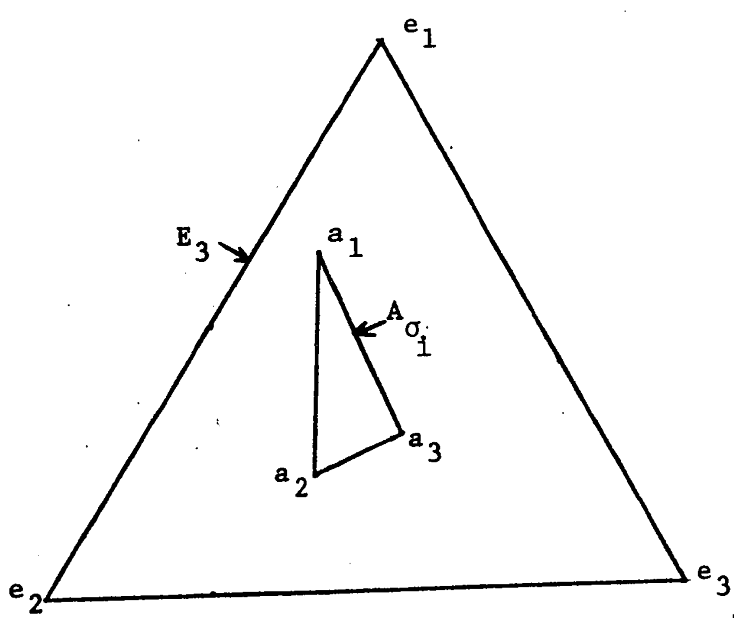

Fig.3.1 Range and domain of the mapping M_{σ_i} .

CHAPTER IV

SERIAL COMBINATION AND DECOMPOSITION OF FINITE STOCHASTIC AUTOMATA

4.1 Multiplication of stochastic matrices.

For a stochastic matrix, we mean a matrix with real nonnegative entries and with unity row sums for each row. Multiplication of stochastic matrices results in another stochastic matrix, the set of stochastic matrices is closed under the operation of multiplication. Interesting properties related to the multiplication of a certain class of square stochastic matrices have been found and applied to the ergodic theorem of Markov chains [22,35,47]. Any stochastic matrix can be regarded as a mapping with a domain and a range as mentioned in Section 3.5. In this section, we shall investigate, based on this concept, certain properties related to the ranges of the product matrix and those of its component matrices.

Let A be an $n \times \ell$ stochastic matrix and B be an $\ell \times m$ stochastic matrix. Let the mappings corresponding to A and B be denoted as M_A and M_B respectively. The domain of M_A is E_n, where $E_n = [e_1, e_2, \ldots\ldots, e_n]$. The range of M_A is R_A, $R_A = [a_1, a_2, \ldots\ldots, a_n]$, where a_i is the ith row vector of A. The domain of M_B is E_ℓ, $E_\ell = [e_1, e_2, \ldots\ldots, e_\ell]$. The range of M_B is R_B, $R_B = [b_1, b_2, \ldots\ldots, b_\ell]$, where b_i is the ith row vector of B. Let C be the product of A and B, i.e.

$$C = A \cdot B.$$

Then C is an $n \times m$ stochastic matrix. The corresponding mapping of C is denoted as M_C. The domain of M_C is E_n, while the range of M_C is R_C, $R_C = [c_1, c_2, \ldots\ldots, c_n]$, where c_i is the ith row vector of C.

Lemma 4.1:

Let A, B, C, R_A, R_B, and R_C be as defined above. If

$C = A \cdot B,$

then $R_B \supseteq R_C$.

Proof:

Let

$$A = \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \cdots\cdots & a_{1\ell} \\ a_{21} & a_{22} \cdots\cdots & a_{2\ell} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & a_{n\ell} \end{bmatrix}$$

$$B = \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ b_\ell \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} \cdots\cdots & b_{1m} \\ b_{21} & b_{22} \cdots\cdots & b_{2m} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ b_{\ell 1} & b_{\ell 2} & b_{\ell m} \end{bmatrix}$$

Then

$$C = A \cdot B = \begin{pmatrix} \sum_{i=1}^{\ell} a_{1i} b_{i1} & \sum_{i=1}^{\ell} a_{1i} b_{i2} & \cdots & \sum_{i=1}^{\ell} a_{1i} b_{im} \\ \sum_{i=1}^{\ell} a_{2i} b_{i1} & \sum_{i=1}^{\ell} a_{2i} b_{i2} & \cdots & \sum_{i=1}^{\ell} a_{2i} b_{im} \\ \vdots & \vdots & & \\ \sum_{i=1}^{\ell} a_{ni} b_{i1} & \sum_{i=1}^{\ell} a_{ni} b_{i2} & \cdots & \sum_{i=1}^{\ell} a_{ni} b_{im} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{\ell} a_{1i} \quad b_i \\ \sum_{i=1}^{\ell} a_{2i} \quad b_i \\ \vdots \\ \sum_{i=1}^{\ell} a_{ni} \quad b_i \end{pmatrix}$$

where $b_i = (b_{i1}, b_{i2}, b_{i3}, \ldots\ldots\ldots\ldots, b_{im})$.

Thus every row vector of C is a convex combination of the row vectors of B, i.e. every determining point of R_C is a convex combination of the determining points of R_B. By Theorem 3.1 all determining points of R_C are contained in R_B. Therefore

$$R_B \supseteq R_C .$$

Q. E. D.

As R_C is always contained in R_B, R_C preserves the parallelism and projectivity of R_B, i.e. if R_B is in a simplex Λ_V and is parallel (or projective) to a face Λ_J of Λ_V, then so will be R_C. On the other hand, as

$$R_C = M_B (R_A) ,$$

and as M_B is affine because B is stochastic, R_C will preserve the linearity of R_A.

<u>Lemma 4.2:</u>

Let B and C be stochastic matrices, and A be a matrix.

If $\quad A \cdot B = C,$

then A is stochastic if and only if $R_B \supseteq R_C$.

Proof:

If A is stochastic, then by Lemma 4.1, $R_B \supseteq R_C$.

On the other hand, since $A \cdot B = C$, every row of A can be regarded as the coordinates of a determining point of R_C with respect to the determining points of R_B. If $R_B \supseteq R_C$, every determining point of R_C is in R_B, and by Theorem 3.1, it can be expressed as a convex combination of the determining points of R_B. Thus A is stochastic since every row of it is stochastic.

Q. E. D.

Theorem 4.1:

Let B and C be stochastic matrices. Let R_B and R_C, the ranges of mappings B and C respectively, be in the same space, i.e. B and C have same number of columns. Let b_{Mj}, b_{mj} and c_{Mj}, c_{mj} be the maximum and minimum components of the jth column vectors of B and C respectively.

If $R_B \supseteq R_C$

$$\text{then } \left.\begin{aligned} b_{Mj} &\geq c_{Mj} \\ b_{mj} &\leq c_{mj}, \end{aligned}\right\} \text{ for all } j.$$

Proof:

Let c_i be any determining point of R_C. Since $R_B \supseteq R_C$, c_i can be expressed as a convex combination of the determining points of R_B, i.e. there exists a stochastic row vector d with number of components equal to the number of rows of B such that

$$d \cdot B = c_i$$

Let c_{ij} be the jth component of c_i and b^j be the jth column vector of B. Then we have

$$d \cdot b^j = c_{ij} .$$

Let b^j_M and b^j_m be the vectors obtained from b^j by replacing all components by b_{Mj} and b_{mj} respectively, then

$$d \cdot b^j_M = b_{Mj}$$

$$d \cdot b^j_m = b_{mj} .$$

Since all components of b^j are non-negative

$$d \cdot b^j_m \leq d \cdot b^j \leq d \cdot b^j_M$$

thus $b_{mj} \leq c_{ij} \leq b_{Mj}$.

This is true for all i.

Since c_{ij} is the ith row and j-th column entry of the stochastic matrix C, we have

$$\left. \begin{aligned} b_{mj} &\leq c_{mj} \\ b_{Mj} &\geq c_{Mj} \end{aligned} \right\} \qquad (4.1)$$

For the same reason, Eq. (4.1) is true for all j. Q. E. D.

The multiplication of matrices does not commute in general except in some special cases. We now define a special kind of symmetric matrix called totally symmetric matrix as follows:

Definition 4.1:

A totally symmetric matrix A is a symmetric matrix such that

$$a_{ij} = \alpha \quad \text{for} \quad i = j$$

$$a_{ij} = \beta \quad \text{for} \quad i \neq j$$

and $\alpha > \beta$.

Any symmetric matrix is naturally a square matrix.

The inverse of a totally symmetric matrix is still totally symmetric. Thus for any order n, the set of all totally symmetric matrices is a commutative group under the operation of multiplication and with I as the identity.

Definition 4.2:

A totally symmetric affine matrix A is a totally symmetric matrix such that $\sum_i a_{ij} = 1$ for all i.

Since the inverse of an affine matrix is affine, for any order n, the set of all totally symmetric affine matrices is a commutative group under the operation of multiplication and with I as the identity.

Definition 4.3:

A totally symmetric stochastic matrix, A is a totally symmetric affine matrix such that $0 \leq a_{ij} \leq 1$ for all i, j.

Physically, a totally symmetric stochastic matrix could be the case of a symmetric noisy communication channel.

With I as the identity, for any order n, the set of all totally symmetric stochastic matrices is a commutative monoid under the operation of multiplication .

Some interesting properties of a square stochastic matrix have been found by Hajnal and Wolfowitz [22,47] .

Definition 4.4 (Hajnal) :

A scrambling matrix is a square stochastic matrix $A = \{a_{ij}\}$ such that for any two rows a_i and a_j of A, both $a_{ik} > 0$ and $a_{jk} > 0$ at least for one k.

Lemma 4.3 (Hajnal):

If A and B are stochastic matrices and either of them is a scrambling matrix, so is $C = A \cdot B$.

A square stochastic matrix A is called indecomposable and a-periodic (SIA) if

$$C = \lim_{n \to \infty} A^n$$

exists and all rows of C are the same. Two stochastic matrices are said to be of the same type if they have zero elements and positive elements in the same places.

Lemma 4.4 (Wolfowitz):

If A is an SIA matrix and $A \cdot B$ and A are of the same type, then A is a scrambling matrix.

Lemma 4.5 (Wolfowitz):

Let $A_1, A_2, \ldots, A_k$ be square stochastic matrices of the same order.

Let ℓ be the number of different types of all SIA matrices of the same order as the A_i's, then all words in the A_i's of length $\geq \ell + 1$ are scrambling matrices.

4.2 Serial combination of finite stochastic automata.

In this section, we shall investigate cases when systems are formed by combining subsystems which are operating serially. By serial combination of two automata, we mean the case that one of the automata is operating after the other, in other words, one automaton is taking the output of the other as its input.

Assume that for any automaton, feeding other automaton has no effect on the operation of the automaton itself. Let Ω_1 and Ω_2 be two stochastic automata combined in series as shown in Fig. 4.1. The output of Ω_1 is S_1. It may be used as output of the combined system as well as the input of Ω_2. They are denoted as Z_1 and Σ_2 in Fig. 4.1, Σ_2 and Z_1 may be the same or totally different, or one might contain the other, or both might have some part the same and some part different of each other. In all cases $\Sigma_2 \neq \phi$ while Z_1 may. Besides Σ_2, there might be another input Σ_3 to the second automaton Ω_2, as shown in Fig. 4.1.

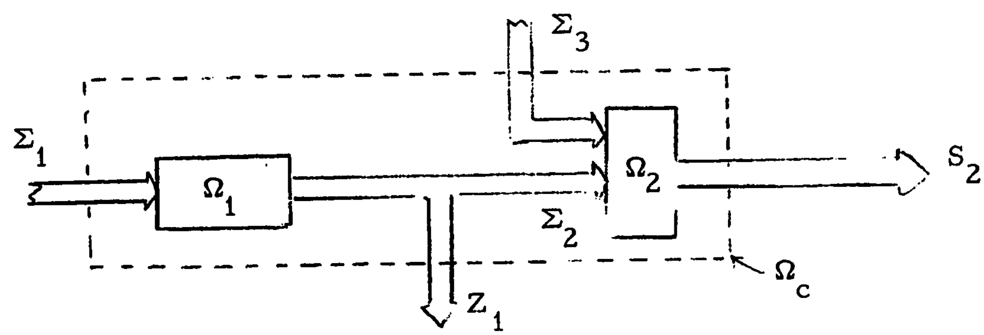

Figure 4.1 Serial combination of Ω_1 and Ω_2.

For simplicity, in the following, we assume that Z_1 and Σ_3 are empty, i.e. there are no output and input from and to the combined system between Ω_1 and Ω_2.

A. Consider different cases for different models of Ω_1 and Ω_2 which are connected as shown in Fig. 4.2. In all cases, let $S_1 = \Sigma_2$, and $s_i = \rho_i$ for all $s_i \in S_1$ and all $\rho_i \in \Sigma_2$.

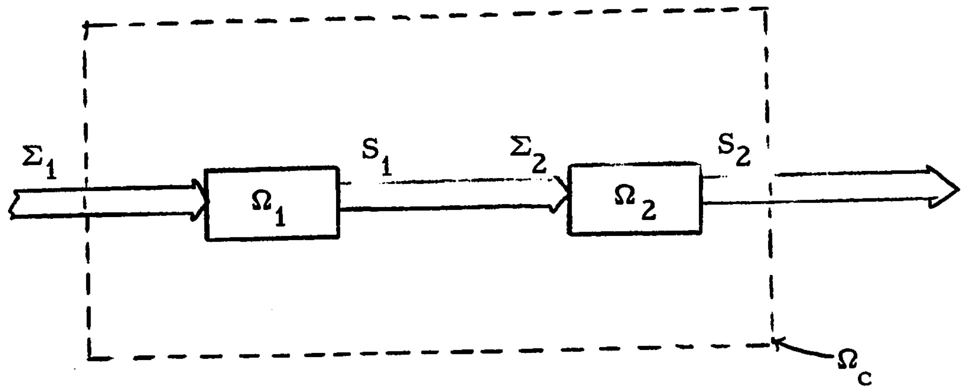

Fig. 4.2 A simple case of serial combination of Ω_1 and Ω_2.

(a) Both Ω_1 and Ω_2 are stochastic automata without memory:

Let $\Omega_1 = < \Sigma_1, S_1, W_1, M_1 >$

$$\Omega_2 = < \Sigma_2, S_2, W_2, M_2 > \qquad (4.2)$$

If $\#(\Sigma_1) = n, \quad \#(S_1) = \#(\Sigma_2) = \ell, \quad \#(S_2) = m,$

then M_1 and M_2 are characterized by stochastic matrices A and B respectively. The order of A is $n x \ell$, and that of B is $\ell x m$. The combined system Ω_c is still a stochastic automaton without memory.

$$\Omega_c = < \Sigma_1, S_2, W_2, M_c >$$

where M_c is characterized by a stochastic matrix C which is the product of A and B. As mentioned in Section 4.1, R_C will preserve the parallel and projective properties of R_B. Thus while Ω_1 is operating independently in Ω_c, the output of Ω_c retains the properties of that of Ω_2. For instance, if for Ω_2, the ratio of probabilities of occurrence over a subset of S_2 is fixed irrespective of the input, then this will be true for Ω_c. Or, if for Ω_2, the sum of the probabilities of occurrence over a subset of S_2 is fixed irrespective of the input, then this will be true for Ω_c also. In general, Ω_c operates simply as Ω_2 with a stochastic input which is generated by Ω_1.

(b) Ω_1 is a stochastic automaton while Ω_2 is a stochastic automaton without memory:

Let $\Omega_1 = < \Sigma_1, S_1, \omega_o, W_1, M_1 >$ (4. 3)

while Ω_2 be as defined in Eq. (4. 2).

Let $\#(S_1) = \#(\Sigma_2) = \ell$.

Any element A_{σ_k} of the set of stochastic matrices $\{A_{\sigma_i}\}$, $\sigma_i \in \Sigma_1$, characterizing M_1 in Eq. (4. 3) is a square matrix of order ℓ . Starting from w_o, the initial state probability distribution of Ω_1, the output probability distribution of Ω_c after an input tape

$$x = \sigma_1 \sigma_2 \cdots\cdots \sigma_n$$

is applied is

$$w_o \cdot A_{\sigma_1} \cdot A_{\sigma_2} \cdots\cdots\cdots A_{\sigma_n} \cdot B = w_o \cdot A_x \cdot B.$$

Let $c_x = w_o \cdot A_x \cdot B.$

The combined system Ω_c is thus a stochastic automaton without feedback, i. e. there is no feedback of information in the system, however, the system has infinite memory of the input.

$$\Omega_c = < \Sigma_1, S_2, W_2, M_c >$$

where Σ_1 and S_2 are as in Eq. (4. 3) and Eq. (4. 2) respectively, and M_c is such a mapping:

$$\Sigma_1^* \xrightarrow{M_c} W_2$$

$$M_c(x) = c_x \quad \text{for all } x \in \Sigma_1^* .$$

If $\#(\Sigma_1) = 1$, then Ω_1 is a Markov chain, The input Σ_2 to Ω_2 is now a Markov process generated by Ω_1. In this case, Ω_c is simply Ω_2 with a Markov process input.

(c) Ω_2 is a stochastic automaton while Ω_1 is a stochastic automaton without memory:

$$\text{Let } \Omega_1 = \langle \Sigma_1, S_1, W_1, M_1 \rangle$$
$$\Omega_2 = \langle \Sigma_2, S_2, u_o, W_2, M_2 \rangle \qquad (4.4)$$

where u_o is the initial state-probability distribution of Ω_2.

Let the stochastic matrix characterizing M_1 of Eq. (4.4) be A; and the set of square stochastic matrices characterizing M_2 of Eq. (4.4) be $\{B_{\rho_i}\}$, $\rho_i \in \Sigma_2$. Let

$$\#(\Sigma_1) = n, \quad \#(S_1) = \#(\Sigma_2) = \ell, \text{ and } \#(S_2) = m.$$

Then the order of A is $n \times \ell$ and that of B_{ρ_i} is m

Let

$$\mathcal{B} = \begin{bmatrix} B_{\rho_1} \\ B_{\rho_2} \\ \vdots \\ B_{\rho_\ell} \end{bmatrix}$$

be an ℓ-components column vector where the ith component is B_{ρ_i}, for $\rho_i \in \Sigma_2$.

$$A = \begin{bmatrix} a_{11} & a_{12} \cdots\cdots\cdots & a_{1\ell} \\ a_{21} & a_{22} \cdots\cdots\cdots & a_{2\ell} \\ & & \\ a_{n1} & a_{n2} \cdots\cdots\cdots & a_{n\ell} \end{bmatrix}$$

where the ith row vector corresponds to the transition probability of Ω_1 when σ_i is applied, for $\sigma_i \in \Sigma_1$.

Let

$$A \cdot \mathcal{B} = \begin{bmatrix} \sum_i a_{1i} \cdot B_{\rho_i} \\ \sum_i a_{2i} \cdot B_{\rho_i} \\ \vdots \\ \sum a_{ni} \cdot B_{\rho_i} \end{bmatrix} = \begin{bmatrix} B_{\sigma_1} \\ B_{\sigma_2} \\ \vdots \\ B_{\sigma_n} \end{bmatrix}$$

where

$B_{\sigma_j} = \sum_i a_{ji} \cdot B_{\rho i}$

for all j.

Thus the combined system Ω_c is a stochastic automaton.

$$\Omega_c = < \Sigma_1, S_2, u_o, W_2, M_c >$$

where Σ_1, S_2, u_o, W_2 are as in Eq. (4.4). M_c is a set of mappings characterized by a set of stochastic matrices $\{B_{\sigma_i}\}$, $\sigma_i \in \Sigma_1$. Since Σ_1 is finite, $\{B_{\sigma_i}\}$ is finite. Star ting from u_o, the state probability distribution of Ω_c after an input tape

$$x = \sigma_1 \sigma_2 \cdots\cdots \sigma_n$$

is applied

$$u_o \cdot B_{\sigma_1} \cdot B_{\sigma_2} \cdots\cdots B_{\sigma_n} = u_o \cdot B_x .$$

(d) Both Ω_1 and Ω_2 are stochastic automata:

Let
$$\Omega_1 = < \Sigma_1, S_1, w_o, W_1, M_1 >$$
$$\Omega_2 = < \Sigma_2, S_2, u_o, W_2, M_2 > \qquad (4.5)$$

Then Ω_c is a stochastic automaton with infinite memory of the input.

Let $\#(S_1) = \#(\Sigma_2) = \ell, \#(\Sigma_1) = n,$ and $\#(S_2) = m$.

M_1 in Eq. (4.5) is thus characterized by $\{A_{\sigma_i}\}$, $\sigma_i \in \Sigma_1$, which is a set of square stochastic matrices of order ℓ; and M_2 is characterized by $\{B_{\rho_i}\}$, $\rho_i \in \Sigma_2$, which is a set of square stochastic matrices of order m. Starting from w_o, for any input tape x to Ω_1, the state-probability-distribution of Ω_1 after x is applied is $w_o \cdot A_x$. The state transition matrix of Ω_2 after a tape x is applied to Ω_1 will be B_x,

$$B_x = w_o \cdot A_x \cdot \begin{bmatrix} B_{\rho_1} \\ B_{\rho_2} \\ \vdots \\ B_{\rho_\ell} \end{bmatrix} = w_o \cdot A_x \cdot \mathcal{B} .$$

where $\mathcal{B}$ is an ℓ-components column vector with B_{ρ_j} as its jth component, for $B_{\rho_j} \in \{B_{\rho_i}\}$.

Therefore

$$\Omega_c = <\Sigma_1, S_2, u_o, W_2, M_c >$$

where Σ_1, S_2, u_o, and W_2 are as in Eq. (4.5) and M_c is characterized by the set of square stochastic matrices $\{B_x\}$, $x \in \Sigma_1^*$. As Σ_1^* is infinite, $\{B_x\}$ is infinite. Let

u_x denote the state-probability-distribution of Ω_c starting from u_o, and after an input tape x is applied. If

$$x = \sigma_1 \sigma_2,$$

then $$u_x = u_o \cdot B_{\sigma_1} \cdot B_{\sigma_1 \sigma_2}.$$

In general,

$$u_{x\sigma_i} = u_x \cdot B_{x\sigma_i}, \quad \text{for all } x \in \Sigma_1^* \text{ and all } \sigma_i \in \Sigma_1 .$$

B. Now consider Ω_1 and Ω_2 feeding each other as shown in Fig. 4.3

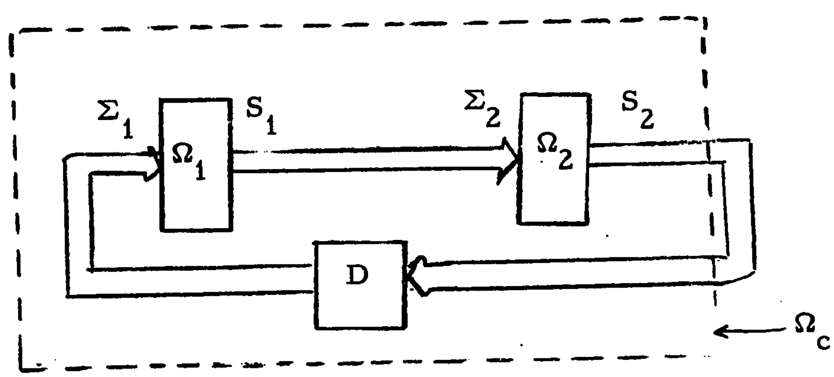

Figure 4.3 Serial combination of Ω_1 and Ω_2 feeding each other.

We shall assume that

$$S_1 = \Sigma_2, \quad \#(S_1) = \#(\Sigma_2) = m,$$

$$s_i = \rho_i, \quad \text{for all } s_i \in S_1 \quad \text{and all } \rho_i \in \Sigma_2;$$

and that $$S_2 = \Sigma_1, \quad \#(S_2) = \#(\Sigma_1) = n,$$

$$s_j = \sigma_j \quad \text{for all } s_j \in S_2 \quad \text{and all } \sigma_j \in \Sigma_1 .$$

In Fig. 4.3, S_1 is directly feeding Ω_2, but S_2 is feeding back to Ω_1 through some sort of delay such that it would allow the transformation of information in Ω_1 and Ω_2 to be completed before any changing of Σ_1 can take place. Parallel to Part (A), we discuss different cases for different models of Ω_1 and Ω_2 as follows.

(a') Both Ω_1 and Ω_2 are finite stochastic automata without memory:

Let
$$\left.\begin{aligned} \Omega_1 &= < \Sigma_1, S_1, W_1, M_1 > \\ \Omega_2 &= < \Sigma_2, S_2, W_2, M_2 > \end{aligned}\right\} \quad (4.6)$$

where $\Sigma_1 \xrightarrow{M_1} W_1$ and $\Sigma_2 \xrightarrow{M_2} W_2$.

M_1 is thus characterized by a stochastic matrix A of order nxm and M_2 is characterized by a stochastic matrix B of order mxn. The combined system Ω_c is a Markov chain with transition matrix C of order n, where

$$C = A \cdot B .$$

(b') Ω_1 is a finite stochastic automaton and Ω_2 is a finite stochastic automaton without memory:

Let

Ω_2 be as defined in Eq. (4.6) and

$$\Omega_1 = < \Sigma_1, S_1, w_o, W_1, M_1 > \quad (4.7)$$

where M_1 is characterized by a set of square stochastic matrices $\{A_{\sigma_i}\}$, $\sigma_i \in \Sigma_1$. The order of each matrix of $\{A_{\sigma_i}\}$ is m. Let the state-probability-distribution of Ω_1 at time t be w(t), $w(t) \in W_1$; and that of Ω_2 at time t be u(t), $u(t) \in W_2$.

Let

$$\mathcal{A} = \begin{bmatrix} A_{\sigma_i} \\ A_{\sigma_i} \\ \vdots \\ A_{\sigma_n} \end{bmatrix}$$

The state transition matrix of Ω_1 at time t will be A(t),

$$A(t) = u(t) \cdot \mathcal{A} . \tag{4.8}$$

After a unit time of operation, the next state-probability-distribution of Ω_1 will be w(t+1),

$$w(t+1) = w(t) \cdot A(t). \tag{4.9}$$

The next state-probability-distribution of Ω_2 will be u(t+1),

$$u(t+1) = w(t+1) \cdot B$$

From Eq. (4.8) and Eq. (4.9), we have

$$u(t+1) = w(t) \cdot u(t) \cdot \mathcal{A} \cdot B$$

$$= u(t) \cdot \begin{bmatrix} w(t) \cdot A_{\sigma_1} \\ w(t) \cdot A_{\sigma_2} \\ \vdots \\ w(t) . A_{\sigma_n} \end{bmatrix} . B = u(t) \cdot C(t). \quad \text{where} \quad C(t) = \begin{bmatrix} w(t) \cdot A_{\sigma_1} \\ w(t) \cdot A_{\sigma_2} \\ \vdots \\ w(t) \cdot A_{\sigma_n} \end{bmatrix} \cdot B$$

is a square stochastic matrix of order n.

Thus the combined system Ω_c is a time-varying Markov chain, i.e. the transition matrix is a function of time.

$$\Omega_c = < S_2, u(o), W_2, M_c (t) >$$

where S_2, and W_2 is as in Eq. (4.6), u(o) is the initial probability distribution over S_2, $u(o) \in W_2$, and $M_c(t)$ is a time-varying function which is characterized by a time-varying square stochastic matrix C(t). At time t, the next state-probability-distribution of Ω_c is given by

$$u(t+1) = u(t) \cdot C(t).$$

(c') Ω_1 is a finite stochastic automaton without memory while Ω_2 is a finite stochastic automaton:

Let Ω_1 as in Eq. (4.6) and

$$\Omega_2 = < \Sigma_2, S_2, u_o, W_2 \; M_2 > \qquad (4.10)$$

where u, is the initial state-probability-distribution, $u_o \in W_2$, and M_2 is the mapping characterized by a set of square stochastic matrices $\{B_{\rho_i}\}$, $\rho_i \in \Sigma_2$.

$$(\Sigma_2 \times W_2) \xrightarrow{M_2} W_2 .$$

Let $\mathcal{B}$ be a column vector whose ith component is $\mathbf{B}_{\rho_i}$, for $\rho_i \in \Sigma_2$, $i = 1, 2, \ldots\ldots, m$. As the mapping M_1 of Ω_1 is characterized by the stochastic matrix A of order nxm. Thus $A \cdot \mathcal{B}$ is an n-component column vector whose ith component is B_{σ_i} where

$$B_{\sigma_i} = a_i \cdot \mathcal{B}$$

and a_i is the ith row vector of A corresponding to the transition probabilities when σ_i appears at the input lead of Ω_1.

Let the state-probability-distribution of Ω_1 at time t be w(t), and that of Ω_2 be u(t). At time t, the state transition matrix of Ω_2 is

$$B(t) = w(t) \cdot \mathcal{B}$$

The next state-probability-distribution of Ω_2 will be

$$u(t+1) = u(t) \cdot B(t) = u(t) \cdot w(t) \cdot \mathcal{B} \ .$$

But $w(t) = u(t) \cdot A$.

Thus $u(t+1) = u(t) \cdot u(t) \cdot A \cdot \mathcal{B}$

where $A \cdot \mathcal{B}$ is independent of time.

Let $C(t) = u(t) \cdot A \cdot \mathcal{B}$.

Then we have

$$u(t+1) = u(t) \cdot C(t)$$

where $C(t)$ is a time-varying square stochastic matrix of order n. The combined system Ω_c is thus a time-varying Markov chain.

$$\Omega_c = < S_2, u(o), W_2, M_c(t) >$$

where $u(o)$ is the initial state-probability-distribution of Ω_c, and $M_c(t)$ is characterized by $C(t)$.

(d') Both Ω_1 and Ω_2 are finite stochastic automata:

Let Ω_1 be as defined in Eq. (4.7) and Ω_2 be as defined in Eq. (4.10). Since both Ω_1 and Ω_2 are stochastic automata, operation will take place only after a change of input occurs.

Let

$$\mathcal{A} = \begin{bmatrix} A_{\sigma_1} \\ A_{\sigma_2} \\ \vdots \\ A_{\sigma_n} \end{bmatrix} \qquad \sigma_i \in \Sigma_1$$

$$\mathcal{B} = \begin{bmatrix} B_{\rho_1} \\ B_{\rho_2} \\ \vdots \\ B_{\rho_m} \end{bmatrix} \qquad \rho_i \in \Sigma_2$$

Let the state-probability-distributions of Ω_1 and Ω_2 at time t be $w(t)$ and $u(t)$ respectively. The state transition matrix of Ω_1 at time t is $A(t)$,

$$A(t) = u(t) \cdot \mathcal{A} .$$

The state transition matrix of Ω_2 will be $B(t)$,

$$B(t) = w(t) \cdot A(t) \cdot \mathcal{B}$$

Let $\quad C(t) = B(t).$

Then $\quad u(t+1) = u(t) \cdot B(t) = u(t) \cdot C(t).$

Thus the combined system Ω_c is a time varying Markov chain.

$$\Omega_1 = < S_2, u(o), W_2, M_c(t) >$$

where $M_c(t)$ is characterized by a time-varying square stochastic matrix $C(t)$ which is of order n.

4.3 Serial decomposition of stochastic matrix

Any stochastic matrix can be decomposed into two stochastic matrices combined by multiplication.

Example 4.2:

Let it be required to find two stochastic matrices A and B such that

$$A \cdot B = C,$$

where C is given as in Eq. (4.11)

$$C = \begin{bmatrix} \frac{2}{3} & \frac{1}{9} & \frac{2}{9} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{6} & \frac{1}{3} \end{bmatrix} \tag{4.11}$$

We notice that the domain of M_c is

$$E_3 = [e_1, e_2, e_3]$$

and the range of M_c is

$$R_c = [c_1, c_2, c_3]$$

where
$$c_1 = (2/3 \quad 1/9 \quad 2/9)$$
$$c_2 = (1/4 \quad 1/2 \quad 1/4)$$
$$c_3 = (1/2 \quad 1/6 \quad 1/3)\ .$$

To find R_B, note that by Theorem 3. 2 and Lemma 4. 2,

$$E_3 \supseteq R_B \supseteq R_c\ .$$

Thus R_B must contain R_c in order that A be stochastic. Suppose we choose

$$R_B = [b_1,\ b_2,\ b_3]$$
$$b_1 = (5/6 \quad 1/18 \quad 2/18)$$
$$b_2 = (1/4 \quad 1/2 \quad 1/4)$$
$$b_3 = (1/4 \quad 0 \quad 3/4)\ .$$

The transition matrix A can be obtained by finding out the bary-centric coordinates of c_1, c_2, and c_3 in R_B. Since the order of C is 3, these coordinates of c_1, c_2, and c_3 with respect to vertices b_1, b_2, and b_3 can be found geometrically to be

$$a_1 = (\ 5/7 \quad 1/7 \quad 1/7\ )$$

$$a_2 = (\ 0 \quad 1 \quad 0\ )$$

$$a_3 = (\ 3/7 \quad 2/7 \quad 2/7\ ).$$

Thus

$$A = \begin{bmatrix} \frac{5}{7} & \frac{1}{7} & \frac{1}{7} \\ 0 & 1 & 0 \\ \frac{3}{7} & \frac{2}{7} & \frac{2}{7} \end{bmatrix} \qquad B = \begin{bmatrix} \frac{5}{6} & \frac{1}{18} & \frac{1}{9} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & 0 & \frac{3}{4} \end{bmatrix}$$

and $A \cdot B = C.$

Given a stochastic matrix C of order nxm, there are infinite ways of decomposing it into two, combined by multiplication. The method of finding A and B such that

$$A \cdot B = C$$

is to choose an R_B containing R_C. This will guarantee that the matrix A is stochastic. For B to be stochastic, R_B must be contained in E_m. The lower bound of the number of determining points of R_B depends on the dimension of R_C. If all the determining points of R_C are linearly independent, then the lower bound of the number of determining points of R_B is the number of determining points of R_C. In general, if the dimension of R_C is ℓ, then the bound is $\ell + 1$. For instance, if the dimension of R_C is 1 then each of the n determining points of R_C can be expressed as a convex combination of two extreme points in m-dimensional space. Thus C can be decomposed into two stochastic matrices A and B, where A is of order nx2 and B is of order 2xm, and

$$A \cdot B = C.$$

If the dimension of R_C is less than m-1, then it is possible to choose R_B such that it is parallel or projective to some faces of E_m, or we might choose R_A such that it is parallel or projective to some faces of E_ℓ, where ℓ is the number of determining points of R_B. However, if the dimension of R_c is equal to m-1, then it is impossible to have any R_B which will be projective or parallel to any face of E_m, or any R_A which will be parallel or projective to any face of E_ℓ . (Excluding the trivial case that any set of points in E_m is projective to any vertex of E_m.)

The number of determining points of R_B has no upper bound. So far as the condition

$$E_m \supseteq R_B \supseteq R_c$$

is fulfilled, the choice of R_B is arbitrary. The assurance that R_B contains R_C can not be obtained geometrically for high dimensional space. The method used in Example 4.2 is valid up to 3-dimensional space. When R_C is in a space of dimension higher than 3, it is impossible to draw a R_B containing R_C and measure the barycentric coordinates geometrically. Therefore, another method is needed.

Theorem 4.2:

Let $G = [d_1, d_2, \ldots\ldots\ldots, d_n]$

be a polyhedral convex set.

Let $\bar{o}_{\hat{i}} = \frac{1}{n-1} (d_1 + d_2 + \ldots\ldots + d_{i-1} + d_{i+1} + \ldots\ldots + d_n)$

$$d_i' = \alpha d_i - \beta \bar{o}_{\hat{i}}, \quad \alpha > 0, \ \beta \geq 0, \ \alpha - \beta = 1 \qquad (4.12)$$

and $G' = [d_1, d_2, \ldots\ldots, d_{i-1}, d_i', d_{i+1}, \ldots\ldots, d_n]$

Then $G' \supseteq G$.

Proof:

From Eq. (4.12)

$$d_i = \frac{1}{\alpha} d_i' + \frac{\beta}{\alpha} \bar{o}_{\hat{i}}$$

$$= \frac{\beta}{(n-1)\alpha} d_1 + \frac{\beta}{(n-1)\alpha} d_2 + \ldots\ldots + \frac{\beta}{(n-1)\alpha} d_{i-1} + \frac{1}{\alpha} d_i' + \frac{\beta}{(n-1)\alpha} d_{i+1} + \ldots\ldots + \frac{\beta}{(n-1)\alpha} d_n$$

$$= a_{i1} d_1 + a_{i2} d_2 + \ldots\ldots + a_{i1} d_i' + \ldots\ldots + a_{in} d_n \qquad (4.13)$$

and $\sum_j^n a_{ij} = \frac{\beta}{\alpha} + \frac{1}{\alpha} = \frac{\beta + 1}{\alpha} = 1, \qquad 0 \leq a_{ij} \leq 1$, for all j .

By Theorem 3.1, d_i is contained in G'. Since every determining point of G is contained in G',

$G' \supseteq G$. Q.E.D.

Theorem 4.2 provides us an easy way to construct a convex set containing a given set G. Note that in constructing G' to be containing G, $\bar{o}_{\hat{i}}$ in Eq. (4.12) can be replaced by any other single point contained in $G_{\hat{i}} = [\ , d_2, \ldots\ldots, d_{i-1}, d_{i+1}, \ldots\ldots, d_n]$. G here corresponds to R_C in Example 4.2, and G' corresponds to R_B. The matrix A is easily found to be composed of row vectors $e_1, e_2, e_3, \ldots, a_i, \ldots\ldots, e_n$ where a_i is determined by Eq. (4.13) to be

$$a_i = (a_{i1}, a_{i2}, \ldots\ldots, a_{in})$$

The above procedures of replacing a determining point of a polyhedral convex set G can be repeatedly applied on other determining points so that finally we shall have matrix A where none of its determining point is a unit vector.

Example 4.3:

Given a stochastic matrix

$$C = \begin{bmatrix} 0.2 & 0.3 & 0.4 & 0.1 \\ 0.5 & 0.2 & 0.2 & 0.1 \\ 0.3 & 0.2 & 0.1 & 0.4 \\ 0.1 & 0.5 & 0.1 & 0.3 \end{bmatrix}$$

Assume it is required to find two stochastic matrices A and B such that

$$A \cdot B = C.$$

From C, we have

$$G = R_C = [c_1, c_2, c_3, c_4]$$

where $c_1 = (0.2, \; 0.3, \; 0.4, \; 0.1)$

$c_2 = (0.5, \; 0.2, \; 0.2, \; 0.1)$

$c_3 = (0.3, \; 0.2, \; 0.1, \; 0.4)$

$c_4 = (0.1, \; 0.5, \; 0.1, \; 0.3)$

Now we construct R_B :

Since $\bar{o}_{\hat{1}} = \frac{1}{3} c_2 + \frac{1}{3} c_3 + \frac{1}{3} c_4$

$= (0.3, \; 0.3, \; \frac{0.4}{3}, \; \frac{0.8}{3})$

$c'_1 = \alpha c_1 - \beta \bar{o}_{\hat{1}}$

Choose $\alpha = 10/9, \; \beta = 1/9$

Then $c'_1 = (\frac{1.7}{9}, \; 0.3, \; \frac{11.6}{27}, \; \frac{2.2}{27})$

Thus $R_B = [b_1, b_2, b_3, b_4]$

where $b_1 = c'_1$

$b_2 = c_2$

$b_3 = c_3$

$b_4 = c_4$

$$A_1 = \begin{bmatrix} 0.9 & \frac{0.1}{3} & \frac{0.1}{3} & \frac{0.1}{3} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad B_1 = \begin{bmatrix} c'_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix}$$

and $A_1 \cdot B_1 = C$.

Next we decompose B_1 :

Choosing $c'_2 = \frac{5}{4} c_2 - \frac{1}{4} \bar{o}_{\hat{2}}$

where $\bar{o}_{\hat{2}} = \frac{1}{3} c'_1 + \frac{1}{3} c_3 + \frac{1}{3} c_4$

we have $c'_2 = \left(\frac{62.2}{108}, \frac{0.5}{3}, \frac{16}{81}, \frac{9.7}{162}\right)$

$$A_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{0.2}{3} & 0.8 & \frac{0.2}{3} & \frac{0.2}{3} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad B_2 = \begin{bmatrix} c'_1 \\ c'_2 \\ c_3 \\ c_4 \end{bmatrix}$$

and $A_2 \cdot B_2 = B_1$.

Now we decompose B_2:

Choosing $c'_3 = \frac{9}{7} c_3 - \frac{2}{7} \bar{o}_{\hat{3}}$

Where $\bar{o}_{\hat{3}} = \frac{1}{3} c'_1 + \frac{1}{3} c'_2 + \frac{1}{3} c_4$

we have $c'_3 = \left(\frac{172}{567}, \frac{10.4}{63}, \frac{100.9}{1701}, \frac{803.3}{1701}\right)$

$$A_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \frac{2}{27} & \frac{2}{27} & \frac{7}{9} & \frac{2}{27} \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad B_3 = \begin{bmatrix} c'_1 \\ c'_2 \\ c'_3 \\ c_4 \end{bmatrix}$$

and $A_3 \cdot B_3 = B_2$.

Finally we decompose B_3 :

Choosing $c'_4 = \frac{9}{8} c_4 - \frac{1}{8} \bar{o}_{\hat{4}}$

Where $\bar{o}_{\hat{4}} = \frac{1}{3} (c'_1 + c'_2 + c'_3)$

We have $c'_4 = \left(\frac{1850.5}{27216}, \frac{810.7}{1512}, \frac{3425}{40824}, \frac{25468.7}{81648} \right)$

$$A_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{1}{27} & \frac{1}{27} & \frac{1}{27} & \frac{8}{9} \end{bmatrix} \qquad B_4 = \begin{bmatrix} c'_1 \\ c'_2 \\ c'_3 \\ c'_4 \end{bmatrix}$$

and $A_4 \cdot B_4 = B_3$.

Thus $C = A_1 \cdot B_1$

$= A_1 \cdot A_2 \cdot B_2$

$= A_1 \cdot A_2 \cdot A_3 \cdot B_3$

$= A_1 \cdot A_2 \cdot A_3 \cdot A_4 \cdot B_4$

$= A \cdot B$

where

$$A = A_1 \cdot A_2 \cdot A_3 \cdot A_4 = \begin{bmatrix} \frac{54946.04}{6561} & \frac{201.52}{6561} & \frac{190.72}{6561} & \frac{222.72}{6561} \\ \frac{162.4}{2187} & \frac{1766.2}{2187} & \frac{119.2}{2187} & \frac{139.2}{2187} \\ \frac{56}{729} & \frac{56}{729} & \frac{569}{729} & \frac{48}{729} \\ \frac{1}{27} & \frac{1}{27} & \frac{1}{27} & \frac{8}{9} \end{bmatrix}$$

$$B = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = B_4 = \begin{bmatrix} c'_1 \\ c'_2 \\ c'_3 \\ c'_4 \end{bmatrix}$$

A special way of decomposing a stochastic matrix would be to decompose it into two such that the first one is a totally summetric one. The following theorem affords us an immediate solution of such a decomposition.

<u>Theorem 4. 3:</u>

If $G = [d_1, d_2, \ldots\ldots, d_n]$

$\bar{o} = \frac{1}{n}(d_1 + d_2 + \ldots\ldots\ldots + d_n)$

$d'_i = \alpha d_i - \beta \bar{o}\ d > 0,\ \beta \geq 0,\ \alpha - \beta = 1,$ for all i (4. 14)

and $G' = [d'_1, d'_2, \ldots\ldots, d'_n]$

then $G' \supseteq G$

Proof: From (4. 14)

$$\bar{o}' = \frac{1}{n} \cdot \sum_{i=1}^{n} d'_i = \frac{1}{n} \cdot \sum_{i=1}^{n} \alpha d_i - \frac{1}{n}(n. \beta. \bar{o})$$

$$= \alpha\bar{o} - \beta\bar{o} = (\alpha - \beta)\bar{o} = \bar{o}$$

By (4. 14)

$$d_i = \frac{1}{\alpha} d'_i + \frac{\beta}{\alpha} \bar{o}$$

$$= \frac{1}{\alpha} d'_i + \frac{\beta}{\alpha} \bar{o}'$$

$$= \frac{1}{\alpha} d'_i + \frac{\beta}{\alpha}\left(\frac{1}{n} \cdot \sum_{i=1}^{n} d'_i\right)$$

$$= \frac{\beta}{\alpha n} d'_1 + \frac{\beta}{\alpha n} d'_2 + \ldots\ldots + \left(\frac{1}{\alpha} + \frac{\beta}{\alpha n}\right) d'_i + \ldots + \frac{\beta}{\alpha n} d'_n$$

$$= \sum_{j=1}^{n} a_{ij}\, d'_j \ .$$

and $$\sum_{j=1}^{n} a_{ij} = \frac{\beta}{an} \cdot n + \frac{1}{a} = \frac{\beta}{a} + \frac{1}{a} = \frac{\beta+1}{a} = 1$$

$$0 \leq a_{ij}. \leq 1, \quad \text{for all } j.$$

By Theoreme 3.1, d_i is contained in G'.
Since every determining point of G is contained in G',

$G' \supseteq G$ Q.E.D.

For special stochastic matrix like totally symmetric one, the decomposition is rather simple. It is readily seen that for a totally symmetric stochastic matrix C, it is always possible to cecompose C into two totally symmetric ones, combining by multiplication. i.e. $C = A \cdot B$.
where B varies from C to I and A varies from I to C.
And it is not difficult to find a totally symmetric stochastic matrix D such that $C = D^2$.

Another question about the serial decomposition of stochastic matrices is that given C and B, is it possible to find a stochastic matrix A such that

$A \cdot B = C.$

The answer to this question is negative if $R_B \subset R_C$. It is possible if and only if $R_B \supseteq R_C$.

<u>Lemma 4.6:</u>

Let B and C be non-singular stochastic matrices of the same order. If $R_B \supseteq R_C$, then there exists unique stochastic matrices A_1 and A_2 such that

$A_1 \cdot B = C$

$B \cdot A_2 = C$

while A_1 and A_2 are similar.

Proof:

In the first case, since R_B contains R_C, it is always possible to express every determining point of R_C in terms of a convex combination of the determining points of R_B. The array of coefficient of such a combination of determining point of R_C will be the required transition matrix A_1. Because B and C are non-singular, all determining point of R_B are linearly independent, and the expression of coordinates of each determining point of R_C with respect to the extreme points of R_B is unique. In the second case, given R_C and a transition matrix B, it is always possible to construct a polyhedral convex set R_{A_2} containing R_C such that the array of coordinates of every determining point of R_C with respect to the determining points of R_{A_2} is B. Since B and C are non-singular, R_{A_2} is unique. Further, as

$$A_1 \cdot B = C \tag{4.15}$$

$$B \cdot A_2 = C$$

$$A_2 = B^{-1} \cdot C = C^{-1} \cdot A_1 \cdot C$$

Let $A_1 \cdot C = C_1$

Thus $A_2 = C^{-1} \cdot C_1$. (4.16)

B and C being non-singular, they must be square matrices. As Eq. (4.15) exists, A_1, B and C all must be square matrices of the same order. Now consider

$$C^{-1} \cdot C = I.$$

Let the order of C be n. Then the range of the mapping M_I is E_n.

C^{-1} can thus be regarded as a transformation which transforms every determining point of R_C into that of E_n.

i.e. $M'_{C^{-1}}(c_i) = e_i$, for i = 1, 2, ..., n.

$M'_{C^{-1}}$ is affine as C^{-1} is affine. Thus

$$M'_{C^{-1}}(R_C) = E_n.$$

From Eq. (4.16)

$$M'_{C^{-1}}(R_{C_1}) = R_{A_2}.$$

As $A_1 \cdot C = C_1$,

from Lemma 4.1

$$R_C \supseteq R_{C_1}.$$

Therefore

$$R_{A_2} = M'_{C^{-1}}(R_{C_1}) \subseteq M_{C^{-1}}(R_C) = E_n,$$

and A_2 is stochastic.

Since $A_1 \cdot B = C$

$B \cdot A_2 = C,$

$A_1 \cdot B = B \cdot A_2$

$A_2 = B^{-1} \cdot A_1 \cdot B.$

B being non-singular by assumption, A_1 and A_2 are therefore similar.

Q. E. D.

4.4 Serial decomposition of finite stochastic automata.

As any stochastic matrix can be decomposed into two combined by multiplication. Any finite stochastic automaton without memory can be decomposed into two combined in serial. i.e. if

$$\Omega_c = < \Sigma_c, S_c, W_c, M_c >$$

is a stochastic automaton without memory, where M_c is characterized by a stochastic matrix C, then Ω_c can be decomposed into Ω_1 and Ω_2 combined in serial. Here,

$$\Omega_1 = < \Sigma_1, S_1, W_1, M_1 >$$

$$\Omega_2 = < \Sigma_2, S_2, W_2, M_2 >$$

where $\Sigma_1 = \Sigma_c$

$$\Sigma_2 = S_1$$

$$S_2 = S_c$$

$$W_2 = W_c .$$

Let the stochastic matrices characterizing M_1 and M_2 be A and B respectively. Then

$$A \cdot B = C.$$

All properties related to the decomposition of stochastic matrix discussed in the previous section is applicable to the decomposition of stochastic automaton without memory as such an automaton is totally characterized by a single stochastic matrix.

If Ω_c is a Markov chain, i.e.

$$\Omega_c = < S_c, w_o, W_c, M_c > \tag{4.17}$$

where M_c is characterized by a stochastic matrix C, then Ω_c can be decomposed into two stochastic automata without memory Ω_1 and Ω_2 as in Fig. 4.3. Let the stochastic matrices characterizing the mappings of Ω_1 and Ω_2 be A and B respectively.

Then $A \cdot B = C.$

Starting from w_o, the state-probability-distribution of Ω_c after n transitions will be

$$w_o \cdot C^n = w_o \cdot (A \cdot B)^n.$$

If C is totally symmetric, A and B can be chosen to be totally symmetric also. In this case, A and B are commutative. Then

$$C^n = (A \cdot B)^n = A^n \cdot B^n .$$

Ω_c can thus be decomposed into Ω_1 and Ω_2 combined in serial where Ω_1 is a Markov chain and Ω_2 is a time-variant stochastic automaton without memory. i.e.

$$\Omega_1 = < S_c, w_o, W_c, M_1 >$$

$$\Omega_2 = < \Sigma_2, S_c, W_c, M_2(t) >$$

where S_c, w_o, and W_c are as in Eq. (4.17), M_1 is characterized by A, and $M_2(t)$ is a time-variant function characterized by B^t. Let the input to Ω_2 at time t be denoted as $w_i(t)$ which is a probability distribution over the elements of Σ_2 while $\Sigma_2 = S_c$. Let the output of Ω_2 be denoted as $w_j(t)$ which is a state-probability-distribution over S_c.

Then $\quad w_j(t) = w_i(t) \cdot B^t$

where $\quad w_i(t) , w_j(t) \in W_c$.

If Ω_c is a stochastic automaton where the mapping M_c is characterized by a set of transition matrices $\{C_{\sigma_i}\}$. Each C_{σ_i} can be decomposed into A_{σ_i} and B combined by multiplication.

Namely $\quad C_{\sigma_i} = A_{\sigma_i} \cdot B, \quad$ for all i.

Thus $C_{\sigma_i \sigma_j} = A_{\sigma_i} \cdot B \cdot A_{\sigma_j} \cdot B.$

If every element of $\{C_{\sigma_i}\}$ is totally symmetric, then A_{σ_i}'s and B can be chosen to be totally symmetric also. Therefore Ω_c^1 can be decomposed into Ω_1 and Ω_2 combined in serial where Ω_1 is a stochastic automaton with set of transition matrices $\{A_{\sigma_i}\}$, and Ω_2 is a time-variant stochastic automaton without memory. The mapping of Ω_2 is characterized by B^t, where t is numerically equal to the length of the tape applied to Ω_c.

CHAPTER V

PARALLEL COMBINATION AND DECOMPOSITION OF FINITE STOCHASTIC AUTOMATA

5.1 Cross product of stochastic matrices.

Definition 5.1

The cross product of two matrices $A = \{a_{ij}\}$ and $B = \{b_{ij}\}$ is defined as

$$C = A \times B = \{a_{ij} \cdot B\} .$$

It is seen that if both A and B are scalars, then C is the scalar product of A and B. If A(B) is scalar while B(A) is a matrix, then C is a matrix which is B(A) multiplied by a scalar. In general, if A is an $r \times n$ matrix and B an $\ell \times m$ matrix, then C is a matrix of order $(r \cdot \ell) \times (n \cdot m)$. For illustration we introduce following example where both A and B are vectors.

Example 5.1:

If $A = \begin{bmatrix} a_{11} & a_{12} \end{bmatrix}$ $B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \end{bmatrix}$

then $C = A \times B = \begin{bmatrix} a_{11} \cdot B & a_{12} \cdot B \end{bmatrix}$

$$= \begin{bmatrix} a_{11} \cdot b_{11} & a_{11} \cdot b_{12} & a_{11} \cdot b_{13} & a_{12} \cdot b_{11} & a_{12} \cdot b_{12} & a_{12} \cdot b_{13} \end{bmatrix}$$

If $A = \begin{bmatrix} a_{11} & a_{12} \end{bmatrix}$

$$B = \begin{bmatrix} b_{11} \\ b_{12} \\ b_{13} \end{bmatrix}$$

then $C = A \times B =$

$a_{11} \cdot B$	$a_{12} \cdot B$
$a_{11} \cdot b_{11}$	$a_{12} \cdot b_{11}$
$a_{11} \cdot b_{12}$	$a_{12} \cdot b_{12}$
$a_{11} \cdot b_{13}$	$a_{12} \cdot b_{13}$

Any stochastic matrix can be regarded as a mapping. Let A and B be stochastic matrices of orders rxn and ℓxm respectively. The mapping associated with A is M_A and that with B is M_B. The range of M_A is R_A which is contained in E_n while the range of M_B is R_B which is contained in E_m. If

$A \times B = C$

C then is a matrix which is also stochastic. Associated with C, there is mapping M_C. The range of M_C which is contained in $E_{(n \cdot m)}$. The properties of parallelism and projectiveness of R_A (or R_B) will be preserved in R_C. Namely, if R_A (R_B) is projective to a j-dimensional face of E_n (E_m), then R_C will be projective to m(n) disjoint faces of $E_{(n \cdot m)}$ each with dimension j. And if $R_A(R_B)$ is parallel to a j-dimensional face of $E_n(E_m)$, then R_C will be parallel to a face of $E_{(n \cdot m)}$ where the dimension of the face is $(j+1) \cdot m(\text{or } n)-1$. By Lemma 3.10, the dimension of R_C in the first case is at most $n \cdot m - j \cdot m(\text{or } n) - 1$. And the dimension of R_C in the second case is at most $n \cdot m - 2$ by Lemma 3.11.

Moreover, if R_A is projective to a j-dimensional face of E_n and R_B is projective to a k-dimensional face of E_m, then R_C will be projective to three sets of faces of $E_{(n \cdot m)}$. The first set contains $(m-k-1)$ elements each with dimension j, the second set contains $(n-j-1)$ elements each with dimension k, and the last set contains only one element with dimension $j \cdot k+1$. All these faces of $E_{(n \cdot m)}$ are disjoint. If R_A is parallel to a j-dimensional face of E_n and R_B is parallel to a k-dimensional face of E_m, then R will be parallel to three faces of $E_{(n \cdot m)}$ each face with dimension $m.(j+1)-1$, $n \cdot (k+1)-1$, and $j \cdot k+k+j$ respectively. In case R_A (R_B) is a single point set, R_C will be projective (parallel) to m disjoint (n-1)-dimensional faces of $E_{(n \cdot m)}$, and will be parallel (projective) to n disjoint (m-1)-dimensional faces of $E_{(n \cdot m)}$.

5.2 Parallel combination of finite stochastic automata.

By parallel combination of two automata, we mean two automata operating independently as shown in Fig. 5.1

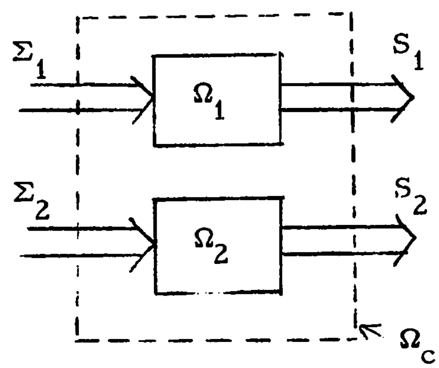

Figure 5.1 Parallel combination of Ω_1 and Ω_2.

In Fig. 5.1, the inputs Σ_1 and Σ_2 may come from the same source. They may be the same or totally different. Or they may have some parts common and some parts different to each other, or one may contain the other. In the following discussion, we assume that they are totally different, i.e. $\Sigma_1 \cap \Sigma_2 = \phi$.

To the combined system Ω_c, the input is Σ_c, the output is S_c, where

$$\Sigma_c = \Sigma_1 \times \Sigma_2 , \qquad \text{and} \qquad S_c = S_1 \times S_2 .$$

Explicitly, if

$$\Sigma_1 = \{\sigma_1, \sigma_2, \ldots\ldots, \sigma_n\}$$

$$\Sigma_2 = \{\rho_1, \rho_2, \ldots\ldots, \rho_m\}$$

then $\Sigma_c = \Sigma_1 \times \Sigma_2 = \{\psi_1 , \psi_2, \ldots\ldots, \psi_{n \cdot m}\}$

where $\psi_k = \sigma_i \ \rho_j$

and $k = \{(i-1) \cdot m\} + j.$

This index relationship is applicable to S_c also.

Now we consider different cases for different models of Ω_1 and Ω_2 connected as in Fig. 5.2.

(a) Both Ω_1 and Ω_2 are finite stochastic automata without memory:

Let $\Omega_1 = < \Sigma_1, S_1, W_1, M_1 >$

$\Omega_2 = < \Sigma_2, S_2, W_2, M_2 >$

where M_1 and M_2 are characterized by the stochastic matrices A and B respectively. The combined system Ω_c is also a stochastic automaton without memory.

$$\Omega_c = < \Sigma_c, S_c, W_c, M_c >$$

where $\Sigma_c = \Sigma_1 \times \Sigma_2$

$S_c = S_1 \times S_2 ,$

W_c is the set of probability distribution over S_c, and M_c is a mapping,

$$\Sigma_c \xrightarrow{M_c} W_c$$

which is characterized by a stochastic matrix C, and

$$C = A \times B.$$

(b) Ω_1 is a stochastic automaton while Ω_2 is a stochastic automaton without memory:

Let $\Omega_1 = < \Sigma_1, S_1, w_o, W_1, M_1 >$

$$\Omega_2 = < \Sigma_2, S_2, W_2, M_2 > \qquad (5.1)$$

where M_1 is characterized by a set of square stochastic matrices $\{A_{\sigma_i}\}$, $\sigma_i \in \Sigma_1$; and M_2 is characterized by a stochastic matrix B. Let b_j be the jth row vector of B corresponding to transition probabilities of Ω_2 when ρ_j is applied, where $\rho_j \in \Sigma_2$. Starting from w_o, if σ_i is applied to Ω_1 and ρ_j is applied to Ω_2, then the output state-probability-distribution is given by $w_o \cdot A_{\sigma_i} \times b_j$.

Thus the combined system Ω_c is a stochastic automaton which stores only a part of the information transformed by the system.

$$\Omega_c = < \Sigma_c, S_c, S_1, w_o, W_1, M_1, W_c, M_c >$$

where

$$\Sigma_c = \Sigma_1 \times \Sigma_2$$

$$S_c = S_1 \times S_2,$$

w_o, S_1, M_1 and W_1 are as in Eq. (5.1); W_c is the set of probability distributions over S_c; and M_c is a mapping,

$$(\Sigma_c \times W_c) \xrightarrow{M_c} W_c$$

which is characterized by a set of stochastic matrices $\{C_{\psi_k}\}$, $\psi_k \in \Sigma_c$.

For $\psi_k = \sigma_i \rho_j$, $\sigma_i \in \Sigma_1$, $\rho_j \in \Sigma_2$

$$C_{\psi_k} = A_{\sigma_i} \times b_j.$$

(c) Ω_2 is a finite stochastic automaton while Ω_1 is a finite stochastic automaton without memory:

Similar to (b).

(d) Both Ω_1 and Ω_2 are finite stochastic automata:

Let $\Omega_1 = < \Sigma_1, S_1, w_o, W_1, M_1 >$

$\Omega_2 = < \Sigma_2, S_2, u_o, W_2, M_2 >$

where M_1 is characterized by a set of square stochastic matrices $\{ A_{\sigma_i} \}$, $\sigma_i \in \Sigma_1$; and M_2 is characterized by a set of square stochastic matrices $\{ B_{\rho_i} \}$, $\rho_i \in \Sigma_2$.

The combined system Ω_c is a stochastic automaton.

$$\Omega_c = < \Sigma_c, S_c, v_o, W_c, M_c >$$

where $\Sigma_c = \Sigma_1 \times \Sigma_2$

$S_c = S_1 \times S_2$

$v_o = w_o \times u_o$

W_c is the set of probability distributions over S_c; and M_c is a mapping,

$$(\Sigma_c \times W_c) \xrightarrow{M_c} W_c$$

which is characterized by a set of square stochastic matrices $\{ C_{\psi_k} \}$, $\psi_k \in \Sigma_c$.

For $\psi_k = \sigma_i \rho_j$, $\sigma_i \in \Sigma_1$, $\rho_j \in \Sigma_2$

$$C_{\psi_k} = A_{\sigma_i} \times B_{\rho_j} .$$

In case when $\#(\Sigma_1) = \#(\Sigma_2) = 1$, i.e. Ω_1 and Ω_2 are finite Markov chains with transition matrices A and B respectively, then Ω_c is also a finite Markov chain with transition matrix C, and $C = A \times B$.

5.3 Parallel decomposition of a stochastic matrix

Let Λ_V be a simplex as defined in Section 3.2. Let π be a uniform parition over V. Denote each block of π as K_i. Then

$$\bigcup_{i \in L} K_i = V$$

$$K_i \cap K_j = \phi \quad , \quad \text{for } i \neq j \quad \text{where } i, j \in L.$$

Definition 5.1:

A component of a point p and a component of another point q are said to be equivalent if they are numerically equal.

Definition 5.2:

Let K_i, K_j, Λ_{K_i} and Λ_{K_j} be as defined before. A point p in Λ_{K_i} is said to be isomorphic to a point q in Λ_{K_j} if corresponding to a component of p, there exists an equivalent component of q, and vice versa.

Definition 5.3:

Let π and Λ_V be as defined before. A point p in Λ_V is said to be symmetric with respect to π if p_{K_i} is isomorphic to p_{K_j} for all $i, j \in L$.

If a point p in Λ_V is symmetric with respect to π, then p can be expressed as a convex combination of all its images p_{K_i}'s. Since all these images are isomorphic, we may put those equivalent components into an equivalent class. Thus we have a new uniform partition δ over V. δ so obtained will be called the induced partition of π. Clearly,

$$\pi \cdot \delta = 0 \quad [12]$$

and $$\#(\pi) \cdot \#(\delta) = \#(V).$$

Definition 5.4:

A point set G in Λ_V is said to be symmetric with respect to π if every point in G is symmetric with respect to π and has the same induced partition of π.

Lemma 5.1:

Let G, π, δ, be as defined above. If G is symmetric with respect to π, then G is also symmetric with respect to δ, the induced partition of π.

Proof:

Let a block of π be K_i, for $i \in L$. For any point p in G, p can be expressed as

$$p = \sum_{i \in L} \alpha_i \cdot p_{K_i}, \quad \text{where} \sum_{i \in L} \alpha_i = 1, \; \alpha_i \geq 0 \quad \text{for all } i \qquad (5.2)$$

A block J_j of δ is formed by collecting from each block K_i of π an element which has the same numerical value β_j as a component of p_{K_i}. The components of p corresponding to J_j will be

$$(\alpha_1 \cdot \beta_j, \alpha_2 \cdot \beta_j, \ldots\ldots\ldots, \alpha_r \cdot \beta_j), \text{ if } \#(L) = r. \qquad (5.3)$$

So p_{J_j}. in Λ_{J_j} is

$$(a_1, a_2, \ldots\ldots\ldots, a_r)$$

which is independent of j. Thus p is symmetric with respect to δ . Evidently, for all points of G, π is an induced partition of δ. Therefore, by Definition 5.4, G is symmetric with respect to δ. Q.E.D.

From Lemma 5.1, we know that π and δ are in fact exist in pair.

Definition 5.5:

Let Λ_V, π, p and G be as defined before. G is said to be symmetrically projective with respect to π if G is symmetric with respect to π and projective to all faces of Λ_V where each face is determined by vertices contained in each block of π .

Definition 5.6:

Let Λ_V, G, V as defined before, if π is a uniform partition of V, then G is said to be parallel with respect to π if G is parallel to Λ_{K_i} , for all $i \in L$, where K_i is a block of π, and

$$\bigcup_{i \in L} K_i = V.$$

Lemma 5.2:

Let Λ_V, G, V, and π as defined before, if G is symmetrically projective with respect~~ive~~ to π, then G is parallel with respect to δ , the induced partition of π.

Proof:

As G is symmetric with respect to π, following the proof of Lemma 5.1, for a point p in G, Eq. (5.2) and Eq. (5.3) hold.

i.e. $$p = \sum_{i \in L} \alpha_i \cdot p_{K_i} \quad , \quad \sum_{i \in L} \alpha_i = 1, \quad \alpha_i \geq 0 \text{ for all } i$$

and the components of p corresponding to a certain block J_j of δ will be

$$(\alpha_1 \cdot \beta_j, \ \alpha_2 \cdot \beta_j, \ldots\ldots\ldots\ldots, \alpha_r \cdot \beta_j)$$

where $r = \#(L)$

and β_j is a component of p_{K_i}. Similarly, for any other point $q \in G$

$$q = \sum_{i \in L} \gamma_i \cdot q_{K_i}, \quad \sum_{i \in L} \gamma_i = 1, \quad \gamma_i \geq 0 \quad \text{for all } i$$

and the components of q corresponding to the block J_j of δ will be

$$(\gamma_1 \cdot \beta'_j, \ \gamma_2 \cdot \beta'_j, \ldots\ldots\ldots\ldots, \gamma_r \cdot \beta'_j)$$

Since G is symmetrically projective with respect to π,

$$p_{K_i} = q_{K_i} \quad , \quad \text{for all } i$$

and $$\beta_j = \beta'_j .$$

The weight of p with respect to Λ_{J_j} will be equal to

$$\sum_{i=1}^{r} \alpha_i \cdot \beta_j = \beta_j \cdot \sum_{i=1}^{r} \alpha_i = \beta_j$$

and that of q will be equal to

$$\sum_{i=1}^{r} \gamma_i \cdot \beta'_j = \beta'_j \cdot \sum_{i=1}^{r} \gamma_i = \beta'_j = \beta_j .$$

Thus any point in G has the same weight with respect to Λ_{J_j}. G is therefore parallel to Λ_{J_j}. Similarly, G is parallel to faces of Λ_V where each face is determined by vertices contained in one block of δ. By Defintiin 5.6, G is parallel with respec to δ.

Q. E. D.

Now for a stochastic matrix C of order nxm, the range of the mapping M_C associated with C is R_C, which is contained in E_m.

Let $V_m = \{e_1, e_2, \ldots\ldots\ldots, e_m\}$.

Let π be a uniform partition over V_m. Then we have

Lemma 5.3:

Let C, V_m be as defined above. C can be decomposed into two stochastic matrices A and B combined in parallel if there exists a uniform partition π over V_m such that R_C is symmetrically projective with respect to π.

Proof:

Let K_i, for $i \in L$, be a block of π.

$$\bigcup_{i \in L} K_i = V_m.$$

As R_C is symmetrically projective with respect to π, R_C has the same image c_{K_i} in each face Λ_{K_i}. Any determining point c_j of R_C can be expressed as a convex combination of c'_{K_i} s, for $i \in L$.

$$c_j = \sum_{i \in L} a_{ji} \cdot c_{K_i} . \qquad (5.4)$$

Let c_{K_i} in Λ_{K_i} be denoted as b_i, for all $i \in L$. As R_C is symmetric with respect to π,

$$b_i = b_j = b \qquad \text{for all } i, j \in L.$$

Let $\#(L) = r$.

Thus c_j in Eq. (5.4) can be expressed as

$$c_j = \boxed{a_{j1} \quad a_{j2} \cdots\cdots\cdots \; a_{jr}} \; x\, b$$

where b is a row vector.

Therefore $C = \{a_{ij}\} \times b = A \times B$

where $A = \{a_{ij}\}$ is an nxr stochastic matrix and $B = b$ is an $1x(m/r)$ stochastic matrix.

Q. E. D.

If in Lemma 5. 3, there exists a uniform partition π' over the set of determining points of R_C such that in each block of π', all points of R_C are the same, then

$$C = A \times B,$$

and A is a stochastic matrix of order $\ell x r$ where ℓ is the number of blocks of π', while B is a stochastic matrix of order $(n/\ell)\, x(m/r)$ where all rows are the same.

<u>Lemma 5. 4:</u>

Let C, V_m, and π as defined in Lemma 5. 3. If there exists two uniform partitions π' and δ', $\pi' \cdot \delta' = 0$, over the set of determining points of R_C such that points in each block of δ' are symmetrically projective with respect to π while those in each block of π' are symmetrically projective with respect to δ, the induced partition of π, then C can be decomposed into two stochastic matrices A and B combined by cross product.

The truth of the above Lemma could be proved following the same line as in the proof of Theorem 5. 1 in the next section.

5.4 Parallel decomposition of finite stochastic automata.

Bacon has proved [1] that a theory similar to Hartmanis' decomposition theory [24] exists for parallel decomposition of stochastic automata. In this section, we shall approach the problem from the geometry concept which we have developed in this thesis.

Let Ω_c be a stochastic automaton without memory,

$$\Omega_c = < \Sigma_c, S_c, W_c, M_c >$$

If $\#(\Sigma_c) = n$

$$\#(S_c) = m,$$

then M_C is characterized by a stochastic matrix C of order nxm. If C can be decomposed into two stochastic matrices A and B combined by cross product, where the order of A is ℓx r and that of B is (n/ℓ)x(m/r), then Ω_c can be decomposed into two stochastic automata without memory combined in parallel. Let Ω_1 and Ω_2 be these two automata.

$$\Omega_1 = < \Sigma_1, S_1, W_1, M_1 >$$

$$\Omega_2 = < \Sigma_2, S_2, W_2, M_2 >$$

where M_1 and M_2 are characterized by A and B respectively, and

$$\#(\Sigma_1) = \ell$$

$$\#(S_1) = r$$

$$\#(\Sigma_2) = n/\ell$$

$$\#(S_2) = m/r.$$

If Ω_c is a Markov chain,

$$\Omega_c = \langle S_c, w_o, W_c, M_C \rangle$$

then M_C is characterized by a square stochastic matrix C which is the transition matrix of the Markov chain. If

$$\#(S_c) = n$$

C will be a square stochastic matrix of order n shown as Eq. (5.5).

$$C = \begin{array}{c|cccc|} & s_1' & s_2' & \cdots\cdots & s_n' \\ \hline s_1 & c_{11} & c_{12} & \cdots\cdots & c_{1n} \\ s_2 & c_{21} & c_{22} & \cdots\cdots & c_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ s_n & c_{n1} & c_{n2} & \cdots\cdots & c_{nn} \\ \hline \end{array} \qquad (5.5)$$

where $s_i \in S_c$, for i=1, 2, ..., n

$s_i' \in S'_c$, for i=1, 2, ..., n

and S'_c is the set of next states of C.

Regard each next state s_i' as a point in $\Lambda_{S'_c}$. Assume all points in S'_C are linearly independent. Then R_C, the range of the mapping M_C, is contained in $\Lambda_{S'_C}$. Let s_i represent the ith determining point of R_C. Then S_c is the set of determining points of R_C. Let π be a uniform partition over S'_c and π' be a uniform partition over S_c. π is said to be <u>equivalent</u> to π' if for $s_i, s_j \in S_c$ and $s_i', s'_j \in S'_c$, s_i' and s'_j contained in the same block of π implies that s_i and s_j are in the same block of π', and vice versa. Again let π be a uniform partition over S'. If R_C is symmetric with respect to π, then by Lemma 5.1, we have an induced partion δ. Next, we can have the equivalent partitions of π and δ over S_c. Let them be π' and δ' respectively. If δ' and π' are such that points in each block of δ' are symmetrically projective with respect to π while those in each block of π' are symmetrically projective with respect to δ, then π is said to be reflexive in C.

Theorem 5.1:

Let S'_c be the set of next states of a transition matrix C. C admits parallel decomposition if and only if there exists a uniform partition π over S'_c such that π is reflexive in C.

Proof:

Necessity:

Let $A = \{a_{ij}\}$ $B = \{b_{ij}\} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_r \end{bmatrix}$ where b_i is a r-component row vector.

Let the orders of A and B be ℓ and r respectively.

If $\ell . r = n$

and $C = A \times B$,

then

$$C = \begin{array}{c|cccc|} & s'_1 & s'_2 & \cdots & s'_n \\ \hline s_1 & a_{11}.B & a_{12}.B & \cdots & a_{1n}.B \\ s_2 & a_{21}.B & a_{22}.B & \cdots & a_{2n}.B \\ \vdots & \vdots & & & \vdots \\ s_n & a_{n1}.B & a_{n2}.B & \cdots & a_{nn}.B \\ \hline \end{array} \quad (5.6)$$

Let

$$\pi = (\overline{s'_1\ s'_2, \ldots, s'_r}\ ;\ \overline{s'_{r+1}, s'_{r+2}, \ldots\ldots, s'_{2r}}\ ; \ldots\ldots$$

$$\ldots\ldots ; \overline{s'_{(\ell-1)\cdot r+1}, s'_{(\ell-1)\cdot r+2}, \ldots\ldots, s'_{(\ell\cdot r)}}\ ) \quad (5.7)$$

Then from Eq. (5.6), we see that R_C is symmetric with respect to π as R_C has the same image R_B in each face Λ_{K_i} of Λ_{S_c}, where each face Λ_{K_i} is determined by vertices in each block of π.

The induced partition δ' of π' will be

$$\delta' = (\overline{s'_1, s'_{r+1}, \ldots\ldots, s'_{(\ell-1)\cdot r+1}}\ ;\ \overline{s'_2, s'_{r+2}, \ldots\ldots, s'_{(\ell-1)\cdot r+2}} ; \ldots$$

$$\ldots\ldots \overline{s'_r, s'_{2r}, \ldots\ldots, s'_{\ell\cdot r}}\ ).$$

Let the equivalent partitions of π and δ be π' and δ' respectively. From Eq. (5.6), we see that points in each block of δ' are symmetrically projective with respect to π because they have the same image b_j, for a certain j, in each face Λ_{K_i}. Similarly, we see that points in each block of π' are symmetrically projective with respect to δ. Thus π is reflexive in C.

Sufficiency: Assume a uniform partion π over S'_c exist while π is reflexive in C. Assume π has ℓ blocks and each block has r elements. Rearrange C so that the next states are in sequence according to π. i.e. rearrange the order of the next states so that Eq. (5.7) holds.

Now partition S_c according to π', the equivalent partition of π. Denote the subset of R_C determined by points contained in one block of π' as G_i. Clearly

$$\bigcup_{i=1}^{n} G_i = R_C$$

Since R_C is symmetric with respect to π, G_i will have the form as

$$G_i = \boxed{\begin{array}{llll}
a^i_{11}(b^i_{11}, b^i_{12}, \ldots, b^i_{1r}) & a^i_{12}(b^i_{11}, b^i_{12}, \ldots, b^i_{1r}) & \ldots & a^i_{1\ell}(b^i_{11}, b^i_{12}, \ldots, b^i_{1r}) \\
a^i_{21}(b^i_{21}, b^i_{22}, \ldots, b^i_{2r}) & a^i_{22}(b^i_{21}, b^i_{22}, \ldots, b^i_{2r}) & \ldots & a^i_{2\ell}(b^i_{21}, b^i_{22}, \ldots, b^i_{2r}) \\
\vdots & \vdots & & \vdots \\
a^i_{r1}(b^i_{r1}, b^i_{r2}, \ldots, b^i_{rr}) & a^i_{r2}(b^i_{r1}, b^i_{r2}, \ldots, b^i_{rr}) & \ldots & a^i_{r\ell}(b^i_{r1}, b^i_{r2}, \ldots, b^i_{rr})
\end{array}} \qquad (5.8)$$

As π is reflexive in C, G_i is symmetrically projective with respect to δ. Thus in Eq. (5.8)

$$a^i_{gk} = a^i_{hk} \qquad \text{for all } g,h = 1\ 2, \ldots\ldots, r$$
$$\text{and for all } k = 1, 2, \ldots\ldots, \ell.$$

Thus Eq. (5.8) can be denoted as

$$G_i = \boxed{a^i_1 \times B^i \quad a^i_2 \times B^i \ \ldots\ldots\ldots\ldots \ a^i_\ell \times B^i} \qquad (5.9)$$

where

$$B^i = \begin{bmatrix} b^i_{11} & b^i_{12} & \cdots & b^i_{1r} \\ b^i_{21} & b^i_{22} & \cdots & b^i_{2r} \\ \vdots & \vdots & & \vdots \\ b^i_{r1} & b^i_{r2} & \cdots & b^i_{rr} \end{bmatrix}$$

Now partition S_c according to δ', the equivalent partion of δ. Denote the subset of R_C determined by points contained in one block of δ' as G'_j. Then G'_j will have the form

$$G'_j = \begin{bmatrix} a^1_{j1}(b^1_{j1}, b^1_{j2}, \ldots, b^1_{jr}) & a^1_{j2}(b^1_{j1}, b^1_{j2}, \ldots, b^1_{jr}) \ldots\ldots & a^1_{j\ell}(b^1_{j1}, b^1_{j2}, \ldots, b^1_{jr}) \\ a^2_{j1}(b^2_{j1}, b^2_{j2}, \ldots, b^2_{jr}) & a^2_{j2}(b^2_{j1}, b^2_{j2}, \ldots, b^2_{jr}) \ldots\ldots & a^2_{j\ell}(b^2_{j1}, b^2_{j2}, \ldots, b^2_{jr}) \\ \vdots & \vdots & \vdots \\ a^\ell_{j1}(b^\ell_{j1}, b^\ell_{j2}, \ldots, b^\ell_{jr}) & a^\ell_{j2}(b^\ell_{j1}, b^\ell_{j2}, \ldots, b^\ell_{jr}) \ldots\ldots & a^\ell_{j\ell}(b^\ell_{j1}, b^\ell_{j2}, \ldots, b^\ell_{jr}) \end{bmatrix}$$

As π is reflexive in C, G'_j is symmetrically projective with respect to π. Thus

$$b^g_{jk} = b^h_{jk} = b_{jk} \quad \text{for all } g, h = 1, 2, \ldots\ldots, \ell \text{ and for all } k = 1, 2, \ldots\ldots, r. \tag{5.10}$$

Since Eq. (5.10) is true for all $j = 1, 2, \ldots\ldots, r$ we have

$$B^g = B^h = B \qquad \text{for all } g, h = 1, 2, \ldots\ldots, \ell .$$

Thus Eq. (5.9) becomes

$$G_i = \boxed{a_1^i \times B \quad a_2^i \times B \; \ldots\ldots\ldots\ldots \; a_\ell^i \times B}$$

and C becomes

$$C = \begin{bmatrix} a_1^1 \times B & a_2^1 \times B & \cdots & a_\ell^1 \times B \\ a_1^2 \times B & a_2^2 \times B & \cdots & a_\ell^2 \times B \\ \vdots & \vdots & & \vdots \\ a_1^\ell \times B & a_2^\ell \times B & \cdots & a_\ell^\ell \times B \end{bmatrix} = \begin{bmatrix} a_1^1 & a_2^1 \cdots & a_\ell^1 \\ a_1^2 & a_2^2 \cdots & a_\ell^2 \\ \vdots & \vdots & \vdots \\ a_1^\ell & a_2^\ell \cdots & a_\ell^\ell \end{bmatrix} \times B$$

$$= A \times B .$$

Q. E. D.

Corollary 5.1:

A Markov chain can be decomposed into two operating in parallel if its transition matrix admits a parallel decomposition.

Corollary 5.2:

Let Ω_c be a finite stochastic automaton,

$$\Omega_c = < \Sigma_c, S_c, w_o, W_c, M_c >$$

where M_C is characterized by a set of transition matrices $\{C_{\sigma_i}\}$, for $\sigma_i \in \Sigma_c$. Then Ω_c can be decomposed into two finite stochastic automata operating in parallel if there exists a partition π over S'_C such that π is reflexive in C_{σ_i} , for all $\sigma_i \in \Sigma_c$.

The conditions for parallel decomposition of a transition matrix C stated in Theorem 5.1 are quite similar to those found by Bacon [1]. The existance of another partition τ other than π such that $\pi \cdot \tau = 0$ in his case is indeed, in our case, in the property of π stated in Lemma 5.1. The conditions that π and τ are independent and both have the substitution property (S. P.) in his case are all inhere . in our case, in the property that π is reflexive in C. Thus instead of looking for partition π and τ which have the S. P. we are looking for a partition π to which R_C is symmetric with respect. We think that such a π is easier to be detected from the transition matrix C. We give the following algorithm to justify our claim.

Algorithm for detecting a partition π with properties stated in Theorem 5.1:

(a) Suppose we are given a transition matrix C with n states. Take a column vector s'_1 from C.

(b) Take another column vector s'_j, (j = 2, 3, ..., n) from C. Take the ratio s'_1 / s'_j componentwise.

(c) Find a uniform partition ξ of the components of s'_1 / s'_j such that in each block all components have the same ratio. (an 0 appearing in the ratio means optional. i. e. it may take any real number.)

If $\xi = 0$ is the only partition of such kind, go to (d).

If there exists more than one such ξ , test each one whether it is reflexive in C.

If any one succeeds, then that is the partition we are looking for.

If all fail, then go to (d).

(d) If $j = n$, then no π exists.

If $j < n$, put $j = j + 1$, go back to (b).

As can be seen, the above process greatly reduces the number of possible partitions to be tested. The reason that the partition π is imbedded in s'_1 / s'_j is that if π exists, then δ exists, then at least for one j, s'_1 and s'_j will be in one block of π or in one block of δ. If s'_1 and s'_j are in one block of π, since π is reflexive in C, the partition ξ of s'_1 / s'_j is the partition δ. If s'_1 and s'_j are in one block of δ, then ξ is π.

Algorithm for decomposing C according to a partition π which is reflexive in C.

(a) Obtain the induced partition δ of π.

(b) Partition the next states of C according to $\pi(\delta)$, and partition the present state of C according to π' (δ'), the equivalent partition of π (δ).

(c) Obtain a submatrix from C by picking any one block of π' (δ') from the present states and a block of π (δ) from the next states.

(d) Normalize each row vector of the matrix obtained in (c). This is the matrix B (A).

The cross product of the matrices A and B obtained in the above process is equivalent to C.

<u>Example 5.1:</u>

Suppose we are given a transition matrix C as shown in Eq. (5.9), and that it is required to find out two transition matrices A and B such that

$$A \times B = C.$$

$C =$

	s'_1	s'_2	s'_3	s'_4	s'_5	s'_6	s'_7	s'_8	s'_9
s_1	.02	.05	.21	.04	.03	.06	.10	.35	.14
s_2	.04	.05	.07	.08	.01	.02	.10	.35	.28
s_3	.15	.05	.18	.06	.30	.12	.02	.03	.09
s_4	.08	.20	.15	.02	.12	.03	.05	.25	.10
s_5	.03	.01	.42	.06	.06	.12	.02	.07	.21
s_6	.12	.04	.30	.03	.24	.06	.01	.05	.15
s_7	.16	.20	.05	.04	.04	.01	.05	.25	.20
s_8	.20	.25	.03	.08	.05	.02	.10	.15	.12
s_9	.10	.25	.09	.04	.15	.06	.10	.15	.06

(5.11)

Firstly, we construct s'_1 / s'_2 which is

$$s'_1 / s'_2 = \begin{bmatrix} 2/5 \\ 4/5 \\ 3/1 \\ 2/5 \\ 3/1 \\ 3/1 \\ 4/5 \\ 4/5 \\ 2/5 \end{bmatrix}$$

The partition ξ is then obtained, which is

$$\xi = (\overline{s_1, s_4, s_9} ; \overline{s_2, s_7, s_8} ; \overline{s_3, s_5, s_6}).$$

Let $\pi' = \xi$ and test π in C. We find that R_C is symmetric with respect to π, for

C =

	s'_1	s'_4	s'_9	s'_2	s'_7	s'_8	s'_3	s'_5	s'_6
s_1	.2 (.1	.2	.7)	.5 (.1	.2	.7)	.3 (.7	.1	.2)
s_2	.4 (.1	.2	.7)	.5 (.1	.2	.7)	.1 (.7	.1	.2)
s_3	.3 (.5	.2	.3)	.1 (.5	.2	.3)	.6 (.3	.5	.2)
s_4	.2 (.4	.1	.5)	.5 (.4	.1	.5)	.3 (.5	.4	.1)
s_5	.3 (.1	.2	.7)	.1 (.1	.2	.7)	.6 (.7	.1	.2)
s_6	.3 (.4	.1	.5)	.1 (.4	.1	.5)	.6 (.5	.4	.1)
s_7	.4 (.4	.1	.5)	.5 (.4	.1	.5)	.1 (.5	.4	.1)
s_8	.4 (.5	.2	.3)	.5 (.5	.2	.3)	.1 (.3	.5	.2)
s_9	.2 (.5	.2	.3)	.5 (.5	.2	.3)	.3 (.3	.5	.2)

(5.12)

The induced partition δ is

$$\delta = (\overline{s'_1, s'_2, s'_5} ; \overline{s'_4, s'_6, s'_7} ; \overline{s'_3, s'_8, s'_9}),$$

Thus

$$\delta' = (\overline{s_1, s_2, s_5} ; \overline{s_4, s_6, s_7} ; \overline{s_3, s_8, s_9})$$

$$\pi' = (\overline{s_1, s_4, s_9} ; \overline{s_2, s_7, s_8} ; \overline{s_3, s_5, s_6})$$

From Eq. (5.12), we find that points in each block of δ' are symmetrically projective with respect to π while those in each block q π' are symmetrically projective with respect to δ. Thus π is reflexive in C.

Now take the block $\overline{s'_1, s'_4, s'_9}$ from the columns of C and the block $\overline{s_1, s_4, s_9}$ from the rows of C. Normalizing each row of this submatrix of C, we obtain

B =

.1	.2	.7
.4	.1	.5
.5	.2	.3

Similarly, by taking $\overline{s'_1, s'_2 \; s'_5}$ and $\overline{s_1, s_2, s_5}$ from the columns and rows of C respectively, we obtain

A =

.2	.5	.3
.4	.5	.1
.3	.1	.6

And A x B will be equivalent to C.

CHAPTER VI

DEFINITE STOCHASTIC AUTOMATA

6.1. Definition

The notion of a definite event was first introduced by Kleene in 1956 [27]. The theory of definite automata was then developed by Rabin and Scott [40], and by Perles et al [38]. Adopting their definition, a finite sequence of symbols on a certain alphabet is called a tape, and a set of tapes is called a definite event if for some integer k, two tapes coinciding on the last k squares are either both in the set or both not in the set. Automata are used for classifying tapes. An automaton defining a definite event is called a definite event is called a definite automaton. The notion of definiteness has been applied by Paz [37] to stochastic matrices where a finite set of stochastic matrices of the same order is called a <u>definite set of matrices of order k,</u> if there exists an integer k such that for a $n \geq k$, any product of n matrices from the set is a matrix with all rows the same. Such a matrix is called a <u>stable matrix</u> by Paz and Reichaw [35].

<u>Definition 6.1</u>:

A definite stochastic automaton is a finite stochastic automaton where the set of stochastic matrices $\{A_{\sigma_i}\}$ characterizing the mapping M is a definite set of order k, for a certain integer k.

A definite stochastic finite automaton Ω of order k so defined will forget its past except for the last k intervals of time. In other words, for x with length $n \geq k$, the state-probability-distribution of the stochastic automaton will depend only on the last k inputs and will be independent of the initial state-probability-distribution. The state-probability-distributions in the first k intervals of time might be different depends on different initial state-probability-distribution when the input tape x begins to apply. However, after a length of k symbols are applied, Ω will reach a situation where the state-probability-distribution can be predicted if only the last k input symbols are known. We call such a state-probability-distribution a stable distribution. All other state-probability-distribution of Ω which are not independent of the initial state-probability-distribution are called transient distribution. Thus similar to the deterministic case where for a definite automaton of order k, a destined state is reached following a sequence of transient states of length k, for a definite stochastic automaton Ω of order k, a stable distribution will be reached through a sequence of transient distributions of length k. The number of stable distribution of Ω is at most equal to $(\#(\Sigma))^k$.

6.2 Conditions of many-to-one mapping.

If an input symbol σ_i is applied to a stochastic automaton Ω, the state transition matrix of Ω will be A_{σ_i}. As mentioned in section 2.5, any transition matrix A_{σ_i} is a mapping to the state-probability-distributions. The range of such a mapping is R_{σ_i} which is a polyhedral convex set determined by points which are the row vectors of the matrix A_{σ_i}. If a tape x is applied, the result is a successive mapping. The range of the resultant mapping is R_x, which is again a polyhedral convex set determined by points which are row vectors of the matrix A_x. If A_x is stable, then R_x is a single point set. Now the interesting problem is under what conditions that R_x will be a single point set. From the theory of Markov chain, we know that if A_{σ_i} is regular, and if

$x = \sigma_i \sigma_i \ldots\ldots \sigma_i = \sigma_i^n$, then R_x approaches to a limiting point as n approaches to infinity. Since we are interested only in the conditions that R_x is a single point set for definite length of x, we consider first the situation when a stochastic matrix is multiplied by another stochastic matrix. Suppose A, B are stochastic matrices and

$$A \,.\, B = C, \tag{6.1}$$

then C is stochastic also. From Lemma 4.1, we know that

$$R_B \supseteq R_C .$$

Thus the dimension of R_C is always not greater than that of R_B. In other words, if there exists a linear relationship among the determining points of R_B, then there exists at least one linear relationship among the determining points of R_C. Still, in Eq. (6.1), we may think of B as a mapping M_B which maps the point set R_A unto the point set R_C. i.e.

$$M_B(R_A) = R_C .$$

Let the ith row vector of A be a_i and that of C be c_i. As defined before, a_i is a determining point of R_A and c_i is a determining point of R_C. Thus if A and C have n rows.

$$M_B(a_i) = c_i \qquad \text{for } i = 1, 2, \ldots\ldots, n.$$

Since B is stochastic, M_B is affine. Therefore, if a linear relationship exists among the determining points of R_A, the same relationship will exist among the corresponding determining points of R_C. Hence, we have only to consider those linearly independent points of R_A in deciding whether or not R_A can be mapped by a mapping or a string of mappings into a single point set.

If this is possible, the upper bound of the length of the string of mappings is determined by the order of A. We quote a theorem from Paz [37] and omit the proof, as follows:

Theorem 6.1: (Paz)

If a finite stochastic automaton Ω is definite, and the number of internal states of Ω is n, then Ω is at most (n-1) - definite.

Lemma 6.1

If $\Omega = (\Sigma, S, w_o, W, M)$ is a definite stochastic automaton, where M is characterized by $\{A_{\sigma_i}\}$, $\sigma_i \in \Sigma$, then none of the matrices in $\{A_{\sigma_i}\}$ is non-singular.

Proof:

If any matrix in $\{A_{\sigma_i}\}$, say A_{σ_k}, is non-singular, by the theory of matrix, the product of non-singular matrices is non-singular, i.e. $A_{\sigma_j}^{2}$ is non-singular. Thus for any definite integer k, $A_{\sigma_j}^{k}$ is not stable. This is contradictory to the assumption that Ω is definite. Q.E.D.

Lemma 6.2:

Let B be an $\ell \times m$ stochastic matrix where all the ℓ row vectors are linearly independent. Let $\{a_i\}$ be a finite set of stochastic ℓ-component row vectors, then points which are linearly independent in the domain of M_B will be linearly independent in the range of M_B.

Proof:

The Lemma is obviously true for $\ell = 1$.

Without loss of generality, suppose now $\ell \geq 3$ and that a_1, a_2, and a_3 are linearly independent. Let

$$M_B(a_i) = c_i \qquad , \quad i = 1, 2, 3 .$$

If c_1, c_2, and c_3 are linearly dependent, then there exist

$$\alpha_1 c_1 + \alpha_2 c_2 + \alpha_3 c_3 = 0 \tag{6.2}$$

where α_1, α_2, and α_3 are real and not all equal to zero.

Let $b_i = \boxed{b_{i1} \quad b_{i2} \quad \cdots\cdots\cdots \quad b_{im}}$

be the ith row vector of B.

Let $a_i = \boxed{a_{i1} \quad a_{i2} \cdots\cdots\cdots\cdots \quad a_{i\ell}}$

$c_i = \boxed{c_{i1} \quad c_{i2} \cdots\cdots\cdots\cdots \quad c_{im}}$, for i = 1, 2, 3.

From Eq. (6.2), we have

$$\sum_{i=1}^{3} \alpha_i \cdot c_{ij} = 0 , \qquad \text{for } j = 1, 2, \ldots\ldots\ldots, m$$

Here $c_{ij} = \sum_{k=1}^{\ell} a_{ik} \cdot b_{kj}$, for $j = 1, 2, \ldots\ldots\ldots, m$

$i = 1, 2, 3$.

Thus $$\sum_{i=1}^{3} \alpha_i \cdot \sum_{k=1}^{\ell} a_{ik} \cdot b_{kj} = 0, \quad \text{for } j = 1, 2, \ldots\ldots, m;$$

$$\sum_{k=1}^{\ell} \left(\sum_{i=1}^{3} \alpha_i \cdot a_{ik} \right) \cdot b_{kj} = 0, \qquad \text{for all } j = 1, 2, \ldots\ldots\, m \tag{6.3}$$

Since all b_i's are linearly independent, Eq. (6. 3) will be true only when

$$\sum_{i=1}^{3} \alpha_i a_{ik} = 0, \qquad \text{for all } k = 1, 2, \ldots\ldots, \ell . \tag{6.4}$$

But Eq. (6. 4) then means a_1, a_2, and a_3 are linearly dependent, a contradiction to the assumption. Q. E. D.

Corollary 6. 1:

Let A be an $n \times \ell$ stochastic matrix and B and $\ell \times m$ stochastic matrix. If all the row vectors of B are linearly independent, then the rank of the product matrix (A . B) is that of A.

Lemma 6. 3:

If a square stochastic matrix B is of order n and rank n-1, then at most two points which are linearly independent in E_n will be mapped by M_B into a single point in E_n.

Proof:

Let $A = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{bmatrix}$ be a non-singular stochastic matrix

where each a_i is an n-component row vector. From matrix theory, the rank of the product matrix (A · B) is n-1. Thus no more than two rows of the matrix (A• 3) will be equal. Q. E. D.

Now back to Eq. (6.1), consider the conditions that linearly independent points of R_A will be mapped by M_B into a single point in R_C. First of all, by the argument mentioned in the proof of Lemma 6.1, the matrix B must be singular. Suppose again that B is a square stochastic matrix of order n and the rank of B is n-1. Let $b_1, b_2, \ldots\ldots, b_n$ be the row vectors of B, then there exists an equation

$$\sum_{i=1}^{n} \alpha_i \cdot b_i = 0$$

where α_i's are real and not all equal to zero. We then may have

$$b_n = \sum_{i=1}^{n-1} \beta_i \cdot b_i .$$

Since B is stochastic,

$$\sum_{i=1}^{n-1} \beta_i = 1.$$

Let

$$e_n' = \sum_{i=1}^{n-1} \beta_i \cdot e_i$$

be called the image of e_n in the hyperplane E_{n-1} determined by $e_1, e_2, \ldots\ldots, e_{n-1}$. We then have:

<u>Theorem 6.2:</u>

Let B, W_n as defined above and let $\overline{L}$ be a line segment in E_n, then all points in $\overline{L}$ will be mapped by M_B into a single point if and only if $\overline{L}$ and the line segment $\overline{e_n e_n'}$ are coplanar and parallel.

Proof:

The trivial cases are $n = 1, 2$.

For $n \geq 3$, by assumption

$$B = \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_{n-1} \\ b_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_{n-1} \\ \sum_{i=1}^{n-1} \beta_i \cdot b_i \end{bmatrix} = \begin{bmatrix} 1 & 0 \ldots\ldots & 0 \\ 0 & 1 \ldots\ldots & 0 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 0 & 0 \ldots\ldots & 1 \\ \beta_1 & \beta_2 \ldots\ldots & \beta_{n-1} \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_{n-1} \end{bmatrix} = T \cdot B'$$

where T is an $n \times (n - 1)$ matrix

B' is an $(n - 1) \times n$ matrix, and

b_i is an n-component row vector.

The mapping M_B can thus be decomposed into M_T and $M_{B'}$ as:

$$M_B = M_{B'} \cdot M_T$$

where $M_{B'}$ is a mapping which will map a point in E_{n-1} into a point in E_n while M_T is an identity mapping for points in E_{n-1} and will map a point in E_n into a point in E_{n-1}. As all the row vectors of B' are linearly independent, by Lemma 6.2, points which are linearly independent in E_{n-1} will be still linearly independent in E_n after having been mapped by $M_{B'}$. Thus in questioning whether points in E_n will be mapped by M_B into a single point in E_n, all we need to consider is whether they are mapped by M_T into a single point in E_{n-1}.

We now prove the sufficiency: Take a point p in E_{n-1}, in the plane H determined by e_n, e'_n and p, draw a line segment $\overline{L}$ parallel to $\overline{e_n e'_n}$. Let the intersection of $\overline{e_n p}$ and $\overline{L}$ be p_1 and that of $\overline{e'_n p}$ and $\overline{L}$ be p_2. Since e'_n and p are in E_{n-1}, p_2 is in E_{n-1}. Similarly, p_1 is in E_n. Being on the line $\overline{e_n p}$ p_1 can be uniquely expressed as

$$p_1 = \gamma_1 e_n + \gamma_2 p, \quad \gamma_1 + \gamma_2 = 1.$$

Similarly, p_2 can be uniquely expressed as

$$p_2 = \gamma'_1 e'_n + \gamma'_2 p, \quad \gamma'_1 + \gamma'_2 = 1.$$

Since $\overline{p_1 p_2}$ is paralled to $\overline{e_n e'_n}$

$$\gamma_1 = \gamma'_1, \quad \gamma_2 = \gamma'_2 .$$

As T is stochastic, M_T is affine. Thus

$$\begin{aligned} M_T(p_1) &= \gamma_1 M_T(e_n) + \gamma_2 M_T(p) \\ &= \gamma_1 e'_n + \gamma_2 p \\ &= p_2 . \end{aligned}$$

Still $M_T(p_2) = p_2$,

as p_2 is in E_{n-1}. p_1 and p_2 being on $\overline{L}$, thus

$$M_T(\overline{L}) = p_2 ,$$

and all points on $\overline{L}$ are mapped by M_T into a single point p_2 in E_{n-1}.

Now we prove the necessity :

Assume $M_T(\overline{L}) = p_2$
where p_2 is in E_{n-1}. In the plane H determined by e_n, e'_n, and p_2 draw a line segment $\overline{L}'$, contained in E_n, passing through p_2, and parallel to $\overline{e_n e'_n}$. From the same argument given in the proof of sufficiency,

$$M_T(\overline{L}') = p_2$$

Now $\overline{L}$ must coincide with $\overline{L}'$, for otherwise points on the plane determined by $\overline{L}$ and $\overline{L}'$ will be mapped by M_T into a single point p_2 in E_{n-1}, which implies that we shall have three linearly independent points in E_n which would be mapped by M_T into p_2. But this is impossible due to Lemma 6. 3. Q. E. D.

Corollary 6. 2:

If B is of order n and rank n-2, and

$$b_{n-1} = \sum_{i=1}^{n-2} \alpha_i \cdot b_i$$

$$b_n = \sum_{i=1}^{n-2} \beta_i \cdot b_i$$

Let
$$e'_{n-1} = \sum_{i=1}^{n-2} \alpha_i \cdot e_i$$

$$e'_n = \sum_{i=1}^{n-2} \beta_i \cdot e_i$$

Let H be a plane segment in E_n, then all points in H will be mapped by M_B into a single point in E_n if and only if H is parallel to the lines $\overline{e_n e'_n}$ and $\overline{e_{n-1} e'_{n-1}}$.

Corollary 6.3:

Let B be of order n, rank n-m, and H be a hyperplane segment in E_n, if $M_B(H)$ is a single point, then the dimension of H is at most m.

Theorem 6.3:

Let Ω be a finite stochastic automaton with n internal states. If Ω is definite and the highest rank of its transition matrix is (n-m), then Ω is k definite, where k is an integer and $k \geq \frac{n-m}{m}$.

Proof :

Obvious by Lemma 6.3 and Corollary 6.3.

6.3. Decision Procedures

By Lemma 6.1, if a stochastic automaton Ω is definite, then every element of its set of stochastic transition matrices must be singular. The most straightforward decision procedure is to form the matrix products of all possible combinations of all elements of the set $\{A_{\sigma_i}\}$ and see whether all these products are stable. If not, then repeat the same procedure after increasing the number of component matrices from two to three, and finally, to n-1, if n is the number of the internal states of Ω. This seems tedious for larger n and large number of elements of Σ. A much simpler testing procedure can be obtained by the following argument.

As shown in the proof of Theorem 6.2, any matrix A_{σ_i} can be decomposed into a matrix T_{σ_i} multiplied by another matrix A'_{σ_i} where all row vectors of A'_{σ_i} are linearly independent.

i.e. $$A_{\sigma_i} = T_{\sigma_i} \cdot A'_{\sigma_i} \qquad (6.5)$$

The operation of <u>" ' " (prime)</u> on a matrix is defined as to keep hold of those row vectors of the matrix which are linearly independent and to strike out those row vectors which are linearly dependent on the above linearly independent vectors. In the case that a matrix is the product of two or more component matrices, the operation of " ' " is to prime the first component matrix whenever it has not been primed.

i.e. if $A_x = A_{\sigma_1} \cdot A_{\sigma_2} \cdots\cdots A_{\sigma_n}$,

then
$$
\begin{aligned}
A'_x &= (A_{\sigma_1} \cdot A_{\sigma_2} \cdots\cdots A_{\sigma_n})' \\
&= (A'_{\sigma_1} \cdot A_{\sigma_2} \cdots\cdots A_{\sigma_n})' \\
&= (A'_{\sigma_1 \sigma_2} \cdots\cdots\cdots A_{\sigma_n})' \\
&\vdots \\
&= (A'_{\sigma_1 \cdots\cdots \sigma_{n-1}} \cdot A_{\sigma_n})' \\
&= A'_{\sigma_1 \sigma_2 \cdots\cdots \sigma_n}
\end{aligned}
$$

The T -matrix in Eq. (6.5) is defined as a transformation which will transform any matrix A'_{σ_i} back to A_{σ_i} . In general

$$A_x = T_x \cdot A'_x \ , \quad \text{for all } x \in \Sigma\, \Sigma^* .$$

Thus if A_x is of order n and rank m, A'_x will be an m x n matrix where all the m row vectors are linearly independent, and T_x will be an n x m matrix. The ith row of T_x is an ith unit vector if the ith row vector in A_x is taken as one of the linearly independent vectors of A_x. If in A_x, the kth row is taken as linearly dependent vector which can be expressed as a linear combination of other vectors which are linearly independent, then the kth row of the matrix T_x is an m-component row vector where the value in the jth component is the coefficient of the jth linearly independent vector of A_x in the expression of the kth row vector. For instance, if A_x is singular such that rows $a_1, a_2, \ldots, a_{n-1}$ are linearly independent, and

$$a_n = \sum_{i=1}^{n-1} \beta_i \cdot a_i$$

then

$$T_x = \begin{bmatrix} 1 & 0 \ldots\ldots & 0 \\ 0 & 1 \ldots\ldots & 0 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 0 & 0 \ldots\ldots & 1 \\ \beta_1 & \beta_2 \ldots\ldots & \beta_{n-1} \end{bmatrix}$$

and

$$A'_x = \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_{n-1} \end{bmatrix}$$

Clearly, if A_x is non-singular, then

$$A_x = A'_x$$

and T_x is the identity matrix I.

For a set of matrices $\{A_{\sigma_i}\}$, $\sigma_i \in \Sigma$, we define a set of matrices called the Q-matrix as

$$Q_{\sigma_i} = A_{\sigma_i} \quad ; \quad \text{for all } \sigma_i \in \Sigma$$

$$Q_{\sigma_i \sigma_j} = Q'_{\sigma_i} \cdot T_{\sigma_j} \quad \text{for all } \sigma_i, \sigma_j \in \Sigma$$

and in general, if

$$x = \sigma_1 \sigma_2 \cdots\cdots \sigma_{n-1} \sigma_n$$

$$Q_x = Q'_{\sigma_1 \sigma_2 \cdots\cdots \sigma_{n-1}} \cdot T_{\sigma_2 \sigma_3 \cdots\cdots \sigma_{n-1} \sigma_n} \qquad (6.6)$$

where the T-matrix is as defined before, and as a matter of fact is an inverse operation to the operation of " ' ",

i.e. $Q_{\sigma_j} = T_{\sigma_j} \cdot Q'_{\sigma_j}$

and $Q_x = T_x \cdot Q'_x$.

<u>Theorem 6.4:</u>

For any tape x, A_x is stable if and only if Q_x is stable.

Proof:

Assume, without loss of generality, that

$$x = \sigma_1 \sigma_2 \cdots\cdots \sigma_{n-1} \sigma_n .$$

Since we are dealing with stochastic matrices, A_x is stable if and only if A'_x is stable. Similarly, Q_x is stable if and only if Q'_x is stable.

But $A'_x = (A_{\sigma_1} \cdot A_{\sigma_2} \cdot A_{\sigma_3} \cdots\cdots A_{\sigma_{n-1}} \cdot A_{\sigma_n})'$,

$= (Q_{\sigma_1} \cdot Q_{\sigma_2} \cdot Q_{\sigma_3} \cdots\cdots Q_{\sigma_{n-1}} \cdot Q_{\sigma_n})'$

$= (Q'_{\sigma_1} \cdot T_{\sigma_2} \cdot Q'_{\sigma_2} \cdots\cdots Q_{\sigma_{n-1}} \cdot Q_{\sigma_n})'$

$= (Q'_{\sigma_1\sigma_2} \cdot Q'_{\sigma_2} \cdot T_{\sigma_3} \cdot Q'_{\sigma_3} \cdots\cdots Q_{\sigma_{n-1}} \cdot Q_{\sigma_n})'$

$= (Q'_{\sigma_1\sigma_2} \cdot Q_{\sigma_2\sigma_3} \cdot Q'_{\sigma_3} \cdot T_{\sigma_4} \cdot Q'_{\sigma_4} \cdots Q_{\sigma_{n-1}} \cdot Q_{\sigma_n})'$

$= (Q'_{\sigma_1\sigma_2} \cdot T_{\sigma_2\sigma_3} \cdot Q'_{\sigma_2\sigma_3} \cdot Q'_{\sigma_3\sigma_4} \cdots Q_{\sigma_{n-1}} \cdot Q_{\sigma_n})'$

$= (Q'_{\sigma_1\sigma_2\sigma_3} \cdot Q'_{\sigma_2\sigma_3} \cdot T_{\sigma_3\sigma_4} \cdot Q'_{\sigma_3\sigma_4} \cdot Q'_{\sigma_4} \cdots Q_{\sigma_{n-1}} \cdot Q_{\sigma_n})'$

$= (Q'_{\sigma_1\sigma_2\sigma_3} \cdot Q_{\sigma_2\sigma_3\sigma_4} \cdot Q'_{\sigma_3\sigma_4} \cdot Q_{\sigma_4\sigma_5} \cdots\cdots Q_{\sigma_{n-1}} \cdot Q_{\sigma_n})'$

.
.
.

$= (Q'_{\sigma_1\sigma_2\cdots\sigma_{n-2}} \cdot T_{\sigma_2\sigma_3\cdots\sigma_{n-1}} \cdot Q'_{\sigma_2\cdots\sigma_{n-1}} \cdot Q'_{\sigma_3\cdots\sigma_{n-1}} \cdots$

$\cdots\cdots Q'_{\sigma_{n-1}} \cdot Q_{\sigma_n})'$

$= (Q'_{\sigma_1\cdots\sigma_{n-1}} \cdot Q'_{\sigma_2\cdots\sigma_{n-1}} \cdot Q'_{\sigma_3\cdots\sigma_{n-1}} \cdots$

$\cdots Q'_{\sigma_{n-1}} T_{\sigma_n} \cdot Q'_{\sigma_n})'$

$= (Q'_{\sigma_1\cdots\sigma_{n-1}} \cdot Q'_{\sigma_2\cdots\sigma_{n-1}} \cdot Q'_{\sigma_3\cdots\sigma_{n-1}}$

$\cdots Q_{\sigma_{n-1}\sigma_n} \cdot Q'_{\sigma_n})'$

.
.
.

$$= (Q'_{\sigma_1 \ldots \sigma_{n-1}} \cdot T_{\sigma_2 \ldots \sigma_n} \cdot Q'_{\sigma_2 \ldots \sigma_n} \cdot Q'_{\sigma_3 \ldots \sigma_n} \cdot \ldots$$

$$\ldots . Q'_{\sigma_{n-1} \sigma_n} \cdot Q'_{\sigma_n})'$$

$$= Q'_{\sigma_1 \ldots \sigma_n} \cdot \quad Q'_{\sigma_2 \ldots \sigma_n} \cdot Q'_{\sigma_3 \ldots \sigma_n} \cdot Q'_{\sigma_4 \ldots \sigma_n} \cdot \ldots$$

$$\ldots . Q'_{\sigma_{n-1} \sigma_n} Q'_{\sigma_n} \qquad (6.7)$$

Since a stable matrix multiplied by any matrix is still stable, A'_x is stable if Q'_x is stable. On the other hand, since in Eq. (6.7), all row vectors in every matrix following Q'_x are linearly independent, the rank of Q'_{σ_n} is not less than that of $Q'_{\sigma_{n-1}\sigma_n}$. In general, the rank of any matrix in Eq. (6.7) is not greater than that of any other matrix in its right and not less than that of any other matrix in its left. By Lemma 6.2, if Q'_x is not stable, A'_x is not stable. Q.E.D.

Thus in order to check whether A_x is stable, all we need to do is check whether Q'_x is stable. The advantage of introducing the Q-matrices is that they are simpler in manipulation, for by Eq. (6.6)

$$
\begin{aligned}
Q_x &= Q'_{\sigma_1 \ldots \sigma_{n-1}} \cdot T_{\sigma_2 \ldots \sigma_{n-1} \sigma_n} \\
&= Q'_{\sigma_1 \ldots \sigma_{n-2}} \cdot T_{\sigma_2 \ldots \sigma_{n-1}} \cdot T_{\sigma_2 \ldots \sigma_n} \\
&\vdots \\
&= Q'_{\sigma_1} \cdot T_{\sigma_2} \cdot T_{\sigma_2 \sigma_3} \cdot \ldots\ldots \cdot T_{\sigma_2 \ldots \sigma_n},
\end{aligned}
$$

and the T-matrices are much simpler than those A-matrices.

Given a stochastic automaton Ω with a set of transition matrices $\{A_{\sigma_i}\}$, $\sigma_i \in \Sigma$, the decision procedure for the defintenes s of Ω is:

(i) Put $k = 1$.

(ii) Construct Q_x and T_x for all $x \in (\Sigma^+)^k$, where Σ^+ is Σ without the empty element.
Check whether all of the Q_x are stable.
If yes, Ω is definite.
If no, then

(iii) Check whether any of the matrix T_x is an identity matrix I.
If yes, Ω is not definite.
If no, check whether $k = n - 1$, where n is the number of internal states of Ω.
If yes, Ω is not definite.
If no, put $k = k + 1$, go back to (ii).

Example 6.1:

Let $\Sigma = \{0, 1\}$

Let a stochastic automaton Ω has a set of transition matrices $\{A_0, A_1\}$ where

$$A_0 = \begin{bmatrix} .25 & .225 & .225 & .3 \\ .1 & .25 & .45 & .2 \\ .2 & .2 & .2 & .4 \\ .2 & .225 & .275 & .3 \end{bmatrix}$$

$$A_1 = \begin{bmatrix} .2 & .3 & .1 & .4 \\ .1 & .35 & .35 & .2 \\ .3 & .35 & .15 & .2 \\ .2 & .325 & .175 & .3 \end{bmatrix}$$

Following the decision procedure described above, firstly, we have $Q_0 = A_0$, $Q_1 = A_1$

$$Q_0 = T_0 \cdot Q'_0 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1/2 & 1/4 & 1/4 \end{bmatrix} \cdot \begin{bmatrix} .25 & .225 & .225 & .3 \\ .1 & .25 & .45 & .2 \\ .2 & .2 & .2 & .4 \end{bmatrix}$$

$$Q_1 = T_1 \cdot Q'_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1/2 & 1/4 & 1/4 \end{bmatrix} \cdot \begin{bmatrix} .2 & .3 & .1 & .4 \\ .1 & .35 & .35 & .2 \\ .3 & .35 & .15 & .2 \end{bmatrix}$$

Next we have

$$Q_{00} = Q'_0 \cdot T_0 = \begin{bmatrix} .4 & .3 & .3 \\ .2 & .3 & .5 \\ .4 & .3 & .3 \end{bmatrix}$$

$$Q_{01} = Q'_0 \cdot T_1 = \begin{bmatrix} .4 & .3 & .3 \\ .2 & .3 & .5 \\ .4 & .3 & .3 \end{bmatrix} = Q_{00}$$

$$Q_{10} = Q'_1 \cdot T_0 = \begin{bmatrix} .4 & .4 & .2 \\ .2 & .4 & .4 \\ .4 & .4 & .2 \end{bmatrix}$$

$$Q_{11} = Q'_1 \cdot T_1 = \begin{bmatrix} .4 & .4 & .2 \\ .2 & .4 & .4 \\ .4 & .4 & .2 \end{bmatrix} = Q_{10}$$

Thus $T_{00} = T_{01} = T_{10} = T_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$

Now $Q_{000} = Q'_{00} \cdot T_{00} = \begin{bmatrix} .4 & .3 & .3 \\ .2 & .3 & .5 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} .7 & .3 \\ .7 & .3 \end{bmatrix}$

$Q_{001} = Q'_{00} \cdot T_{01} = Q'_{00} \cdot T_{00} = Q_{000}$

$Q_{010} = Q'_{01} \cdot T_{10} = Q'_{00} \cdot T_{00} = Q_{000}$

$Q_{011} = Q'_{01} \cdot T_{11} = Q'_{00} \cdot T_{00} = Q_{000}$

$Q_{100} = Q'_{10} \cdot T_{00} = \begin{bmatrix} .4 & .4 & .2 \\ .2 & .4 & .4 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} .6 & .4 \\ .6 & .4 \end{bmatrix}$

$Q_{101} = Q'_{10} \cdot T_{01} = Q_{100}$

$Q_{110} = Q'_{11} \cdot T_{10} = Q_{100}$

$Q_{111} = Q'_{11} \cdot T_{11} = Q_{100}$

Since all Q_x, for $x \in (\Sigma^+)^3$ are stable, Ω is 3-definite.

CHAPTER VII

CONCLUSION

We have introduced the concept of treating a stochastic matrix of a finite stochastic automaton as a point set in a convex polyhedron. Several properties that such a point set might possess have been studied. The exploitation of convex geometry is not just for mathematical interest. The geometric induction performed in this thesis either leads to the solution of decomposing a finite stochastic automaton or to the establishment of the decision procedures of definite stochastic automata. Practical methods have been obtained thereupon. The application of the concept is by no means limited to the problem of decomposition and decision of definiteness. Other aspects of finite stochastic automata can be treated with the same concept. Better insight, unified viewpoint, simpler methods, and new interesting properties might be obtained. All of these remain to be done.

BIBLIOGRAPHY

[1] Bacon, Glenn C., "The Decomposition of Stochastic Automata", Information and Control, 7, 1964, pp. 320 - 339.

[2] Bacon, Glenn C., " Minimal-State Stochastic Finite State Systems", IEEE Trans. on Circuit Theory, June 1964, pp. 307 - 308.

[3] Braines, S. N., and Svechincky, V. B., "Matrix Structure in Simulation of Learning". I. R. E. Trans. Inform. Theory IT-8, No. 5, 1962, pp. 186 - 190.

[4] Bush, R. R. and F. Mosteller, Stochastic Models for Learning, John Wiley, 1955.

[5] Bruce, G. D., and K. S. Fu, "A Model for Finite-State Probabilistic Systems". First Annual Allerton Conference on Circuit and System Theory, University of Illinois, Urbana, 1963, pp. 632 - 651.

[6] Carlyle, J. W., "Reduced Forms for Stochastic Sequential Machine." J. of Mathematical Analysis and Applications. 7, 1963, pp. 167 - 175.

[7] Carlyle, J. W., "On the External Probability Structure of Finite-State Channels" Information and Control 7, 1964, pp. 385 - 397.

[8] Carlyle, J. W., "Equivalent Stochastic Sequential Machines", Electronics Research Lab., U. C. Berkeley, Institute of Eng. Research, Series No. 60, Issue No. 415, Nov. 1961.

[9] Chen, I-Ngo and C. L . Sheng, " A geometrical approach to the decomposition of Markov Chains", Tech. Report No. 67-11, Dept. of Electrical Engineering, University of Ottawa, Ottawa, Canada, Oct. 1967.

[10] Chen, I-Ngo and C. L. Sheng, "Decision Problems of Definite Stochastic Automata", Tech. Report No. 68-8 , Dept. of Electrical Engineering, University of Ottawa, Ottawa, Canada, May , 1968.

[11] Chandrasekaran, C., and Shen, David W. C., "On Expediency and Convergence in Variable-Structure Automata", IEEE Trans. on System Sc. and Cybernetics SSC-4, No. 1, March 1968, pp. 52 - 60.

[12] Davis, W. A., "Contributions to Structure Theory of Sequential Machines", Doctoral dissertation in Electrical Engineering Department, University of Ottawa, Oct. 1966.

[13] Feller, W. An Introduction to Probability Theory and Its Application, Vol. 1, Wiley, New York, 1958.

[14] Fu, K. S. and G. J. McMurtry, "A Study of Stochastic Automata as Models of Adaptive and Learning Controllers", Purdue University, Lafayette, Ind., Rept. TR-EE 65-8, 1965.

[15] Fu, K. S. and R. W. Mclaren, "An application of Stochastic Automata to the Synthesis of Learning Systems", Purdue University, Lafayette, Ind., Rept. TR-EE 65-17, 1965.

[16] Gantmacher, F.R., The Theory of Matrices, Vol. 2, Chelsea, New York, 1959.

[17] Gill, Arthur, "Time-Varying Sequential Machines", Jour. of Franklin Inst. Dec. 1963, pp. 519 - 539.

[18] Ginsburg, S. "Set of Tapes Accepted by Different Types of Automata", J. Assoc. Computer March. 8, No. 1, 1961.

[19] Ginzburg, A. and Yoeli, M. "Products of Automata and The Problem of Covering". Trans. Am. Math. Soc. 116, 1965, pp. 253 - 266.

[20] Glinski, G.S. and M. Inghel, "A Probabilistic Approach to the Working Phase of A Two-Mode Threshold Larning Process", Technical Report No. 67-5, Dept. of Electrical Engineering, University of Ottawa, March, 1967.

[21] Glinski, G.S., "Information Theory Problems of Unreliable Finite Automata", Presented at Symposium on Relay Systems Theory and Finite Automata, Moscow, USSR, Sept. 24 - Oct. 2, 1962.

[22] Hajnal, J., "The Ergodic Properties of Non-Homogenous Finite Markov Chains", Proc. Comb. Phil. Soc. 52, 1956, pp. 67 - 77.

[23] Hajnal, J., "Weak Ergodicity in Non-Homogeneous Finite Markov Cahins", Proc. Cambridge Philos. Soc. 54, 1958, pp 233 - 246.

[24] Hartmanis, J., "Loop-free Structure of Sequential Machines", Information Control 5, 1962, pp. 25 - 43.

[25] Kemeny, J. G., and J. L. Snell, Finite Markov Chains, D. Van Nostrand Company, Princeton, N. J., 1960.

[26] Kemeny, J. G., H. Mirkil, J. L. Snell, G. L. Thompson, Finite Mathematical Structures, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1963.

[27] Kleene, S. C., "Representation of events in nerve nets and finite automata, " in Automata Studies, Annals of Mathematics Studies, vol. 34, Princeton University Press, Princeton, N. J., 1956, pp. 3 - 41 .

[28] Levin, V. I., "Probability and Reliability Analysis of Finite Automata", Engineering Cybernetics, No. 5, Sept. - Oct., 1965, pp. 53 - 61.

[29] Mealy, G. H. "Method for Synthesizing Sequential Circuits", Bell System Tech. Jour., Vol. 34, 1955, pp. 1045 - 1079.

[30] Modenov, P. S. and A. S. Parkhomenko, Geometric Transformations, Vol. 1, Academic Press, N. Y. and London, 1965.

[31] Moore, E. F., "Gendanken-Experiments on Sequential Machines", in Automata Studies C. E. Shannon and J. McCarthy, eds., Princeton, N. J.: Princeton University Press. 1956.

[32] Page, Carl V., "Equivalences between Probabilistic and Deterministic Sequential Machines", Information and Control 9, 1963, pp. 469 - 520.

[33] Paz, A., "A Finite Set of n x n Stochastic Matrices Generating All n-Dimensional Probability Vectors Whose Coordinates Have Finite Binary Expansion". SIAM J. Control, Vol. 5, No. 4, 1967, pp. 545 - 554.

[34] Paz, A., "Minimization Theorems and Techniques for Sequential Stochastic Machines", Information and Control, 11, 1967, pp. 155 - 166.

[35] Paz, A., and M. Reichaw, "Ergodic Theorems for Sequences of Infinite Stochastic Matrices" Proc. Camb. Phil. Soc. 63, 1967, pp. 777 - 784.

[36] Paz, A., "Some Aspects of Probabilistic Automata", Information and Control, 9, Feb. 1966, pp. 26 - 60.

[37] Paz, A., "Definite and Quasidefinite Sets of Stochastic Matrices", Proc. Amer. Math. Soc. 16, 1965, pp. 634 - 641.

[38] Perles, M., M.O. Rabin, and E. Shamir, "The Theory of Definite Automata", IEEE Transaction on Electronics Computers, EC-12, June 1963, pp. 233 - 243.

[39] Rabin, M. O., "Probabilistic Automata", Information and Control 6, 1963, pp. 230 - 245.

[40] Rabin, M.O., and Scott, D. "Finite Automata and Their Decision Problems". I.B.M. J. Res. Develop. 3, No. 2, 1959, pp. 114 - 125.

[41] Sklansky, J., "Learning Systems for Automatic Control", IEEE Trans. on Automatic Control, AC-11, No. 1, January, 1966, pp. 16 - 19.

[42] Tsertsvadze, G.N., "Certain Properties of Stochastic Automata and Certain Methods for Synthesizing Them", Avtomatika i Telemekhanika, Vol. 24, No. 3, March 1963, pp. 341 - 352.

[43] Tsetlin, M.L., "On the behavior of Finite Automata in Random Media," Avtomat, i Telemekh., Vol 22, 1961.

[44] Udagawa and Inagaki, "Probabilistic Studies of Logical Circuits and the Synthesis of Reliable Circuits using Probabilistic Logical Matrix", Jour. I.E.C.E. Japan, 46, 11, Nov. 1963, p. 1639.

[45] Udagawa, K. and Y. Inagaki, "Probabilistic Studies of Sequential Circuits and Estimation of Their Reliability Using Markov Chain", EACJ (Electronics and Communications in Japan Vol. 47, No. 4, April 1964, pp. 276 - 286.

[46] Varshavskii, V.I., and I.P. Vorontsona, "On the Behavior of Stochastic Automata With A Variable Structure", Avtomatika i Telemekhanika, Vol. 24, No. 3, March 1963, pp. 353 - 360.

[47] Wolfowitz, J., "Products of Indecomposable, Aperiodic, Stochastic Matrices", Proc. Amer. Soc. 14, 1963, pp. 733 - 737.

[48] Yoeli, M., "The Cascade Decomposition of Sequential Machines," IRE Trans. on Electronic Computers, EC-10, 1961, pp. 587 - 592.

[49] Yoeli, M., "Cascade - Parallel Decomposition of Sequential Machines", IEEE Trans. on Electronic Computers, EC-12, 1963, pp. 322 - 324.

VITA

Name: I-Ngo Chen

Born: 13 July, 1934. Canton, China.

Educated:

Primary: Canton, China.

Junior High School: Canton, China.

Senior High School: Taiwan, China.

University:

(a) National Taiwan University,
Taipei, Taiwan, China.
1955 Bachelor of Science in
Engineering (B. S. E)
Electrical Engineering.

(b) National Chiao Tung University,
Hsinchu, Taiwan, China.
1963 Master of Science (M. Sc.)
Electronic Engineering.