HOMOMORPHISMS BETWEEN STOCHASTIC SEQUENTIAL MACHINES AND RELATED PROBLEMS

by

A. Paz

Memorandum No. ERL-M221

23 August 1967

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Homomorphisms Between Stochastic Sequential Machines and Related Problems*

A. Paz†
Electronics Research Laboratory
and
Department of Electrical Engineering
and Computer Sciences
University of California, Berkeley

ABSTRACT

The problem of finding a stochastic sequential machine with minimal number of states and homomorphic to a given machine is studied in various aspects. The methods used for investigating the above problem are based upon the properties of a certain convex polyhedron associated with the given machine.

* The research reported herein was supported by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force under AFOSR Grant AF-AFOSR-639-67.

† On leave from Technion, Israel Institute of Technology, Haifa, Israel.

INTRODUCTION

The problem of reducing the number of states of a given stochastic-sequential machine has been studied by several authors [1]-[3]. A full and effective solution to the problem of finding a machine with minimal number of states in an equivalence class of machines has been given in the above-mentioned papers, although the algorithms devised do not lead in general to a unique solution to this problem. In the previous year it occurs to both the author and to Ott [5]-[7] that further reduction is possible if the equivalence restriction is weakened in a certain sense.

The study here has been prompted by the pioneering work of Ott [5],[6], who investigates the above questions very thoroughly. Unfortunately the work of Ott is obscured by the introduction of too many newly-defined entities and theorems of minor importance. Some of the most important theorems of Ott are reproduced here (Theorems 1 through 4 in Part II and Sec. IV-B), and, as the organization of our work differs from the organization of Ott's, new proofs are provided to these theorems.

The content of our work is as follows: The first part, the exposition, contains most of the known facts and theorems on the basis of which the consequent parts are developed. The first part also contains an explanation of the problems to be considered in the consequent parts. In parts II and III we study some problems introduced by Ott and some new, related problems, in several aspects. In the last part the uniqueness of minimal machines is studied in the light of the concepts developed in the previous parts.

An attempt has been made to simplify matters as much as possible and to put the whole work under a common geometrical frame.

The work is self contained but familiarity with one of the references [2], [6] or [7] is recommended.

I. EXPOSITION

A. The Set K^M

Definition 1. A stochastic sequential machine (S. S. M) is a quadruple $M = (S, X, Y, \{A^M(y/x)\})$ where S, X and Y are finite sets (the internal states, the inputs and the outputs respectively) and $\{A^M(y/x)\}$ is a finite set containing $|X| \times |Y|$ square matrices of order $|S|$ ($|U|$ denotes the number of elements in a set U) such that $A^M(y/x) = [a_{ij}^M(y/x)]$ with $m_{ij}(y/x) \geq 0$ and

$$\sum_{y \varepsilon Y} \sum_{j=1}^{|S|} m_{ij}^M(y/x) = 1$$

The superscript M is used for identifying the specific machine M and will be omitted from here on if context is clear.

The pair (v, u) denotes any output-input pair of words over Y and X respectively, such that the length of u equals the length of v

($\ell(u) = \ell(v)$), e.g., $u = x_1 \ldots, x_k$, $v = y_1, \ldots, y_k$) $x_i \varepsilon X$, $y_i \varepsilon Y$. We assume that the pairs (v, u) are ordered in a way such that $\ell(v_1) < \ell(v_2)$ => $(v_1, u_1) < (v_2, u_2)$ and that this order is kept fixed.

When two or more machines are under consideration, we assume that all of them have the same X and Y sets, and the pairs (v, u) are ordered in all of them by the same order.

With every S. S. M., M, we associate an infinite set of n-dimensional ($n = |S|$) column vectors K^M defined as follows:

$$K^M = \{\eta(v, u) : v \varepsilon Y^*, \ u \varepsilon X^*\}$$

where X^* and Y^* are the sets of all words over X and Y, respectively, $\eta = \eta(\lambda, \lambda)$ is a vector all the entries of which are equal to 1 (λ denotes the empty word) and

$$\eta(yv, xu) = A(y/x)\eta(v, u) \cdot$$

The vectors $\eta(v, u)$ are thus defined recursively and we assume that they are ordered according to the order of the corresponding pairs (v, u).

Let K^M and K^{M*} be two sets as shown above corresponding to two S. S. M.'s, M and M* respectively. K^M is <u>stochastically homomorphic</u> to K^{M*} ($K^M \overset{\sim}{\rightarrow} K^{M*}$) if there is a stochastic matrix B such that

$$B\, \eta^M(v, u) = \eta^{M*}(v, u) \qquad \text{for all } (v, u). \tag{1}$$

Theorem 1. The relation (1) is equivalent to the following relation (2) (provided that B has the due dimensions).

$$B\,A^{M}(y|x)\,\eta^{M}(v, u) = A^{M^*}(y|x)\,B\,\eta^{M}(v, u) \text{ for all } (v, u) \text{ and all } (y, x) \tag{2}$$

Proof: Consider Fig. 1.

The figure shows that (1) implies (2). If (2) holds then:

$B\,\eta^{M}(\lambda, \lambda) = \eta^{M^*}(\lambda, \lambda)$, by the fact that both $\eta^{M}(\lambda, \lambda)$ and $\eta^{M^*}(\lambda, \lambda)$ have all entries equal to 1 and B is stochastic.

Furthermore, if (1) holds for the pair (v, u) then:

$B\,\eta^{M}(yv, xu) = B\,A^{M}(y|x)\,\eta^{M}(v, u) = A^{M^*}(y|x)\,B\,\eta^{M}(v, u) =$
$= A^{M^*}(y|x)\,\eta^{M^*}(v, u) = \eta^{M^*}(yv, xu)$; by the definitions and by (2). The theorem follows by induction.

B. The Matrix H^M

With every S. S. M., M, we associate a finite matrix H^M having the following properties:

The first column of H^M is η^M. (3.1)

All the columns of H^M are elements of K^M, are linearly independent and every other element of K^M is a linear combination of them. (3.2)

The (v, u) pairs corresponding to the columns of H^M are the smallest pairs (in the order of the (v, u) pairs) such that H^M satisfies the above conditions (3.1) and (3.2). (3.3)

It has been shown by Carlyle that given M one can find effectively H^M (see [2],[5].

<u>Theorem 2.</u> The relation (2) is equivalent to the following relation:

$$B A^M(y|x) H^M = A^{M^*}(y|x) B H^M \tag{4}$$

<u>Proof</u>: Clearly (2) implies (4). That (4) implies (2) follows from the fact that every vector $\eta^M(v, u)$ is a linear combination of the columns of H^M.

C. <u>The Covering Property</u>

Let M be a given S. S. M.. With every n-dimensional ($n = |S|$) probabilistic row vector π we associate a function $p_\pi^M : \{(v, u)\} \to [0, 1]$ defined as follows:

$$p_\pi^M(v, u) = \pi\eta^M(v, u)$$

Two functions as above, p_π^M and $p_\pi^{M^*}$, are equal if

$$p_{\pi}^{M}(v, u) = p_{\pi'}^{M^*}(v, u) \quad \text{for all pairs } (v, u) \tag{5}$$

This requirement is equivalent to:

$$\pi \eta^{M}(v, u) = \pi' \eta^{M^*}(v, u) \text{ for all pairs } (v, u) \tag{6}$$

If the functions are induced by the same machine M then (6) is equivalent to:

$$\pi H^{M} = \pi' H^{M} \tag{7}$$

because of the property (3.2) of H^M.

Definition 2. For a given S. S. M., M, we define the set of functions $\mathcal{P}^M$ induced by M as

$$\mathcal{P}^M = \{p_{\pi}^{M}\}$$

Definition 3. Let M and M^* be two given S. S. M.'s. The machine M covers the machine M^* ($M \geq M^*$) if $\mathcal{P}^M \supseteq \mathcal{P}^{M^*}$.

Theorem 3. $M \geq M^*$ if and only if $K^M \rightrightarrows K^{M^*}$.

Proof: This is lemma 1 of Ott [6].

D. Two Covering Problems

The following problem has been investigated by Ott [5], [6]. (See also Paz [7]).

Problem 1. Given an S. S. M., M, find a machine $M^* \geq M$ which has the least number of states possible.

Ott's investigation, although giving deep insight into the problem, is far from providing a general effective solution to it. Also the example given by Ott, in order to show that the problem is not trivial, is a degenerate example, and one may ask whether there are nondegenerate machines which admit a solution to this problem.

Our concern here is to suggest and investigate the following related problem.

Problem 2. Given an S. S. M., M, find a machine M^* which has the least number of states possible such that $M \geq M^*$.

We shall show by examples that this second problem is not trivial and not equivalent to the first problem, specifically:

1. There is an S. S. M., M, which admits a solution to the second problem, with M^* having less states than M, but no M^* having less states than M is such that $M^* \geq M$.

2. There is an S. S. M., M, which admits a solution to the first problem, with M^* having less states than M, but no M^* having less states than M is such that $M \geq M^*$.

3. There is an S. S. M., M, which admits a solution to both the first and second problems, with M^* having less states than M.

4. There is an S. S. M., M, such that there are equivalent machines to M with the same number of states, but there is no machine

M^* having less states than M such that $M \geq M^*$, or $M^* \geq M$.

All the examples, to be shown, are nondegenerate and this will provide an answer to the above posed question concerning the first problem. Furthermore, we shall show that the second problem is simpler, in a certain sense, than the first one, and we shall generalize a previous result of the author [Ref. 7, Theorem 6].

E. Reduced and Minimal State Form

Another question to be considered arises from the following considerations:

Definition 4. Two S. S. M.'s, M and M^*, are equivalent if $\mathcal{P}^M = \mathcal{P}^{M^*}$ ($M \geq M^*$ and $M^* \geq M$ or $K^M \precsim K^{M^*}$ and $K^{M^*} \precsim K^M$).

Definition 5. An S. S. M., M, is in reduced form if no two extremal functions induced by M are equal (an extremal function is a function p_π^M such that π is an extremal vector, i.e., a vector having one entry equal to one, all the other entries being equal to zero).

Remark 1. The above definition and (7) imply that M is in reduced form iff no two rows of H are equal.

Definition 6. An S. S. M., M, is in minimal state form if no extremal function of M is a convex combination of other extremal functions of M.

<u>Remark 2.</u> The above definition and (6) imply that M is in minimal form iff no row of H^M is a convex combination of its other rows.

The problem of checking whether a given S. S. M. is in reduced form or in minimal state form has been solved effectively in the literature [1] - [3], [5] and algorithms have been found for reducing a given S. S. M. to an equivalent machine in one of these forms [1] - [3]. The resulting machines have, in general, less states than the source machines. On the other hand, the reduced form or the minimal state form of a given machine (these forms differ in many cases [1], [2]) is not unique [3] in general. In the last section of this paper we shall give a decidable condition which implies the existence of a unique minimal (reduced) form of a given machine.

II. GEOMETRICAL INTERPRETATION

A. Machines of Common Rank

<u>Notation.</u> Let M and H^M be an S. S. M. and its related H^M matrix. Then m = rank M = rank H^M = # of columns in H^M (as these columns are independent).

In the sequel we shall need the following theorem of Ott.

<u>Theorem 4.</u> Let M and M^* be S. S. M.'s of common rank such that $M \geq M^*$ with $\eta^{M^*}(v, u) = B\,\eta^M(v, u)$ for every pair (v, u). Then $H^{M^*} = B\,H^M$.

Proof: Let J^{M^*} be the matrix such that $J^{M^*} = B H^M$. Denote the columns of H^M by $\eta_1 \ldots \eta_m$ and the corresponding columns in J^{M^*} by $\eta_1^* \ldots \eta_m^*$. Let $\eta^M(v, u)$ be any vector in K^M. Then

$$\eta^M(v, u) = \sum_{i=1}^{m} a_i \eta_i \quad \text{(property (3.2) of } H^M\text{)}.$$

This implies that

$$\eta^{M^*}(v, u) = B \eta^M(v, u) = \sum a_i B \eta_i = \sum a_i \eta_i^* . \tag{8}$$

It follows that any vector in K^{M^*} is a linear combination of vectors in J^{M^*} and therefore, as M and M^* have common rank, rank J^{M^*} = rank H^M = rank H^{M^*}. Furthermore, the columns of H^{M^*} must also be columns in J^{M^*} (arranged in the same order), for if this is not true, then let $\eta^{M^*}(v_o, u_o)$ be the first column in H^{M^*} which is not a column of J^{M^*}. Then $\eta^M(v_o, u_o)$, the corresponding vector in K^M, is not in H^M and therefore is a linear combination of the columns of H^M preceding the vector $\eta^M(v_o, u_o)$ in K^M (property (3.3) of H^M). It is implied by (8) that $\eta^{M^*}(v_o, u_o)$, a column of H^{M^*}, is a linear combination of other columns in H^{M^*} contrary to the property (3.2) of H^{M^*}. This proves the theorem, for the columns of H^{M^*} cannot be a proper subset of the columns in J^{M^*} (as rank J^{M^*} = rank H^{M^*}) and therefore $J^{M^*} = H^{M^*} = B H^M$.

B. Geometrical Interpretation of H^M

Let M and H^M be, as before, an S. S. M. and its related H^M matrix,and assume that H^M has m columns and n rows ($m \leq n$ necessarily). The rows of H^M will be considered as points in m-dimensional space and the notation $h_1 \ldots h_n$ will be used to refer to these points. All the machines under consideration from here on will be assumed in minimal state form (this is not a restriction as one can construct effectively an equivalent minimal state form machine for any given machine). It follows from remark 2 in the previous section that iff M is in minimal state form then the points $h_1 \ldots h_n$ are vertices of a convex polyhedron.

A substochastic vector $\pi = (\pi_1 \ldots \pi_n)$ is a vector such that $\pi_i \geq 0$ and $\sum \pi_i \leq 1$. With any such nonzero vector we associate a point in m-dimensional space defined by $\frac{\pi}{\sum \pi_i} H$ and denoted by $h(\pi)$. Clearly, $h(\pi) \varepsilon \operatorname{conv}(h_1 \ldots h_2)$. If A is a matrix with substochastic rows A_i then $H(A)$ denotes the set of nonzero points $\{h(A_i)\}$.

Two substochastic vectors π and ρ are similar ($\sim$) relative to H^M if $h(\pi) = h(\rho)$ and they are equivalent ($\simeq$), relative to H^M if $\pi H^M = \rho H^M$. Note that $\pi \sim \rho$ iff there is a constant α such that $\pi \simeq \alpha \rho$ (provided that π and ρ are nonzero vectors).

Two substochastic matrices A and B are similar ($\sim$), relative to H^M if their corresponding nonzero rows are similar rows and they are equivalent ($\simeq$), relative to H^M if their corresponding rows are equivalent.

C. Comparison of the Two Covering Problems

We have seen in the previous sections that $M \geq M^*$ for two given machines is equivalent to the following conditions:

There is a stochastic matrix B such that

$$B\, \eta^{M}(v, u) = \eta^{M^*}(v, u) \qquad \text{for every pair } (v, u) \tag{9}$$

or, equivalently, such that

$$B\, A^{M}(y/x)\, H^{M} = A^{M^*}(y/x)\, B\, H^{M} \qquad \text{for all pairs } (y, x) \tag{10}$$

Assume now that the machine M^* is given and it is required to find a machine M having less states such that $M \geq M^*$. If no more information is given about the machine M then the relation (10) is of little help for the solution to the above problem, as the matrix H^M is not known.

If we assume that rank M = rank M^* then, because of theorem 4 and the relation (9) (with B a stochastic matrix), one can begin with any matrix H^M such that

$$\text{conv}\,(h_1^{M}, \ldots, h_n^{M}) \supseteq (h_1^{M^*}, \ldots, h_{n^*}^{M^*}), \quad n < n^*$$

and having the usual properties ($0 \leq h_{ij}^{M} \leq 1$, $h_{11}^{M} = h_{21}^{M} \ldots h_{n1}^{M} = 1$) and then try to find nonnegative matrices which will satisfy Eq. (10) (B is determined by the choice of H^M). If one H^M fails to provide a solution then another H^M can be assumed and so on. Note, however, that even if no solution can be found under the assumption that rank M = rank M^*, this does not mean that no solution exists, for there is no reason to believe that the above assumption is necessary for the existence of a solution.

Assume now that the machine M is given and it is required to find a machine M^* having less states such that $M \geq M^*$.

Considering again the relation (10) we see that a solution to this problem depends upon the finding of a stochastic matrix B having less than n rows such that nonnegative matrices $A^{M^*}(y/x)$ can be found which will satisfy the relation (10). Any matrix B, as above, can serve as a starting point for calculation and this does not depend upon the assumed rank of M^*. This is why we claimed that the second problem is simpler than the first one. Note, however, that any solution to the second problem is restrictive in the sense that not all the functions p_π^M induced by M have equal functions induced by M^*. On the other hand, it will be shown that this problem is not trivial. Moreover, there are machines to which only this problem admits a solution.

D. The Second Covering Problem

The purpose of the following theorem is to provide a geometrical interpretation to the second covering problem. The meaning of this interpretation as well as its possible uses will be illustrated by examples in the following sections.

Theorem 5. Let M be an n-state S. S. M. There exists a machine M^* with $n^* < n$ states such that $M \geq M^*$ iff there exists a stochastic $n^* \times n$ matrix B with n^* rows such that

$$\bigcup_{(y, x)} H^M(B A^M(y/x)) \subset \text{conv } H^M(B) \tag{11}$$

A machine M^* as above can be constructed effectively if a matrix B satisfying (11) is given.

Proof: Assume that $M \geq M^*$ with $n^* < n$. Then (10) is satisfied by a stochastic matrix B with n^* rows. Let $\xi = (\xi_1 \ldots \xi_{n^*})$ be any non-zero row of $A^*(y/x)$ for a given pair (y, x). Then $\frac{\xi}{\sum \xi_i}$ B H is a convex combination of the rows of B H. This means that $\frac{\xi}{\sum \xi_i} B H \in$ conv H^M (B). Let $\rho = (\rho_1 \ldots \rho_n)$ be the row corresponding to ξ in the matrix $B A^M(y/x)$. It follows from (10), and from the fact that B is stochastic and because the first column of H^M is a column of ones, that $\rho H = \xi B H \Longrightarrow \sum \rho_i = \sum (\xi B)_i = \sum \xi_i$. Therefore

$$h(\rho) = \frac{\rho}{\sum \rho_i} H = \frac{\xi}{\sum \xi_i} B H \in \text{Conv } H^M(B)$$

$$\therefore \bigcup_{(y, x)} H^M(B A^M(y, x)) \subseteq \text{Conv } H^M(B) .$$

Assume now that there is a stochastic $n^* \times n$ matrix B with $n^* < n$ such that $\bigcup_{(y,x)} H^M(B\,A^M(y,x)) \subseteq \text{Conv}\ H^M(B)$ then any point in the left-hand side is a convex combination of the points in $H^M(B)$. The points in the left-hand side are of the form $\alpha \rho H$ where α is a normalizing constant and ρ is a row in a matrix $B\,A^M(y, x)$ for some pair (y, x).

We have therefore

$$\alpha \rho H = \pi B H \qquad (12)$$

where π is a stochastic vector.

It is easy to see that the relation (10) will be satisfied if the matrices $A^{M^*}(y, x)$ are defined as follows:

(a) If a row in $B\,A^M(y, x)$ is a zero row, then the corresponding row in $A^{M^*}(y, x)$ is a zero row.

(b) Let ρ be a nonzero row in $B\,A^M(y, x)$. Then the corresponding row in $A^{M^*}(y, x)$ will be $\frac{1}{\alpha}\pi$ where π and α are as in (12).

The theorem is thus proved.

Example 1. Let M be the 4-state machine described as follows ($X = Y = \{0, 1\}$):

$$A(0/0) = \begin{bmatrix} \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & 0 \\ \frac{3}{8} & \frac{3}{16} & \frac{3}{16} & 0 \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad A(1/0) = \begin{bmatrix} 0 & \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & \frac{1}{8} & \frac{1}{8} & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

$$A(0/1) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{8} & \frac{1}{8} \\ 0 & \frac{3}{4} & \frac{3}{16} & \frac{3}{16} \\ 0 & \frac{1}{4} & \frac{1}{8} & \frac{1}{8} \end{bmatrix} \qquad A(1/1) = \begin{bmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{4} & 0 & 0 & \frac{1}{4} \\ \frac{1}{8} & 0 & 0 & \frac{1}{8} \\ \frac{1}{4} & 0 & 0 & \frac{1}{4} \end{bmatrix}$$

A matrix H^M for this machine is:

$$H^M = \begin{bmatrix} 1 & \frac{1}{2} & 0 \\ 1 & \frac{3}{4} & \frac{1}{2} \\ 1 & \frac{1}{2} & \frac{3}{4} \\ 1 & 0 & \frac{1}{2} \end{bmatrix}$$

Let B be the stochastic matrix:

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Consider now Fig. 2. The rows of H^M, h_1, h_2, h_3 and h_4 are represented in that figure as two-dimensional points (as the first coordinate of these points is equal to 1, one can consider the two-dimensional subspace as having the first coordinate equal to 1).

The set of points $H(B)$ is represented by little circles, while the set of points $\bigcup_{(y,x)} H^M(B\, A(y, x))$ is represented by crosses (the points are easily computed). It is seen that the condition of the theorem holds for this case and therefore a 3-state machine M^* exists such that $M \geq M^*$. Using the procedure suggested in that theorem we find the matrices $A^{M^*}(y, x)$ which are the following:

$$A^{M^*}(0/0) = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & 0 \\ \frac{5}{16} & \frac{5}{16} & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad A^{M^*}(1/0) = \begin{bmatrix} 0 & \frac{1}{2} & 0 \\ 0 & \frac{3}{8} & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

$$A^{M^*}(0/1) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{5}{16} & \frac{5}{16} \\ 0 & \frac{1}{4} & \frac{1}{4} \end{bmatrix} \qquad A^{M^*}(1/1) = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{3}{16} & 0 & \frac{3}{16} \\ \frac{1}{4} & 0 & \frac{1}{4} \end{bmatrix}$$

It can be shown that there is no machine covered by M^* and having less states (see the examples which follow). On the other hand, we shall show now that there is a machine M^+ having less states than M and covering it.

We first observe that if M^+ as above exists then, as in the proof of Theorem 4, the columns of H^M must also be columns in $B\, H^{M^+}$. H^M has 3 columns and therefore H^{M^+} must have at least 3 columns and at least 3 rows (the columns of H^{M^+} are independent). Now M^+ has less states than M (which has 4 states) so that H^{M^+} must have 3 rows and 3 columns (otherwise the columns will not be independent). Thus $B\, H^{M^+} = H^M$. As B is stochastic it must be true that $(h_1^M, \ldots h_4^M) \subset \operatorname{conv}(h_1^{M^+}, h_2^{M^+}, h_3^{M^+})$ and, as seen in Fig. 2, the only possible choice is

$$H^{M^+} = \begin{bmatrix} 1 & \frac{1}{2} & 0 \\ 1 & 1 & 1 \\ 1 & 0 & \frac{1}{2} \end{bmatrix}$$

The machine M^+ is now easily found using a procedure described by Ott [5]

$$A^{M^+}(0/0) = \begin{bmatrix} \frac{5}{16} & \frac{1}{8} & \frac{1}{16} \\ \frac{5}{8} & \frac{1}{4} & \frac{1}{8} \\ 0 & 0 & 0 \end{bmatrix} \qquad A^{M^+}(1/0) = \begin{bmatrix} \frac{1}{8} & \frac{1}{4} & \frac{1}{8} \\ 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix}$$

$$A^{M^+}(0/1) = \begin{bmatrix} 0 & 0 & 0 \\ \frac{1}{4} & \frac{3}{8} & \frac{3}{8} \\ \frac{1}{8} & \frac{3}{16} & \frac{3}{16} \end{bmatrix} \qquad A^{M^+}(1/1) = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} \end{bmatrix}$$

From the above examples two facts can be learned:

(a) There is a nondegenerate machine M (i.e., a machine such its next state is not a deterministic function of its present state and input-output pair) which can be covered by a machine M^* having less states.

(b) The configuration $M^* \leq M \leq M^+$, where all three machines are nondegenerate but M^* and M^+ have less states than M, is possible.

E. A Sufficient Condition for the Second Problem

Theorem 6. Let M be an n-state machine. Let $h_1^* \ldots h_{n^*}^*$ be a set of $n^* < n$ points in m-dimensional space such that (m = # of columns of H^M)

$$\bigcup_{(y,x)} H^M(A^M(y/x)) \subset \text{conv}\,(h_1^* \ldots h_{n^*}^*) \subset \text{conv}\,(h_1 \ldots h_n) \tag{13}$$

Then there is a matrix B with n^* rows satisfying the condition (11) of Theorem 5.

Proof: If B is any stochastic matrix then any point in the set $\bigcup_{(y,x)} H^M(B\,A^M(y/x))$ is a convex combination of points in the set $\bigcup_{(y,x)} H^M(A^M(y/x))$. Let B be the matrix such that $B\,H^M = \begin{pmatrix} h_1^* \\ \vdots \\ h_{n^*}^* \end{pmatrix}$.

Then clearly $\text{conv}\,(h_1^* \ldots h_{n^*}^*) = \text{conv}\,H^M(B)$. This shows that the condition of Theorem 5 is satisfied so that this theorem holds true.

To make the content of the above theorem clear, consider again the example in the previous section and Fig. 2. It is seen in that figure that $\text{conv}(h_1 \ldots h_4)$ is the simplex whose vertices are the four points $h_1 \ldots h_4$ and $\bigcup_{(y,x)} H^M(A^M(y/x))$ is the set containing the four crossed points denoted by $u_1 \ldots u_4$. The choice $h_1^* = h_1$, $h_2^* = \frac{1}{2}(h_2 + h_3)$ $h_3^* = h_4$ (the only possible choice for this example) will satisfy the relation (13) and the required matrix B is determined by this choice.

It is thus seen that Theorem 6 may be useful in the case where the set conv $\bigcup_{(y,x)} H^M(A^M(y/x))$ is a proper subset of the set conv $(h_1 \ldots h_n)$, a condition which is decidable. Unfortunately, even in this case, to find the required matrix B which will satisfy the conditions of Theorem 5 one must be able to solve a nontrivial geometrical problem, namely:

Given two convex polyhedra V_1 and V_2 such that $V_1 \subset \text{Int } V_2$. Find a polyhedron V_3 with minimal number of vertices such that $V_1 \subseteq V_3 \subset \text{Int } V_2$.

As far as we know, this problem is not yet generally solved, but any solution to this problem in some particular case (as in the previous example) with $V_1 = \text{conv} \bigcup_{(y,x)} H^M(A^M(y/x))$, $V_2 = \text{conv } (h_1 \ldots h_n)$ and $V_3 = \text{conv } (h_1^* \ldots h_{n*}^*)$ will lead to a matrix B as required, provided that conv V_1 is a proper subset of conv V_2.

F. <u>The Second Covering Problem is Independent</u>

The purpose of the following example is to show that the conditions of Theorem 6 are not necessary conditions. Specifically, we shall show that a machine M exists such that conv $\bigcup_{(y,x)} H(A(y/x)) =$ conv$(h_1, \ldots, h_n)$ but there is $M^* \leq M$ such that M^* has less states than M.

We will also show, using the same example, that there is an n-state machine M such that there is an n^*-state machine $M^* \leq M$ with $n^* \leq n$, but no machine with less than n-state covers M. This will demonstrate that the second covering problem does not depend on the first one.

Example 2. Let M be the 5-state machine described as follows ($X = \{0, 1\}$, $Y = \{0, 1, 2\}$):

$$A(0/0) = \begin{bmatrix} 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad A(1/0) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad A(2/0) = \begin{bmatrix} 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$A(0/1) = \begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad A(1/1) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad A(2/1) = \begin{bmatrix} 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{bmatrix}$$

An H^M matrix for this machine is the following:

$$\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 1 & \frac{1}{2} & 0 & \frac{1}{2} \\ 1 & 0 & 0 & \frac{1}{2} \\ 1 & 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 & 1 \end{bmatrix} \quad \begin{matrix} h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \end{matrix}$$

One finds easily that for this machine

$$\bigcup_{(y,x)} H(A(y/x)) = \{h_1, h_2, h_3, h_4, h_5, \tfrac{1}{2}(h_2 + h_4), \tfrac{1}{2}(h_1 + h_3), \tfrac{1}{2}(h_2 + h_5)\}$$

and therefore:

$$\operatorname{conv} \bigcup_{(y,x)} H(A(y/x)) = \operatorname{conv}(h_1, \ldots, h_5).$$

Yet a machine $M^* \leq M$ having only 4 states can be found. The machine M^* is simply the machine which is obtained from M by deleting the 5th column and the 5th row in all the matrices of M. This fact is easily verified using the transformation matrix:

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

We shall now show that there is no machine $M^* \geq M$ having less than 5 states.

Using an argument similar to that used in the analysis of the previous Example 1, one can show that if there is an $M^+ \geq M$ with less than 5 states, then it must have 4 states and necessarily rank M^+ = rank M. This will lead to a single possible choice for H^{M^+} and B, namely:

$$H^{M^+} = \begin{bmatrix} 1 & 1 & 0 & \frac{1}{2} \\ 1 & 0 & 0 & \frac{1}{2} \\ 1 & 0 & 1 & \frac{1}{2} \\ 1 & 0 & 0 & 1 \end{bmatrix} \qquad B = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

with the consequence that the required matrices A^{M^+} and A^M satisfy the relation (10) with H^{M^+} and B as above. Let us check this relation for the matrix $A^M(1/0)$. We have

$$\begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad A^{M^+}(1/0)\, H^{M^+} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 1 & \frac{1}{2} & 0 & \frac{1}{2} \\ 1 & 0 & 0 & \frac{1}{2} \\ 1 & 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

The first, second and fifth rows in the right-hand side are zero rows. This implies that all the rows in $A^{M^+}(1/0)\, H^{M^+}$ are zero rows (all the entries are nonnegative and every row of $A^{M^+}(1/0)\, H^{M^+}$ contributes to the formation of the first, second and fifth rows of the right-hand side). But this is impossible as there are nonzero rows in the right-hand side.

G. The Second Covering Problem is Not Trivial

The purpose of the example in this section is to show that not every n-state machine M has a machine $M^* \leq M$ with $n^* < n$ states, even if rank $M < n$. Moreover, the same example will show that there is an n-state machine M such that there is another machine $M^+ \geq M$ with $n^+ < n$ states, but no machine M^* with $n^* < n$ states is such that $M \geq M^*$.

Example 3. Let M be the 4-state machine defined as follows ($X = Y = \{0, 1\}$):

$$A(0/0) = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{3}{8} & \frac{3}{8} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & 0 \end{bmatrix} \qquad A(1/0) = \begin{bmatrix} \frac{1}{4} & 0 & \frac{1}{4} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{8} & 0 & \frac{1}{8} & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} & 0 \end{bmatrix}$$

$$A(0/1) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & \frac{3}{8} & \frac{3}{8} \end{bmatrix} \qquad A(1/1) = \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & \frac{1}{8} & 0 & \frac{1}{8} \end{bmatrix}$$

An H^M matrix for this machine is:

$$H^M = \begin{bmatrix} 1 & \frac{1}{2} & 0 \\ 1 & \frac{3}{4} & \frac{1}{2} \\ 1 & \frac{1}{2} & \frac{3}{4} \\ 1 & 0 & \frac{1}{2} \end{bmatrix}$$

Thus, rank $M = 3 < 4 =$ # of states of M. Using a procedure described by Ott [5] one finds easily that the following 3-state machine M^+ is such that $M^+ \geq M$ where the defining matrices of M^+ are

$$A^{M^+}(0/0) = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \quad A^{M^+}(1/0) = \begin{bmatrix} \frac{3}{8} & 0 & \frac{1}{8} \\ \frac{3}{4} & 0 & \frac{1}{4} \\ 0 & 0 & 0 \end{bmatrix}$$

$$A^{M^+}(0/1) = \begin{bmatrix} 0 & 0 & 0 \\ \frac{1}{8} & \frac{1}{8} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix} \quad A^{M^+}(1/1) = \begin{bmatrix} 0 & \frac{3}{4} & \frac{1}{4} \\ 0 & \frac{3}{8} & \frac{1}{8} \\ 0 & 0 & 0 \end{bmatrix}$$

To prove our second assertion (that there is no $M^* \leq M$ with less than 4 states), we shall first display the set $\bigcup_{(y,x)} H^M(A^M(y/x))$ in the form of a table T as follows:

T	$H^M(A^M(0/0))$	$H^M(A^M(1/0))$	$H^M(A^M(0/1))$	$H^M(A^M(1/1))$
s_1	$\frac{1}{2}(h_1 + h_2)$	$\frac{1}{2}(h_1 + h_3)$	O	$\frac{1}{2}(h_2 + h_4)$
s_2	O	$\frac{1}{2}(h_1 + h_3)$	$\frac{1}{2}(h_3 + h_4)$	$\frac{1}{2}(h_2 + h_4)$
s_3	$\frac{1}{2}(h_1 + h_2)$	$\frac{1}{2}(h_1 + h_3)$	$\frac{1}{2}(h_3 + h_4)$	$\frac{1}{2}(h_2 + h_4)$
s_4	$\frac{1}{2}(h_1 + h_2)$	$\frac{1}{2}(h_1 + h_3)$	$\frac{1}{2}(h_3 + h_4)$	$\frac{1}{2}(h_2 + h_4)$

where $s_1 \ldots s_4$ are the states of M and $h_1 \ldots h_4$ are the rows of H^M (the table also contains the points corresponding to zero rows in the matrices A^M). Let B be any stochastic matrix with 4 columns and $m \leq 3$ rows. The table T' corresponding to $\bigcup_{(y,x)} H^M(B\,A^M(y/x))$, will have only m rows. A nonzero entry in any box, in a column of T', will be a convex combination of the nonzero entries in the corresponding column of T. As all the nonzero entries in the same column of T are equal, any convex combination of these entries will result in an entry having the same value as the combined entries. If the matrix B has nonzero entries in three different, or in all its columns, then the table T' will have nonzero entries in all its columns which will be equal to the nonzero entries in the corresponding columns of T (this follows from the above remarks and from the definitions). This implies that, for this case

$$\bigcup_{(y,x)} H^M(B\,A^M(y/x)) = \bigcup_{(y,x)} H^M(A^M(y/x)) \qquad (14)$$

On the other hand, as B has $m \leq 3$ rows, we have that $H^M(B)$ has only m points which are seen in Fig. 2 (the different points in $\bigcup H^M(A^M(y/x))$ are denoted by $u_1 \ldots u_4$ in the figure), no set containing less than 4 points in the interior of conv $(h_1 \ldots h_4)$ can have the set $u_1 \ldots u_4$ in the interior of its convex closure, so that the relation (11) cannot hold true in this case (because of (14)).

If the matrix B has nonzero entries in two columns or one column only, then the table T' will have nonzero entries in at least 3 columns. In this case, the set $\bigcup_{(y,x)} H^M(B\,A^M(y/x))$ will contain three of the four points $u_1 \ldots u_4$, with any three of these points not collinear. On the other hand, the set $H^M(B)$ will contain at most three collinear points so that the relation (11) cannot hold for this case either. Our assertion is thus proved.

H. Two Additional Examples

The following machine M (this example has been used by Even in [3] for a different purpose) is a 5-state machine such that no machine M^* exists with less than 5 states and such that $M^* \geq M$ or $M^* \leq M$, but rank $M = 4 < 5$. The method for proving these assertions is similar to the one used in the previous examples (although the actual proof is more complicated) and is left to the reader.

<u>Example 4:</u> The machine M with the above-mentioned properties is defined as follows ($X = \{0, 1\}$, $Y = \{0, 1, 2\}$):

$$A(0/0) = \begin{bmatrix} 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad A(1/0) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad A(2/0) = \begin{bmatrix} 0 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \end{bmatrix}$$

$$A(0/1) = \begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad A(1/1) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad A(2/1) = A(2/0)$$

$$H^M = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 1 & \frac{1}{2} & 0 & \frac{1}{2} \\ 1 & 0 & 0 & \frac{1}{2} \\ 1 & 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

It is worth mentioning that the above machine is strongly connected and it has equivalent monisomorphic strongly connected machines. (Its minimal state form is not unique.)

Our next example is concerned with the degenerate case of deterministic machines. The following theorem has been proved by Ott [6].

<u>Theorem</u>. If M is a reduced deterministic S. S. M., then no S. S. M. which covers M has fewer states. The example below shows that if the word "covers" in the above theorem is replaced by the sentence "is covered by," then the transformed implication is false.

<u>Example 5.</u> Let M be the (deterministic) 4-state machine defined as follows ($X = \{0, 1, 2\}$ $Y = \{0, 1\}$):

$$A(0/0) = \begin{bmatrix} 1 & \\ 1 & 0 \\ 1 & \\ 0 & \end{bmatrix} \quad A(0/1) = \begin{bmatrix} 1 & \\ 1 & 0 \\ 0 & \\ 0 & \end{bmatrix} \quad A(0/2) = \begin{bmatrix} 1 & \\ 0 & 0 \\ 0 & \\ 0 & \end{bmatrix}$$

$$A(1/0) = \begin{bmatrix} 0 & \\ 0 & 0 \\ 0 & \\ 1 & \end{bmatrix} \quad A(1/1) = \begin{bmatrix} 0 & \\ 0 & 0 \\ 1 & \\ 1 & \end{bmatrix} \quad A(1/2) = \begin{bmatrix} 0 & \\ 1 & 0 \\ 1 & \\ 1 & \end{bmatrix}$$

For this machine the matrix H^M is:

$$H^M = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

Thus rank $M = 4 = \#$ of states of M. Nevertheless, it is easy to see that with $B = (1000)$, the 1-state trivial machine M^*, which associates the probability 1 with every input-output pair, is such that $M \geq M^*$.

III. INITIALIZED STOCHASTIC SEQUENTIAL MACHINES

Following Ott [5], we define an initialized stochastic sequential machine (I. S. S. M.) as an S. S. M. together with an a priori fixed initial distribution π. Thus, when we consider an I. S. S. M., (M, π), we are interested only in the designated function P_π^M induced by M with initial distribution π.

In his thesis, Ott considered the following problem: Given an I. S. S. M., (M, π), find an I. S. S. M., (M^*, π^*), with a minimal number of states such that $p_\pi^M = p_{\pi^*}^{M^*}$. He showed that this problem can be reduced to the first covering problem (Sec. I-C here). This fact, although providing insight into the nature of the problem, does not lead to a solution, as the first covering problem does not have as yet an efficient solution either. One is therefore compelled to seek other approaches to this problem, which may lead effectively to a full or partial solution (in some particular cases). One such approach has been made by the author elsewhere [Ref. 7-Theorem 6] and our scope here is to generalize that result. This is done in the following:

Theorem 7. Let (M, π) be a given I. S. S. M. with n-states. If there is a stochastic matrix B with $n^* < n$ rows such that:

(a) $\bigcup_{(y,x)} H^M(B\, A^M(y/x)) \subset \text{conv } H(B)$

(b) $h(\pi) \,\varepsilon\, \text{conv } H(B)$

Then there is an I. S. S. M., (M^*, π^*) with n^* states, such that $p_\pi^M = p_{\pi^*}^{M^*}$. (M, π) can be effectively constructed if a matrix B, as above, is given.

Proof: Condition (a) is identical to condition (10) and it implies therefore that there is an S. S. M., M^* with $n^* < n$ states such that $M \geq M^*$, or $B\, \eta^M(v, u) = \eta^{M^*}(v, u)$ for all pairs (v, u). If B is given, then M can be constructed effectively as in Theorem 5. Condition (b) implies that the point πH^M is a convex combination of the rows of $B\, H^M$, which means that there is a probabilistic vector π^* such that $\pi H^M = \pi^* B\, H^M$ and, as the columns of H^M are a basis for all vectors of the form $\eta^M(v, u)$, we have $\pi\eta^M(v, u) = \pi^* B\eta^M(v, u)$ for all pairs (v, u). Combining the two results we have

$$p_\pi^M(v, u) = \pi\eta^M(v, u) = \pi^* B\eta^M(v, u) = \pi^* \eta^{M^*}(v, u) = p_{\pi^*}^M(v, u)$$

as required. The theorem is thus proved.

Remark: It is easy to see that Theorem 6 in [7] is a particular case of this theorem with B a degenerate stochastic matrix (having one entry equal to 1 in every row, all the other entries being equal to zero).

Corollary 8. Let (M, π) be a given I. S. S. M. with n-states. Let $h_1^* \ldots h_{n^*}^*$ be a set of $n^* < n$ points in m-dimensional space (m = # of columns of H^M) such that (a') $\bigcup H^M(A^M(y, x)) \subset \text{conv}\,(h_1^* \ldots h_{n^*}^*) \subset \text{conv}\,(h_1 \ldots h_n)$

(b') $h(\pi) \in \text{conv}\,(h_1^* \ldots h_n^*)$

Then there is an I. S. S. M., (M^*, π^*) with n^* states such that $p_\pi^M = p_{\pi^*}^M$.

Proof: As in the proof of Theorem 6, the conditions (a') and (b') of the corollary imply the conditions (a) and (b) of Theorem 7 for a matrix B uniquely determined by the set of points $(h_1^* \ldots h_n^*)$.

The following example will illustrate an application to the above corollary.

Example 6. Let (M, π) be the I. S. S. M. defined as follows $(X = Y = \{0, 1\})$: $\pi = (\frac{1}{8}\ 0\ \frac{7}{8}\ 0)$

$$A(0/0) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad A(1/0) = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} & 0 \end{bmatrix}$$

$$A(0/1) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & 0 & \frac{1}{2} & \frac{1}{4} \end{bmatrix} \qquad A(1/1) = \begin{bmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

A matrix H^M for this machine is:

$$H^M = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

It is seen that the convex body whose vertices are the rows of H^M in two-dimensional space is the unit square (the first coordinate of all the rows of H^M is omitted) so that there is no $M^* \geq M$ (with less states).

The points in $\bigcup_{(y,x)} H(A(y, x))$ are:

$(\frac{1}{2}\ 0)$, $(\frac{3}{4}\ \frac{1}{2})$, $(\frac{1}{2}\ \frac{3}{4})$, $(0\ \frac{1}{2})$ and $h(\pi) = (\frac{7}{8}\ \frac{7}{8})$.

The conditions of Corollary 8 will be satisfied if we choose $h_1^* = (\frac{1}{2}\ 0)$ $h_2^* = (1, 1)$ and $h_3^* = (0\ \frac{1}{2})$. The resulting matrix B will be:

$$B = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{bmatrix}$$

and the required I. S. S. M. is found to be (using the methods of the previous sections): $\pi^* = (\frac{1}{12}\ \frac{5}{6}\ \frac{1}{12})$

$$A^{M^*}(0/0) = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad A^{M^*}(1/0) = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

$$A^{M^*}(0/1) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{4} & \frac{1}{4} \end{bmatrix} \quad A^{M^*}(1/1) = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} \end{bmatrix}$$

A closer examination of the above example will show that Theorem 6 in [7] is not applicable to it so that Theorem 7 and Corollary 8 are a proper generalization of that theorem.

IV. UNIQUENESS OF REDUCED FORM AND MINIMAL STATE FORM

A. Reduced Form Machines

Given an S. S. M., one can construct an equivalent reduced form and an equivalent minimal state form. In general, the resulting machine is not unique and may depend upon the sequence of reductions applied to the source machine. The problem of the existence of nonisomorphic equivalent reduced forms, or minimal state forms, has been discussed by Carlyle [2], Even [3] and others. In the following sections we give a geometric interpretation, in the light of the previous sections, to most known, and some unknown common properties of classes of machines, and to the uniqueness problem.

The final section points to the Gale-transform method of investigation which may have uses in the theory of stochastic sequential machines.

Proposition 9. Let M and M^* be two equivalent S. S. M.'s with n and n^* states, respectively. Then rank M = rank M^*, conv $(h_1^M, \ldots, h_n^M)$ = conv $(h_1^{M^*}, \ldots, h_{n^*}^{M^*})$ and there are stochastic matrices B and B^* such that $H^{M^*} = B\, H^M$ and $H^M = B^* H^{M^*}$.

Proof: $M \cong M^*$ implies that $M \geq M^*$ and $M^* \geq M$. By Theorem 3 there are stochastic matrices B and B^* such that $\eta^M(v, u) = B^* \eta^{M^*}(v, u)$ and $\eta^{M^*}(v, u) = B\, \eta(v, u)$ for all pairs (v, u). This implies that rank M $\geq$ rank M^* and rank $M^* \geq$ rank M or rank M = rank M^*. By Theorem 4 $H^{M^*} = B\, H^M$ and $H^M = B^*\, H^{M^*}$. The meaning of the above equations is that every row of H^M is a convex combination of rows of H^{M^*} and conversely, i.e., conv $(h_1^M \ldots h_n^M)$ = conv $(h_1^{M^*}, \ldots, h_{n^*}^{M^*})$.

Definition 7. Two machines M and M^* are state equivalent if for every extremal function in $\mathcal{P}^M$ there is an equal <u>extremal</u> function in $\mathcal{P}^{M^*}$ (see Definition 5) and conversely.

It is easy to show that state equivalence implies equivalence. A consequence of the Definition 7 above is the following:

Proposition 10. Let M and M^* be two state equivalent S. S. M's with n and n^* states, respectively. Then $\{h_1^M, \ldots, h_n^M\} = \{h_1^{M^*}, \ldots, h_{n^*}^{M^*}\}$ (i.e., the set of rows of H^M equals the set of rows of H^{M^*}).

Proof: The entries in a row of the form h_i^M are values of a function p_π^M with π an extremal vector (as the columns in H^M are columns in K^M). The columns of K^{M^*} corresponding to the columns of H^M in

K^M are the columns of H^{M^*} (as the machines are equivalent, Proposition 9 applies here). As M^* is state equivalent to M there is an extremal vector π^* with $p_\pi^M = p_{\pi^*}^{M^*}$. Thus the values of $p_{\pi^*}^{M^*}$ corresponding to those of p_π^M in h_i^M are in some $h_j^{M^*}$ (π^* is extremal), and are equal to the corresponding values in h_i^M. The same argument holds true if M is replaced by M^* and the proposition is proved.

Proposition 11. Let M and M^* be two state equivalent and reduced S. S. M's with n and n^* states, respectively. Then $n = n^*$, the rows of H^M are a permutation of the rows of H^{M^*} and, if $A^M(y/x)$ and $A^{M^*}(y/x)$ are corresponding matrices of M and M^*, respectively, then $A^M(y/x) \simeq A^{M^*}(y/x)$, relative to H^M (see Sec. II-B for definition of $\simeq$).

Proof: It follows from the definitions (see Remark 1 in Sec. I-E) that no two rows of H^M and no two rows of H^{M^*} are equal (M and M^* are reduced). By the previous proposition $\{h_1^M, \ldots, h_n^M\} = \{h_1^{M^*}, \ldots, h_{n^*}^{M^*}\}$. Combining these facts we have $n = n^*$, and the ordered set of rows of H^M is a permutation of the ordered corresponding set of rows of H^{M^*}. Let $A^M(y/x)$ and $A^{M^*}(y/x)$ be two corresponding matrices, and assume that $H^M = H^{M^*}$ (this is not a restriction as the above argument implies that the equality will hold true if the states of H^{M^*} are ordered in such a way that corresponding states will have the same index), then, as the machines are state equivalent, and the equivalence is one-one, we have:

$\eta^M(v/u) = \eta^{M^*}(v/u)$ for all pairs (v, u) so that

$A^M(y/x) \simeq A^{M^*}(y/x)$ relative to H^M (or to H^{M^*}),

for $\eta(yv, xu) = A(y/x) \cdot \eta(v/u)$.

Remark. The last implication of Proposition 11 is due to Carlyle [2].

Definition 8. Let M and H^M be an n-state S. S. M. and its corresponding H^M matrix. A substochastic n-dimensional vector π is simplicial if $h^M(\pi)$ is a point in a face of conv $(h_1, \ldots, h_n)$ which is a simplex.

Theorem 12. Let M be a reduced S. S. M. There is a reduced S. S. M., M^* which is state-equivalent but not isomorphic to M iff there is a row $\xi_i(y/x)$ in a matrix $A^M(y/x)$ which is not simplicial (two machines are isomorphic if they are equal up to a permutation of states).

Proof: Assume first that all the rows in the matrices $A^M(y/x)$ are simplicial. If there is a reduced machine M^* which is state-equivalent to M then, by Proposition 11, we have $A^M(y/x)\, H^M = A^{M^*}(y/x)\, H^M$ (after a proper rearrangement of states), for all pairs (y, x). As the rows of $A^M(y/x)$ are simplicial, this is possible only if $A^M(y/x) = A^{M^*}(y/x)$ up to isomorphism (for a point in the interior of a simplex has a unique representation as a combination of its vertices). Thus M is isomorphic to M^*. Assume now that there is a row $\xi_i(y/x)$ in a matrix $A^M(y/x)$ which is not simplicial. This means that $h(\xi_i(y/x)) = \sum \alpha_i h_i$ where the h_i corresponding to nonzero coefficients α_i are not a simplex. This implies, by the classical theorem of Rado on convex bodies, that there is a set of coefficients (β_i) different from the set (α_i) such that the combination $\sum \beta_i h_i$ is convex and equals the combination $\sum \alpha_i h_i$. Thus, there is a substochastic vector ρ_i which differs from $\xi_i(y/x)$

such that $h^M(\xi_i(y/x)) = h^M(\rho)$. Let M^* be a machine derived from M by replacing the vector $\xi_i(y/x)$ in $A^M(y/x)$ by the vector ρ_i. For any pair (v, u) we have $p_\pi^{M^*}(v, u) = p_\pi^M(v, u)$ as the columns $\eta^M(v, u)$ are linear combinations of the columns of H^M and the above replacement does not affect the columns of H^M. Thus M^* is a machine which is state-equivalent but not isomorphic to M.

Theorem 13. Let M be an S. S. M. The construction of a reduced form through the procedure of merging equivalent states may lead to resultant machines which are not isomorphic only if there are two rows $\xi_i(y/x) \neq \xi_j(y/x)$ in a matrix A(y/x) which are not simplicial such that $h_i = h_j$ and $h(\xi_i(y/x)) = h(\xi_j(y/x))$.

Proof: In the merging procedure, some rows are deleted and the corresponding columns are added to other columns. After the reduced machine M^* is obtained, a row $\xi_i^*(y/x)$ in a matrix $A^{M^*}(y/x)$ will represent the same point in conv $(h_1^{M^*}, \ldots, h_{n^*}^{M^*}) = \text{conv}\,(h_1^M, \ldots, h_n^M)$ as the corresponding row in $A^M(y/x)$ of M. If in the merging procedure the original row $\xi_i(y/x)$ is deleted then another row, say $\xi_j(y/x)$ such that the states i and j are equivalent, is kept for M^*, and $\xi_j^*(y/x)$ will represent that same point in conv $(h_1^{M^*}, \ldots, h_{n^*}^{M^*})$. Thus $\xi_j^*(y/x)$ is simplicial if $\xi_i(y/x)$ is. If $\xi_j^*(y/x)$ is simplicial then its representation as a convex combination of the distinct points $(h_1^{M^*}, \ldots, h_{n^*}^{M^*})$ is unique and is therefore independent (up to isomorphism) upon the state which is chosen to remain or the states which are chosen to be deleted.

If two states i and j are equivalent but for some (y, x), $\xi_i(y/x)$ and $\xi_j(y/x)$ are not simplicial and the condition of the theorem does not hold. Then $\xi_i(y/x) = \xi_j(y/x)$ so that the resulting machine M^* is independent (up to isomorphism) upon which of the two states are deleted. It follows that if the condition of the theorem does not hold, then the resulting reduced machine M^* is unique up to isomorphism.

B. Minimal State Form Machines

Let M and M^* be two equivalent minimal state form machines with n and n^* states respectively. It follows from the definitions (see Remark 2 in Sec. I-E) that the rows of H^M and H^{M^*} are distinct vertices of convex polyhedra. By Proposition 9 (M is equivalent to M^*) we have conv $(h_1^M, \ldots, h_n^M) =$ conv $(h_1^{M^*}, \ldots, h_{n^*}^{M^*})$. Combining these facts together we have $n = n^*$ and $(h_1^M, \ldots, h_2^M) = (h_1^{M^*}, \ldots, h_{n^*}^{M^*})$ with all elements in both sets distinct. This implies that there are permutation $n \times n$ matrices B and B^* such that $H^{M^*} = B\,H^M$ and $H^M = B^*\,H^{M^*}$ (this is another proof of a theorem of Ott [5]). If only one of the two machines, say M, is in minimal state form, then $n < n^*$ for the set $(h_1^M, \ldots, h_n^M)$ is the set of vertices of conv $(h_1^{M^*}, \ldots, h_{n^*}^{M^*})$ (this is another proof of a theorem of Bacon [1-Corollary 1] . If both M and M^* are minimal state forms, then there are permutation $n \times n$ matrices B and B' such that $\eta^{M^*}(v, u) = B\,\eta^M(v, u)$ and $\eta^M(v, u) = B^*\,\eta^{M^*}(v, u)$, for (see Proposition 9) rank M = rank M^* and there are stochastic matrices B and B^* satisfying these equalities (by definition, as $M \geq M^*$ and $M^* \geq M$), this implies (Theorem 4) that $H^M = B^* H^{M^*}$ and $H^{M^*} = B H^M$, so that B and B^* must be permutation matrices (the rows of H^M and H^{M^*} are distinct and the rows of one matrix are a permutation of the rows of the second). The equalities $\eta^{M^*}(v, u) = B \eta^M(v, u)$, $\eta^M(v, u) = B^* \eta^{M^*}(v, u)$ with B and B^* permutation matrices imply that M is state equivalent to M^* as is easy to see (this is another proof of a known theorem of Bacon [1-Theorem 1]). We have thus shown that if M and M^* are minimal form and equivalent, then they are state equivalent. It is clear that M and M^* are also in reduced form (the minimal state form is a restriction of the reduced form). We have as a consequence that (Proposition 11) if $A^M(y/x)$ and $A^{M^*}(y/x)$ are corresponding matrices of M and M^* respectively, then $A^M(y/x) \simeq A^{M^*}(y/x)$ relative to H^M (or to H^{M^*}). On the basis of the above properties two theorems, parallel to Theorems 12 and 13, can be proved (the proof is left to the reader), namely:

Theorem 14. Let M be a minimal state S. S. M. There is a minimal state M^* which is equivalent but not isomorphic to M iff there is a row $\xi_i(y/x)$ in a matrix $A^M(y/x)$ which is not simplicial.

Theorem 15. Let M be an S. S. M. The construction of a minimal state form machine equivalent to M may lead to resultant machines which are not isomorphic only if there are two rows $\xi_i(y/x) \neq \xi_j(y/x)$ in a matrix $A(y/x)$ which are not simplicial such that $h_i = h_j$, $h_i(h_j)$ is a vertex of $\operatorname{conv}(h_1, \ldots, h_n)$ and $h(\xi_i(y/x)) = h(\xi_j(y/x))$.

Note that the additional requirement, that $h_i(h_j)$ is a vertex of conv $(h_i, \ldots, h_n)$, is necessary because all rows $\xi_i(y/x)$ such that h_i is not a vertex, will be deleted in the minimization process (see Bacon [1]).

Remarks: Carlyle introduced a particular class of sequential machines called observer/state-calculable machines [2]. These machines have the property that every row in any matrix $A^M(y/x)$ has at most one nonzero entry. Carlyle has shown that the merging of states procedure for such machines leads to a unique reduced equivalent machine. This property of the above class of machines is a result of the fact that the merging procedure preserves the membership in the class. On the other hand, an observer/state-calculable machine may have a minimal state-equivalent form which is not observer/state-calculable (and is not unique). Nevertheless, if in a particular case a machine in the above class is such that its equivalent minimal state form remains in the class, independently of the sequence of reductions, then this minimal state-equivalent form machine is unique. This follows from the fact that all rows in all matrices of a minimal state form, observer/state-calculable machine, are simplicial (all the rows of H are vertices and all nonzero rows of the matrices have only one nonzero entry).

C. A Decision Procedure

We have seen in the previous sections that the uniqueness of the reduced form or minimal state form of a given machine M depends upon the nature of the points $h^M(\xi(y/x))$, where $\xi(y/x)$ in a row in a matrix $A^M(y/x)$. To find out the nature of these points one must be able to extract from the set of points $(h_1^M \ldots, h_n^M)$ to be denoted by V, all the subsets W such that conv (W) is a face of conv (V). After this is achieved one must be able to decide whether the faces, conv (W), are simplexes or not. A decision procedure for these questions, based on a method introduced by Gale, has been pointed out by Perles (see Grünbaum [4]). Here we shall give the theorems upon which the decision procedure is based, without proof. The reader is referred to [4] for proof.

Let M be an S. S. M. with corresponding $n \times m$ matrix H^M. With H^M we associate a new matrix $\bar{H}^M$ such that the columns of $\bar{H}^M$ form a basis for the null space of the columns of H^M (i.e., if $\bar{\eta}$ is a column of $\bar{H}^M$, then $tr(\bar{\eta})\, H = \vec{0}$, and any vector ξ such that $\xi H^M = \vec{0}$ is a linear combination of the columns of $\bar{H}^M$). Clearly, $\bar{H}^M$ is an $n \times (n - m)$ matrix. If V is the n-tuple $V = (h_1^M, \ldots, h_n^M)$ of points in m-space, then $\bar{V} = (\bar{h}_1^M, \ldots, \bar{h}_n^M)$, the n-tuple of points in (n - m)-space, which are the rows of $\bar{H}^M$, is called the Gale-transform of V. Many geometric properties of the n-tuple V have as counterparts meaningful geometric properties of its Gale-transform $\bar{V}$. The properties relevant to the decision problem described above are listed below.

Let $J = \{i_1, \ldots, i_k\}$ be a subset at the set $N = \{1, \ldots, n\}$. We shall denote by V(J) the k-tuple $V(J) = (h_{i_1}^M, \ldots, h_{i_k}^M)$ and similarly $\bar{V}(J)$ denotes the k-tuple $\bar{V}(J) = (\bar{h}_{i_1}, \ldots, \bar{h}_{i_k})$. If $W = V(J)$, then $V \sim W$

stands for $V(N - J)$. A k-tuple $W = V(J)$ is a coface of V if $\text{conv}(V \sim W)$ is a face of conv(V) (we shall say also that $V \sim W$ is a face of V).

1. $W = V(J) \subset V$ is a coface of V iff either $W = \phi$ or $\vec{0} \in$ relative interior of $\text{conv}(\overline{V}(J))$. (The convex polyhedron conv(V) as a whole is also considered as a face of itself.)

2. Let $V(J) = W$ be a face of V. Then this face is a simplex iff the dimension of $\text{conv}(\overline{V}(J_1))$ equals the dimension of $\text{conv}\,\overline{V}(J)$ for every set $J_1 \subset J$ such that $V(J_1)$ is a nonempty coface of $V(J)$.

ACKNOWLEDGMENTS

The author wishes to express his gratitude to the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, California, and to the Air Force Office of Scientific Research, Office of Aerospace Research, U. S. Air Force, which provided him with ideal working conditions, and also to his colleagues with whom he had many fruitful discussions.

REFERENCES

[1] G. C. Bacon, "Minimal-state stochastic finite-state systems," IEEE Trans. Circuit Theory CT-11, June 1964, pp. 307-308.

[2] J. W. Carlyle, Stochastic finite-state system theory, " in System Theory, L. A. Zadeh and E. Polak, Editors, Ch. 10, McGraw-Hill (in press).

[3] S. Even, "Comments on the minimization of stochastic machines," IEEE Transactions on E. C., Vol. 14, No. 4, August 1965, pp. 634-637.

[4] B. Grünbaum, Convex Polytopes, Interscience, August 1967.

[5] E. H. Ott, "Theory and applications of stochastic sequential machines," Sperry Rand Research Center, Report No. SRRC-RR-66-39, May 1966.

[6] E. H. Ott, "Reconsider the state minimization problem for stochastic finite state systems," IEEE conference record of the seventh annual symposium on switching and automata theory, October 1966, pp. 267-273..

[7] A. Paz, "Minimization theorems and techniques for sequential stochastic machines," to appear in the special issue of Inf. and Control on the Asilomar Conference.

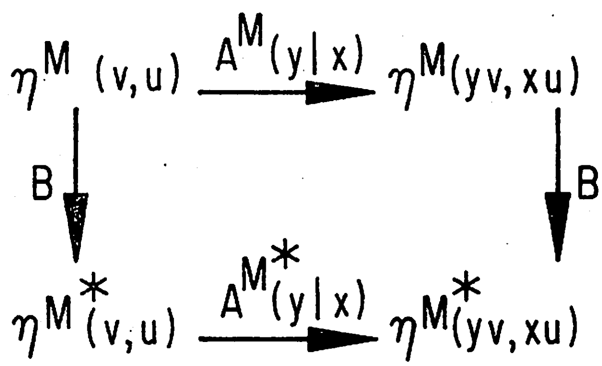

Fig. 1. Mapping B from machine M to machine M^*.

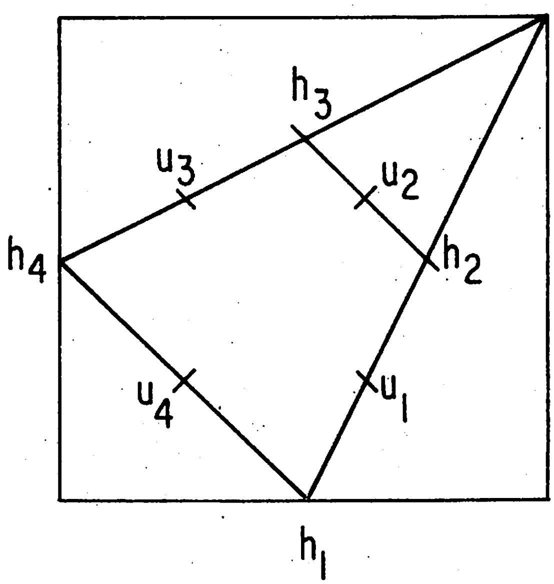

Fig. 2. Geometrical representation of H^M.