Systems and Control in the Twenty-First Century

Christopher I. Byrnes
Biswa N. Datta
Clyde F. Martin
David S. Gilliam
Editors

Birkhäuser
Boston • Basel • Berlin

Contents

New Developments in the Theory of Positive Systems

B.D.O. Anderson[1]

1 Introduction

This paper deals with some special finite-dimensional linear systems problems, broadly speaking ones where the underlying matrices in state-variable descriptions of the systems considered contain nonnegative or positive entries.

The problems tend to be difficult, for a number of reasons. These include the fact that one of the common tools of linear system theory, that of replacing a triple $\{A, b, c\}$ realizing a transfer function matrix $H(z) = c^T(zI - A)^{-1}b$ by $\{TAT^{-1}, Tb, (T^{-1})^T c\}$ for an arbitrary nonsingular T, is in general not available, since such a transformation in general will destroy the nonnegativity.

The paper actually focuses on three problems: the so-called positive linear system realization problem, the problem of exponential forgetting of initial conditions and an associated smoothing issue in hidden Markov models, and the problem of realizing a hidden Markov model, given the collection of probabilities of output strings. The latter problem very much draws together ideas from the first two problems.

We begin however with two motivational sections: one sets out some examples of positive systems and the other records some broad questions associated with such systems. We then briefly review the Perron-Frobenius theory on the eigenstructure of nonnegative matrices before tackling the three problems above. For the latter two problems, a sort of time-varying generalization of the Perron-Frobenius theory is needed; we use the treatment of [1] as a base.

In the final section, we record some open problems.

The following individuals helped greatly in the development of the ideas of this paper: Manfred Deistler, Subhrakanti Dey, Lorenzo Farina, Hajime Maeda, John Moore and Louis Shue.

[1]The author wishes to acknowledge the funding of the activities of the Cooperative Research Centre for Robust and Adaptive Systems by the Australian Commonwealth Government under the Cooperative Research Centres Program.

2 Some Examples of Positive Systems

For convenience, we shall restrict attention to discrete-time systems. In nearly all instances, to a discrete-time result there corresponds a continuous-time result.

2.1 Deterministic Systems

Consider the equation

$$x(k+1) = Ax(k) + bu(k). \tag{2.1}$$

This equation is a discrete-line Leslie population model [2], [3] when each entry of $x(k)$ corresponds to members of a population in a given age cohort at time instant k. In a closed population, $b = 0$ and entries of A are nonnegative (most are in fact zero). If immigration is allowed, then b has nonnegative entries and $u(\cdot)$ is nonnegative. Obviously all entries of $x(\cdot)$ are nonnegative, and the total population is given by

$$y(k) = cx(k), \tag{2.2}$$

where $c = [1 \dots 1]$. Variants of this model can be used to distinguish subsets of the population, e.g. males and females.

A second class of positive deterministic systems is exemplified by compartmental models, [4, 5] where the entries of x corresponds to the quantities of different entities (chemicals, water, heat, telephone calls, etc.)

As a final example, we cite *charge routing networks*, which are a special form of realization of a digital filter using MOS technology, [6]. The entries of A, b, c, x, y and u are necessarily nonnegative.

2.2 Markov Chains and Hidden Markov Models

If $X_k, k = 0, 1, \dots$ is a state moving at discrete time instants between one of a finite number of states $1, 2, \dots, N$ in a Markov manner, it is termed a finite state Markov chain, [7]. If $Pr\,[X_{k+1} = i \mid X_k = j] = a_{ij}$ is independent of k, it is stationary. Let

$$\pi(k) = \text{ vector with i-th entry } Pr[X_k = i]. \tag{2.3}$$

Then

$$\pi(k+1) = A\pi(k). \tag{2.4}$$

Here, entries of A and π are nonnegative.

In a hidden Markov model [8, 9] in addition to the state process, there is an observation or measurement process Y_k, assuming values in the set

$\{1, 2, \ldots, M\}$, with Y_k depending probabilistically on X_k, via an $M \times N$ matrix C with $c_{ij} = Pr[Y_k = i \mid X_k = j]$. Let

$$\sigma(k) = \text{ vector with j-th entry } Pr[Y_k = j]. \tag{2.5}$$

Then

$$\sigma(k) = C\pi(k). \tag{2.6}$$

Now (2.4) and (2.5) together define a positive system.

2.3 Multidimensional Positive Systems

Equations of the form

$$\begin{aligned} x(h+1,k+1) &= Ax(h,k+1) + Bx(h+1,k) \\ &\quad + Cu(h,k+1) + Du(h+1,k) \\ y(h,k) &= Hx(h,k) + Ju(h,k), \end{aligned}$$

in which all matrices and vectors are nonnegative, have been used in the modeling of river pollution, gas absorption, and the diffusion and advection of biological material,[10]. Of course, h, k refer to two spatial variables.

Equations of the form

$$x(h+1,k+1) = Ax(h,k+1) + Bx(h+1,k),$$

where x is a vector of probabilities, and $A = \alpha P, B = (1-\alpha)Q$ with $0 \le \alpha \le 1$ and P^T, Q^T stochastic matrices have been used to model two-dimensional Markov chains, [11]. Such chains can be used as signal models in designing image processing algorithms.

3 Some Broad Questions

Based on standard ideas of linear systems, deterministic and stochastic, there are a number of ideas which present themselves in relation to positive systems. Some of these are as follows:

(i) Is there a realization theory, i.e. a way of passing from an external (input/output or transfer function) description to an internal (state variable) description of a deterministic positive system?

(ii) Is there a realization theory for hidden Markov models, i.e. a way of passing from the collection of joint probabilities associated with the output process to a HMM description?

(iii) How may one approximate a high order positive system by a low order positive system?

the poles of the transfer function. It is remarkable that if A is primitive, *this is not the case:*

Lemma 5.1 Let $H(z)$ be a nonzero rational transfer function with nonnegative realization $c^T(zI - A)^{-1}b$ in which A is primitive. Then $\rho(A)$ is necessarily a pole of $H(z)$.

Proof. By primitivity, there exist positive v and w for which $Av = \rho(A)v, w^T A = w^T \rho(A)$; since $\rho(A)$ is the only eigenvalue of A of this magnitude, it is easily checked that $\rho(A)^{-k}A^k \to vw^T$ as $k \to \infty$. Hence

$$\rho(A)^{-k}c^T A^k b \to c^T vw^T b.$$

Since $H(z)$ is nonzero, c and b are not identically zero. Hence $c^T v > 0$, $w^T b > 0$ and $c^T vw^T b > 0$. Thus $\rho(A)^k$ shows up in $h(k)$, i.e. $\rho(A)$ is a pole of $H(z)$.

What if we drop the assumption that A is primitive? Then a simple argument shows that either $\rho(A)$ is a pole of $H(z)$, or with no loss of nonnegativity, uncontrollable and/or unobservable states can be removed to construct a smaller dimension nonnegative realization. We give the unobservable case in detail.

Lemma 5.2. Let $H(z)$ be a nonzero rational transfer function with nonnegative realization $c^T(zI - A)^{-1}b$. Then either $\rho(A)$ is observed or unobservable states can be removed without losing nonnegativity.

Proof. Suppose $Av = \rho(A)v$, and suppose states are ordered so that $c^T = (c_1^T \quad 0 \quad 0)$ and $v^T = (0 \quad 0 \quad v_3^T)$ with c_1, v_3 positive. Then the nonnegativity/
positivity constraints and the equation

$$\begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ v_3 \end{pmatrix} = \rho(A) \begin{pmatrix} 0 \\ 0 \\ v_3 \end{pmatrix}$$

imply $A_{13} = 0, A_{23} = 0$, and a lower dimension realization of $H(z)$ is provided by

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}, \begin{pmatrix} c_1 \\ 0 \end{pmatrix}.$$

We have an immediate consequence.

Theorem 5.1 If a rational $H(z)$ with nonnegative impulse response has a nonnegative realization, the poles of $H(z)$ of maximum modulus must be a subset of those which are allowed eigenvalues of maximum modulus of a nonnegative matrix, and include a positive real pole.

Example. (Rational $H(z)$, nonnegative $h(k)$ with no finite dimensional nonnegative realization). Consider $h(k) = (\frac{1}{2})^k sin^2 k$ for which

$$H(z) = \frac{1}{2}\left[\frac{\frac{1}{2}}{z-\frac{1}{2}} - \frac{z(\frac{1}{2}cos2) - \frac{1}{4}}{z^2 - zcos2 + \frac{1}{4}}\right].$$

The maximum modulus poles of $H(z)$ are $\frac{1}{2}, \frac{1}{2}exp(\pm 2j)$, and these are not of the form $\frac{1}{2}\omega_i$, where ω_i is a k-th root of unity for some integer k. So no finite dimensional nonnegative realization exists.

A major advance was obtained by reformulating the realization problem using cones. Let $X = \{x_i, i = 1,2\ldots\}, x_i \epsilon R^n$. Then $\mathcal{X}$ = cone $X = \{\sum \alpha_i x_i, \alpha_i \geq 0, x_i > 0$ for finitely many $i\}$. The dual of $\mathcal{X}$, is $\mathcal{X}^* = \{y \mid y^T x \geq 0 \, \forall \, x \epsilon \mathcal{X}\}$. For a matrix $P, \mathcal{P}$ = cone P is the cone generated by the columns of P.

Ohta, Maeda and Kodama [14] proved

Theorem 5.2 Let an n-th degree $H(z)$ with nonnegative impulse response have a minimal realization $h^T(zI - F)^{-1}g$. Let $\mathcal{R}$ = cone $[g, Fg, F^2 g, \ldots]$. Then $H(z)$ has a nonnegative realization of dimension N if and only if there exists an $n \times N$ P and $\mathcal{P}$ = cone P with

$$\mathcal{R} \subset \mathcal{P} \qquad F\mathcal{P} \subset \mathcal{P} \qquad h \epsilon \mathcal{P}^*$$

Note that if P is known, construction of nonnegative A, b and c is easy: $\mathcal{R} \subset \mathcal{P} \Rightarrow g \epsilon \mathcal{P} \Rightarrow g = Pb$ for some $b \geq 0$; $F\mathcal{P} \subset \mathcal{P} \Rightarrow FP = PA$ for some $A \geq 0$; $h \epsilon \mathcal{P}^* \Rightarrow h^T P = c^T$ for $c \geq 0$. Then $h^T F^k g = h^T F^k P b = h^T P A^k b = c^T A^k b$.

When does P exist? How may it be found? Answers are to be found in [13].

Theorem 5.3 Suppose an n-th degree $H(z)$ with nonnegative impulse response and minimal realization $h^T(zI - F)^{-1}g$ has just one pole of maximum modulus, and this pole is simple and positive real. Then a P as in Theorem 5.2 can be found.

The idea behind constructing P is as follows:

(i) Normalize $g, Fg, F^2 g \ldots \epsilon R^n$ to unit length (the normalized vectors span the same cone)

(ii) Let $Fv = \rho(F)v, v > 0$. Then $v/\|v\|$ is the limit of the normalized vectors in (i).

(iii) Let $\bar{B}_\epsilon$ be a box of side length 2ϵ symmetrically positioned about $v/\|v\|$. Let B_ϵ be the cone formed by vectors from the origin to the corners of $\bar{B}_\epsilon$. Then for certain finite M_1, M_2 one can take

$$\mathcal{P} = \text{ cone } [g \quad Fg \ldots F^{M1} g \quad B_\epsilon \quad FB_\epsilon \ldots F^{M2} B_\epsilon].$$

The above result assumes there is just a single and simple maximum modulus pole. Extending to the case of a single multiple pole is easy, [13]. With more work, the case of more than one pole of maximum modulus can be treated. Partial results are in [13] and the complete results are in unpublished work of Farina, Maeda and Kitano, [15, 16]. The key to the complete results is the following theorem.

Theorem 5.4 Suppose an n-th degree $H(z)$ has a nonnegative impulse response $h(k)$. For any integer K define K impulse responses obtained by sampling $h(k)$ at K time units apart:

$$\begin{aligned} h_1(k) &= h(kK) \\ h_2(k) &= h(kK+1) \\ &\vdots \\ h_K(k) &= h(hK+\overline{K-1}). \end{aligned}$$

Then $h(k)$ has a nonnegative realization if and only if $h_1(k),\ldots,h_K(k)$ all have nonnegative realizations.

Outline of Proof. If $H(z) = c^T(zI - A)^{-1}b$ with nonnegative A, b, c then $H_i(z) = c^T(zI - A^K)^{-1}A^{i-1}b$. Also, if $H_i(z) = c_i^T(zI - A_i)^{-1}b_i$ for nonnegative A_i, b_i, c_i with impulse response $h_i(k)$, define $\bar{H}_i(z)$ to have impulse response $\{0,\ldots,0,h_i(1),0,0,\ldots h_i(2),\ldots\}$ with $h_i(1)$ the response at the i-th time, $h_i(2)$ the response at the $(K+i)$-th time etc. Then $\bar{H}_i(\bar{z}) = \bar{c}_i^T\ (zI - \bar{A}_i)^{-1}\bar{b}_i$ where

$$\bar{A}_i = \begin{bmatrix} 0 & I & 0 & \ldots & 0 \\ 0 & 0 & I & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \\ & & & & I \\ A_i & 0 & 0 & \ldots & 0 \end{bmatrix}, \qquad \bar{b}_i = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ b_i \end{bmatrix},$$

$$\bar{c}_i^T = [0\ldots c_i^T 0\ldots 0],$$

with c_i^T occurring in block $K - i + 1$. Finally $h(k) = \sum \bar{h}_i(k), H(z) = \sum \bar{H}_i(z)$ and a nonnegative realization is immediate.

How is this applied? To test for and obtain a nonnegative realization first refer to Theorem 5.1. Assume the necessary conditions of Theorem 5.1 are fulfilled.

If $H(z)$ has a single maximum modulus pole, Theorem 5.3 gives realizability just when the pole is positive real. If $H(z)$ has more than one maximum modulus pole, a necessary condition for realizability (by Theorem 5.1) is that for some integer K, λ^K for any maximum modulus pole λ

is positive real. Now each transfer function associated with $h_1(k), \ldots h_K(k)$ either has a single maximum modulus pole at λ^K, in which case Theorem 5.3 gives a positive realization, or there is one or more maximum modulus poles at a smaller value than λ^K and *at the same time* a smaller degree than $H(z)$. (This second possibility will arise if the maximum modulus poles of $H(z)$ are not observed in a particular set of K-spaced impulse response samples).

If the transfer function resulting in this second case has more than one maximum modulus pole, realizability is either ruled out on account of lack of rotational symmetry of the maximum modulus poles of the spectrum (Theorem 5.1), or subsampling is used again. The whole process must terminate in view of the degree reduction which occurs if a subsampled transfer function has more than one pole of maximum modulus.

It is not straightforward to define the minimal dimension of nonnegative realizations, nor to characterize the set of (possibly minimal) nonnegative realizations in terms of one, nor to approximate (with a nice error formula) one nonnegative realization by a lower dimension one. It is in principle possible to answer the question: "does there exist a realization of a particular dimension $N \geq n$?" using Tarski-Seidenberg Theory [17]. For the question can be restated as: do there exist "$a_{ij}, b_i, c_j, i = 1, 2, \ldots, N, j = 1, 2, \ldots N$ such that $a_{ij} \geq 0, \quad b_i \geq 0, \quad c_j \geq 0, \quad h_k = c^T A^{k-1} b, \quad k = 1, 2, \ldots, 2N$"? This is an existence question involving real solutions of polynomial equalities and inequalities, which is the subject of the Tarski-Seidenberg theory, [17].

6 Initial Condition Forgetting and Smoothing in Hidden Markov Models

In Section 2.2, we defined a hidden Markov model. *Filtering* for a hidden Markov model is the process of computing recursively the filtered probability vector $\pi_{k/k}$ with i-th entry $Pr[X_k = i \mid Y_k = y_k, Y_{k-1} = y_{k-1}, \ldots]$. It is also useful to work with $\pi_{k+1/k}$, the one-step-ahead prediction probability vector, with i-th entry $Pr[X_{k+1} = i \mid Y_k = y_k, Y_{k-1} = y_{k-1} \ldots]$.

6.1 Initial Condition Forgetting

The first question we want to consider is: as $k \to \infty$, does $\pi_{k/k}$ become independent of $\pi_{0/0}$ (initial condition) or equivalently, old measurements? Why is this an important question? First, in practical applications of an HMM problem, there will be many situations in which initial condition data is very poor; one would like to know that the filter can "recover" from such inadequacy. Second, if there is not some form of exponential forgetting, there is a risk that round-off and quantization errors present in the numerical implementation of a filter may accumulate to the point where for large time, computational accuracy is lost and the filter is of no use.

To treat the exponential forgetting question, we shall explain how $\pi_{k/k}$ evolves [8]. Let

$$C_m = diag[c_{m1}, c_{m2}, \ldots, c_{mN}],$$

and let C_{y_k} denote that matrix in the set $C_1, C_2, \ldots, C_M$ resulting when Y_k assumes the value y_k. Bayes' rule leads to

$$\pi_{k+1/k} = A\pi_{k/k}(\text{time update}) \tag{6.1}$$

$$\pi_{k+1/k+1} = \frac{C_{y_{k+1}}\pi_{k+1/k}}{[1 \quad 1 \ldots 1]C_{y_{k+1}}\pi_{k+1/k}}(\text{ measurement update}). \tag{6.2}$$

If we let $\bar{\pi}_{k+1/k}$ and $\bar{\pi}_{k+1/k+1}$ denote positively scaled versions of $\pi_{k+1/k}$, $\pi_{k+1/k+1}$ (the latter have entries summing to unity), we have unnormalized update equations:

$$\bar{\pi}_{k+1/k} = A\bar{\pi}_{k/k} \qquad \bar{\pi}_{k+1/k} = C_{y_{k+1}}\bar{\pi}_{k+1/k}$$

or

$$\bar{\pi}_{k+1/k+1} = (C_{y_{k+1}}A)\bar{\pi}_{k/k}. \tag{6.3}$$

Evidently,

$$\bar{\pi}_{k/k} = E_k E_{k-1} \ldots E_1 \bar{\pi}_{0/0}, \tag{6.4}$$

where each E_i is drawn from the set of matrices $\{C_1A, C_2A, \ldots, C_MA\}$.

Products of nonnegative and positive matrices exhibit some properties like those of powers of nonnegative or positive matrices. An extensive analysis has been presented by Seneta [1] using a tool called the Birkhoff contraction coefficient. The key result is that if A, C are both positive, then the columns of $D(k) = E_k E_{k-1} \ldots E_1$ tend to proportionality exponentially fast as $k \to \infty$, i.e.

$$D(k) \to \begin{bmatrix} d_{11}(k) & v_2 d_{11}(k) & \ldots & v_N d_{11}(k) \\ \vdots & \vdots & & \vdots \\ d_{N1}(k) & v_2 d_{N1}(k) & \ldots & v_N d_{N1}(k) \end{bmatrix} = \begin{bmatrix} d_{11}(k) \\ \vdots \\ d_{N1}(k) \end{bmatrix} [1 \quad v_2 \ldots v_N].$$

(The convergence is also provable if A is primitive with C positive, or if A is positive and C is nonnegative, see [18]). It follows that

$$\bar{\pi}_{k/k} \to \begin{bmatrix} d_{11}(k) \\ \vdots \\ d_{N1}(k) \end{bmatrix} [1 \quad v_2 \ldots v_N]\bar{\pi}_{0/0},$$

or that

$$\pi_{k/k} \to \frac{1}{\sum_j d_{j1}(k)} \begin{bmatrix} d_{11}(k) \\ \vdots \\ d_{N1}(k) \end{bmatrix},$$

and so there is exponential forgetting of $\bar{\pi}_{0/0}$.

A related result can be found in [19], obtained by a different but similar tool, perhaps more suited to considering products of stochastic matrices, rather than general nonnegative matrices. Full details of the above argument are in [18], which also indicates how convergence rate bounds are computable, and demonstrates that forgetting occurs at least as fast as the underlying Markov state process forgets its initial probability vector.

As noted above, the convergence results allow A to be only nonnegative as long as it is primitive, or C to be nonnegative, but apparently not both relaxations simultaneously. Indeed, one can construct examples of primitive A and nonnegative C where in the worst case, there is no forgetting of an initial condition, although on average there is a forgetting.

These results are similar to those obtained in Kalman filtering problems, where exponential forgetting of initial conditions is a consequence of detectability and stabilizability assumptions [20]; without exponential forgetting, the numerical behaviour of both an HMM filter and a Kalman filter is likely to be unreliable.

6.2 Fixed-Lag Smoothing of Hidden Markov Models

The vector $\pi_{k/k}$ sums up what the measurements up till time k tell us about X_k. For some fixed Δ (and variable k), the vector $\pi_{k/k+\Delta}$ with i-th entry $Pr[X_k = i \mid Y_j = y_j, j \leq k + \Delta]$ (termed a fixed lag smoothed probability vector) sums up what the measurements up till k *and* then on till $k + \Delta$ tell us about X_k. The vector cannot be computed until time $k + \Delta$, but it should tell us more about X_k than $\pi_{k/k}$. Provided the delay in availability is acceptable, $\pi_{k/k+\Delta}$ (which is not difficult to compute) is a more useful vector than $\pi_{k/k}$, because it will be more informative about X_k. A key question is: how does $\pi_{k/k+\Delta}$ behave with Δ. Work in [18] shows that for a fixed k, $\pi_{k/k+\Delta}$ approaches a limit as $\Delta \to \infty$, at an exponentially fast rate that is the same as that associated with the forgetting of initial conditions (and is certainly uniform in k). Thus when Δ is taken as several times the time constant of this exponential rate, virtually all the improvement which smoothing offers over filtering is extracted.

These results verify conjectures (which were bolstered with simulation data made more than two decades ago, [21]), and parallel results in Kalman filter theory, [22].

7 The HMM Realization Problem

For convenience, we shall work with a slightly modified definition of an HMM in this section. As before, X_k is a stationary Markov state process which can assume N discrete values, $1, 2, \ldots, N$. We shall assume that the output process Y_k can assume M output levels, and that outputs are associated with states according to a set of stationary conditional probabilities:

$$a_{ij}(y) = Pr(X_{k+1} = j, Y_{k+1} = y \mid X_k = i). \tag{7.1}$$

It follows that

$$A = A(1) + \ldots + A(M) \tag{7.2}$$

is a stochastic matrix, with $\pi^T A = \pi^T$ for some nonnegative vector π and $Ae = e$ where $e = [1 \quad 1 \ldots 1]^T$. It is reasonable to assume that A is irreducible, so that π is positive.

Let $y_{\alpha 1} y_{\alpha 2} \ldots y_{\alpha q}$ be an output string (starting with $y_{\alpha 1}$ and ending with $y_{\alpha q}$). Then it is not hard to verify that $Pr(Y_t = y_{\alpha 1}, Y_{t+1} = y_{\alpha 2}, \ldots Y_{t+q-1} = y_{\alpha q})$, written $Pr(y_{\alpha 1} y_{\alpha 2} \ldots y_{\alpha q})$ is given by

$$Pr(y_{\alpha 1} y_{\alpha 2} \ldots y_{\alpha q}) = \pi^T A(y_{\alpha 1}) A(y_{\alpha 2}) \ldots A(y_{\alpha q}) e. \tag{7.3}$$

The realization problem is: given the set of quantities $Pr(y_{\alpha 1} y_{\alpha 2} \ldots y_{\alpha q})$, find $\pi > 0$, $A(1), \ldots, A(M)$ with $A = \sum_i A(i)$ stochastic, and $\pi^T A = \pi^T$ such that the formula (7.3) holds. (Embedded within the problem is of course the question of existence: what conditions on the collection of $Pr(y_{\alpha 1} \; y_{\alpha 2} \ldots y_{\alpha q})$ allow solvability of the problem? Obviously, if the quantities $Pr(y_{\alpha 1} y_{\alpha 2} \ldots y_{\alpha q})$ are known to result from some *unknown* HMM, then the existence question is bypassed).

The HMM problem is a complication of the normal linear system realization problem in several respects:

(i) It is a multivariable problem. In the usual linear system realization problem, one is given a sequence $c^T b, c^T A b, c^T A^2 b, \ldots$ corresponding to the terms in a Laurent series expansion of $c^T (zI - A)^{-1} b$ and one has to find c, A, b. Here, in case there are just two output levels, one is given $c^T b$, $c^T A_1 b$, $c^T A_2 b$, $c^T A_1^2 b$, $c^T A_1 A_2 b$, $c^T A_2 A_1 b$, $c^T A_2^2 b$ etc., these terms corresponding to coefficients in a Laurent series expansion of $c^T (I - z_1^{-1} A_1 - z_2^{-1} A_2)^{-1} b$. More generally, one has coefficients in the expansion of $c^T \left[I - \sum_{i=1}^{M} z_i^{-1} A(i) \right]^{-1} b$, and one has to find the $A(i)$, b and c

(ii) It is a *nonnegative* realization problem, just like the problem of Section 5.

(iii) There are special *constraints*: $A = \sum_{i=1}^{M} A(i)$ must be stochastic, $Ae = e$ and $\pi^T A = \pi^T$.

The problem has been studied for almost three decades, see e.g. [23, 30]. The most comprehensive results are to be found in [31]. We shall give only an outline of the results here.

We shall first describe how the multivariable nature of the problem is handled.

7.1 The Generalized Hankel Matrix

Let Y^* denote the set of finite strings of Y, let Y_t^+ denote $\{Y_{t+1}, Y_{t+2}, \ldots\}$ and let Y_t denote $\{\ldots, Y_{t-1}, Y_t\}$.

Let us enumerate the strings u_i of Y^* lexicographically, ordering entries of a string from left to right, such that the length $|u_i|$ increases monotonically with i. Include the empty sequence as the first element of the enumeration. Thus the ordering is $\{\phi, 0, 1, 00, 10, 01, 11, 000, 100, \ldots\}$. Further, position u_i along the time axis so that the right most (or last occurring) symbol occurs at time t. Thus $u_i \epsilon Y_t^-$.

Consider a second enumeration $\{v_j\}$, identical save that string entries are ordered from right to left; thus the ordering is $\phi, 0, 1, 00, 01, 10, 11, 000, 100, \ldots$ Further, v_j is so positioned that the left most symbol (the first occurring in time) occurs at time $t+1$. Thus $v_j \epsilon Y_t^+$.

The generalized Hankel matrix H is defined to have $i-j$ element $p(u_i v_j)$, where u_i, v_j are the $i-th$ and $j-th$ elements of the two enumerations.

Thus with two output symbols the top left corner of H will appear as

$$
\begin{array}{c}
 \\ \phi \\ 0 \\ 1 \\ 00 \\ 10 \\ 01 \\ 11 \\ \\
\end{array}
\begin{array}{c|cc|ccccc}
\phi & 0 & 1 & 00 & 01 & 10 & 11 & \\
\hline
1 & p(0) & p(1) & p(00) & p(01) & p(10) & p(11) & \cdot \\
\hline
p(0) & p(00) & p(01) & p(000) & p(001) & p(010) & p(011) & \cdot \\
p(1) & p(10) & p(11) & p(100) & p(101) & p(110) & p(111) & \cdot \\
\hline
p(00) & p(000) & p(001) & p(0000) & p(0001) & p(0010) & p(0011) & \cdot \\
p(10) & p(100) & p(101) & p(1000) & p(1001) & p(1010) & p(1011) & \cdot \\
p(01) & p(010) & p(011) & p(0100) & p(0101) & p(0110) & p(0111) & \cdot \\
p(11) & p(110) & p(111) & p(1100) & p(1101) & p(1110) & p(1111) & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot &
\end{array}
$$

We shall let H_{KL} denote the block matrix of H given by

$$[p(u_i v_j) / \, |u_i| = K, |v_j| = L].$$

The top left corner of H above is partitioned as

$$H = \begin{bmatrix} H_{00} & H_{01} & H_{02} & \cdot \\ H_{10} & H_{11} & H_{12} & \cdot \\ H_{20} & H_{21} & H_{22} & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix}.$$

Conventional Hankel matrices have a structure associated with blocks parallel to the anti-diagonal. This is true also of H. Consider

$$H_{K,L+1} = [H_{K,L}(y_1) \;\; H_{K,L}(y_2) \ldots H_{K,L}(y_m)],$$

with the definition

$$H_{K,L}(y_1) = p(u_i y_1 v_j), \text{with } |u_i| = K, |v_j| = L.$$

(The enumeration scheme for the columns is relevant here).

Because of the row enumeration scheme, we can also evidently write

$$H_{K+1,L} = \begin{bmatrix} H_{K,L}(y_1) \\ \vdots \\ H_{K,L}(y_n) \end{bmatrix}.$$

Thus although $H_{K+1,L} \neq H_{K,L+1}$ (as is normal for Hankel matrices), the identity is true provided same rearrangement of entries is permitted.

It is then possible to show

Theorem 7.1 Let H be the infinite generalized Hankel matrix associated with a hidden Markov model with N states. Then rank $H \leq N$.

Suppose now we were simply presented with an infinite generalized Hankel matrix H with finite rank N, say. We can proceed a modest distance towards obtaining an HMM, by taking care of the multivariable finite-dimensional realization problem without taking account of nonnegativity. This is done in the following theorem, which states how vectors x_ϕ, y_ϕ and matrices F_i can be defined satisfying an analog of (7.3), but without a nonnegativity constraint on the entries.

Theorem 7.2 Let H be the infinite generalized Hankel matrix of output string probabilities $Pr(u_i v_j)$ associated with an unknown hidden Markov model with an unknown but finite number of states. Let rank $H = N$, and factor H as

$$H = XY = \begin{bmatrix} X_\phi \\ X_1 \\ X_2 \\ \vdots \end{bmatrix} [Y_\phi \;\; Y_1 \;\; Y_2 \ldots], \tag{7.4}$$

where X_K and Y_L correspond to u_i and v_j of length K and L respectively, and have N columns and rows respectively. Let $\bar{Y}_N$ denote the submatrix of $[Y_\phi \;\; Y_1 \;\; Y_2 \ldots]$ containing the first N linearly independent columns, indexed by strings $\bar{v}_1, \bar{v}_2, \ldots \bar{v}_N$. Define $\bar{Y}_{iN}$ to be those columns of Y indexed by $y_i\bar{v}_1, y_i\bar{v}_2, \ldots, y_i\bar{v}_N$ and set

$$F_1 = \bar{Y}_{1N}\bar{Y}_N^{-1} \ldots F_M = \bar{Y}_{MN}\bar{Y}_N^{-1}. \qquad (7.5)$$

Let x_ϕ^T, y_ϕ denote the first row and first column of X and Y.

Then

$$p(u_i v_j) = x_\phi^T F_1^{\alpha_0} \ldots F_M^{\xi_0} F_1^{\alpha_1} \ldots y_\phi, \qquad (7.6)$$

where $u_i v_j$ consists of α_0 ones, followed by β_0 twos, etc. (Any of the α_i, β_i etc. may be zero).

It is obvious that x_ϕ, y_ϕ and the F_i may not be nonnegative, let alone satisfy further special constraints; e.g. $\sum F_i$ is stochastic with x_ϕ^T and y_ϕ left and right eigenvectors. Also, if H comes from a HMM, the number of states in the HMM may exceed N - in fact (just as in the positive systems realization problem) there may be no HMM realizing H with precisely N states.

7.2 Introducing Nonnegativity via Cones

The result of Ohta et al allowing reformulation of the positive system realization problem has its parallel in the HMM problem, in that it gives the key for introducing nonnegative quantities in place of x_ϕ, y_ϕ and the F_i in (7.6) above, [3, 8, 14, 29].

Theorem 7.3 Adopt the hypotheses of Theorem 6. Then a necessary and sufficient condition for the existence of nonnegative c, b and $A_i, i = 1, \ldots m$ for which

$$p(u_i v_j) = c^T A_1^{\alpha_0} \ldots A_M^{\xi_0} A_1^{\alpha_1} \ldots b, \qquad (7.7)$$

is that there exists a matrix R with associated $\mathcal{R} =$ cone R with

$$F_i \mathcal{R} \subset \mathcal{R} \qquad (7.8)$$

$$y_\phi \epsilon \mathcal{R} \qquad (7.9)$$

$$x_\phi \epsilon \mathcal{R}^* \qquad (7.10)$$

7.3 Cone Existence and Construction

Theorem 7 reformulates, rather than solves, a multivariable nonnegative realization problem. In this subsection, we indicate sufficient (but very broad) conditions for the existence (and constructability) of a cone, [31].

Suppose first that H of finite rank N comes from an (unknown) HMM in which all the probabilities $Pr(X_{k+1} = j, Y_{k+1} = y \mid X_k = i)$ are positive.

- Consider the submatrix of the generalized Hankel matrix comprising the first N linearly independent rows. It has an infinite number of columns.
- Consider a factorization of this N rowed matrix as $\bar{X}_N Y$, where

$$Y = [Y_0 \quad Y_1 \quad Y_2 \ldots] = [y_\phi \,\vdots\, F_1 y_\phi \ldots F_M y_\phi \,\vdots\, F_1^2 y_\phi \quad F_1 F_2 y_\phi \ldots]$$

  and $\bar{X}_N$ is a certain set of N linearly independent rows of X with

$$X^T = [x_\phi \,\vdots\, F_1^T x_\phi \ldots F_M^T x_\phi \,\vdots\, (F_1^T)^2 x_\phi \ldots].$$

  Suppose that $x_\phi, F_i, i = 1, \ldots, M$ and y_ϕ have been identified, as in Theorem 7.2.
- Change the coordinate basis so that in the new basis $X_N^T = I$. Then $x_\phi^T = [1 \quad 0 \ldots 0]$ and Y is simply an N-rowed submatrix of the Hankel matrix.
- Evidently, the infinite set of columns of Y span a cone containing y_ϕ, x_ϕ is in the dual cone, and the cone is F_i–invariant for $i = 1, 2, \ldots, M$. If the cone is polyhedral, we would be done. However, *given that the cone has an infinite set of generators, it may not be polyhedral.*
- It turns out that a large but finite number of columns of Y, together with perturbations around the columns, do define a polyhedral cone containing all the columns of Y that is also F_i-invariant for $i = 1, 2, \ldots, M$.
- The reason is that columns of the generalized Hankel matrix corresponding to very long strings obey (in the limit) some alignment properties. The same will be true of subvectors of these columns containing N entries. The subvectors will then be in a cone defined using early subvectors of the infinite sequence, together with perturbations thereof.

The alignment property just referred to is a consequence of assuming positivity of $Pr(X_{k+1} = j, Y_{k+1} = y \mid X_k = i)$ i.e. of the matrices $A(i), i = 1, 2, \ldots, M$ and actually holds under weaker conditions; as we know, long products of positive and some nonnegative matrices approach a rank one matrix as the number of terms in the product approaches infinity. This idea was used in Section 6 to establish an exponential forgetting property of an HMM filter. Here, it can be used to establish the forgetting property

$$\begin{bmatrix} p(\bar{u}_1 \mid v\bar{v}) \\ \vdots \\ p(\bar{u}_N \mid v\bar{v}) \end{bmatrix} \rightarrow \begin{bmatrix} p(\bar{u}_1 \mid v) \\ \vdots \\ p(\bar{u}_N \mid v) \end{bmatrix},$$

when $|v| \to \infty$, and $|\bar{v}| = 1$. The convergence is exponential with $|v|$, and uniform in $v, \bar{v}$. By Bayes' rule, $p(\bar{u}_i \mid v\bar{v}) = p(\bar{u}_i v\bar{v})/p(v\bar{v})$ and $p(\bar{u}_i \mid v) = p(\bar{u}_i v)/p(v)$. Hence the two conditional probability vectors above are scaled versions of two columns of the N-rowed submatrix of H, corresponding to the columns indexed by v and $v\bar{v}$.

7.4 Satisfying the Special Constraints

At this stage, let us suppose there has been constructed nonnegative c, b and $A_i, i = 1, 2, \ldots, M$ so that (7.7) holds. We now consider how the special constraints (stochasticity and eigenvector properties) can be achieved. By way of a preliminary calculation, let Ψ^k denote a don't care sequence of length k. If there is an underlying (but unknown) hidden Markov model with positive transition probabilities, it is straightforward to show that

$$Pr(u\Psi^k v) \to Pr(u)Pr(v) \text{ as } k \to \infty. \tag{7.11}$$

(This result is hardly unexpected). Appealing to the formula (7.6), one can show that

$$(\sum_{i=1}^{M} F_i)^k \to y_\phi x_\phi^T. \tag{7.12}$$

Additionally, we have

$$\sum_{j=1}^{M} p(\bar{u}_i y_j) = p(\bar{u}_i) \qquad i = 1, 2, \ldots, N,$$

or

$$\bar{X}_N[F_1 y_\phi + \ldots + F_M y_\phi] = \bar{X}_N y_\phi,$$

whence

$$(\sum_{i=1}^{M} F_i) y_\phi = y_\phi. \tag{7.13}$$

Similarly,

$$x_\phi^T (\sum_{i=1}^{M} F_i) = x_\phi^T. \tag{7.14}$$

These facts establish that x_ϕ^T, y_ϕ are the only left and right eigenvectors of $\sum_{i=1}^{M} F_i$ corresponding to eigenvalue 1, and all other eigenvalues are of lesser magnitude.

These facts can be combined with the cone formula $F_i\mathcal{R} \subset \mathcal{R}$ etc. (from which $F_i R = RA_i$) to conclude that c^T, b are unique left and right eigenvectors of $\sum_{i=1}^{M} A_i$ corresponding to eigenvalue 1. A positive diagonal basis transformation taking A_i to $\Lambda A_i \Lambda^{-1} = A(i)$, b to $\Lambda b = e$ and c^T to $c^T\Lambda^{-1}$ ensures that $\sum_{i=1}^{M} A_i$ is stochastic. Simply take $\Lambda^{-1} = \text{diag}\,(b)$.

8 Open Problems

There are a considerable number of open problems, some already alluded to. We list some:

(i) For the nonnegative realization problem, how can one simply figure the minimum dimension of nonnegative realizations?

(ii) How are all nonnegatively minimal realizations of rational $H(z)$ with $h_k \geq 0$ connected?

(iii) How can one approximate (systematically and preferably with an easily interpreted error bound) a high state dimension nonnegative realization by a lower dimension nonnegative realization ?

(iv) One can state variations on the above problems for HMMs.

References

[1] E. Seneta. *Nonnegative Matrices and Markov Chains.* New York: Springer Verlag, 1981.

[2] S. Rinaldi and L. Farina. *I Sistemi Lineari Positivi.* Milan: Citta Studi Edizione, 1995.

[3] P. H. Leslie. On the use of matrices in certain population mathematics. *Biometrika* 35 (1945), 183-212.

[4] H. Maeda, S. Kodama and F. Kajiya. Compartmental system analysis: Realization of a class of linear systems with physical constraints. *IEEE Transactions on Circuits and Systems* CAS-24(1) (1977), 8-14.

[5] J.A. Jacquez. *Compartmental Analysis in Biology and Medicine.* New York: Elsevier, 1972.

[6] A. Gersho and B B Gopinath. Charge-routing networks. *IEEE Trans Circuits and Systems* CAS-26 (1979), 81-92.

[7] M. Rosenblatt. *Markov Processes; Structure and Asymptotic Behaviour.* Berlin: Springer Verlag, 1971. 81-92.

[8] L.R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77(2) (1989), 257-285.

[9] R.J. Elliott, L. Aggoun and J.B. Moore. *Hidden Markov Model: Estimation and Control.* New York: Springer Verlag, 1994.

[10] E. Fornasini and M.E. Valcher. On the spectral and combinatorial structure of 2D positive systems. To appear *Linear Algebra and Applications.*

[11] E. Fornasini. 2D Markov chains. *Linear Algebra and its Applications* **140** (1990), 101-127 .

[12] R.A. Horn and C.R. Johnson. *Matrix Analysis.* Cambridge: Cambridge University Press, 1985.

[13] B.D.O. Anderson, M. Deistler, L. Farina and L. Benvenutti. Nonnegative realization of a linear system with nonnegative impulse response. *IEEE Transactions on Circuits and Systems* **4** (1996), 134-142.

[14] Y. Ohta, H. Maeda and S. Kodama. Realizability, observability and realizability of continuous-time positive systems. *SIAM Journal of Control and Optimization* **22** (1984), 171-180.

[15] T. Kitano and H. Maeda. Positive realization by downsampling the impulse response. Preprint, 1996.

[16] L. Farina. On the existence of a positive realization. Preprint, 1996.

[17] N. Jacobson. *Basic Algebra I.* San Francisco: Freeman and Co., 1974.

[18] L. Shue, B.D.O. Anderson and S. Dey. Exponential stability of filters and smoothers for hidden Markov models. Preprint, 1996.

[19] A. Arapostathis and S.I. Marcus. Analysis of an identification algorithm arising in the adaptive estimation of Markov chains. *Mathematics of Control, Signals and Systems* **3** (1990), 1-29.

[20] B.D.O. Anderson and J.B. Moore. Detectability and stabilizability of time-varying discrete-time linear systems. *SIAM Journal of Control and Optimization* **18** (1981), 20-32.

[21] D. Clements and B.D.O. Anderson. A nonlinear fixed-lag smoother for finite-state Markov processes. *IEEE Trans. Information Theory* **2** (1975), 446-452.

[22] B.D.O. Anderson and J.B. Moore. *Optimal Filtering.* Englewood Cliff, N.J.: Prentice Hall, 1981.

[23] D. Blackwell and L. Koopmans. On the identifiability problem for functions of finite Markov chains. *Annals of Mathematical Statitics* **28** (1957), 1011-1015.

[24] E.J. Gilbert. On the identifiability problem for functions of finite Markov chains. *Annals of Mathematical Statistics* **30** (1959), 688-697.

[25] S.W. Dharmadhikari. Functions of finite Markov chains. *Annals of Mathematical Statistics* **34** (1963), 1022-1032.

[26] S.W. Dharmadhikari. Sufficient conditions for a stationary process to be a function of a finite Markov chain. *Annals of Mathematical Statistics* **34** (1963), 1033-1041.

[27] A. Heller. On stochastic processes derived from Markov chains. *Annals of Mathematical Statistics* **36** (1965), 1286-1291.

[28] M. Fliess. Series rationnelles positives et processus stochastiques. *Ann. Inst. Henri Poincare* **11** (1975), 1-21.

[29] G. Picci. On the internal structure of finite state stochastic processes. *Recent Developments in Variable Structure Systems, Economics and Biology.* (R.R. Mohler and A. Rubert, Eds.). Vol. 162 in *Lecture Notes in Economics and Math. Systems.* Berlin: Springer Verlag, 288-304.

[30] H. Ito, S-I. Amani and K. Kobayashi. Identifiability of hidden Markov information sources and their minimum degrees of freedom. *IEEE Transactions on Information Theory* **38** (1992), 324-333.

[31] B.D.O. Anderson. The realization problem for hidden Markov models. Preprint, 1996.

Department of Systems Engineering and Cooperative Research, Centre for Robust and Adaptive Systems Research, School of Information Sciences and Engineering, Australian National University, Canberra, ACT 0200, Australia