Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

The Realization Problem for Hidden Markov Models: The Complete Realization Problem

M. Vidyasagar

***Abstract*— Suppose m is a positive integer, and let $\mathcal{M} := \{1, \ldots, m\}$. Suppose $\{\mathcal{Y}_t\}$ is a stationary stochastic process assuming values in $\mathcal{M}$. In this paper we study the question: When does there exist a hidden Markov model (HMM) that *perfectly* reproduces the *complete* statistics of this process?**

Though HMM's are more than forty years old, no *complete* solution to this problem is available. It is known that a *necessary* condition for the process to have a HMM is that an associated 'Hankel' matrix should have finite rank. It is also known that the condition is *not* sufficient in general. In subsequent work, an algorithm for constructing a HMM for a finite rank process has been given, *assuming at the outset* that the process has a HMM. Hence, to date there are no conditions, either necessary or sufficient, for a process to have a HMM *that can be stated in terms of the process alone, and nothing else*.

Against this background, in the present paper we show the following: (i) Suppose a process has finite Hankel rank. Then there always exists a 'regular quasi-realization' of the process. Moreover, two regular quasi-realizations are related through a similarity transformation. (ii) If in addition the process is α-mixing, every regular quasi-realization has additional features. Specifically, the 'state transition' matrix associated with the quasi-realization satisfies the 'quasi-strong Perron property' (its spectral radius is one, the spectral radius is a simple eigenvalue, and there are no other eigenvalues on the unit circle). (iii) Suppose a process has finite Hankel rank, is both α-mixing as well as 'ultra-mixing' (a property defined here), and in addition satisfies a technical condition. Then it has an irreducible HMM realization (and not just a quasi-realization). Moreover, the Markov process underlying the HMM is either aperiodic (and is thus α-mixing), or else satisfies a 'consistency condition.' In the other direction, suppose a HMM satisfies the consistency condition plus another technical condition. Then the associated output process has finite Hankel rank, is α-mixing and is also ultra-mixing. Taken together, these two results show that, modulo two technical conditions, the finite Hankel rank condition, α-mixing, and ultra-mixing are 'almost' necessary and sufficient for a process to have an irreducible and aperiodic HMM.

I. Introduction

A. General Remarks

Hidden Markov models (HMM's) were originally introduced in the statistics literature as far back as 1957. Subsequently, they were used with partial success in a variety of applications in the engineering world, starting in the late 1970's. Some of these applications include speech processing [28], [25] and source coding. In recent years, HMM's have also been used in some problems in computational biology, such identifying the genes of an organism from its DNA [22], [29], [12] and classifying proteins into a small number of families [21]. The bibliographies of [11], [27] contain many references in this area. In spite of there being so many applications of hidden Markov models, many of the underlying statistical questions remain unanswered. The aim of this paper is to address some of these issues.

The problem under study can be briefly stated as follows. Suppose m is a positive integer and let $\mathcal{M} := \{1, \ldots, m\}$. Suppose $\{\mathcal{Y}_t\}$ is a stationary stochastic process assuming values in $\mathcal{M}$. We are interested in the following kinds of questions:

1) Suppose the complete statistics of the process $\{\mathcal{Y}_t\}$ are known. Under what conditions is it possible to construct a hidden Markov model (HMM) for this process?
2) Suppose it is not possible to construct a HMM for the process. Is it possible to construct at least a 'quasi' HMM for the process? If so, what properties does such a quasi-realization have?
3) How can one construct a 'partial realization' for the process, that faithfully reproduces the statistics of the process only up to some finite order?
4) Suppose one has access not to the entire statistics of the process, but merely several sample paths, each of finite length. How can one compute approximations to the true statistics of the process on the basis of these observations, and what is the confidence one has in the accuracy of these estimates?
5) Suppose one has constructed a partial realization of the process on the basis of a finite length sample path. How are the accuracy and confidence in the estimates of the statistics translated into accuracy and confidence estimates on the parameters in the model?

Ideally, we would like to be able to say something about *all* of these questions. In a 'practical' application, the last three questions are the ones to which we would most like to have an answer. However, these are also the most difficult questions to answer. In this paper, we provide nearly complete answers to the first two questions. In a companion paper, we provide nearly complete answers to the remaining three questions.

The current situation with regard to the existence of a HMM for a stationary stochastic process can be summarized as follows: With every process over a finite alphabet we can associate a 'Hankel' matrix, call it H. Then the process has a HMM *only if* H has finite rank. However, the converse

Tata Consultancy Services, No. 1 Software Units Layout, Madhapur, Hyderabad 500 081, India, sagar@atc.tcs.co.in

0-7803-9568-9/05/$20.00 ©2005 IEEE

is not true in general. If it is known at the outset that the process has a HMM, and if some other assumptions hold, then it is possible to construct another HMM for the process, (usually with a much larger state space than the original).

Against this background, in the present paper we show the following:

1) Suppose a process has finite Hankel rank. Then there *always* exists a 'quasi-realization' of the process. That is, there exist a row vector, a column vector, and a set of matrices, together with a formula for computing the frequencies of arbitrary strings that is similar to the corresponding formula for HMM's. Moreover, the quasi-realization can be chosen to be 'regular,' in the sense that the size of the 'state space' in the quasi-realization can always be chosen to equal the rank of the Hankel matrix. Further, two different regular quasi-realizations of the same process are related through a similarity transformation. Hence, given a finite Hankel-rank process, it is possible to determine whether or not it has a regular HMM in the conventional sense, by testing the feasibility of a nonlinear programming problem.
2) If in addition the process is α-mixing, every regular quasi-realization has additional features. Specifically, a matrix associated with the finitely computable model (which plays the role of the state transition matrix in a HMM) satisfies the 'quasi-strong Perron property' (its spectral radius is one, the spectral radius is also an eigenvalue, and there are no other eigenvalues on the unit circle). A corollary is that if a finite Hankel rank α-mixing process has a regular HMM in the conventional sense, then the associated Markov chain is irreducible and aperiodic. While this last result is not surprising, it does not seem to have been stated explicitly.
3) Suppose a process has finite Hankel rank, is both α-mixing as well as 'ultra-mixing' (a property defined here), and in addition satisfies a technical condition. Then it has an irreducible HMM realization (not just a quasi-realization). Moreover, the Markov process underlying the HMM is either aperiodic (and is thus α-mixing), or else satisfies a 'consistency condition.' In the other direction, suppose a HMM satisfies the consistency condition plus another technical condition. Then the associated output process has finite Hankel rank, is α-mixing and is also ultra-mixing. Taken together, these two results show that, modulo two technical conditions, the finite Hankel rank condition, α-mixing, and ultra-mixing are 'almost' necessary and sufficient for a process to have an irreducible and aperiodic HMM.

The basic ideas of HMM realization theory are more than forty years old. A sufficient condition for the existence of a HMM, involving the existence of a suitable polyhedral cone, was established by Dharmadhikari [14]. So what is new forty years later? In [1], Anderson says that "The use of a cone condition, described by some as providing a solution to the realization problem, constitutes (in this author's opinion) a restatement of the problem than a solution of it. This is because the cone condition is encapsulated by a set of equations involving unknowns; there is no standard algorithm for checking the existence of a solution or allowing construction of a solution;" He then proceeds to give sufficient conditions for the existence of a suitable cone, as well as a procedure for constructing it. However, in order to do this he *begins* with the assumption that the process under study has a HMM; see Assumption 1 on p. 84 of [1]. As a consequence, some of the proofs in that paper make use of the properties of the unknown but presumed to exist HMM realization. In contrast, in the present paper the objective is to state all conditions only in terms of the process under study, and nothing else, and *deduce* the existence of a HMM, rather than to *postulate* it, as in [1]. This objective is achieved. Thus it is believed that the present paper is perhaps the first one in the forty year-old literature on the subject to state a set of conditions purely in terms of the process under study for the existence of a HMM. The fact that the set of necessary conditions and the set of sufficient conditions are so close is an added bonus.

II. Preliminaries

Throughout the paper, we use the definition of a HMM introduced in [1], which may be termed the 'joint Markov process' type of HMM. Suppose $\{\mathcal{Y}_t\}$ is a stationary stochastic process on the finite alphabet $\mathcal{M} := \{1, \ldots, m\}$. We say that the process $\{\mathcal{Y}_t\}$ has a **HMM of the 'joint Markov process' type** if there exists another stationary stochastic process $\{\mathcal{X}_t\}$ over a finite state space $\mathcal{N} := \{1, \ldots, n\}$ such that the joint process $\{(\mathcal{X}_t, \mathcal{Y}_t\}$ is Markov, and in addition,

$$\Pr\{(\mathcal{X}_t, \mathcal{Y}_t)|\mathcal{X}_{t-1}, \mathcal{Y}_{t-1}\} = \Pr\{(\mathcal{X}_t, \mathcal{Y}_t)|\mathcal{X}_{t-1}\}. \quad (1)$$

This definition ensures that the process $\{\mathcal{X}_t\}$ by itself is a Markov process.

Given an integer l, the set $\mathcal{M}^l$ consists of l-tuples. These can be arranged either in first-lexical order (flo) or last-lexical order (llo). First-lexical order refers to indexing the first element, then the second, and so on, while last-lexical order refers to indexing the last element, then the next to last, and so on. For example, suppose $m = 2$ so that $\mathcal{M} = \{1, 2\}$. Then

$$\mathcal{M}^3 \text{ in llo } = \{111, 112, 121, 122, 211, 212, 221, 222\},$$

$$\mathcal{M}^3 \text{ in flo } = \{111, 211, 121, 221, 112, 212, 122, 222\}.$$

Since the process $\{\mathcal{Y}_t\}$ is assumed to be stationary, given any finite string $\mathbf{u} \in \mathcal{M}^*$, we can speak of its frequency $f_{\mathbf{u}}$. Thus, if $|\mathbf{u}| = l$ and $\mathbf{u} = u_1 \ldots u_l$, we have

$$f_{\mathbf{u}} := \Pr\{(\mathcal{Y}_{t+1}, \mathcal{Y}_{t+2}, \ldots, \mathcal{Y}_{t+l}) = (u_1, u_2, \ldots, u_l)\}.$$

Since the process is stationary, the above probability is independent of t.

Given integers $k, l \geq 1$, the matrix $F_{k,l}$ is defined as

$$F_{k,l} = [f_{\mathbf{uv}}, \mathbf{u} \in \mathcal{M}^k \text{ in flo}, \mathbf{v} \in \mathcal{M}^l \text{ in llo}] \in [0,1]^{m^k \times m^l}.$$

Thus the rows of $F_{k,l}$ are indexed by an element of $\mathcal{M}^k$ in flo, while the columns are indexed by an element of $\mathcal{M}^l$ in llo. For example, suppose $m = 2$. Then

$$F_{1,2} = \begin{bmatrix} f_{111} & f_{112} & f_{121} & f_{122} \\ f_{211} & f_{212} & f_{221} & f_{222} \end{bmatrix},$$

whereas

$$F_{2,1} = \begin{bmatrix} f_{111} & f_{112} \\ f_{211} & f_{212} \\ f_{121} & f_{122} \\ f_{221} & f_{222} \end{bmatrix}.$$

Given integers $k, l \geq 1$, we define the matrix $H_{k,l}$ as

$$H_{k,l} := \begin{bmatrix} F_{0,0} & F_{0,1} & \ldots & F_{0,l} \\ F_{1,0} & F_{1,1} & \ldots & F_{1,l} \\ \vdots & \vdots & \vdots & \vdots \\ F_{k,0} & F_{k,1} & \ldots & F_{k,l} \end{bmatrix}.$$

The matrix $H_{k,l}$ resembles a Hankel matrix in the sense that the matrix in the (i, j)-th block consists of frequencies of strings of length $i+j$. Finally, we define H (without any subscripts) to be the infinite matrix of the above form, that is,

$$H := \begin{bmatrix} F_{0,0} & F_{0,1} & \ldots & F_{0,l} & \ldots \\ F_{1,0} & F_{1,1} & \ldots & F_{1,l} & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ F_{k,0} & F_{k,1} & \ldots & F_{k,l} & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}.$$

Through a mild abuse of language we refer to H as the Hankel matrix associated with the process $\{\mathcal{Y}_t\}$.

Suppose the process $\{Y_t\}$ has a HMM. Then there is a very compact formula for computing the frequency of any string $\mathbf{u} \in \mathcal{M}^*$. For $1 \leq i, j \leq n, u \in \mathcal{M}$, define

$$m_{ij}^{(u)} := \Pr\{\mathcal{Y}_{t+1} = u \& \mathcal{X}_{t+1} = j | \mathcal{X}_t = i\}.$$

Define $M^{(u)}$ to be the $n \times n$ matrix whose ij-th entry is $m_{ij}^{(u)}$. Let A denote the state transition matrix of the $\mathcal{X}_t$ process. Then it is clear that $A = \sum_{u \in \mathcal{M}} M^{(u)}$. Let π denote the stationary distribution of the $\mathcal{X}_t$ process, so that $\pi = \pi A$. Let $\mathbf{e}_n$ denote the column vector consisting of n one's. Then, for every $\mathbf{u} \in \mathcal{M}^*$, it can be shown that

$$\begin{aligned} f_{\mathbf{u}} &= \sum_{i=1}^{n} \sum_{j_1=1}^{n} \cdots \sum_{j_l=1}^{n} \pi_i m_{ij_1}^{(u_1)} \cdots m_{j_{l-1}j_l}^{(u_l)} \\ &= \pi M^{(u_1)} \cdots M^{(u_l)} \mathbf{e}_n. \end{aligned} \quad (2)$$

Note that

$$\sum_{l \in \mathcal{M}} M^{(l)} = A, \ \pi \left[\sum_{l \in \mathcal{M}} M^{(l)} \right] = \pi, \quad (3)$$

$$\left[\sum_{l \in \mathcal{M}} M^{(l)} \right] \mathbf{e}_n = \mathbf{e}_n. \quad (4)$$

The formula (3) leads at once to the following result, which can be traced back to the earliest literature in the subject.

Theorem 1: Suppose $\{\mathcal{Y}_t\}$ has a HMM with the associated $\{\mathcal{X}_t\}$ process having n states. Then $\text{Rank}(H) \leq n$.

Thus $\text{Rank}(H)$ being finite is a *necessary* condition for the given process to have a HMM. However, the converse is not true in general, as shown by Fox and Rubins [19] and Dharmadhikari and Nadkarni [17].

The complete realization problem is essentially one of 'inverting' the formula (3). Given $f_{\mathbf{u}}$ for every $\mathbf{u} \in \mathcal{M}^*$, we need to construct a vector π and matrices $M^{(u)}, u \in \mathcal{M}$ such that (3) holds. Let us refer to the process $\{\mathcal{Y}_t\}$ as 'having finite Hankel rank' if $\text{Rank}(H) < \infty$.

We conclude this subsection by recalling a negative result of Sontag [30], in which he shows that the problem of deciding whether or not a given 'Hankel' matrix has finite rank is undecidable.

III. Existence of Regular Quasi-Realizations for Finite Hankel Rank Processes

In this section, we study processes whose Hankel rank is finite, and show that it is *always* possible to construct a 'regular quasi-realization' of such a process. The results of this section are not altogether surprising. Given that the infinite matrix H has finite rank, it is clear that there *must exist* recursive relationships between its various elements. Earlier work, most notably [13], [10], contains some such recursive relationships. However, the present formulae are the cleanest, and also the closest to the conventional formula (3). Moreover, the above formulae are the basis for the construction of a 'true' (as opposed to quasi) HMM realization in subsequent sections.

Some notation is introduced to facilitate the subsequent proofs. Suppose k, l are integers, and $I \subseteq \mathcal{M}^k, J \subseteq \mathcal{M}^l$; thus every element of I is a string of length k, while element of J is a string of length l. Then we define

$$F_{I,J} := \begin{bmatrix} f_{\mathbf{i}_1\mathbf{j}_1} & f_{\mathbf{i}_1\mathbf{j}_2} & \ldots & f_{\mathbf{i}_1\mathbf{j}_{|J|}} \\ f_{\mathbf{i}_2\mathbf{j}_1} & f_{\mathbf{i}_2\mathbf{j}_2} & \ldots & f_{\mathbf{i}_2\mathbf{j}_{|J|}} \\ \vdots & \vdots & \vdots & \vdots \\ f_{\mathbf{i}_{|I|}\mathbf{j}_1} & f_{\mathbf{i}_{|I|}\mathbf{j}_2} & \ldots & f_{\mathbf{i}_{|I|}\mathbf{j}_{|J|}} \end{bmatrix}. \quad (5)$$

Thus $F_{I,J}$ is a submatrix of $F_{k,l}$ and has dimension $|I| \times |J|$. In the same spirit, if I is a subset of $\mathcal{M}^k$ and l is an integer, we use the 'mixed' notation $F_{I,l}$ to denote $F_{I,\mathcal{M}^l}$. Finally, given any string $\mathbf{u} \in \mathcal{M}^*$, we define

$$F_{k,l}^{(\mathbf{u})} := [f_{\mathbf{iuj}}, \mathbf{i} \in \mathcal{M}^k \text{ in flo}, \mathbf{j} \in \mathcal{M}^l \text{ in llo}], \quad (6)$$

$$F_{I,J}^{(\mathbf{u})} := [f_{\mathbf{iuj}}, \mathbf{i} \in I, \mathbf{j} \in J]. \quad (7)$$

Lemma 1: Suppose H has finite rank. Then there exists a smallest integer k such that

$$\text{Rank}(F_{k,k}) = \text{Rank}(H).$$

Moreover, for this k, we have

$$\text{Rank}(F_{k,k}) = \text{Rank}(H_{k+l,k+s}),\ \forall l, s \geq 0. \tag{8}$$

Definition 1: Suppose a process $\{\mathcal{Y}_t\}$ has finite Hankel rank r. Suppose $n \geq r$, $\mathbf{x}$ is a row vector in $\mathbf{R}^n$, $\mathbf{y}$ is a column vector in $\mathbf{R}^n$, and $C^{(u)} \in \mathbf{R}^{n \times n}\ \forall u \in \mathcal{M}$. Then we say that $\{n, \mathbf{x}, \mathbf{y}, C^{(u)}, \mathbf{u} \in \mathcal{M}\}$ is a **quasi-realization** of the process if three conditions hold. First,

$$f_{\mathbf{u}} = \mathbf{x}C^{(u_1)} \ldots C^{(u_l)}\mathbf{y}\ \forall \mathbf{u} \in \mathcal{M}^*, \tag{9}$$

where $l = |\mathbf{u}|$. Second,

$$\mathbf{x}\left[\sum_{u \in \mathcal{M}} C^{(u)}\right] = \mathbf{x}.$$

Third,

$$\left[\sum_{u \in \mathcal{M}} C^{(u)}\right]\mathbf{y} = \mathbf{y}.$$

We say that $\{n, \mathbf{x}, \mathbf{y}, C^{(u)}, \mathbf{u} \in \mathcal{M}\}$ is a **regular quasi-realization** of the process if $n = r$, the rank of the Hankel matrix.

Note that a regular quasi-realization in some sense completes the analogy with the formulas (3), (3) and (4).

Now consider the matrix $F_{k,k}$, which is chosen so as to have rank r. Thus there exist sets $I, J \subseteq \mathcal{M}^k$, such that $|I| = |J| = r$ and $F_{I,J}$ has rank r. For each symbol $u \in \mathcal{M}$, define

$$\bar{D}^{(u)} := F_{I,J}^{-1}F_{I,J}^{(u)}, D^{(u)} := F_{I,J}^{(u)}F_{I,J}^{-1}.$$

Theorem 2: Suppose the process $\{\mathcal{Y}_t\}$ has finite Hankel rank, say r. Then the process always has a regular quasi-realization. In particular, choose the integer k as in Lemma 1, and choose index sets $I, J \subseteq \mathcal{M}^k$ such that $|I| = |J| =$ and $F_{I,J}$ has rank r. Define the matrices $U, V, D^{(u)}, \bar{D}^{(u)}$ as before. The the following two choices are regular quasi-realizations. First, let

$$\mathbf{x} = \theta := F_{0,J}F_{I,J}^{-1},\ \mathbf{y} = \phi := F_{I,0},\ C^{(u)} = D^{(u)}\ \forall u \in \mathcal{M}. \tag{10}$$

Second, let

$$\mathbf{x} = \bar{\theta} := F_{0,J},\ \mathbf{y} = \bar{\phi} := F_{I,J}^{-1}F_{I,0},\ C^{(u)} = \bar{D}^{(u)}\ \forall u \in \mathcal{M}. \tag{11}$$

This result can be compared to [1], Theorem 1, p. 90 and Theorem 2, p. 92..

Next, it is shown that any two 'regular' quasi-realizations of the process are related through a similarity transformation.

Theorem 3: Suppose a process $\{\mathcal{Y}_t\}$ has finite Hankel rank r, and suppose $\{\theta_1, \phi_1, D_1^{(u)}, u \in \mathcal{M}\}$ and $\{\theta_2, \phi_2, D_2^{(u)}, u \in \mathcal{M}\}$ are two quasi-realizations of this process. Then there exists a nonsingular matrix T such that

$$\theta_2 = \theta_1 T^{-1}, D_2^{(u)} = TD_1^{(u)}T^{-1}\ \forall u \in \mathcal{M}, \phi_2 = T\phi_1.$$

While this theorem is not surprising, it does not seem to have been explicitly stated in the literature.

IV. Spectral Properties of Alpha-Mixing Processes

In this section, we add the assumption that the finite Hankel rank process under study is also α-mixing, and show that the regular quasi-realizations satisfy the so-called 'quasi-strong Perron property.'

Theorem 4: Suppose the process $\{\mathcal{Y}_t\}$ is α-mixing and has finite Hankel rank r. Let $\{r, \mathbf{x}, \mathbf{y}, C^{(\mathbf{u})}, u \in \mathcal{M}\}$ be any regular quasi-realization of the process, and define

$$S := \sum_{u \in \mathcal{M}} C^{(\mathbf{u})}.$$

Then $S^l \to \mathbf{yx}$ as $l \to \infty$, $\rho(S) = 1$, $\rho(S)$ is a simple eigenvalue of S, and all other eigenvalues of S have magnitude strictly less than one.

This theorem can be compared with [1], Theorem 4, p. 94.

Corollary 1: Suppose a stationary process $\{\mathcal{Y}_t\}$ is α-mixing and has a regular realization. Then the underlying Markov chain is aperiodic and irreducible.

V. Ultra-Mixing Processes and the Existence of HMM's

In the previous two sections, we studied the existence of quasi-realizations. In this section, we study the existence of 'true' (as opposed to quasi) realizations. We introduce a new property known as 'ultra-mixing' and show that if a process has finite Hankel rank, and is both α-mixing as well as ultra-mixing, then modulo a technical condition it has a HMM where the underlying Markov chain is itself α-mixing (and hence aperiodic and irreducible) or else satisfies a 'consistency condition.' The converse is also true, modulo another technical condition.

A. The Consistency Condition

Before presenting the sufficient condition for the existence of a HMM, we recall a very important result from [1]. Consider a 'joint Markov process' HMM where the associated matrix A (the transition matrix of the $\{\mathcal{X}_t\}$ process) is irreducible. In this case, it is well known and anyway rather easy to show that the state process $\{\mathcal{X}_t\}$ is α-mixing if and only if the matrix A is aperiodic in addition to being irreducible. If A is aperiodic (so that the state process is α-mixing), then the output process $\{\mathcal{Y}_t\}$ is also α-mixing. See for example [31], Theorem 3.12, p. 110. However, the converse is not always true. It is possible for the output process to be α-mixing even if the state process is not. Theorem 5 of [1] gives necessary and sufficient conditions for this to happen. This condition is referred to here as the 'consistency condition.'

B. The Ultra-Mixing Property

In earlier sections, we studied the spectrum of various matrices under the assumption that the process under study is α-mixing. For present purposes, we introduce a different kind of mixing property.

Definition 2: Given the process $\{\mathcal{Y}_t\}$, suppose it has finite Hankel rank, and let k denote the unique integer defined in Lemma 1. Then the process $\{\mathcal{Y}_t\}$ is said to be **ultra-mixing** if there exists a sequence $\{\delta_l\} \downarrow 0$ such that

$$\left| \frac{f_{\mathbf{iu}}}{f_{\mathbf{u}}} - \frac{f_{\mathbf{iuv}}}{f_{\mathbf{uv}}} \right| \leq \delta_l, \ \forall \mathbf{i} \in \mathcal{M}^k, \mathbf{u} \in \mathcal{M}^l, \mathbf{v} \in \mathcal{M}^*. \quad (12)$$

Note that, the way we have defined it here, the notion of ultra-mixing is defined only for processes with finite Hankel rank.

The ultra-mixing property can be interpreted as a kind of long-term indepedence. It says that the conditional probability that a string begins with $\mathbf{i}$, given the next l entries, is just about the same whether we are given just the next l entries, or the next l entries as well as the still later entries.

In [26], Kalikow defines a notion that he calls a 'uniform martingale,' which is the same as an ultra-mixing stochastic process. He shows that a stationary stochastic process over a finite alphabet is a uniform martingale if and only if it is also a 'random Markov process.'

C. The Main Result

Recall that a set $\mathcal{S} \subseteq \mathbf{R}^r$ is said to be a 'cone' if $\mathbf{x}, \mathbf{y} \in \mathcal{S} \Rightarrow \alpha\mathbf{x} + \beta\mathbf{y} \in \mathcal{S} \ \forall \alpha, \beta \geq 0$. The term 'convex cone' is also used to describe such an object. Given a (possibly infinite) set $\mathcal{V} \subseteq \mathbf{R}^r$, the symbol $\mathrm{Cone}(\mathcal{V})$ denotes the smallest cone containing $\mathcal{V}$, or equivalently, the intersection of all cones containing $\mathcal{V}$. Next, we introduce two cones that play a special role in the proof. Suppose as always that the process under study has finite Hankel rank, and define the integer k as in Lemma 1. Throughout, we use the quasi-realization $\{r, \theta, \phi, D^{(u)}\}$ defined in (10). Now define

$$\mathcal{C}_c := \mathrm{Cone}\{D^{(\mathbf{u})}\phi : \mathbf{u} \in \mathcal{M}^*\},$$

$$\mathcal{C}_o := \{\mathbf{y} \in \mathbf{R}^r : \theta D^{(\mathbf{v})}\mathbf{y} \geq 0, \ \forall \mathbf{v} \in \mathcal{M}^*\}.$$

Note that from (9) and (10) we have

$$\theta D^{(\mathbf{v})} D^{(\mathbf{u})}\phi = f_{\mathbf{uv}} \geq 0, \ \forall \mathbf{u}, \mathbf{v} \in \mathcal{M}^*.$$

Hence $D^{(\mathbf{u})}\phi \in \mathcal{C}_o \ \forall \mathbf{u} \in \mathcal{M}^*$, and as a result $\mathcal{C}_c \subseteq \mathcal{C}_o$. Moreover, both $\mathcal{C}_c$ and $\mathcal{C}_o$ are invariant under $D^{(w)}$ for each $w \in \mathcal{M}$. The key difference between $\mathcal{C}_c$ and $\mathcal{C}_o$ is that the former cone need not be closed, whereas the latter cone is always closed (this is easy to show).

As before, let r denote the rank of the Hankel matrix, and choose subsets $I, J \subseteq \mathcal{M}^k$ such that $|I| = |J| = r$ and $F_{I,J}$ has rank r. For each finite string $\mathbf{u} \in \mathcal{M}^*$, define the vectors

$$\mathbf{p_u} := \frac{1}{f_{\mathbf{u}}} F_{I,0}^{(u)} = [f_{\mathbf{iu}}/f_{\mathbf{u}}, i \in I] \in [0,1]^{r \times 1},$$

$$\mathbf{q_u} := \frac{1}{f_{\mathbf{u}}} F_{0,J}^{(u)} = [f_{\mathbf{uj}} \ f_{\mathbf{u}}, j \in J] \in [0,1]^{1 \times r}.$$

Thus both $\mathbf{p_u}$ and $\mathbf{q_u}$ are probability vectors, in the sense that their components are all nonnegative and add up to one. Now let us consider the countable collection of probability vectors $\mathcal{A} := \{\mathbf{p_u} : \mathbf{u} \in \mathcal{M}^*\}$. Since $\mathbf{p_u}$ equals $D^{(\mathbf{u})}\phi$ within a scale factor, it follows that $\mathcal{C}_c = \mathrm{Cone}(\mathcal{A})$. Moreover, since $\mathcal{A} \subseteq \mathcal{C}_c \subseteq \mathcal{C}_o$ and $\mathcal{C}_o$ is a closed set, it follows that the set of cluster points of $\mathcal{A}$ is also a subset of $\mathcal{C}_o$.[1]

Now we state the main result of this section.

Theorem 5: Suppose the process $\{\mathcal{Y}_t\}$ satisfies the following conditions:

1) It has finite Hankel rank.
2) It is ultra-mixing.
3) It is α-mixing.
4) The cluster points of the set $\mathcal{A}$ of probability vectors are finite in number and lie in the interior of the cone $\mathcal{C}_o$.

Under these conditions, the process has an irreducible 'joint Markov process' hidden Markov model. Moreover the HMM satisfies the consistency conditions of [1], Theorem 5.

Remark: Among the hypotheses of Theorem 5, Conditions 1 through 3 are 'real' conditions, whereas Condition 4 is a 'technical' condition.

The proof of Theorem 5 depends on first establishing the following lemma, which shows that under the stated hypotheses there exists a *polyhedral* cone that satisfies certain invariance properties. The existence of such an invariant polyhedral cone was first *assumed* in [14], and then used to establish the existence of a HMM. However, the present paper is the first one to *deduce* the existence of such a polyhedral invariant cone in terms of properties that can be stated in terms of the process under study alone.

Lemma 2: Suppose the process under study is ultra-mixing, and that the cluster points of the probability vector set $\mathcal{A}$ are finite in number and belong to the interior of the cone $\mathcal{C}_c$. Choose D to be the *minimum norm solution* of the equation $F_{k+1,k} = DF_{k,k}$, and partition D in the usual format. As before, define $\mathbf{b} := \mathbf{e}^t_{m^k}(I - \Pi)$ to be the projection of $\mathbf{e}^t_{m^k}$ onto the row range of $F_{k,k}$. Then there exists a polyhedral cone $\mathcal{P}$ such that

1) $\mathcal{P}$ is invariant under each $D^{(u)}, \mathbf{u} \in \mathcal{M}$.
2) $\mathcal{C}_o \subseteq \mathcal{P} \subseteq \mathcal{C}_c$.
3) $F_{k,0} \in \mathcal{P}$.
4) $\mathbf{b} \in \mathcal{P}^p$.
5) $\phi \in \mathcal{P}$.
6) $\theta^t \in \mathcal{P}^p$.

Theorem 5 gives *sufficient* conditions for the existence of an irreducible HMM that satisfies some consistency conditions in addition. It is therefore natural to ask how close these sufficient conditions are to being necessary. The paper [1] also answers this question.

Theorem 6: Suppose a HMM satisfies the consistency condition of [1], Theorem 5. Suppose in addition that there

[1]Recall that a vector $\mathbf{y}$ is said to be a 'cluster point' of $\mathcal{A}$ if there exists a sequence in $\mathcal{A}$, no entry of which equals $\mathbf{y}$, converging to $\mathbf{y}$. Equivalently, $\mathbf{y}$ is a cluster point if $\mathcal{A}$ if every neighbourhood of $\mathbf{y}$ contains a point of $\mathcal{A}$ not equal to $\mathbf{y}$.

exists an index $q \leq s$ such that the following property holds: For every string $\mathbf{u} \in \mathcal{M}^q$ and every integer r between 1 and p, every column of the product $M_r^{(u_1)} M_{r+1}^{(u_2)} \ldots M_{r+q-1}^{(u_q)}$ is either zero or else is strictly positive. In this computation, any subscript M_i is replaced by i mod p if $i > p$. With this property, the HMM is α-mixing and also ultra-mixing.

For a proof, see [1], Lemma 2.

Thus we see that there is in fact a very small gap between the sufficiency condition presented in Theorem 5 and the necessary condition discovered earlier in [1]. If the sufficient conditions of Theorem 5 are satisfied, then there exists an irreducible HMM that also satisfies the consistency conditions of [1], Theorem 5. Conversely, if an irreducible HMM satisfies the consistency conditions, and one other technical condition, then it satisfies three out of the four hypotheses of Theorem 5, the only exception being the technical condition about the cluster points lying in the interior of the cone $\mathcal{C}_c$.

VI. Conclusions and Future Work

In this paper, we have advanced the theory of constructing a stochastic model for a process assuming values in a finite alphabet. In particular, it has been shown that *every* process having the finite Hankel rank property always has a regular 'quasi-realization,' whether or not it has a regular realization as a HMM. A new notion called "ultra-mixing" has been introduced. It has been shown that if a finite Hankel rank process is both α-mixing and ultra-mixing, and if an additional technical condition is satisfied, then the process has an irreducible HMM and satisfies a consistency condition. There is a near converse: If a finite Hankel rank process has an irreducible HMM and satisfies a consistency condition, and also satisfies another technical condition, then the process is both α-mixing as well as ultra-mixing.

Acknowledgements

The author thanks Prof. Probal Chaudhuri of the Indian Statistical Institute, Kolkata, and Prof. Rajeeva Karandikar of the Indian Statistical Institute, Delhi, for several helpful discussions. He also thanks Prof. Isaac Meilijson of Tel Aviv University for drawing his attention to the paper [26].

References

[1] B. D. O. Anderson, "The realization problem for hidden Markov models," *Mathematics of Control, Signals, and Systems*, 12(1), 80-120, 1999.
[2] B. D. O. Anderson, M. Deistler, L. Farina and L. Benvenuti, "Nonnegative realization of a system with a nonnegative impulse response," *IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications*, 43, 134-142, 1996.
[3] P. Baldi and S. Brunak, *Bioinformatics: A Machine Learning Approach*, (Second Edition), MIT Press, Cambridge, MA, 2001.
[4] L. E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains,"*Annals of Mathematical Statistics*, 37, 1554-1563, 1966.
[5] L. E. Baum, T. Petrie, G. Soules and N. Weiss, "A maximization technique occuring in the statistical analysis of probabilistic functions of Markov chains," *Annals of Mathematical Statistics*, 41(1), 164-171, 1970.
[6] L. Benvenuti and L. Farina, "A tutorial on the positive realization problem," *IEEE Transactions on Automatic Control*, 49, 651-664, 2004.
[7] A. Berman and R. J. Plemmons, *Nonnegative Matrices*, Academic Press, New York, 1979.
[8] D. Blackwell and L. Koopmans, "On the identifiability problem for functions of finite Markov chains," *Annals of Mathematical Statistics*, 28, 1011-1015, 1957.
[9] J. W. Carlyle, "Identification of state-calculable functions of finite Markov chains," *Annals of Mathematical Statistics*, 38, 201-205, 1967.
[10] J. W. Carlyle, "Stochastic finite-state system theory," in *System Theory*, L. Zadeh and E. Polak (Eds.), Chapter 10, McGraw-Hill, New York, 1969.
[11] S. E. Cawley, A. L. Wirth and T. P. Speed, "Phat – a gene finding program for *Plasmodium falciparum*," *Molecular & Biochemical Parasitology*, 118, 167-174, 2001.
[12] A. L. Delcher, D. Harmon, S. Kasif, O. White and S. L. Salzberg, "Improved microbial gene identification with GLIMMER," *Nucleic Acids Research*, 27(23), 4636-4641, 1999.
[13] S. W. Dharmadhikari, "Functions of finite Markov chains," *Annals of Mathematical Statistics*, 34, 1022-1031, 1963.
[14] S. W. Dharmadhikari, "Sufficient conditions for a stationary process to be a function of a Markov chain," *Annals of Mathematical Statistics*, 34, 1033-1041, 1963.
[15] S. W. Dharmadhikari, "A characterization of a class of functions of finite Markov chains," *Annals of Mathematical Statistics*, 36, 524-528, 1965.
[16] S. W. Dharmadhikari, "A note on exchangeable processes with states of finite rank," *Annals of Mathematical Statistics*, 40(6), 2207-2208, 1969.
[17] S. W. Dharmadhikari and M. G. Nadkarni, "Some regular and non-regular functions of finite Markov chains," *Annals of Mathematical Statistics*, 41(1), 207-213, 1970.
[18] M. Fliess, "Series rationelles positives et processus stochastique," *Annales de l'Institut Henri Poincaré, Section B*, XI, 1-21, 1975.
[19] M. Fox and H. Rubin, "Functions of processes with Markovian states," *Annals of Mathematical Statistics*, 39, 938-946, 1968.
[20] E. J. Gilbert, "The identifiability problem for functions of Markov chains," *Annals of Mathematical Statistics*, 30, 688-697, 1959.
[21] A. Krogh, M. Brown, I. S. Mian, K. Sjölander and D. Haussler, "Hidden Markov models in computational biology: Applications to protein modeling," *J. Mol. Biol.*, 235, 1501-1531, 1994.
[22] A. Krogh, I. S. Mian and D. Haussler, "A hidden Markov model that finds genes in *E. coli* DNA," *Nucleic Acids Research*, 22(22), 4768-4778, 1994.
[23] J. M. van den Hof, "Realization of continuous-time positive linear systems," *Systems and Control Letters*, 31, 243-253, 1997.
[24] J. M. van den Hof and J. H. van Schuppen, "Realization of positive linear systems using polyhedral cones," *Proceedings of the 33rd IEEE Conference on Decision and Control*, 3889-3893, 1994.
[25] F. Jelinek, *Statistical Methods for Speech Recognition*, MIT Press, Cambridge, MA, 1997.
[26] S. Kalikow, "Random Markov processes and uniform martingales," *Israel Journal of Mathematics*, 71(1), 33-54, 1990.
[27] W. H. Majoros and S. L. Salzberg, "An empirical analysis of training protocols for probabilistic gene finders," *BMC Bioinformatics*, available at `http://www.biomedcentral.com/1471-2105/5/206`, 21 December 2004.
[28] L. W. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, 77(2), 257-285, February 1989.
[29] S. L. Salzberg, A. L. Delcher, S. Kasif and O. White, "Microbial gene identification using interpolated Markov models," *Nucleic Acids Research*, 26(2), 544-548, 1998.
[30] E. D. Sontag, "On certain questions of rationality and decidability," *Journal of Computer and System Science*, 11, 375-381, 1975.
[31] M. Vidyasagar, *Learning and Generalization with Applications to Neural Networks*, Springer-Verlag, London, 2003.
[32] M. Vidyasagar, *Nonlinear Systems Analysis*, SIAM Publications, Philadelphia, PA, 2003.