Adv. App. Prob. **11**, 118–133 (1979)
Printed in N. Ireland

FORWARDS AND BACKWARDS MODELS FOR FINITE-STATE MARKOV PROCESSES

B. D. O. ANDERSON,* *University of Newcastle, N.S.W.*
T. KAILATH,** *Stanford University*

Abstract

The construction and properties of reversible and dynamically reversible models for finite-state Markov processes are studied. Certain results on approximating processes with rational power spectra with dynamically reversible finite-state models are also obtained.

REVERSIBLE PROCESS; FINITE STATE PROCESS; POWER SPECTRA; MARKOV PROCESS

1. Introduction

A commonly considered problem is that of spectral factorization: given the power spectrum of a second-order stationary process, find a time-invariant linear system such that excitation of the linear system by white noise results in the output process having the prescribed spectrum. In this paper, we are still concerned with the generation of a rational power spectrum, but now via a different procedure. We conceive of a finite-state Markov process as generating the power spectrum, following ideas in for example [3].

Our main task in this paper is to investigate the applicability of several ideas developed in [10] on the time-reversal of spectrum models. More precisely, in [10] it was shown that given a finite-dimensional linear system driven by white noise running, as usual, forwards in time, there exists an associated system running backwards in time generating the same covariance. The initial state in the forwards model is independent of future noise, while the terminal state in the backwards model is independent of past noise. The first question that then arises in connection with finite-state Markov models is whether, given a forward-time model generating a certain power spectrum, there exists an associated backward-time finite-state Markov model, generating the same

---

Received 9 May 1978.

* Postal address: Department of Electrical Engineering, University of Newcastle, Newcastle, N.S.W. 2308, Australia.

** Postal address: Stanford University, Information Systems Laboratory, Stanford, CA 94305, U.S.A.

This work was supported by the U.S. Army Research Office, Grant DAAG29-77-C-0042, by the Australian Research Grants Committee, National Science Foundation under U.S.–Australian Cooperative Science Program, and by the Air Force Office of Scientific Research, Air Force Systems Command, under Contract AF44-620-74-C-0068.

spectrum. The answer is yes, the backwards model in effect being obtainable by obtaining the reverse transition probabilities associated with the forwards model; the formulas bear an interesting resemblance to those obtained in [10].

In [1], there is also investigated the question of what covariance may be generated by passing white noise into a linear system such that the state process of the linear system has a reversibility property. The same question is considered for finite-state Markov processes in this paper, and we conclude that internal reversibility can be achieved for precisely the same class of covariances. Earlier workers, see e.g. [6], [7], have in fact shown that if the internal reversibility property holds, then the covariance necessarily has a certain structure. The new idea here is thus the minor one of establishing sufficiency (via a constructive procedure yielding an irreducible process) in addition to the known necessity.

Now any scalar zero-mean stationary second-order process $y(\cdot)$ is second-order reversible, in the sense that $E[y(t)y(t+T)]=E[y(t)y(t-T)]$, but not every second-order reversible process has a model with the state process possessing a reversibility property. However, we find on using a definition of Whittle [12] of *dynamic* reversibility for finite-state Markov processes that from any finite-state Markov model which is not internally reversible nor internally dynamically reversible, we can usually construct from it another model generating the same power spectrum which does have the property of internal dynamic reversibility. The Whittle definition stands in clear parallel to that used in discussing second-order processes in a statistical mechanics context [4] and employed in [1], and the result we obtain parallels one of the results of [1].

The results are all valid for discrete and continuous-time models. By and large, we shall restrict consideration to continuous-time models.

2. Process models

Following the notation of Brockett [3] we assume the state process takes one of the values $e_i=[0\cdots 0\quad 1\quad 0\cdots 0]'$ (1 being in the ith position of the n-vector). With $p_i(t)=\Pr[x(t)=e_i]$ and $p=[p_1\quad p_2\cdots p_n]'$ there exists a constant matrix A such that

$$\frac{dp}{dt}=Ap. \tag{2.1}$$

The entries a_{ij} of A satisfy for all i, j

$$a_{ij}\geqq 0 \qquad (i\neq j) \qquad a_{ij}\leqq 0 \qquad \sum_{i=1}^{n} a_{ij}=0. \tag{2.2}$$

We further assume that the nullspace of A is one-dimensional, and there is a unique steady-state probability distribution π for (2.1) with each entry positive:

$$A\pi = 0 \qquad \pi > 0 \qquad 1_n'\pi = 1. \tag{2.3}$$

(Here, 1_n is a vector with 1's in every position.) Under these conditions the finite-state Markov process is irreducible. It is easy to check that these additional requirements also force

$$a_{ii} < 0. \tag{2.4}$$

(For if $a_{ii} = 0$, (2.2) implies $a_{ij} = 0$ for all j and then $Ae_i = 0$, $e_i \not> 0$.)

If either (2.1) operates from $t_0 = -\infty$, or from some finite t_0 with $p(t_0) = \pi$, it follows that $p(t) = \pi$ for all t, and

$$\Pi = E[x(t)x'(t)] = \operatorname{diag}[\pi_i]. \tag{2.5}$$

More generally, it can readily be checked that

$$E[x(t)x'(s)] = e^{A(t-s)}\Pi 1(t-s) + \Pi e^{A'(s-t)}1(s-t) \tag{2.6}$$

where $1(\cdot)$ denotes the unit step function.

An output process y is defined by $y = c'x$ with c an n-vector. It follows that

$$E[y(t)y(s)] = c'e^{A(t-s)}\Pi c 1(t-s) + c'\Pi e^{A'(s-t)}c 1(s-t) \tag{2.7}$$

and

$$\operatorname{Cov}[y(t), y(s)] = E[y(t)y(s)] - c'\pi\pi'c \tag{2.8}$$

since

$$E[y(t)] = c'E[x(t)] = c'\pi. \tag{2.9}$$

3. Backwards time models

A backwards time model is defined in the obvious way. With state vector $x^b(t)$ residing in the same state-space as for the forwards time model of the preceding section, we take $p_i^b = \Pr[x_i^b(t) = e_i]$ and

$$\frac{dp^b}{dt} = -A^b p^b. \tag{3.1}$$

The entries of A^b satisfy (2.2) with a_{ij} replaced by a_{ij}^b, we assume there exists a unique $\pi^b > 0$ with $1_n'\pi^b = 1$ and $A^b\pi^b = 0$, and $a_{ii}^b > 0$ then follows. We conceive of (3.1) operating from $t_f = +\infty$ or from some finite t_f in the region $t \leqq t_f$. If $p^b(t_f) = \pi^b$, then $p^b(t) = \pi^b$ for all t. Furthermore,

$$\Pi^b = E[x^b(t)x^{b\prime}(t)] = \operatorname{diag}[\pi_i^b]$$

and

$$E[x^b(t)x^{b\prime}(s)]=\Pi^b[\exp A^{b\prime}(t-s)]1(t-s) + [\exp A^b(s-t)]\Pi^b 1(s-t). \tag{3.2}$$

The expressions for $E[y^b(t)y^b(s)]$ and $\text{Cov}\,[y^b(t), y^b(s)]$ are immediate.

We are interested in obtaining $E[y^b(t)y^b(s)]=E[y(t)y(s)]$, and $\text{Cov}\,[y^b(t), y^b(s)]=\text{Cov}\,[y(t), y(s)]$. It turns out that we can achieve equality in fact of $E[x^b(t)x^{b\prime}(s)]$ and $E[x(t)x'(s)]$, as well as the associated covariances.

Theorem 3.1. With the forward process model as defined in Section 2, take

$$A^b=\Pi A'\Pi^{-1}. \tag{3.3}$$

Then A^b satisfies the infinitesimal stochastic matrix constraints of (2.2) with a_{ij} replaced by a_{ij}^b; there is a unique π^b such that $A^b\pi^b=0$, viz.

$$\pi^b=\pi>0 \tag{3.4}$$

so that

$$\Pi^b=\Pi. \tag{3.5}$$

Moreover, with $x^b(\cdot)$ and $y^b(\cdot)$ as defined above

$$\begin{aligned} E[x^b(t)x^{b\prime}(s)]&=E[x(t)x'(s)]\\ \text{Cov}\,[x^b(t), x^b(s)]&=\text{Cov}\,[x(t), x(s)] \end{aligned} \tag{3.6}$$

and similarly for the y and y^b processes.

Proof. That $a_{ij}^b\geqq 0$, $i\neq j$, and $a_{ii}^b\leqq 0$, in fact $a_{ii}^b<0$, is trivial from the definitions. Then

$$\begin{aligned} \sum_{i=1}^n a_{ij}^b&=1_n'\Pi A'\Pi^{-1}e_j\\ &=\pi' A'\Pi^{-1}e_j\\ &=0. \end{aligned}$$

Since rank $A=\text{rank}\,A^b$, the nullspace of A^b is one-dimensional. Further, $\Pi A'\Pi^{-1}\pi=\Pi A'1_n=0$, in view of the equality part of (2.2). The claims relating to (3.4) and (3.5) are thus verified. Equations (3.6) and the corresponding equations for y and y_b follow easily from (3.3).

Some intuitive insight into this result is more easily obtained from the discrete-time version; in this context, we study

$$p(t+1)=A^d p(t) \tag{3.7}$$

with suitable conditions on A^d ensuring the existence of a unique π such that

$\pi = A^d\pi$, $\pi > 0$, $1_n'\pi = 1$. Now a_{ij}^d is the quantity $\Pr\{x(t+1) = e_i \mid x(t) = e_j\}$. The quantity a_{ij}^{db} is obtained from the same equation (3.3) as applies for the continuous case:

$$
\begin{aligned}
a_{ij}^{db} &= \pi_i a_{ji}^d \pi_j^{-1} \\
&= \Pr\{x(t) = e_i\} \Pr\{x(t+1) = e_j \mid x(t) = e_i\} \Pr^{-1}\{x(t+1) = e_j\} \qquad (3.8)\\
&= \Pr\{x(t) = e_i \mid x(t+1) = e_j\}
\end{aligned}
$$

where we have used stationarity and Bayes' rule. This formula exposes the time-reversibility clearly. Theorem 3.1 is providing an infinitesimal version of the second-order aspects of this result, and in the light of (3.8), it is not surprising that we can prove a stronger version of that part of Theorem 3.1 relating to the $x(\cdot)$, $x^b(\cdot)$ processes.

Theorem 3.2. With notation as for Theorem 3.1, let $t_1 < t_2 < \cdots < t_k$ be an arbitrary set of times, and $e_{i_1}, \cdots, e_{i_k}$ an arbitrary selection of unit n-vectors. Then

$$\Pr[x^b(t_1) = e_{i_1}, \cdots, x^b(t_k) = e_{i_k}] = \Pr[x(t_1) = e_{i_1}, \cdots, x(t_k) = e_{i_k}]. \qquad (3.9)$$

To prove this result, it is enough, in view of the Markovian nature of the $x(\cdot)$ and $x^b(\cdot)$ processes, to show that

$$\Pr[x^b(t_1) = e_{i_1} \mid x^b(t_2) = e_{i_2}] = \Pr[x(t_1) = e_{i_1} \mid x(t_2) = e_{i_2}]$$

and

$$\Pr[x^b(t_k) = e_{i_k}] = \Pr[x(t_k) = e_{i_k}].$$

Both equalities are easily established, the first using (3.3). Since $y(t_j) = c_{i_j}$ if and only if $x(t_j) = e_{i_j}$, it is evident that (3.9) implies a similar relation between the y and y^b processes.

It is also interesting to note the parallel with results on backwards time models for a process which is the output of a linear system excited by white noise as studied by Ljung and Kailath in [10] (see also [9], [11]). Restricting attention to the stationary scalar case only, a forward model is one of the form

$$\dot{x} = Ax + bu \qquad y = c'x \qquad (3.10)$$

in which $\operatorname{Re} \lambda_i(F) < 0$, $u(\cdot)$ is a zero mean white gaussian process with $E[u(t)u(s)] = \delta(t-s)$. The model either operates forward in time from a finite t_0 in which case $x(t_0)$ is a zero mean gaussian random variable with covariance Π, defined by

$$\Pi A' + A\Pi = -bb' \qquad (3.11)$$

or it can operate from $t_0 = -\infty$. The excitation $u(\cdot)$ is independent of $x(t_0)$, or

$\lim_{t_0 \to -\infty} x(t_0)$. The above set-up yields

$$E[x(t)x'(s)] = e^{A(t-s)}\Pi 1(t-s) + \Pi e^{A'(s-t)} 1(s-t) \tag{3.12}$$

and

$$E[y(t)y(s)] = c'e^{A(t-s)}\Pi c 1(t-s) + c'\Pi e^{A'(s-t)} c 1(s-t). \tag{3.13}$$

Note that $E[x(t)x'(t)] = \Pi$. The parallel with (2.5) through (2.7) should also be observed.

For convenience, we shall assume that in (3.10) the pair $[A, b]$ is completely controllable [2]. This can be shown to imply that Π is positive definite. A reverse time model can then be defined as

$$-\dot{x}^b = \Pi A' \Pi^{-1} x_b + b u_b$$

$$y_b = c' x_b$$

where $u_b(\cdot)$ is a zero mean white gaussian process with $E[u_b(t)u_b'(s)] = \delta(t-s)$. The system operates backward in time from some $x(t_f)$, a zero mean gaussian random variable with $E[x(t_f)x'(t_f)] = \Pi$, and $u(t)$ for $t < t_f$ is independent of $x(t_f)$. One can let $t_f \to \infty$ if desired. Straightforward calculations show that $E[x_b(t)x_b'(s)] = E[x(t)x'(s)]$ and $E[y_b(t)y_b(s)] = E[y(t)y(s)]$, and the parallel with the result of Theorem 3.1 should be evident.

4. Models with reversible state processes

We say that a stationary model of the type considered in the previous sections has a reversible state process if for all t and integer i, j

$$\Pr[x(t) = e_i \mid x(0) = e_j]\pi_j = \Pr[x(t) = e_j \mid x(0) = e_i]\pi_i. \tag{4.1}$$

This idea goes back to Kolmogorov [8]. Since $\Pr[x(t) = e_i \mid x(0) = e_j] = e_i' e^{At} e_j$, (4.1) is equivalent to

$$e_i' e^{At} \pi_j e_j = e_j' e^{At} \pi_i e_i = e_i' \pi_i e^{A't} e_j$$

or

$$e^{At}\Pi = \Pi e^{A't}$$

or

$$A = \Pi A' \Pi^{-1}. \tag{4.2}$$

Comparing (4.2) with (3.3), we see that models with reversible state processes are those for which, in terms of the ideas of the previous section, the backwards model is precisely the forwards time model run in the reverse time direction. Note also that (4.1) is equivalent to $\Pr[x(t)=e_i, x(0)=e_j]=\Pr[x(t)=e_j, x(0)=e_i]$ which is in turn equivalent to $\Pr[y(t)=c_i, y(0)=c_j]=\Pr[y(t)=c_j, y(0)=c_i]$ and the $y(\cdot)$ process is reversible.

The spectral properties of the $y(\cdot)$ process are obtained in for example Keilson's work, [6], [7]. One form of the result is as follows.

Theorem 4.1. Consider a model of the form described in Section 2 with the additional property that (4.2) holds. Then

$$\operatorname{Cov}[y(t), y(s)]=\sum_{i=1}^{n-1} \alpha_i \exp[-\beta_i |t-s|] \tag{4.3}$$

for positive constants α_i, β_i with the β_i distinct. Equivalently, the spectrum of y is $\sum_{i=1}^{n-1} 2\alpha_i\beta_i(\omega^2+\beta_i^2)^{-1}$.

The basic idea behind the proof of the theorem is to observe from (4.2) that $\Pi^{1/2}A\Pi^{-1/2}$ is symmetric, causing A to have real eigenvalues. A modal decomposition of A provides (4.3).

It is interesting to note that covariances of the form (4.3) are precisely those which are achievable as the output covariance of a linear system excited by white noise with the state process of that linear system reversible [1]. Our main purpose here however is to establish the converse of the above theorem.

Theorem 4.2. Consider a covariance of the form (4.3) with $\alpha_i, \beta_i>0$. Then there exists a finite-state Markov process of the form described in Section 2 generating the covariance, and such that the state process is reversible.

The proof of this theorem will be via a constructive procedure, building substantially on theory associated with passive resistor–capacitor networks. The existence of certain connections between such networks and finite-state Markov processes has been observed already by Keilson [6]: equations of the type (2.1) have been noted to parallel the equations for the capacitor voltages in an undriven resistor–capacitor network with a certain topology, viz. one in which all capacitors have one end connected to a ground node.

We recall the following result from the theory of network synthesis [5].

Lemma (Cauer). Let

$$z(s)=\frac{\alpha_n}{s}+\sum_{i=1}^{n-1}\frac{\alpha_i}{s+\beta_i}, \qquad \alpha_i>0, \qquad \beta_i>0,$$

with the β_i distinct. Then there exists a resistor–capacitor network (obtainable from a continued fraction expansion of $z(s)$) with driving point impedance $z(s)$ and with the topology of Figure 1.

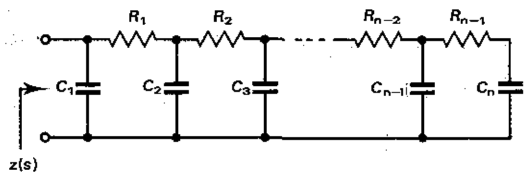

Figure 1
Cauer resistor–capacitor synthesis

Now let $q_i(t)$ be the change on capacitor C_i at time t, let the input current be $u(t)$ and let the response voltage be $v(t)$. Application of Kirchhoff's law leads to

$$\begin{aligned}
\dot{q}_1 &= u - R_1^{-1}(C_1^{-1}q_1 - C_2^{-1}q_2)\\
\dot{q}_2 &= R_1^{-1}(C_1^{-1}q_1 - C_2^{-1}q_2) - R_2^{-1}(C_2^{-1}q_2 - C_3^{-1}q_3)\\
&\;\;\cdot\\
&\;\;\cdot\\
&\;\;\cdot\\
\dot{q}_{n-1} &= R_{n-2}^{-1}(C_{n-2}^{-1}q_{n-2} - C_{n-1}^{-1}q_{n-1}) - R_{n-1}^{-1}(C_{n-1}^{-1}q_{n-1} - C_n^{-1}q_n)\\
\dot{q}_n &= R_{n-1}^{-1}(C_{n-1}^{-1}q_{n-1} - C_n^{-1}q_n)\\
v &= C_1^{-1}q_1
\end{aligned}$$

which we may rearrange, with $G_i = R_i^{-1}$ and $q = [q_1 \quad q_2 \cdots q_n]'$, as

$$\dot{q} = Fq + gu \qquad v = h'q$$

where

$$F = \begin{bmatrix} -G_1 & G_1 & 0 & \cdots & \cdot \\ G_1 & -(G_1+G_2) & G_2 & \cdots & \cdot \\ 0 & G_2 & -(G_2+G_3) & & \\ & & \ddots & & \\ \cdot & \cdot & & & \\ \cdot & \cdot & -(G_{n-1}+G_n) & G_n & \\ \cdot & \cdot & G_n & -G_n & \end{bmatrix} (\operatorname{diag} C_i^{-1}) \quad g = \begin{bmatrix} 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}$$

$h' = C_1^{-1}[1 \quad 0 \cdots 0]$.

It follows (on taking Laplace transforms, or as a standard result of linear system theory) that $z(s) = h'(sI - F)^{-1}g$.

As observed by Keilson [6], one has a parallel between the free (unforced) behaviour of this network (according to $\dot{q} = Fq$) and an n-state process governed by $\dot{p} = Ap$. The parallel for this network however runs deeper: we can, in fact, show correspondence between the network output, and the stochastic process output $c'x$: moreover, if each of the network resistors is viewed as having a white noise source in series (modelling the resistor thermal noise) one can check that the network output has a power spectrum $z(j\omega)+z(-j\omega) = \sum_{i=1}^{n-1} 2\alpha_i\beta_i(\omega^2+\beta_i^2)^{-1}$ to within a scaling constant. This is the same as the spectrum of the output obtained from the finite-state Markov process.

Proof of Theorem 4.2. Let α_n be an arbitrary positive constant. (Further significance than its mere appearance in the expression for $z(s)$ will be given later.) From the quantities α_i, β_i, $1 \leqq c \leqq n-1$ specified in the statement of Theorem 4.1 and from α_n, find the matrix F and vector h according to the procedure described above. We claim that a model for the process y can be defined by

$$A = F \qquad c = \frac{\left(\sum_{i=1}^{n} C_i\right)^{1/2}}{C_1}[1 \quad 0 \cdots \quad 0]'. \tag{4.4}$$

To establish this claim, we must check that A is an infinitesimal stochastic matrix with one-dimensional nullspace and a positive nullvector π, that A has the reversibility property (4.2) and that Cov [y(t), y(s)] is as required by (4.3). The various requirements on A can be checked, in part using the network ideas. First Equation (2.2) is immediately checked. Second from the fact that $z(s) = h'(sI-F)^{-1}g$ and the form of $z(s)$, it follows that $|sI-A| = s\prod_{i=1}^{n-1}(s+\beta_i)$, showing that A has $(n-1)$ non-zero eigenvalues and one zero eigenvalue; thus the nullspace is one-dimensional. Third, with

$$\pi = (\textstyle\sum C_i)^{-1}(\operatorname{diag} C_i)1_n \tag{4.5}$$

one has by direct calculation $A\pi = 0$, $1_n'\pi = 1$ and $\pi > 0$. Last, using (4.5) one can easily check that $A = \Pi A'\Pi^{-1}$.

The correlation of y for $t \geqq s$ is

$$c'e^{A(t-s)}\Pi c = C_1^{-1}e_1'e^{F(t-s)}e_1 = h'e^{F(t-s)}g.$$

Since

$$z(s) = h'(sI-F)^{-1}g = \alpha_n s^{-1} + \sum_{i=1}^{n-1}\alpha_i(s+\beta_i)^{-1},$$

this means that

$$E[y(t)y(s)] = \alpha_n + \sum_{i=1}^{n-1}\alpha_i \exp[-\beta_i |t-s|].$$

It will then follow that

$$\mathrm{Cov}\,[y(t), y(s)] = \sum_{i=1}^{n-1} \alpha_i \exp[-\beta_i\,|t-s|]$$

as required, if we can show that $E[y(t)] = \alpha_n^{1/2}$.

The quantity α_n is most easily obtained from network theoretic calculations as the inverse of the total capacitance across the network terminals when resistors are short-circuited:

$$\alpha_n = (\sum C_i)^{-1}$$

while $E[y(t)] = c'\pi = (\sum C_i)^{-1/2} = \alpha_n^{1/2}$.

We see now the significance of α_n as defining the mean of $y(t)$, and the above procedure as one of specifying a reversible finite-state process with prescribed covariance *and* mean. Of course, if $E[y(t)]$ is negative, one simply changes the sign in (4.4). If $E[y(t)]$ is to be zero, one replaces the definition (4.4) of c by

$$c = \frac{\left(\sum_{i=1}^{n} C_i\right)^{1/2}}{C_1}[1 \quad 0 \cdots 0]' - (\sum C_i)^{-1/2}[1 \quad 1 \cdots 1]'. \tag{4.6}$$

The fact that this yields the same covariance for y is not hard to check, while the fact that $E[y(t)] = 0$ is virtually immediate. Note that this definition draws us somewhat away from the network analogy.

5. Models with dynamically reversible state processes

In the previous section, we have shown that, given a covariance with certain properties, there exists a reversible process modelling this covariance. Now note that Theorem 4.1 shows that the poles of the spectrum must lie along the real axis, and thus periodic or quasiperiodic behaviour is excluded, which is puzzling in view of our knowledge of lossless (conservative) physical systems. Also, any second-order stationary process $y(\cdot)$ has a second-order reversibility property, $\mathrm{Cov}\,[y(t), y(t+T)] = \mathrm{Cov}\,[y(t+T), y(t)]$, which is a trivial consequence of the stationarity and scalar nature of the process, and accordingly, this makes it relevant to seek some form of reversibility in any arrangement generating $y(\cdot)$.

As observed by Whittle [12], one approach to obtaining a form of reversibility for a finite-state Markov process is to introduce a generalized notion of dynamic reversibility, as is frequently done in the rather different models of statistical mechanics [4] and can be done in seeking reversible models of a process produced by exciting a linear system with white gaussian noise [1]. The statistical mechanics thinking runs as follows: Let $x=[x_1' \quad x_2']'$ in which the entries of x_1 can be thought of as analogs of position and the entries of x_2 as analogs of velocity. Then x is dynamically reversible if the statistics of the reversed x process are identical with those of $[x_1' \quad -x_2']'$. (Reversibility in the sense of the previous section then corresponds to x_2 evanescing.)

The modification suggested by Whittle [12] to cope with finite-state Markov processes is as follows.

We suppose that to each state j there corresponds a conjugate state j', with $(j')'=j$; the process is dynamically reversible if its statistics are unchanged by the double operation of time reversal and state conjugation. For finite state processes dynamic reversibility can be characterised in terms of the infinitesimal stochastic matrix of the system. Whittle's result is as follows. (A rough explanation follows the theorem statement.)

Theorem 5.1. Consider a Markov process with $2n$ states $e_i, i=1,\cdots,2n$ with the underlying probability equation

$$\dot{p}=Ap. \tag{5.1}$$

Suppose further that (2.2) through (2.4) hold, with n replaced by $2n$. Then the process is dynamically reversible if and only if

$$\pi=\begin{bmatrix}\pi_1\\ \pi_1\end{bmatrix} \tag{5.2}$$

$$A=\begin{bmatrix}F & Q\\ R & F'\end{bmatrix}\begin{bmatrix}\Pi_1^{-1} & 0\\ 0 & \Pi_1^{-1}\end{bmatrix} \tag{5.3}$$

for some $\pi_1>0, F, Q=Q'$ and $R=R'$. Here, $\Pi_1=\text{diag }\pi_1$.

We remark that the matrix A^b defining the associated backward model is easily seen from (3.3) to be

$$A^b=\begin{bmatrix}F' & R\\ Q & F\end{bmatrix}\begin{bmatrix}\Pi_1^{-1} & 0\\ 0 & \Pi_1^{-1}\end{bmatrix}. \tag{5.4}$$

To get a feel for this result, suppose that all states are equiprobable, i.e. $\Pi=(1/2n)I$. Then the states $e_1,\ldots,e_n$ are conjugate to $e_{n+1},\cdots,e_{2n}$; time reversal sends A to A^b, and permuting the first and second block rows and columns, which corresponds to conjugation, recovers A.

Our aim in this section is to construct a model of a prescribed covariance with an irreducible, dynamically reversible, finite-state model. There is at once a major difficulty: it is unknown whether every rational covariance has a finite-state Markov model, let alone one with dynamic reversibility. However, as shown by Brockett [3], it is possible to approximate any rational covariance with a finite-state Markov model (the better the approximation, the more states may have to be used), and so it is reasonable to ask whether an approximation is possible with a dynamically reversible finite-state Markov model. We answer this question in the following theorem.

Theorem 5.2. Consider an n-state Markov model of the type described in Section 2. Then if all entries of A are non-zero there exists a dynamically reversible $2n$-state Markov model satisfying the obvious modification of (2.1) through (2.3) and generating the same covariance. Such a model is obtainable from an infinitesimal stochastic matrix

$$B=\begin{bmatrix} A-\varepsilon\pi 1_n' & \varepsilon\pi 1_n' \\ \varepsilon\pi 1_n' & \Pi A'\Pi^{-1}-\varepsilon\pi 1_n' \end{bmatrix} \tag{5.5}$$

(where ε is a suitably small positive constant) and B has the property that its nullspace is of rank 1,

$$B\begin{bmatrix} \frac{1}{2}\pi \\ \frac{1}{2}\pi \end{bmatrix}=0 \tag{5.6}$$

and

$$c'e^{At}\Sigma c=[c' \quad c']e^{Bt}\begin{bmatrix} \frac{1}{2}\Pi & 0 \\ 0 & \frac{1}{2}\Pi \end{bmatrix}\begin{bmatrix} c \\ c \end{bmatrix}. \tag{5.7}$$

We remark that with $\varepsilon=0$ above, a reducible model results; the idea of constructing a reducible model this way came in a roundabout way from network theory, and is linked to a construction of a dynamically reversible model of a covariance using a linear system excited by white gaussian noise [1]. The device of introducing the non-zero ε to obtain irreducibility follows Brockett [3].

Proof. Because all entries of A are non-zero, it follows that for sufficiently small ε, all entries in the diagonal blocks of B have the same sign as when $\varepsilon=0$, i.e. $b_{ii}<0$ and $b_{ij}>0$ for $i\neq j$. Also, examination of the off-diagonal blocks of B shows $b_{ij}>0$ for $i\neq j$. Moreover, we easily verify that $\sum_{i=1}^{2n} b_{ij}=0$ for all j, so that B is an infinitesimal stochastic matrix. Next, Equation (5.6) is immediate. Let us check that the nullspace of B is one-dimensional. Assume that $B[\alpha_1' \quad \alpha_2']=0$. Then

$$A\alpha_1+\varepsilon\pi 1_n'(\alpha_2-\alpha_1)=0$$

implying that $A\alpha_1$ is a non-zero multiple of π or $\alpha_1=\alpha_2$. Suppose $A\alpha_1=k\pi$, $k\neq 0$. Then $0=1_n'A\alpha_1=k1_n'\pi=k$, a contradiction. Therefore, $\alpha_1=\alpha_2$, whence it is easily established that $\alpha_1=\alpha_2=k\pi$. This establishes the nullspace property.

Dynamic reversibility, see (5.2) and (5.3), is easily checked.

It remains to check (5.7). Observe that

$$\begin{bmatrix} -\varepsilon\pi 1_n' & \varepsilon\pi 1_n' \\ \varepsilon\pi 1_n' & -\varepsilon\pi 1_n' \end{bmatrix}\begin{bmatrix} A & 0 \\ 0 & \Pi A'\Pi^{-1} \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & \Pi A'\Pi^{-1} \end{bmatrix}\begin{bmatrix} -\varepsilon\pi 1_n' & \varepsilon\pi 1_n' \\ \varepsilon\pi 1_n' & -\varepsilon\pi 1_n' \end{bmatrix} = 0$$

so that, in view of this commutativity,

$$e^{Bt} = \exp\left\{\begin{bmatrix} \varepsilon\pi 1_n' & \varepsilon\pi 1_n' \\ \varepsilon\pi 1_n' & \varepsilon\pi 1_n' \end{bmatrix} t\right\}\begin{bmatrix} e^{At} & 0 \\ 0 & e^{\Pi A'\Pi^{-1}t} \end{bmatrix}.$$

Now with

$$T = \frac{1}{\sqrt{2}}\begin{bmatrix} I & I \\ I & -I \end{bmatrix} = T^{-1}$$

for any X one has

$$T\begin{bmatrix} -X & X \\ X & -X \end{bmatrix}T^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & -2X \end{bmatrix}.$$

Thus

$$\begin{aligned}
&[c' \quad c']\exp\begin{bmatrix} -\varepsilon\pi 1_n' & \varepsilon\pi 1_n' \\ \varepsilon\pi 1_n' & -\varepsilon\pi 1_n' \end{bmatrix} t \\
&= [c' \quad c']\frac{1}{\sqrt{2}}\begin{bmatrix} I & I \\ I & -I \end{bmatrix}\exp\begin{bmatrix} 0 & 0 \\ 0 & -2\varepsilon\pi 1_n' t \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} I & I \\ I & -I \end{bmatrix} \\
&= [\sqrt{2}c' \quad 0]\begin{bmatrix} 1 & 0 \\ 0 & \exp(-2\varepsilon\pi 1_n' t) \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} I & I \\ I & -I \end{bmatrix} \\
&= [c' \quad c'].
\end{aligned}$$

Now (5.7) follows easily:

$$\begin{aligned}
[c' \quad c']e^{Bt}\begin{bmatrix} \tfrac{1}{2}\Pi & 0 \\ 0 & \tfrac{1}{2}\Pi \end{bmatrix}\begin{bmatrix} c \\ c \end{bmatrix} &= [c' \quad c']\begin{bmatrix} e^{At} & 0 \\ 0 & e^{\Pi A'\Pi^{-1}t} \end{bmatrix}\begin{bmatrix} \tfrac{1}{2}\Pi & 0 \\ 0 & \tfrac{1}{2}\Pi \end{bmatrix}\begin{bmatrix} c \\ c \end{bmatrix} \\
&= \tfrac{1}{2}c'e^{At}\Pi c + \tfrac{1}{2}c'e^{\Pi A'\Pi^{-1}t}\Pi c \\
&= \tfrac{1}{2}c'e^{At}\Pi c + \tfrac{1}{2}c'\Pi e^{A't}c \\
&= c'e^{At}\Pi c.
\end{aligned}$$

Equality of covariance follows from equality of the correlations and the easily checked equality of means.

The above result is not as complete as one would like in view of the requirement that all entries of A be non-zero. It is not known to us whether the first part of the theorem holds without this restriction. If we drop the restriction, the best we can do at this point is a result involving approximation.

Theorem 5.3. Consider an n-state Markov model of the type described in Section 2. Then there exists a dynamically reversible $2n$-state model with

$$B_a = \begin{bmatrix} A + \varepsilon\pi 1_n' - 2\varepsilon I & \varepsilon\pi 1_n' \\ \varepsilon\pi 1_n' & \Pi A'\Pi^{-1} + \varepsilon\pi 1_n' - 2\varepsilon I \end{bmatrix} \tag{5.8}$$

where ε is a suitably small positive constant, B_a has nullspace of rank 1,

$$B_a \begin{bmatrix} \frac{1}{2}\pi \\ \frac{1}{2}\pi \end{bmatrix} = 0 \tag{5.9}$$

and

$$[c' \quad c'] e^{B_a t} \begin{bmatrix} \frac{1}{2}\Pi & 0 \\ 0 & \frac{1}{2}\Pi \end{bmatrix} \begin{bmatrix} c \\ c \end{bmatrix} = = e^{-2\varepsilon t} c' e^{At} \Pi c + (1 - e^{-2\varepsilon t}) c' \pi\pi' c. \tag{5.10}$$

The proof of all claims with the possible exception of (5.10) follows as in Theorem 5.2. We outline now the proof of (5.10). Write B_a as

$$B_a = \begin{bmatrix} A & 0 \\ 0 & \Pi A'\Pi^{-1} \end{bmatrix} + \begin{bmatrix} -2\varepsilon I & 0 \\ 0 & -2\varepsilon I \end{bmatrix} + \begin{bmatrix} \varepsilon\pi 1_n' & \varepsilon\pi 1_n' \\ \varepsilon\pi 1_n' & \varepsilon\pi 1_n' \end{bmatrix}$$

and recognize that each pair of summands commutes. Also, proceeding along lines similarly to those used in Theorem 5.2, we have

$$\exp \begin{bmatrix} \varepsilon\pi 1_n' & \varepsilon\pi 1_n' \\ \varepsilon\pi 1_n' & \varepsilon\pi 1_n' \end{bmatrix} = \frac{1}{2} \begin{bmatrix} I & I \\ I & I \end{bmatrix} \begin{bmatrix} \exp 2\varepsilon\pi 1_n' t & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} I & I \\ I & -I \end{bmatrix}$$

and a series expansion yields

$$\exp 2\varepsilon\pi 1_n' t = I - \pi 1_n' (1 - e^{2\varepsilon t}).$$

All this means that

$$e^{B_a t} = \frac{1}{2} \begin{bmatrix} I & I \\ I & I \end{bmatrix} \begin{bmatrix} I - \pi 1_n'(1 - e^{2\varepsilon t}) & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} I & I \\ I & -I \end{bmatrix} \begin{bmatrix} e^{At} & 0 \\ 0 & e^{\Pi A'\Pi^{-1} t} \end{bmatrix} e^{-2\varepsilon t}.$$

Equation (5.10) is more obtainable, using the facts that $1_n' e^{At} = 1_n'$ and $1_n' e^{\Pi A'\Pi^{-1}} = 1_n'$, which may be verified by a series expansion of the matrix exponential.

By taking ε arbitrarily small, we can evidently make the approximation as good as desired. We are prevented from taking $\varepsilon = 0$ by the fact that the resulting Markov state process would become reducible.

The discrete-time versions of the results of this section are easily obtainable. In fact, with A_d a stochastic matrix governing the discrete-time equation $p(k+1) = A_d p(k)$ for the evolution of probabilities we again form (5.5) for suitably small ε with A_d replacing A. (Of course, we have $A_d \pi = \pi$ now, rather than $A_d\pi = 0$). Then if all entries of A_d are positive, B_d will be a stochastic matrix.

We remark that if A_d is derived by sampling of $\dot{p} = Ap$ with A an infinitesimal stochastic matrix, i.e. if

$$A_d = e^{A\Delta}$$

for some fixed Δ, the sampling interval, then the conditions on A listed in Section 2 guarantee that all entries of A_d are positive. For this reason, the conditions of Theorem 5.2 appear less restrictive in discrete time.

6. Conclusions

The paper has been concerned with making several points. First, given a forward finite-state process model, it is easy to obtain a corresponding backwards model, with the construction showing a close parallel with a construction applicable to processes generated by passing white noise into a linear system. Second, we have shown the correspondence between models with a reversible state process and a power spectrum with all real poles, again observing a parallel between this result and one applying when white noise is passed into a linear system. Third and finally, we have described how, almost always, an arbitrary finite-state Markov model can be expanded to be dynamically reversible, using a notion of dynamic reversibility that stems from statistical mechanics ideas and that has also had application in studying processes produced by passing white noise into a linear system.

References

[1] ANDERSON, B. D. O. AND KAILATH, T. (1978) Forward, backward and dynamically reversible models for second order processes. In *Proc. IEEE Internat. Symp. Circuits and Systems*, IEEE, New York, 981–986.

[2] BROCKETT, R. W. (1970) *Finite Dimensional Linear Systems*. Wiley, New York.

[3] BROCKETT, R. W. (1977) Stationary covariance generation with finite state Markov processes. In *Proc.* 1977 *Joint Auto. Control Conf., San Francisco*, 1057–1060.

[4] DE GROOT, S. R. (1966) *Thermodynamics of Irreversible Processes*. North-Holland, Amsterdam.

[5] GUILLEMIN, E. A. (1957) *Synthesis of Passive Networks*. Wiley, New York.

[6] KEILSON, J. (1965) A review of transient behaviour in regular diffusion and birth–death processes, II. *J. Appl. Prob.* **2**, 405–428.

[7] KEILSON, J. (1971) On the structure of covariance functions and spectral density functions for processes reversible in time. Report CSS-71-03, University of Rochester.

[8] KOLMOGOROV, A. N. (1936) Zur Theorie der Markoffschen Ketten. *Math. Ann.* **112,** 155–160.

[9] LAINIOTIS, D. G. (1976) General backwards Markov models. *IEEE Trans. Auto. Control* AC-**21,** 595–598.

[10] LJUNG, L. AND KAILATH, T. (1976) Backwards Markovian models for second-order stochastic processes. *IEEE Trans. Inf. Theory* IT-**22,** 488–491.

[11] SIDHU, G. S. AND DESAI, U. B. (1976) New smoothing algorithms based on reversed-time lumped models. *IEEE Trans. Auto Control* AC-**21,** 538-541.

[12] WHITTLE, P. (1975) Reversibility and acyclicity. In *Perspectives in Probability and Statistics,* ed. J. Gani, distributed by Academic Press, London, for the Applied Probability Trust, Sheffield, 217–224.