SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD A050344

DDC FILE COPY

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR-TR- 78-0184	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): STATE-ESTIMATION OF PARTIALLY-OBSERVED MARKOV CHAINS: DECOMPOSITION, CONVERGENCE, AND COMPONENT IDENTIFICATION	5. TYPE OF REPORT & PERIOD COVERED: Interim rept.
	6. PERFORMING ORG. REPORT NUMBER: ESL-P-790
7. AUTHOR(s): LOREN K. PLATZMAN	8. CONTRACT OR GRANT NUMBER(s): AFOSR-77-3281
9. PERFORMING ORGANIZATION NAME AND ADDRESS: MASSACHUSETTS INSTITUTE OF TECHNOLOGY, ELECTRONIC SYSTEMS LABORATORY, CAMBRIDGE, MASS. 02139	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F 2304/A1
11. CONTROLLING OFFICE NAME AND ADDRESS: AIR FORCE OFFICE OF SCIENTIFIC RESEARCH/NM, DEPARTMENT OF THE AIRFORCE, BOLLING AIR FORCE BASE, D.C. 20332	12. REPORT DATE: 15 Dec 1977
	13. NUMBER OF PAGES: 41
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): AFOSR TR-78-0184	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A partially-observed Markov chain (S,Y) consists of an N-state Markov chain S, along with a process Y of noisy observations of the transitions of S. A metric on stochastic N-vectors and a generalized ergodic coefficient on the transition probability matrices of (S, Y) are defined, resulting in a notion (similar to weak ergodicity) of deteriorating dependence on initial value in a process of distributions of the state (of S) conditioned on past observations. If (S, Y) is stationary, then S may be decomposed into M≤N

next page

127 200 JOB

Best
Available
Copy

20. components, where M=1 neither implies nor is implied by ergodicity of S, such that conditional state distributions within each component geometrically approach an initial-value-independent process in the manner described above, and one or more equivalent components eventually dominate the others. A method for drift-free finite-memory approximation (or realization) of this process is also introduced.

AFOSR-TR- 78-0184

December 15, 1977 ESL-P-790

STATE-ESTIMATION OF PARTIALLY-OBSERVED MARKOV CHAINS: DECOMPOSITION, CONVERGENCE, AND COMPONENT IDENTIFICATION***

by

Loren K. Platzman

*This research was conducted at the MIT Electronic Systems Laboratory with support extended by the Air Force Office of Scientific Research under grants AFOSR-72-2273 and 77-3281.

Paper submitted to Annals of Probability

STATE-ESTIMATION OF PARTIALLY-OBSERVED MARKOV CHAINS: DECOMPOSITION, CONVERGENCE, AND COMPONENT IDENTIFICATION

BY

LOREN K. PLATZMAN

BELL TELEPHONE LABORATORIES

NAPERVILLE, ILLINOIS 60540

SUMMARY

A partially-observed Markov chain $(\underline{S},\underline{Y})$ consists of an N-state Markov chain $\underline{S}$, along with a process $\underline{Y}$ of noisy observations of the transitions of $\underline{S}$. A metric on stochastic N-vectors and a generalized ergodic coefficient on the transition probability matrices of $(\underline{S},\underline{Y})$ are defined, resulting in a notion (similar to weak ergodicity) of deteriorating dependence on initial value in a process of distributions of the state (of $\underline{S}$) conditioned on past observations. If $(\underline{S},\underline{Y})$ is stationary, then $\underline{S}$ may be decomposed into $M \leq N$ components, where $M=1$ neither implies nor is implied by ergodicity of $\underline{S}$, such that conditional state distributions within each component geometrically approach an initial-value-independent process in the manner described above, and one or more equivalent components eventually dominate the others. A method for drift-free finite-memory approximation (or realization) of this process is also introduced.

Submitted for Publication in the Annals of Probability

1. Introduction

Consider a probability space $(\Omega, \underline{F}, \underline{P})$. For ease of notation, random variables (i.e., symbols that implicitely denote measurable functions of $(\Omega, \underline{F})$) will be shown in boldface.

Let $\underline{S} = \{\underline{s}(k) \epsilon S : k \epsilon I\}$ and $\underline{Y} = \{\underline{y}(k) \epsilon Y : k \epsilon I\}$ be stochastic processes where $S = \{1, \ldots, N\}$, Y is a finite set, and $I = \{\ldots,-1,0,1,\ldots\}$. Also let the following "splitting" condition be satisfied:

For any $k \epsilon I$, the "past" $\{\underline{s}(k'), \underline{y}(k') : k' < k\}$
and the "future" $\{\underline{y}(k'-1), \underline{s}(k') : k' > k\}$
are (conditionally) independent given $\underline{s}(k)$.

Then $\underline{S}$ is a <u>finite-state Markov chain</u> (FSMC) (the state process), and $\underline{Y}$ is a process of noisy observations of the transitions of $\underline{S}$ (the observation process). The pair $(\underline{S}, \underline{Y}) = \{(\underline{s}(k), \underline{y}(k)) : k \epsilon I\}$ is called a <u>partially-observed Markov chain</u> (POMC).

Since $(\underline{S}, \underline{Y})$ is itself a FSMC, the probability measure $\underline{P}$ may be completely specified by a finite array of time-indexed transition probabilities. We will adopt a notation associated with probabilistic automata, $\{P^k(y) : y \epsilon Y, k \epsilon I\}$, where $P^k(y) = [P^k_{ij}(y)]$ is an $N \times N$ matrix having entries

$$(1.1) \qquad P^k_{ij}(y) = Pr\{\underline{s}(k+1)=j, \underline{y}(k)=y \mid \underline{s}(k)=i\} \qquad i,j \epsilon S,\ y \epsilon Y,\ k \epsilon I.$$

Thus $P^k(y)$ is a substochastic matrix and $\Sigma_{y\epsilon Y} P^k(y)$ is a stochastic matrix, the transition probability matrix of $\underline{S}$.

Consider an observer who wishes to characterize the "future" $\{\underline{s}(k'),\underline{y}(k') : k'\geq k\}$, given a finite string of most recent observations

(1.2) $\underline{y}(k_o,k) = \underline{y}(k_o)\ \underline{y}(k_o+1) \ldots \underline{y}(k-1).$

This is accomplished by computing the horizontal N-vector $\underline{\eta}(k_o,k)$ having entries

(1.3) $\underline{\eta}_i(k_o,k) = Pr\{\underline{s}(k)=i \mid \underline{y}(k_o,k)\}$ $\qquad i\epsilon S$

and lying in the simplex

(1.4) $\Pi = \{\pi : \pi_i\geq 0\ \forall i\epsilon S,\ \Sigma_{i\epsilon S}\pi_i=1\}.$

Clearly, $\underline{\eta}(k_o,k)$ is the a posteriori probability distribution of $\underline{s}(k)$ given observations $\underline{y}(k_o,k)$. Moreover, in the sense that each state in S is associated with an extreme point of Π, $\underline{\eta}(k_o,k)$ is a conditional state expectation, and hence a least-squares state estimate, at time k, given observations $\underline{y}(k_o,k)$.

Let $\nu = (1,1,\ldots,1)^t$ and define

(1.5) $T(\pi,P) = \pi P/\pi P\nu$ $\qquad \pi P\nu>0.$

The state estimate may now be recursively updated according to the identity:

(1.6) $\underline{\eta}(k_0,k+1) = T(\underline{\eta}(k_0,k),\ P^k(\underline{y}(k)))$.

We note that $\{\underline{\eta}(k_0,k) : k=k_0,k_0+1,\ \ldots\}$ is therefore a Markov process and may, itself, be viewed as the "state" in an alternate realization of $\underline{Y}$.

Multiple-step transitions are similarly expressed. Let $z = y_1 \cdots y_n$ and define the matrix product:

(1.7) $P^k(z) = P^k(y_1) \cdot P^{k+1}(y_2) \cdot \ldots \cdot P^{k+n-1}(y_n)$.

It is easily seen that

(1.8) $\underline{\eta}(k_0,k+n) = T(\underline{\eta}(k_0,k),\ P^k(\underline{y}(k,k+n)))$.

We also define a process of state estimate approximations constructed as in (1.8), but from an initial value $\underline{\pi}$:

(1.9) $\underline{\eta}(\underline{\pi};\ k_0,k) = T(\underline{\pi},\ P^k(\underline{y}(k_0,k)))$.

For any $k \in I$, the process $\{\underline{\eta}(k-n,k) : n = 0,1,\ \ldots\}$ is a uniformly bounded martingale. Hence, by Proposition V-2-6 of Neveu [9], there exists a vector $\underline{\eta}(k)$ satisfying

(1.10) $\underline{\eta}(k) = \lim_{k_0 \to -\infty} \underline{\eta}(k_0,k)$ a.s.

Since $T(\cdot,P)$ is continuous, (1.8) and (1.9) yield

(1.11) $\underline{\eta}(k) = T(\underline{\eta}(k_0),\ P^{k_0}(\underline{y}(k_0,k))) = \underline{\eta}(\underline{\eta}(k_0);\ k_0,k)$.

This paper examines the evolution of the state estimate approximation process $\{\underline{\eta}(\underline{\pi};\ k_0,k) : k=k_0,k_0+1,\ \ldots\}$. Specifically, we establish conditions under which $\underline{\eta}(\underline{\pi};\ k_0,k) \to \underline{\eta}(k)$, in some sense, as $k\to\infty$. To this end, a metric Δ on Π and a generalized ergodic coefficient α on substochastic matrices are defined, in Sections 3 and 4, respectively, so that

(1.12) $$\tfrac{1}{2}\ \Sigma_{i\varepsilon S}\ |\pi_i - \pi_i'| \leq \Delta[\pi,\pi'] \leq 1,$$

(1.13) $$\alpha[PQ] \leq \alpha[P]\ \alpha[Q] \leq 1,$$

(1.14) $$\Delta[T(\pi,P),\ T(\pi',P)] \leq \alpha[P]\ \Delta[\pi,\pi'].$$

It follows that $\Delta[\underline{\eta}(\underline{\pi};\ k_0,k),\ \underline{\eta}(k)] \leq \Pi_{k'=k_0}^{k-1}\ \alpha[P^{k'}(\underline{y}(k'))]$, and that both sides of the inequality are nonincreasing in k, a.s. Sufficient conditions for this bound to vanish as $k\to\infty$ are given by Theorems (5.3) and (5.4). Theorem (6.10) shows that if $(\underline{S},\underline{Y})$ is stationary and certain conditions (concerning future-dependence of $\underline{\pi}$ and state-minimality of $(\underline{S},\underline{Y})$) are satisfied, then

(1.15) $$\lim_{k\to\infty}\ \Sigma_{i\varepsilon S}\ |\underline{\eta}_i(\underline{\pi};\ k_0,k) - \underline{\eta}_i(k)| = 0 \quad \text{a.s.}$$

although it may simultaneously hold that

(1.16) $$\Delta[\underline{\eta}(\underline{\pi};\ k_0,k),\ \underline{\eta}(k)] = 1 \quad \text{a.s.},\ k=k_0,k_0+1,\ldots\ .$$

Section 6 also introduces a stochastic process $\{\underline{\Sigma}(k) : k\varepsilon I\}$ which, under these conditions, displays the following properties (made explicit by (6.1), (6.2) and (6.7) respectively):

(a) Decomposition: $\underline{\Sigma}(k)$ is a collection of disjoint subsets of S and $\underline{s}(k)$ is contained by some element $\underline{\sigma}(k)$ of $\underline{\Sigma}(k)$ a.s.

(b) Recursive Calculability: $\underline{\sigma}(k)$ may be expressed as a (deterministic) function of $\underline{\sigma}(k_0)$ and $\underline{y}(k_0,k)$.

(c) Convergence: $\Delta[\underline{\eta}(\underline{\pi};\ k_0,k),\ \underline{\eta}(k)]\to 0$, geometrically in k, as $k\to\infty$, provided that $\{i\ :\ \underline{\pi}_i > 0\} \subseteq \underline{\sigma}(k_0)$.

Thus $\underline{\eta}(\underline{\pi};\ k_0,k)$ may be decomposed into $\underline{M} = \#\underline{\Sigma}(k_0)$ components, one of which may converge (in Δ) to $\underline{\eta}(k)$ as $k\to\infty$. The convergent component is identified by maximum likelihood on the basis of past or future observations. A necessary and sufficient condition for $\{\underline{\eta}(k)\}$ itself to be a FSMC is given by Theorem (6.14).

In analogy to (finite-dimensional) linear system theory, the ergodic classes of $\underline{S}$ in S may be likened to controllable subspaces, whereas the subsets $\underline{\Sigma}(k_0)$ of S may be likened to observable subspaces. It is consistent with this analogy that $M = \#\underline{\Sigma}(k) = 1$ neither implies, nor is implied by, ergodicity of $\underline{S}$.

POMCs have been studied by many authors, notably Blackwell [2], Drake [5] Astrom [1], Sondik [13], (also see [12]), and Rudemo [11]. Using a result of Furstenberg and Kesten [6], Kaijser [8] has obtained a weaker version of Theorem (5.3), and conjectures Theorem (6.10). A similar result has been derived by Devore [4]. Ergodic coefficients are discussed at length in Chapter 5 of Isaacson and Madsen [7]. A POMC is also a free probabilistic automaton; see Paz [10].

The analysis given here may be generalized to processes having denumerable state-output sets, to certain continuous-time processes (following Rudemo [11]), and to semi-Markov processes (following White [14]).

2. Notations

Sets: If A and B are sets, then A-B is the set of elements in A that are not contained in B. #A is the number of elements in A. $\emptyset$ is the null set.

Strings: $Y^k(n)$ is the set of strings $z = y_1 y_2 \cdots y_n$ satisfying $P^k(z) \neq 0$. $Y^k(*) = \bigcup_{n=1}^{\infty} Y^k(n)$. If $z = y_1 \ldots y_n$ and $\tilde{z} = \tilde{y}_1 \ldots \tilde{y}_m$ then $z\tilde{z}$ denotes the concatenation $y_1 \cdots y_n \tilde{y}_1 \cdots \tilde{y}_m$.

Scalars: $a \div b$ denotes integer quotient rounded down; i.e., $q = a \div b$ is the integer of largest magnitude satisfying $|qb| \leq |a|$ and $\mathrm{sgn}(q) = \mathrm{sgn}(a/b)$. Also $(a)^+ = \max(0,a)$

Vectors: R_N is the Euclidean space of vertical N-vectors. For $\sigma \subseteq S$, e^σ is a row N-vector with entries $e_i^\sigma = 1$ if $i \in \sigma$, $e_i^\sigma = 0$ otherwise. For $i \in S$, e^i denotes the transpose of $e^{\{i\}}$, the i^{th} unit stochastic vector in Π.

Faces of Π: For any subset σ of S, $\Pi(\sigma)$ denotes the set of stochastic vectors π satisfying:

$$\pi_i > 0 \iff i\epsilon\sigma, \qquad\qquad \forall i\epsilon S.$$

Bayes' Operator: For $\pi\epsilon\Pi$, $w\epsilon R_N$, with $w_i \geq 0$ $\forall i\epsilon S$ and $\pi w > 0$, let $\pi \circ w$ denote the vector in Π having entries

$$(\pi \circ w)_i = \pi_i w_i / \pi w.$$

This may be interpreted as follows: Let π represent a priori probabilities of some random variable, s, on sample space S. Consider an event A occuring with conditional probability w_i provided that s = i. Then $\pi \circ w$ is the vector of a posteriori probabilities of random variable s, given A.

Stochastic Vector Restriction: For $\pi\epsilon\Pi$ and $\sigma \subseteq S$ with $\Sigma_{i\epsilon\sigma} \pi_i > 0$, define

$$[\pi|\sigma] = \pi \circ e^{\sigma}.$$

Subrectangularity: An NxN substochastic matrix P is M-subrectangular if there exist collections $\{I_1, \ldots, I_M\}$ and $\{J_1, \ldots, J_M\}$, of mutually disjoint, nonempty subsets of S such that

$$\{(i,j) : P_{ij} > 0\} = (I_1 \times J_1) \cup \cdots \cup (I_M \times J_M).$$

A 1-subrectangular matrix is also called subrectangular.

__State Estimate Approximations__: $\Pi(k)$ denotes the set of random variables $\underline{\pi}$ having values in Π such that $\underline{\pi}$ is (conditionally) independent of the "future" $\{\underline{y}(k'-1), \underline{s}(k') : k'>k\}$ given $\underline{s}(k)$, and $\underline{\pi}_{\underline{s}(k)} > 0$, a.s. $\Pi(k)$ contains each of the following: ν/N, $\underline{\pi}(k_o,k)$, $\underline{\pi}(k)$, $e^{\underline{s}(k)}$, and $\underline{\pi}(\underline{\pi}; k_o,k)$ if $\underline{\pi}\epsilon\Pi(k_o)$.

__Stopping Times__: Let $\underline{Y}(k_o,k)$ denote the sigma-algebra generated by $\{\underline{y}(k_o,k)\}$. Then $\Theta_f(k_o)$ is the set of random variables $\underline{n}$ having positive integer values and satisfying $\{\underline{n} = k-k_o\}\epsilon\underline{Y}(k_o,k)$ $\forall k>k_o$; and $\Theta_b(k)$ is the set of random variables $\underline{n}$ having positive integer values and satisfying $\{\underline{n} = k-k_o\}\epsilon\underline{Y}(k_o,k)$ $\forall k_o<k$.

3. __A Metric on Π__

This section introduces a metric that is used to measure the "closeness" of state estimate approximations.

(3.1) __Definition__: For $\pi,\pi'\epsilon\Pi$, define

(a) $|\pi - \pi'| = \Sigma_{i\epsilon S} |\pi_i - \pi_i'|$;

(b) $\delta[\pi,\pi'] = \Sigma_{i\epsilon S} (\pi_i-\pi_i')^+$;

(c) $\Delta[\pi,\pi'] = \sup\{\delta[\pi\cdot w, \pi'\cdot w] : w\epsilon R_N,\ w_i\geq 0\ \forall i\epsilon S,\ \pi w>0,\ \pi' w>0\}$.

(3.2) <u>Lemma</u>: $|\cdot - \cdot|$, δ and Δ are metrics on Π, and

$$0 \leq \tfrac{1}{2}|\pi - \pi'| = \delta[\pi,\pi'] \leq \Delta[\pi,\pi'] \leq 1.$$

<u>Proof</u>: Trivial.

(3.3) <u>Proposition</u>: (Evaluation of Δ). For $\pi,\pi'\epsilon\Pi$, define:

$$c_1 = \min\{\pi_i'/\pi_i : \pi_i>0\};$$
$$c_2 = \min\{\pi_i/\pi_i' : \pi_i'>0\}.$$

Then

$$\Delta[\pi,\pi'] = \frac{1 - \sqrt{c_1 c_2}}{1 + \sqrt{c_1 c_2}}$$

<u>Proof</u>: The proof is given in Section 3'.

The metric δ, also known as the <u>Hajnal measure</u>, has many applications in the theory of ergodic Markov chains [7]. Informally, $\delta[\pi,\pi']$ is the (minimal) "quantity of probability" that would have to be "reassigned" in order to transform probability distribution π into probability distribution π'. Similarly, $\Delta[\pi,\pi']$ is the least upper bound on the quantity of conditional probability by which π and π' might differ if they were conditioned on identical observations.

The distinction between δ and Δ is also illuminated by an examination of the topologies they induce on $\mathbb{X}$: the topology induced by δ is connected, but Δ causes $\mathbb{X}$ to be separated into the 2^N-1 faces $\{\mathbb{X}(\sigma),\ \emptyset \subset \sigma \subseteq S\}$.
In the computation of state estimates, this separation can be significant. Consider a POMC in which $\underline{\eta}(0) = (1-\epsilon, \epsilon)$, $\epsilon<<1$, but it is desired to approximate $\underline{\eta}(0)$ by e^1. In a δ sense, $\underline{\eta}(0)$ is "near" the approximation e^1; the unconditional expectation of a bounded function of the initial state will be little affected by this approximation. Suppose, however, that every output that subsequently evolves corresponds to transition probabilities $Q = \begin{bmatrix} .1 & 0 \\ 0 & .9 \end{bmatrix}$. Now $\underline{\eta}(\underline{\eta}(0);\ 0,k) = T(\underline{\eta}(0),[Q]^k)$ tends to e^2; yet if the approximation $\underline{\eta}(0) = e^1$ is used, then $\underline{\eta}(e^1;\ 0,k) = T(e^1,[Q]^k) = e^1$ is obtained.
Thus an initial error, of δ-sense magnitude $\epsilon<<1$, may lead to an eventual error of δ-sense magnitude close to 1, albeit after conditioning on an event of probability less than ϵ. Proposition (4.3) will show that such increase in Δ-sense approximation error cannot occur.

3'. <u>Proof of Proposition (3.3)</u>

Clearly $\Delta[\pi,\pi] = 0$. If $\{i : \pi_i>0\} \neq \{i : \pi_i'>0\}$ then $\Delta[\pi,\pi'] = 1$. To see this, assume without loss of generality that $i\epsilon S$ is such that $\pi_i > 0$ and $\pi_i' = 0$, and define $w^m=(1-1/m)e^i+(1/m)\nu/N$.
Then $\{w^m\}$ is a sequence in Π for which $\lim_{m\to\infty} \delta[\pi\circ w^m,\pi'\circ w^m] = 1$; since $(\pi\circ w^m)_i \to 1$ as m increases, but $(\pi'\circ w^m)_i = 0$.
By (3.2), the sequence $\{w^m\}$ is supremal.

The case $\{i : \pi_i>0\} = \{i : \pi_i'>0\}$, $\pi \neq \pi'$, remains.
Assume without loss of generality that $\pi_i>0$ and $\pi_i'>0$, $\forall i\epsilon S$.
Now $0 < c_1 < 1$ and $0 < c_2 < 1$; hence $0 < c_1 < c_2^{-1} < \infty$. Define:

$$\Delta_\zeta[\pi,\pi'] = \sup\{\delta[\pi\circ w,\pi'\circ w] : w\epsilon R_N,\ w_i\geq 0\ \forall i\epsilon S,\ \pi' w/\pi w=\zeta\}$$

$$= \sup\{\Sigma_{i\epsilon S}\ \pi_i w_i - \frac{\pi_i' w_i}{\zeta}^{+} : w\epsilon R_N,\ w_i\geq 0\ \forall i\epsilon S,\ \pi w=1,\ \pi' w=\zeta\}$$

which exists for all $c_1\leq\zeta\leq c_2^{-1}$. Clearly

$$\Delta[\pi,\pi'] = \sup\{\Delta_\zeta[\pi,\pi'] : c_1\leq\zeta\leq c_2^{-1}\}.$$

But $\Delta_\zeta[\pi,\pi']$ may be expressed as the solution of a linear program

$$\Delta_\zeta[\pi,\pi'] = \left[\begin{array}{ll} \text{max:} & aw \\ \text{subject to:} & \pi w = 1 \\ & \tilde{\pi}' w = 1 \\ & w \geq 0 \end{array}\right]$$

where $a_i = (\pi_i - \tilde{\pi}_i')^+$, $\tilde{\pi}_i' = \pi_i'/\zeta$. Any optimal basic w that solves this linear program has at most two non-zero entries. Let their indices be denoted (i,j) and assume without loss of generality that

$$(i,j) \in A = \{(i,j) : (\pi_i'/\pi_i) < (\pi_j'/\pi_j)\}.$$

The optimal index pair (i,j) satisfies $a_i > 0$ and $a_j = 0$; for otherwise one of the following contradictions is implied

(i) $a_i = 0,\ a_j = 0 \Longrightarrow \Delta_\zeta[\pi,\pi'] = 0$;

(ii) $a_i > 0,\ a_j > 0 \Longrightarrow \Delta_\zeta[\pi,\pi'] = a_i w_i + a_j w_j = (\pi_i - \tilde{\pi}_i')w_i + (\pi_j - \tilde{\pi}_j')w_j = 1 - 1 = 0$; or

(iii) $a_i = 0,\ a_j > 0 \Longrightarrow (i,j) \notin A$.

Hence ζ is such that $(\pi_i'/\pi_i) \leq \zeta \leq (\pi_j'/\pi_j)$. The basic feasible solution with indices (i,j) is now seen to take the form:

$$w_i^* = \frac{\tilde{\pi}_j' - \pi_j}{\pi_i\tilde{\pi}_j' - \pi_j\tilde{\pi}_i'} \geq 0, \qquad w_j^* = \frac{\pi_i - \tilde{\pi}_i'}{\pi_i\tilde{\pi}_j' - \pi_j\tilde{\pi}_i'} \geq 0 .$$

Now

$$\Delta_\zeta[\pi,\pi'] = \max\{\Delta_{\zeta,i,j}[\pi,\pi'] : (\pi_i'/\pi_i) \leq \zeta \leq \pi_j'/\pi_j)\}$$

where

$$\Delta_{\zeta,i,j}[\pi,\pi'] = aw^* = a_i w_i^* = \frac{\pi_i\pi_j' + \pi_j\pi_i' - \zeta\pi_i\pi_j - \zeta^{-1}\pi_i'\pi_j'}{\pi_i\pi_j' - \pi_j\pi_i'} .$$

Consequently

$$\Delta[\pi,\pi'] = \sup\{\Delta_\zeta[\pi,\pi'] : c_1 \leq \zeta \leq c_2^{-1}\}$$

$$= \sup\{\Delta_{\zeta,i,j}[\pi,\pi'] : (i,j)\in A,\ (\pi_i'/\pi_i) \leq \zeta \leq (\pi_j'/\pi_j)\}$$

$$= \max_{(i,j)\in A}\ \sup\{\Delta_{\zeta,i,j}[\pi,\pi'] : (\pi_i'/\pi_i) \leq \zeta \leq (\pi_j'/\pi_j)\}.$$

Since $\Delta_{\zeta,i,j}[\pi,\pi']$ is <u>concave</u> in ζ, it achieves a unique maximum at

$\zeta = \zeta^* = \sqrt{\dfrac{\pi_i'\pi_j'}{\pi_i\pi_j}}$. Thus

$$\Delta[\pi,\pi'] = \max_{(i,j)\in A}\{\Delta_{\zeta^*,i,j},[\pi,\pi']\}$$

$$= \max_{(i,j)\in A}\left\{\frac{1-\sqrt{\dfrac{\pi_i'\pi_j}{\pi_i\pi_j'}}}{1+\sqrt{\dfrac{\pi_i'\pi_j}{\pi_i\pi_j'}}}\right\}$$

$$= \frac{1-\sqrt{c_1c_2}}{1+\sqrt{c_1c_2}}\ .$$

4. The Generalized Ergodic Coefficient

It is well known that if P is a stochastic matrix and

$$a[P] = \max_{i,j \in S} \delta[e^i P, e^j P] < 1$$

then, for any $\pi, \pi' \in \Pi$,

$$\delta[\pi P, \pi' P] \leq a[P] \; \delta[\pi, \pi'],$$

i.e., the transformation $f[\pi] = \pi P$ is a contraction in Π. One consequence of this property of P is that $\{\pi(P)^n\}$ approaches a unique limit as $n \to \infty$. The rate of covergence $a[P]$ is called the ergodic coefficient of the stochastic matrix P. This section generalizes the concept of an ergodic coefficient to substochastic matrices; a corresponding contraction property of $T(\cdot, P)$ is also established.

(4.1) **Definition**: If P is a nonzero substochastic matrix, then define

$$\alpha[P] = \max\{\Delta[T(e^i,P), T(e^j,P)] : e^i P \neq 0, e^j P \neq 0\}$$

Remark: The evaluation of $\alpha[P]$ by (3.3) requires N^3 operations. This is comparable to the effort expended when multiplying two NxN matrices.

The generalized ergodic coefficient $\alpha[P]$ has the following properties:

(4.2) __Lemma__: (a) $0 \leq \alpha[P] \leq 1$ for all substochastic matrices $P \neq 0$.

(b) $\alpha[P] < 1 \iff P$ is subrectangular.

(c) $\alpha[P] = 0 \iff \text{rank}[P] = 1$.

__Proof:__ Trivial.

(4.3) __Proposition__: (Contraction Property of T)

$$\Delta[T(\eta,P),\ T(\eta',P)] \leq \alpha[P]\ \Delta[\eta,\eta'], \qquad \eta P \neq 0,\ \eta' P \neq 0.$$

__Proof:__ The proof is given in Section 4'.

(4.4) __Corollary__: $\alpha[P] = \sup\ \{\Delta[T(\eta,P),\ T(\eta',P)] : \eta P \neq 0,\ \eta' P \neq 0\}$.

(4.5) __Corollary__: $\alpha[PQ] \leq \alpha[P]\ \alpha[Q]$.

This analysis may be modified to obtain a generalized ergodic coefficient for M-subrectangular matrices.

(4.6) **Definition:** If P is a nonzero substochastic matrix, define:

$$\alpha_M[P] = \begin{cases} \max\{\Delta[T(e^i,P),T(e^j,P)) : P_{ik}>0 \text{ and } P_{jk}>0, \text{ some } k\in S\}, & \text{if } P \text{ is } M\text{-subrectangular} \\ 1, & \text{otherwise} \end{cases}$$

Remark: $\alpha[P] = \alpha_1[P]$.

(4.7) **Lemma:** If $P^1,\dots,P^M$ are subrectangular and $P = \Sigma_{m=1}^{M} P^m$ is M-subrectangular, then $\alpha_M[P] = \max_m\{\alpha[P^m]\}$.

(4.8) **Corollary:** (a) $0 \leq \alpha_M[P] \leq 1$

(b) $\alpha_M[P] < 1 \iff P$ is M-subrectangular.

(c) $\alpha_M[P] = 0 \iff \text{rank}[P] = M$ and P is M-subrectangular.

Proof: Trivial; compare (4.2).

(4.9) **Corollary:** If PQ is M-subrectangular then

$$\alpha_M[PQ] \leq \alpha_M[P]\,\alpha_M[Q].$$

Proof: Trivial; compare (4.5)

4′. Proof of Proposition (4.3)

We require the following well-known Lemma:

(4′.1) **Lemma:** If $w \in R_N$ and $\pi, \pi' \in \Pi$ then $|\pi w - \pi' w|$
$\leq \delta[\pi,\pi'] \cdot \max_{i,i' \in S}[w_i - w_{i'}]$.

Proof: By (3.2), $\Sigma_{j\in S}\,(\pi_j-\pi'_j)^+ = \Sigma_{j\in S}\,(\pi'_j-\pi_j)^+$.

Now $\pi w-\pi' w = \Sigma_{j\in S}\,(\pi_j-\pi'_j)\,w_j = \Sigma_{j\in S}[(\pi_j-\pi'_j)^+ - (\pi'_j-\pi_j)^+]w_j$

$\leq \{[\Sigma_{j\in S}\,(\pi_j-\pi'_j)^+]\cdot[\max_{i\in S}\,w_i]\} - \{[\Sigma_{j\in S}(\pi'_j-\pi_j)^+]\cdot[\min_{i'\in S}\,w_{i'}]\}$

$= \delta[\pi,\pi'] \cdot [\max_{i\in S}\,w_i - \min_{i\in S}\,w_i] = \delta[\pi,\pi'] \cdot [\max_{i,i'\in S}\,w_i-w_{i'}]$.

By a similar argument $\pi w-\pi' w \geq \delta[\pi,\pi'] \cdot [\min_{i\in S}\,w_i - \max_{i\in S}\,w_i]$

$= \delta[\pi,\pi'] \cdot [-\max_{i,i'\in S}\,w_i-w_{i'}]$.

In order to prove (4.3) define:

$\pi^i = T(e^i,P)$, $\qquad e^iP\neq 0$;

$W = \{w\in R_N : w_i\geq 0\ \forall i\in S,\ \pi P w>0,\ \pi' P w>0\}$;

$I(w) = \{i : e^iPw > 0\}$.

Since it contains ν, the N-vector of one's, W is nonempty. Also, if $w \in W$, then $Pw \neq 0$ and $I(w)$ is nonempty. Now

$$\Delta[T(\eta,P),\ T(\eta',P)]$$

$$= \sup_{w \in W} \left\{ \Sigma_{j \in S} \left(\frac{\Sigma_{i \in I(w)} \eta_i P_{ij} w_j}{\eta P w} - \frac{\Sigma_{i \in I(w)} \eta'_i P_{ij} w_j}{\eta' P w} \right)^+ \right\}$$

$$= \sup_{w \in W} \max_{J \subseteq S} \left\{ \Sigma_{j \in J} \left(\frac{\Sigma_{i \in I(w)} \eta_i P_{ij} w_j}{\eta P w} - \frac{\Sigma_{i \in I(w)} \eta'_i P_{ij} w_j}{\eta' P w} \right) \right\}$$

$$= \sup_{w \in W} \max_{J \subseteq S} \left\{ \Sigma_{i \in I(w)} \left(\frac{\Sigma_{j \in J} \eta_i P_{ij} w_j}{\eta P w} - \frac{\Sigma_{j \in J} \eta'_i P_{ij} w_j}{\eta' P w} \right) \cdot \left(\frac{\Sigma_{j \in S} P_{ij} w_j}{\Sigma_{j \in S} P_{ij} w_j} \right) \right\}$$

$$= \sup_{w \in W} \max_{J \subseteq S} \left\{ \Sigma_{i \in I(w)} \left(\frac{\Sigma_{j \in S} \eta_i P_{ij} w_j}{\eta P w} - \frac{\Sigma_{j \in S} \eta'_i P_{ij} w_j}{\eta' P w} \right) \cdot \left(\frac{\Sigma_{j \in J} P_{ij} w_j}{\Sigma_{j \in S} P_{ij} w_j} \right) \right\} .$$

Application of (4'.1) now yields

$$\Delta[T(\eta,P),\ T(\eta',P)]$$

$$\leq \sup_{w\in W} \max_{J\subseteq S} \left\{ \left[\Sigma_{i\in I(w)} \left(\frac{\Sigma_{j\in S}\, \eta_i P_{ij} w_j}{\eta P w} - \frac{\Sigma_{j\in S}\, \eta_i' P_{ij} w_j}{\eta' P w} \right)^{+} \right] \right.$$

$$\left. \cdot \left[\max_{i,i'\in I(w)} \left(\frac{\Sigma_{j\in J}\, v_j^i\, w_j}{v^i w} - \frac{\Sigma_{j\in J}\, v_j^{i'} w_j}{v^{i'} w} \right) \right] \right\}$$

$$\leq \Delta[\eta,\eta'] \cdot \alpha[P].$$

5. Convergence of Information Vector Approximations

(5.1) **Definition:** The approximation closeness is defined by

$$\underline{\xi}(\underline{z};\ k_0,k) = \Delta[\underline{\eta}(\underline{z};\ k_0,k),\ \underline{\eta}(k)], \qquad \underline{z}\in H(k_0).$$

(5.2) **Lemma:**

$$\tfrac{1}{2}\,|\underline{\eta}(\underline{z};\ k_0,k) - \underline{\eta}(k)| \le \underline{\xi}(\underline{z};\ k_0,k) \le \underline{\xi}(\underline{z};\ k_0,\ k_1)\cdot\alpha[P(\underline{y}(k_1,k))],$$

$$\underline{z}\in H(k_0),\ k_0 \le k_1 \le k \in I.$$

Proof: Apply (1.11), (3.2) and (4.3).

The following theorems give easily verified bounds on $\underline{\xi}(\cdot;\cdot,\cdot)$.

(5.3) **Theorem:** (Strong Geometric convergence). Define

$$\alpha(n) = \sup_{k\in I}\{\max_{z\in Y^k(n)}\{\alpha[P^k(z)]\}\}.$$

Then

$$\alpha(n\cdot m) \le \alpha(n)\alpha(m)$$

and

$$\underline{\xi}(\underline{z};\ k_0,k) \le \alpha(k-k_0) \le [\alpha(n)]^{(k-k_0)\div n}, \qquad \text{a.s. } \forall \underline{z}\in H(k_0).$$

Moreover, if $\alpha(n^*) < 1$, for some positive integer n^*, then

$$\lim_{k\to\infty} \underline{\xi}(\underline{z};\ k_0,k) = 0 \qquad \text{a.s. } \forall\ \underline{z}\in H(k_0)$$

Proof: Immediate, using (4.5) and (5.2).

(5.4) Theorem (Weak Geometric convergence). Define

$$\bar{\alpha}(n) = \sup_{k\in I}\{\max_{i\in S} E\{\alpha[P^k(\underline{y}(k,k+n))] \mid \underline{s}(k)=i\}\}.$$

Then

$$\bar{\alpha}(n+m) \le \bar{\alpha}(n)\bar{\alpha}(m)$$

and

$$E\{\xi(\underline{\pi};\, k_0,k) \mid \underline{s}(k_0)\} \le \bar{\alpha}(k-k_0) \le [\bar{\alpha}(n)]^{(k-k_0)\div n} \quad \text{a.s.} \quad \forall \underline{\pi}\in H(k_0).$$

Moreover, if $\bar{\alpha}(n^*) < 1$, for some positive integer n^*, then

$$\lim_{k\to\infty} \xi(\underline{\pi};\, k_0,k) = 0 \quad \text{a.s.} \ \forall\, \underline{\pi}\in H(k_0)$$

Proof:

$$\bar{\alpha}(n+m) \le \sup_{k\in I} \{\max_{i\in S} E\{\alpha[P^k(\underline{y}(k,k+n))] \cdot \alpha[P^{k+n}(\underline{y}(k+n,k+n+m))] \mid \underline{s}(k)=i\}\}$$

$$\le \sup_{k\in I} \{\max_{i\in S} E\{\alpha[P^k(\underline{y}(k,k+n))] \cdot E\{\alpha[P^{k+n}(\underline{y}(k+n,k+n+m))] \mid \underline{s}(k)=i,\ \underline{y}(k,k+n)\} \mid \underline{s}(k)=i\}\}$$

$$= \sup_{k\in I}\{\max_{i\in S} E\{\alpha[P^k(\underline{y}(k,k+n))] \cdot [\Sigma_{j\in S} E\{\alpha[P^{k+n}(\underline{y}(k+n,k+n+m))] \mid \underline{s}(k+n)=j\} \Pr\{\underline{s}(k+n)=j \mid \underline{s}(k)=i,\underline{y}(k,k+n)\}] \mid \underline{s}(k)=i\}\}$$

$$\le \sup_{k\in I}\{\max_{i\in S} E\{\alpha[P^k(\underline{y}(k,k+n))] \cdot \bar{\alpha}(m) \mid \underline{s}(k)=i\}$$

$$\le \bar{\alpha}(n)\ \bar{\alpha}(m).$$

The remainder is immediate, using (4.5) and (5.2).

Theorems (5.3) and (5.4) give conditions under which $\underline{\eta}(k)$ may be arbitrarily closely approximated by a finite-state automaton. This automaton is of course time-varying if $(\underline{S},\underline{Y})$ is nonstationary. In (5.3), assume $\alpha(n)<1$, for some n. Then for any $\epsilon>0$ there is a $n(\epsilon)$ such that $\alpha(n(\epsilon))\leq\epsilon$. Construct an automaton whose state at time k is $\underline{x}(k) = \underline{y}(k-n(\epsilon),k)$ and associate with $\underline{x}(k)$ any element $\hat{\underline{\eta}}(k)$ in the range of $T(\cdot,P^{k-n(\epsilon)}(\underline{x}(k)))$. Now $\frac{1}{2}\,|\hat{\underline{\eta}}(k)-\underline{\eta}(k)| \leq \epsilon\ \forall k\epsilon I$. Defining $t(n) = \log \alpha(n)\ /\ \log \#Y(n)$, (where the logarithm may be taken to any desired base,) the automaton having Q states achieves an error bounded above by $Q^{t(n)}/\alpha(n)$. Similarly, defining $\bar{t}(n) = \log \bar{\alpha}(n)\ /\ \log \#Y(n)$, the automaton having Q states achieves an average error bounded above in the sense of (5.4) by $Q^{\bar{t}(n)}/\bar{\alpha}(n)$. If $t(n)$ or $\bar{t}(n) \ll -1$, then relatively small values of Q can be made to yield extremely close approximations of the information vector. This type of approximation enjoys the advantage, over grid techniques, that it is not subject to drift.

If $\alpha(n)=0$ for some n, then $\underline{\eta}(k)$ may be exactly determined by a finite-state automaton; for $\{\underline{y}(k-n,k) : k\epsilon I\}$ is now a FSMC; and hence $\{\underline{\eta}(k)\}$ is FSMC. A sufficient condition for $\{\underline{\eta}(k)\}$ to be a FSMC is given by Theorem (6.14) in Section 6.

6. Analysis of Stationary POMCs

In this section, we study POMCs that satisfy the following assumptions:

(A1) Time-Invariance: $P^k(y)=P^{k'}(y)\ \forall k,k'\in I,\ y\in Y$.
In view of (A1), the superscript "k" will be omitted throughout this section.

(A2) State Nontransitivity: If $P_{ij}(z)>0$, then there is a $\tilde{z}\in Y(*)$ such that $P_{ji}(\tilde{z})>0$.

(A3) Output Recurrence: For any $i\in S$, $z\in Y(*)$, there is a $\tilde{z}\in Y(*)$ such that $e^i P(\tilde{z}z) \neq 0$. This may be satisfied by considering, one at a time, appropriate combinations of recurrent state classes in S.

The elimination of redundant states (following Paz[10]), although desirable in certain applications, will not be required.

Let F denote a matrix whose columns span the linear space generated by $\{P(z)\nu : z\in Y(*)\}$ or, equivalently, by $\{P(z)\nu : z\in Y(N)\}$; the equivalence follows from Theorem I.B.2.1 of Paz [10]. Two stochastic vectors satisfying $\pi F = \pi' F$ are said to be equivalent since
$\pi P(z)\nu = \pi' P(z)\nu\ \forall z\in Y(*)$, and thus they convey the same information about future observations.

Central to the analysis of stationary POMCs is the concept of a recursively computable decomposition of the state set into components whose transition probabilities are asymptotically subrectangular in the sense of (5.4).

(6.1) **Theorem**: Assume (A1), (A2), (A3) and let $\underline{\Sigma}(k_0,k)$ denote the collection of subsets of S defined by

$$\underline{\Sigma}(k_0,k) = \{\{j : P_{ij}(\underline{y}(k_0,k))>0\} : i\epsilon S\} - \emptyset.$$

Then there exist a stochastic process $\{\underline{\Sigma}(k) : k\epsilon I\}$ and an integer $M \leq N$ such that

(a) $\#\underline{\Sigma}(k) = M$, a.s. $\forall k\epsilon I$.

(b) The elements of $\underline{\Sigma}(k)$ are disjoint, a.s. $\forall k\epsilon I$.

(c) $\underline{s}(k)$ is contained by a single element $\underline{\sigma}(k)$ of $\underline{\Sigma}(k)$, a.s. $\forall k\epsilon I$.

(d) For any $k\epsilon I$, there is a random variable $\underline{n}(k)\epsilon\Theta_b(k)$, $E\{\underline{n}(k)\}<\infty$, such that $\underline{\Sigma}(k_0,k) = \underline{\Sigma}(k) \iff k_0 \leq k - \underline{n}(k)$, a.s.

(e) For any $k_0\epsilon I$, there is a random variable $\underline{m}(k_0)\epsilon\Theta_f(k_0)$, $E\{\underline{n}(k_0)\}<\infty$, such that $\underline{\Sigma}(k_0,k) = \underline{\Sigma}(k) \iff k \geq k_0 + \underline{m}(k_0)$, a.s.

(f) $\underline{\Sigma}(k_0,k) = \underline{\Sigma}(k) \iff P(\underline{y}(k_0,k))$ is M-subrectangular, a.s.

Proof: The proof is given in Section 6'.

If $\underline{\sigma}(k_0)$ is known to the observer, then subsequent values of $\underline{\sigma}(k)$ may be computed according to the identity

(6.2) $\underline{\sigma}(k) = \tau(\underline{\sigma}(k_0), \underline{y}(k_0,k))$, $\tau(\sigma,z) = \{j : P_{ij}(z)>0, i\epsilon\sigma\}$.

Similarly, (6.2) provides a means for determining $\underline{\Sigma}(k)$ from $\underline{\Sigma}(k_0)$ and $\underline{y}(k_0,k)$.

We now show that $\underline{\eta}(\underline{\pi}; k_0,k)$ may be decomposed into components $[\underline{\eta}(\underline{\pi}; k_0,k)|\sigma]$, $\sigma\epsilon\underline{\Sigma}(k)$, each of which converges in a manner similar to (5.3) or (5.4). This decomposition is preserved by (1.9) in the sense that

(6.3) $$\underline{\eta}([\underline{v}|\sigma];\ k_0,k) = [\underline{\eta}(\underline{v};\ k_0,k)\ |\ \tau(\sigma,\underline{y}(k_0,k))]$$

$$\text{a.s.}\quad \forall\ \underline{v}\epsilon H(k_0),\ \sigma\epsilon\underline{\Sigma}(k_0),\ \underline{v}e^{\sigma}>0.$$

The convergence properties are stated as follows. In analogy to (5.1) define

$$\underline{\xi}_M(\underline{v};\ k_0,k)$$

$$= \max\{\Delta[\underline{\eta}([\underline{v}|\sigma];\ k_0,k),\ \underline{\eta}([\underline{\eta}(k_0)|\sigma];\ k_0,k)]\ :\ \sigma\epsilon\underline{\Sigma}(k_0),\ \underline{v}e^{\sigma}>0,\ \underline{\eta}(k_0)e^{\sigma}>0\}$$

(6.4) $$= \max\{\Delta[[\underline{\eta}(\underline{v};\ k_0,k)|\sigma],\ [\underline{\eta}(k)|\sigma]]\ :\ \sigma\epsilon\underline{\Sigma}(k),\ \underline{\eta}(\underline{v};\ k_0,k)e^{\sigma}>0,\ \underline{\eta}(k)e^{\sigma}>0\},$$

$$\underline{v}\epsilon H$$

so that, in analogy to (5.2),

(6.5) $$\tfrac{1}{2}|\underline{\eta}(\underline{v};\ k_0,k)-\underline{\eta}(k)| \le \underline{\xi}_M(\underline{v};\ k_0,k) \le \underline{\xi}_M(\underline{v};\ k_0,\ k_1)\cdot\alpha_M[P(\underline{y}(k_1,k))],$$

$$\underline{v}\epsilon H(k_0),\ k_0 \le k_1 \le k \epsilon I.$$

The immediate consequences of (5.3) and (5.4) are now:

(6.6) __Theorem__: (Strong ~~exponential~~ GEOMETRIC convergence within a component.)

Assume (A1), (A2), (A3), and define

$$\alpha_M(n) = \max_{z\epsilon Y(n)}\{\alpha_M[P(z)]\}.$$

Then

$$\alpha_M(n+m) \le \alpha_M(n)\ \alpha_M(m)$$

and

$$\underline{\xi}_M(\underline{v};\ k_0,k) \le \alpha_M(k-k_0) \le [\alpha_M(n)]^{(k-k_0)\div n},\qquad \text{a.s.}\quad \forall\ \underline{v}\epsilon H(k_0).$$

Moreover, if $\alpha_M(n^*) < 1$, for some positive integer n^*, then

$$\lim_{k\to\infty}\ \underline{\xi}(\underline{v};\ k_0,k) = 0,\qquad \text{a.s.}\ \forall\ \underline{v}\epsilon H(k_0)$$

(6.7) **Theorem:** (Weak GEOMETRIC convergence within a component.)

Assume (A1), (A2), (A3), and define

$$\bar{\alpha}_M(n) = \max_{i \in S} E\{\alpha_M[P(\underline{y}(0,n))] \mid \underline{s}(0)=i\}.$$

Then

$$\bar{\alpha}_M(n^*) < 1 \text{ for some positive integer } n^*;$$

$$\bar{\alpha}_M(n+m) \leq \bar{\alpha}_M(n)\, \bar{\alpha}_M(m);$$

$$E\{\underline{\xi}_M(\underline{x};\, k_0,k) \mid \underline{s}(k_0)\} \leq \bar{\alpha}_M(k-k_0) \leq [\bar{\alpha}_M(n)]^{(k-k_0)\div n}, \quad \text{a.s. } \forall\, \underline{x} \in H(k_0)$$

$$\lim_{k\to\infty} \underline{\xi}_M(\underline{x};\, k_0,k) = 0, \quad \text{a.s. } \forall\, \underline{x} \in H(k_0).$$

Finally, we show that a weighted sum of equivalent components of $\underline{n}(\underline{x};\, k_0,k)$ converges to $\underline{n}(k)$. Define:

$$\pi_i(\hat{\Sigma}) = \Pr\{\underline{s}(0)=i \mid \underline{\Sigma}(0) = \hat{\Sigma}\}$$

$$f(\hat{\sigma},\underline{\Sigma}(k)) = \bigcup \{\sigma \in \underline{\Sigma}(k) \;:\; [\pi(\underline{\Sigma}(k))|\sigma]F = [\pi(\underline{\Sigma}(k))|\hat{\sigma}]F\}$$

(6.8) $$\underline{\Sigma}^*(k) = \{f(\sigma,\underline{\Sigma}(k)) \;:\; \sigma \in \underline{\Sigma}(k)\}$$

$$\underline{\sigma}^*(k) = f(\underline{\sigma}(k),\, \underline{\Sigma}(k))$$

$$\psi(\hat{\sigma},\hat{\sigma}^*,\hat{\Sigma}) = [\pi(\hat{\Sigma})|\hat{\sigma}^*]\, e^{\hat{\sigma}}$$

We may interpret $f(\hat{\sigma},\, \underline{\Sigma}(k))$ as the union of elements in $\underline{\Sigma}(k)$ that are indistinguishable from $\hat{\sigma}$; likewise $\underline{\Sigma}^*(k)$ is the collection of indistinguishable unions in $\underline{\Sigma}(k)$ and $\underline{\sigma}^*(k)$ is the element of $\underline{\Sigma}^*(k)$ that contains $\underline{\sigma}(k)$.

(6.9) **Lemma:** (Component Identification) Assume (A1), (A2), (A3) and consider the events

$$A = \{\underline{I}(k_o) = \hat{I}\}$$

$$B = A \cap \{\underline{\sigma}(k_o) = \hat{\sigma}\}$$

where $\Pr\{A\} \geq \Pr\{B\} > 0$. Then

$$\lim_{\substack{k'-k\to\infty \\ k\leq k_o\leq k'}} \Pr\{B \mid A, \underline{y}(k,k')\} = \psi(\hat{\sigma}, \underline{\sigma}^*(k_o), \hat{I}) \quad \text{a.s.}$$

Proof: Throughout this proof, consider all random variables and events to be conditioned on A. For $n \in I$ define

$$\underline{k}(n) = \begin{cases} \min\{\underline{k} > \underline{k}(n-1) : \underline{I}(\underline{k}) = \hat{I},\ \tau(\sigma, \underline{y}(k_o, \underline{k})) = \sigma \quad \forall \sigma \in \hat{I}, & n > 0 \\ k_o, & n = 0 \\ \max\{\underline{k} < \underline{k}(n+1) : \underline{I}(\underline{k}) = \hat{I},\ \tau(\sigma, \underline{y}(\underline{k}, k_o)) = \sigma \quad \forall \sigma \in \hat{I}, & n < 0 \end{cases}$$

$$\underline{s}'(n) = \underline{s}(\underline{k}(n))$$

$$\underline{y}'(n,n') = \underline{y}(\underline{k}(n), \underline{k}(n'))$$

For $\sigma \in \hat{I}$, also define

$$\underline{\lambda}_\sigma(n,n') = \ln(\Pr\{\underline{y}'(n,n') \mid B\} / \Pr\{\underline{y}'(n,n') \mid \underline{\sigma}(0) = \sigma\}).$$

Now $\underline{\lambda}_\sigma(n,n')$ and $\underline{\lambda}_\sigma(0,n'-n)$ are identically distributed, so

$$E\{\underline{\lambda}_\sigma(n,n') \mid B\} = \bar{\lambda}_\sigma(n'-n).$$

By convexity of the $-\ln$ function, $\bar{\lambda}_\sigma(n)$ is nondecreasing in n. Moreover, $\underline{s}'(n)$ is a FSMC whose transition probability matrix P′ satisfies $P'_{ij} > 0 \ \forall i,j \in \sigma \in \hat{I}$; hence by a standard argument of convergence of powers of positive stochastic matrices, there is, for every $\varepsilon > 0$, a function d^ε satisfying $\ln(\Pr\{\underline{y}'(d^\varepsilon(n)-n, d^\varepsilon(n)) \mid \underline{y}'(0,n), B\}$ $/ \Pr\{\underline{y}'(d^\varepsilon(n)-n, d^\varepsilon(n)) \mid \underline{y}'(0,n), \underline{\sigma}(0) = \sigma\} \geq (1-\varepsilon)\ \underline{\lambda}_\sigma(d^\varepsilon(n)-n, d^\varepsilon(n))$ a.s.

Now consider $n_1 = 1$, $n_{i+1} = d^\varepsilon(n_i)$, and $\underline{\phi}_\sigma(i) = \ln(\Pr\{\underline{y}'(0,n_i),$ $\underline{y}'(n_{i+1}-n_i, n_{i+1}) \mid B\} / \Pr\{\underline{y}'(0,n_i), \underline{y}'(n_{i+1}-n_i, n_{i+1}) \mid \underline{\sigma}(0) = \sigma\}).$

By convexity of the -ln function,

$$\underline{\lambda}_\sigma(0,n_i) + (1-\epsilon)\ \overline{\lambda}_\sigma(n_i)$$
$$\leq E\{\phi_\sigma(i)\ |\ \underline{\lambda}_\sigma(0,n_i),B\}$$
$$\leq E\{\underline{\lambda}_\sigma(0,n_{i+1})\ |\ \underline{\lambda}_\sigma(0,n_i),B\}$$

Hence, either $\underline{\lambda}_\sigma(n,n')=0$ a.s. in B $\Longleftrightarrow$ $\underline{Y}$ and $\underline{\sigma}(k_s)$ are conditionally independent given $\{\underline{\sigma}(k_s) = \sigma \text{ or } \hat{\sigma}\}$ $\Longleftrightarrow$ $\sigma \subseteq f(\hat{\sigma},\hat{\Sigma})$; or else $\lim_{n'-n\to\infty} \underline{\lambda}_\sigma(n,n') = \infty$ a.s. in B $\Longleftrightarrow$ $\lim_{\substack{k'-k\to\infty \\ k'\leq k_s\leq k}} Pr\{\underline{s}(k_s)\epsilon\sigma\ |\ A,\underline{y}(k,k')\}$ $= 0$ a.s. in B $\Longleftrightarrow$ $\sigma \subseteq S-f(\hat{\sigma},\hat{\Sigma})$. The desired result follows trivially.

(6.10) <u>Theorem</u>: (Overall convergence) Assume (A1), (A2), (A3). Then

(a) $\underline{\eta}(k) = [\underline{\eta}(k)|\underline{\sigma}^*(k)]$

(b) $\underline{\eta}(k)e^\sigma = \psi(\sigma,\underline{\sigma}^*(k),\underline{\Sigma}(k))$ $\qquad \forall\sigma\epsilon\underline{\Sigma}(k)$

(c) $[\underline{\eta}(k)|\sigma]F = \underline{\eta}(k)F$ $\qquad \forall\sigma\epsilon\underline{\Sigma}(k),\ \sigma \subseteq \underline{\sigma}^*(k)$

and for all $\underline{\pi}\epsilon H(k_s)$:

(a′) $\lim_{k\to\infty} \underline{\eta}(\underline{\pi};\ k_s,k)\ e^{\underline{\sigma}^*(k)} = 1$ a.s.

(b′) $\lim_{k\to\infty} |\ \Sigma_{\sigma\epsilon\underline{\Sigma}(k)}\ (\Sigma_{\sigma^*\epsilon\underline{\Sigma}^*(k)}\ \psi(\sigma,\sigma^*,\underline{\Sigma}(k))\ \underline{\eta}(\underline{\pi};\ k_s,k)e^{\sigma^*})$
$\qquad\qquad \cdot\ [\underline{\eta}(\underline{\pi};\ k_s,k)|\sigma]\ -\ \underline{\eta}(k)\ | = 0$ a.s.

(c′) $\lim_{k\to\infty} |\underline{\eta}(\underline{\pi};\ k_s,k)F\ -\ \underline{\eta}(k)F| = 0$ a.s.

<u>Remark</u>: If $\underline{I}^*(k) = \underline{I}(k)$ a.s., e.g., if F has rank N, then (c′) becomes

$$\lim_{k\to\infty} | \underline{\eta}(\underline{v};\ k_0,k) - \underline{\eta}(k)| = 0 \quad \text{a.s.}$$

which establishes the conjecture of Kaijser [8]. This convergence can occur under numerous conditions readily derived from (b′); for example, it is implied by $\underline{v} = v(\underline{I}(k_0))$.

<u>Proof</u>: Parts (a), (b), (c) are the consequence of (6.9); parts (a′), (b′), (c′) then follow from (1.11) and (6.7).

Hence, the state estimates may be approximated arbitrarily closely (in a time average or mean $|\cdot|$ sense) by a finite-state automaton. If $\underline{\sigma}^*(k_0)$ is known initially, the automaton is constructed as discussed in Section 5. Memory-efficient approximations to $\underline{\eta}(\underline{v};\ k_0,k)$ e^{σ^*}, $\sigma^* \epsilon \underline{I}^*(k)$, may be obtained by generalized hypothesis testing procedures; see Section 2.2 of [3].

The state estimates for a POMC satisfying (A1), (A2), and (A3) can be <u>exactly</u> computed by a (properly initialized) finite-state automaton if $\{\underline{\eta}(k)\}$ is a FSMC, i.e., if there is a finite set $G \subset \Pi$ such that $\underline{\eta}(k) \epsilon G\ \forall k \epsilon I$. In this case $\{(\underline{\eta}(k), \underline{y}(k)) : k \epsilon I\}$ is a state-calculable POMC. This condition may be verified in the manner described below.

Let Γ denote the partition of S into ergodic classes; hence for $i\epsilon\gamma\epsilon\Gamma$ there holds $\gamma = \{j : P_{ij}(z)>0, z\epsilon Y(*)\}$. For each $\gamma\epsilon\Gamma$, select $\hat{\sigma}_\gamma \epsilon \hat{I}_\gamma$, $\hat{z}_\gamma \epsilon Y(*)$ with $\hat{\sigma}^*_\gamma = f(\hat{\sigma}_\gamma, \hat{I}_\gamma)$, $M_\gamma = \#\{\sigma\epsilon\hat{I}_\gamma : \sigma \subseteq \hat{\sigma}^*_\gamma\}$ such that

$$\hat{\sigma}_\gamma \subseteq \gamma;$$

$$\Pr\{\underline{\sigma}(0)=\hat{\sigma}_\gamma, \underline{I}(0)=\hat{I}_\gamma\} > 0;$$

$$\sigma = \tau(\sigma, \hat{z}_\gamma) \quad \forall\, \sigma\epsilon\hat{I}_\gamma,\ \sigma \subseteq \hat{\sigma}^*;$$

(6.11)
$$\hat{P}^\gamma_{ij} = \left\{\begin{matrix} P_{ij}(\hat{z}_\gamma), & \text{if } i\epsilon\hat{\sigma}^*_\gamma \\ 0, & \text{otherwise} \end{matrix}\right\} \quad \text{is } M_\gamma\text{-subrectangular.}$$

Now $\alpha_{M_\gamma}[\hat{P}^\gamma]<1$; hence by (4.9) there exists a unique limit

(6.12) $$\hat{\pi}^\gamma = \lim_{n\to\infty} T([\pi(\hat{I}_\gamma)|\hat{\sigma}^*_\gamma], (\hat{P}^\gamma)^n).$$

Finally, construct

$$G(0) = \{\hat{\pi}^\gamma : \gamma\epsilon\Gamma\};$$

(6.13) $$G(m+1) = G(m) \cup \{T(\pi,P(y)) : \pi\epsilon G(m),\ y\epsilon Y,\ \pi P(y)\neq 0\};$$

$$\hat{G}(m) = \{\pi F : \pi\epsilon G(m)\}.$$

We may now demonstrate:

(6.14) __Theorem__: Assume (A1), (A2), (A3). Then

(a) $\{\underline{\eta}(k)\}$ is a FSMC iff $G(m^*) = G(m^*+1)$ for some positive integer m^*. Moreover, if such an m^* exists, then $G=G(m^*)$ is the minimal subset of Π satisfying $\underline{\eta}(k)\epsilon G$ a.s. $\forall k\epsilon I$.

(b) $\{\underline{\eta}(k)\}$ is equivalent to a FSMC iff $\hat{G}(m^*) = \hat{G}(m^*+1)$ for some positive integer m^*. Moreover, if such an m^* exists, then $\hat{G}=\hat{G}(m^*)$ is the minimal set satisfying $\underline{\eta}(k)F \epsilon \hat{G}$ a.s. $\forall k\epsilon I$; the equivalent FSMC must assume at least $\#\hat{G}$ distinct values.

Remark: This proposition provides the basis for an algorithm that decides whether $\{\underline{\eta}(k)\}$ is (or is equivalent to) a FSMC having, at most, a specified number of states. It cannot, however, be used to decid whether $\{\underline{\eta}(k)\}$ is (or is equivalent to) a FSMC when the size of G (resp. $\hat{G}$) is unconstrained.

Proof: Only the proof of (a) is given. The proof of (b) is similar.

(Necessity) Suppose $G(m^*) = G(m^*+1)$. Then for every $i \in S$ there is a $\tilde{\pi}^i \in G(m^*)$ such that $\tilde{\pi}^i_i > 0$; now $\underline{\tilde{\eta}}(0) = \tilde{\pi}\underline{s}(0) \in H(0)$.
Define $\underline{\tilde{\eta}}(k) = \underline{\eta}(\underline{\tilde{\eta}}(0); 0,k)$. By (6.7) $\lim_{k\to\infty} |\underline{\tilde{\eta}}(k) - \underline{\eta}(k)| = 0$.
But $\underline{\tilde{\eta}}(k) \in G(m^*+k) = G(m^*)$, and $\{\underline{\eta}(k)\}$ is stationary.
Hence $\underline{\eta}(k) \in G(m^*)$ a.s. $\forall k \in I$.

(Sufficiency) Assume that $\{\underline{\eta}(k)\}$ is a FSMC and let G be the minimal set satisfying $\underline{\eta}(k) \in G$ a.s. $\forall k \in I$.
Clearly $G \cap \Pi(\hat{\sigma}^*_\gamma)$ is nonempty, $\forall \gamma \in \Gamma$. Thus $G(0) \subseteq G$;
for otherwise some $\gamma \in \Gamma$ satisfies $\hat{\pi}^\gamma \notin G \Rightarrow \min_{\pi \in G} \frac{1}{2} |\hat{\pi}^\gamma - \pi| = \epsilon > 0$;
but $\alpha_{M_\gamma}[\hat{P}^\gamma]^n < \epsilon$ (some n); now the concatenation z^* of $\hat{z}^\gamma$ with itself n times and $\pi^* = T(\pi, P(z^*)$(some $\pi \in G \cap \Pi(\hat{\sigma}_\gamma)$) satisfy $\pi^* \in G$ and $|\hat{\pi}^\gamma - \pi^*| < \epsilon$, which is a contradiction.
Since $\{T(\pi,P(y)) : \pi \in G,\ y \in Y,\ \pi P(y) \neq 0\} = G$, it follows that $G(m) \subseteq G \Rightarrow G(m+1) \subseteq G$. Now $G(0) \subseteq G(1) \subseteq \ldots \subseteq G$.
Since G is a finite set, m^* exists and $G(m^*) = G(m^*+1) \subseteq G$.
But now $G(m^*) \subseteq G$ by the argument given above; thus $G(m^*) = G$.

6′ Proof of Theorem (6.1)

Before giving the proof, we introduce some new terminology. Let S^* denote the power set (collection of subsets) of S; S^{**} in turn will be the power set of S^*. Now define $\mathbf{A} : S^{**} \to S^{**}$, $\mathbf{U} : S^{**} \to S^{**}$, $R : S \times Y(*) \to S^*$, $\Sigma : Y(*) \to S^{**}$, and $U : Y(*) \to S^{**}$, as follows:

$$\mathbf{A}[\Sigma] = \{\sigma\varepsilon\Sigma : \text{no element of } \Sigma \text{ is properly contained in } \sigma\} \subseteq \Sigma$$

$$\mathbf{U}[\Sigma] = \{\sigma_1 \cup \sigma_2 \cup \ldots \cup \sigma_n : \sigma_1, \ldots, \sigma_n \varepsilon \Sigma,\ n>0\}$$

$$B_{ij}(z) = 1 \text{ if } P_{ij}(z)>0,\ 0 \text{ if } P_{ij}(z) = 0$$

$$R(i,z) = \{j : P_{ij}(z)>0\}$$

$$\Sigma(z) = \{R(i,z) : i\varepsilon S\} - \{\emptyset\}$$

$$U(z) = \mathbf{U}[\Sigma(z)]$$

$$M = \min\{\#\Sigma(z) : z\varepsilon Y(*)\}$$

$$M^* = \min\{\#U(z) : z\varepsilon Y(*)\}.$$

Note that $\underline{\Sigma}(k_0,k) = \Sigma(\underline{y}(k_0,k))$. Also define the partial order "<" on $Y(*)$:

$$\tilde{z} < z \text{ if } \tilde{z} = z'z \text{ for some } z'\varepsilon Y(*).$$

Finally, define

$$E = \{z\varepsilon Y(*) : U(\tilde{z}) = U(z)\ \forall \tilde{z}<z\}.$$

(6′.1) Lemma: (a) $U(z'z) \subseteq U(z)$

(b) $\#U(z'z) \leq \#U(z')$

Proof: By definition, $U(z) = \{\bigcup_{i\varepsilon C} R(i,z) : C\varepsilon S^*\} - \{\emptyset\}$; and $R(i,z'z) = \bigcup_{j\varepsilon R(i,z')} R(j,z)$; so $U(z'z) = \{\bigcup_{j\varepsilon C} R(j,z) : C\varepsilon U(z')\} - \{\emptyset\}$.

(6′.2) Corollary: $E = \{z : \#U(z)=M^*\}$.

Proof: By (6′.1)(a), $\#U(z)=M^*$ implies $z \in E$. Now suppose $\#U(z)>M^*$. By definition of M^*, we may select z^* such that $\#U(z^*)=M^*$, and by (A3), we may select $\tilde{z}$ such that $z^*\tilde{z}z \in Y(*)$. Now $\#U(z^*\tilde{z}z) = M^*$ by (6′.1) (b); so $z \notin E$.

(6′.3) Corollary: For any $k \in I$, there is a random variable $\underline{n}(k) \in \theta_b(k)$, $E\{\underline{n}(k)\}<\infty$, such that $\underline{y}(k_0,k) \in E \iff k_0 \leq k-\underline{n}(k)$

Proof: Select $\underline{n}(k) = \min\{\underline{n} : \#U(\underline{y}(k-\underline{n},k))=M^*\}$. $E\{\underline{n}(k)\} < \infty$ because $\{[\underline{s}(k-n), B(\underline{y}(k-n,k)), U(\underline{y}(k-n,k))] : n=0, \ldots\}$ is a FSMC whose recurrent states $[s,B,U]$ have the property $\#U=M^*$, by (A2), (A3) and (6′.1) (b), and since $\underline{n}(k)$ is therefore bounded above by the number of transitions until this FSMC enters a recurrent state, a random variable having finite expectation.

(6′.4) Lemma: $\Sigma(z) = \mathbf{A}[U(z)]$, $\forall z \in E$.

Proof: (By contradiction): Clearly $\Sigma(z) \supseteq \mathbf{A}[U(z)]$.
Now assume that $z \in E$, $\sigma \in \Sigma(z)$, $\sigma' \in U(z)$ and $\sigma' \subsetneq \sigma$. We may select $\sigma'' \in \mathbf{A}[U(z)]$ so that $\sigma'' \subseteq \sigma'$. Let i be any element of σ'' and pick j so that $R(j,z)=\sigma$; now $P_{ji}(z)>0$. By (A?), there exists $\tilde{z} \in Y(*)$ satisfying $P_{ij}(\tilde{z})>0$. Now construct $z_0=z$ and $z_n = z_{n-1}\,\tilde{z}\,z_{n-1}$, $n \in S$. Trivially $P_{ji}(z_n)>0$ so $z_n \in Y(*)$, and hence $z_n < z$. But $z \in E$, so $\sigma'' \in \mathbf{A}\,[U(z)] = \mathbf{A}\,[U(z_n)] \subseteq \Sigma(z_n)$, and there is a k_n such that $R(k_n, z_n) = \sigma''$. Moreover, by construction of z_n, $R(k_i, z_n) \supseteq \sigma \supset \sigma''$, $i=1, \ldots, n-1$, so $k_n \neq k_{n'}$ when $n \neq n'$. But now $R(i,z_n) \supseteq \sigma \supset \sigma''$, $\forall\, i \in S$ and $\sigma'' \notin U(z_n) = U(z) \supseteq \mathbf{A}[U(z)]$, which contradicts the original assumption that $\sigma'' \in \mathbf{A}[U(z)]$.

(6'.5) <u>Corollary</u>: $\tilde{z} < z \in E \longrightarrow \Sigma(\tilde{z}) = \Sigma(z)$

(6'.6) <u>Lemma</u>: The elements of $\Sigma(z)$ are disjoint, $\forall\, z \in E$.

<u>Proof</u>: Suppose (i) $\hat{z} < z \in E$ and (ii) $B(\hat{z}\hat{z}) = B(\hat{z})$.

Then (i) implies $\hat{z} \in E$ and hence, by (6'.5), $\Sigma(\hat{z}) = \Sigma(z)$.

Now, by (ii), $R(j,\hat{z}) = \bigcup_{i \in R(j,\hat{z})} R(i,z)$.

So $k \in R(i,\hat{z}) \cap R(j,\hat{z})$ implies $R(k,\hat{z}) \subseteq R(i,\hat{z})$ and $R(k,\hat{z}) \subseteq R(j,\hat{z})$. Hence, by (6'.4), $R(i,\hat{z}) = R(j,\hat{z})$, and the elements of $\Sigma(\hat{z})$ are disjoint It only remains to show that there exists a $\hat{z}$ satisfying (i) and

Pick i,j so $P_{ji}(z) > 0$ and pick $\tilde{z}$ so $P_{ij}(\tilde{z}) > 0$; $\tilde{z}$ exists by (A2). Define $z_1 = \tilde{z}\, z$ and $z_{n+1} = z_n \tilde{z}\, z$. By construction $z_n z_m = z_{n+m}$. Moreover $P_{ii}(z'z) > 0$ so $P(z_n) \neq 0$ and $z_n \in E$. Since $\{B(z_n) : n=1,2,\ldots\}$ is a finite set, there exist $n,m>0$ such that $B(z_n) = B(z_{n+m})$. Hence,

$$B(z_n) = B(z_{n+m}) = B(z_m)\, B(z_n) = B(z_m)\, B(z_{n+m}) = B(z_{n+2m}) = \ldots = B(z_{n+nm})$$

and similarly $B(z_{(m-1)n}\, z_n) = B(z_{(m-1)n}\, z_{n+nm})$; hence $B(z_{nm}) = B(z_{nm}\, z_{nm})$. Identify $\hat{z} = z_{nm}$ to complete the proof.

(6'.7) <u>Corollary</u>: $E = \{z : \#\Sigma(z) = M\}$.

<u>Proof</u>: By (6'.2) and (6'.6), $M^* = \#U(z) = 2^{\#\Sigma(z)} - 1$, $z \in E$, and $M^* < \#U(z) \leq 2^{\#\Sigma(z)} - 1$, $z \notin E$.

Proof of (6.1): Define $\underline{n}(k)$ as in (6'.3), and let $\underline{\Sigma}(k) = \underline{\Sigma}(k-\underline{n}(k),k)$.

Now $\underline{\Sigma}(k) = \Sigma(\underline{y}(k-\underline{n}(k), k))$, where $\underline{y}(k-\underline{n}(k),k) \in E$.

(a) Follows from (6'.7).

(b) Follows from (6'.6).

(c) If j is contained by no element of $\underline{\Sigma}(k) = \underline{\Sigma}(k-\underline{n}(k),k)$, then $P_{ij}(\underline{y}(k-\underline{n}(k),k)) = 0 \ \forall i \in S$, by definition of $\underline{\Sigma}(\cdot,\cdot)$, and hence $\Pr\{\underline{s}(k)=j \mid \underline{y}(k-n(k),k)\} = 0$.
Thus $\underline{s}(k)$ is contained by an element $\underline{\sigma}(k)$ of $\underline{\Sigma}(k)$ a.s.
Uniqueness of $\underline{\sigma}(k)$ follows from (b) above.

(d) Follows from (6'.3) and (6'.5).

(e) Define $\tilde{P}_{ij}(y) = \Pr\{\underline{s}(k-1)=j, \underline{y}(k)=y \mid \underline{s}(k)=i\}$
and $\tilde{P}(y_1 \ldots y_n) = \tilde{P}(y_n) \ldots \tilde{P}(y_1)$.
Then $\{\tilde{P}(y)\}$ defines a POMC with state process $\{\tilde{\underline{s}}(k)=\underline{s}(-k) : k \in I\}$ and observation process $\{\tilde{\underline{y}}(k)=\underline{y}(-k) : k \in I\}$.
Also $\tilde{P}_{ij}(z) > 0 \Longleftrightarrow P_{ji}(z) > 0$. Hence by (f), below, and (6'.3), there is an $\underline{m}(k) \in \theta_f(k)$ such that $k \geq k_o + \underline{m}(k_o)$
$\Longleftrightarrow \tilde{P}(\tilde{\underline{y}}(-k,-k_o))$ is M-subrectangular $\Longleftrightarrow P(\underline{y}(k_o,k))$ is M-subrectangular.

(f) Follows from (6'.6) and (6'.7).

7. **Examples**

We now illustrate by means of simple stationary POMCs the concepts introduced in this paper.

(7.1) **Example**: Suppose that $\underline{S}$ is an ergodic Markov chain with

$$P_{ij}(y) = \begin{cases} P_{ij}, & \text{if } y = j \\ 0, & \text{otherwise} \end{cases}$$

Now $\underline{y}(k) = \underline{s}(k)$ $\forall k \in I$, i.e. there is perfect state observation. $\alpha(1)=0$, and $\underline{\eta}(\underline{z};\ k_o,k) = \underline{\eta}(k) = e^{\underline{s}(k)}$ whenever $k>k_o$ and $\underline{z} \in H(k_o)$. Here $\{\underline{\eta}(k)\}$ is a FSMC.

(7.2) **Example**: Suppose that $\underline{S}$ is a Markov chain with strictly positive transition probability matrix. Let $Y = S$ and

$$P_{ij}(y) = \begin{cases} P_{ij}, & \text{if } y = i-j \text{ or } y = N+i-j \\ 0, & \text{otherwise} \end{cases}$$

Now $\underline{s}(k) = [\underline{s}(k_o) + \Sigma^{k}_{k'=k_o+1}\underline{y}(k')] \bmod N$, so $(\underline{S}, \underline{Y})$ is state-calculable. This result is expressed by (6.1), where $\underline{M}=N$ and $\underline{I}(k) = \{\{1\}, \ldots, \{N\}\}$, the set of singletons in S. Clearly $\underline{\sigma}(k) = \{\underline{s}(k)\}$.

Hence, if no two rows of F coincide, and if $\underline{\pi} \in H(k_0)$, then

$$\underline{\eta}(k) = e^{s(k)};$$

$$\Delta[\underline{\eta}(\underline{\pi};\ k_0,k),\ \underline{\eta}(k)] = 1, \text{ whenever } \underline{\pi} \neq e^{s(k_0)};$$

$$\lim_{k\to\infty} |\underline{\eta}(\underline{\pi};\ k_0,k) - \underline{\eta}(k)| = 0 \text{ a.s.}$$

Note that $|\underline{\eta}(\underline{\pi};\ k_0,k) - \underline{\eta}(k)|$ is not monotone decreasing in k. Again $\{\underline{\eta}(k)\}$ is a FSMC.

(7.3) Example: Suppose

$$P(1) = \begin{bmatrix} 1/4 & 1/4 \\ 0 & 1/2 \end{bmatrix},\quad P(2) = \begin{bmatrix} 1/2 & 0 \\ 1/4 & 1/4 \end{bmatrix}.$$

Now $\underline{\eta}(k+1) = 1/2[(1,0)+\underline{\eta}(k)]$ if $y(k+1) = 1$, and $\underline{\eta}(k+1) = 1/2[(0,1)+\underline{\eta}(k)]$ if $y(k+1) = 2$. Thus it happens that $\frac{1}{2}|\underline{\eta}(\underline{\pi};\ k_0,k) - \underline{\eta}(k)| \leq (.5)^{k-k_0}$. This convergence is predicted by Theorem (5.4) which states that $E\{\Delta[\underline{\eta}(\underline{\pi};\ k_0,k),\ \underline{\eta}(k)]\} \leq \bar{\alpha}(k-k_0)$. Since $\bar{\alpha}(2) = .634$, we have $\bar{\alpha}(n) \leq (.634)^{n \div 2}$. Moreover $\Delta[\underline{\eta}(\underline{\pi}\,k_0,k),\ \underline{\eta}(k)]$ is monotone decreasing in k a.s., by (4.3). Clearly, $\{\underline{\eta}(k)\}$ is not a FSMC. But $\underline{Y}$ is a process of independent identically distributed Bernoulli trials; hence past observations convey no information about future observations. Indeed $F = \nu$ and all information vectors are equivalent. Since $\hat{G}$ contains only the element 1, $\{\underline{\eta}(k)\}$ is equivalent to a one-state FSMC.

8. ACKNOWLEDGEMENT

This paper is based primarily on research conducted at the Decision and Control Sciences Group of the MIT Electronics Systems Laboratory, supported in part by AFOSR under contracts 72-2273 and 77-3281, and reported in the author's doctoral dissertation at MIT.

REFERENCES

1. K. J. Astrom, "Optimal Control of Markov Processes with Incomplete State Information," J. Math. Anal. Appl. 10, pp. 174-205, 1965.

2. D. Blackwell, "The Entropy of Functions of Finite-State Markov Chains," Trans. First Prague Conference Information Theory, Statistical Decision Functions, Random Processes, Prague, 13-20, 1957.

3. T. M. Cover and T. J. Wagner, "Topics in Statistical Recognition," in Digital Pattern Recognition, K. S. Fu Ed., Springer-Verlag, New York, 1976.

4. J. L. Devore, "A Note on the Observation of a Markov Source Through a Noisy Channel," IEEE Trans. Inform. Theory, Vol. IT-20, pp. 762-764, 1974.

5. A. W. Drake, "Observation of a Markov Process through a Noisy Channel," Sc.D. Thesis, Department of Electrical Engineering, M.I.T., Cambridge, MA., 1962.

6. H. Furstenberg and H. Kesten, "Products of Random Matrices", Ann Math Statist. 31, pp. 457-469, 1960.

7. D. L. Isaacson and R. W. Madsen, Markov Chains: Theory and Applications, Wiley, N.Y., 1976.

8. T. Kaijser, "A Limit Theorem for Partially-Observed Markov Chains," Ann. Prob. 3, pp. 677-696, 1975.

9. J. Neveu, Discrete-Parameter Martingales, North-Holland, Amsterdam, 1975.

10. A. Paz, Introduction to Probabilistic Automata, Academic, N.Y., 1971.

11. M. Rudemo, "State-Estimation for Partially Observed Markov Chains," J. Math. Anal. Appl. 44, pp. 581-611, 1973.

12. R. D. Smallwood and E. J. Sondik, "The Optimal Control of Partially Observable Markov Processes over a Finite Horizon," Opns. Res. 21, pp. 1071-1081, 1973.

13. E. J. Sondik, "The Optimal Control of Partially-Observable Markov Processes," Stanford Ph.D. Thesis; also reported in Stanford Information Systems Laboratory Technical Report No. 6252-4, 1971.

14. C. C. White, "Procedures for the Solution of a Finite-Horizon, Partially-Observed, Semi-Markov Optimization Problem," Opns. Res. 24, pp. 348-358, 1976.