

# An innovations approach to the discrete-time stochastic realization problem

M. Gevers (\*) and W. R. E. Wouters (\*\*)

## SUMMARY

In a companion paper we have used the innovations concept to derive linear least squares estimation formulas for discrete-time random processes that are related to an observation process. We have shown that covariance information is all that is needed to derive a predictor for an observation process [1]. In this paper we use these results to compute finite state-space or auto-regressive moving average (ARMA) models for finite-dimensional stationary discrete time processes whose covariance is given. The reason for computing such models is that the predictor formulas become finitely recursive.

## 1. INTRODUCTION

In a companion paper [1] we have discussed the discrete-time linear least-squares problem. Basic to our treatment of the problem was the concept of innovations and the General Innovations Filter (GIF). We showed that with a given finite variance stochastic observation process  $\{y\}$  one can associate a related white noise process  $\{\epsilon\}$ , the innovations process, that can be obtained from  $\{y\}$  by a causal and causally invertible transformation. All that is needed to make this transformation is the knowledge of the covariance function of the process  $\{y\}$ .

Given the innovations process, all kinds of least-squares estimates are easily obtainable, e.g. a predictor for  $\{y\}$  can be built. The problem with the GIF is that, although the filter can be computed recursively from the covariance function, the computations are in general not finitely recursive, so a growing memory is required. However, we showed that, when a finite dimensional model for the process of interest is available, considerable simplifications of the GIF formulas can be obtained, and finitely recursive filters can be designed. Two kinds of models were considered. We looked at so called state-space models where the process  $\{y\}$  is given as a linear projection of a related Markov process  $\{x\}$  corrupted by additive white noise. Also considered were the so called mixed auto-regressive moving-average or ARMA models, where the process  $\{y\}$  is generated from a finite linear combination of past  $y_t$  values (AR) plus a finite linear combination of some past and present

values of a driving white noise process (MA). As we already said in the introduction of [1], because of the tremendous success of the Kalman filter formulation, which uses a state space model, the idea permeated most engineering circles that a lumped (i.e. finite dimensional) model of the observation process was necessary to derive finitely recursive estimation formulas. Obviously, elementary estimation theory tells us that covariance information is all that is needed to solve the estimation problem. The question arising then is whether it is possible to derive finitely recursive estimation formulas from covariance information. The answer is that, yes, for a given class of covariance functions this is indeed possible. If we have a process  $\{y\}$  which can be represented by a finite dimensional white noise driven model, it is indeed possible to infer such a model from the covariance function of  $\{y\}$ . Of course, once we have a finite dimensional model we can, by the methods presented in [1], obtain finitely recursive estimators. The problem then we are going to tackle in this paper is how to obtain a finite dimensional model for a given process  $\{y\}$ , given its covariance function, and assuming of course that such a finite dimensional model does indeed exist. This is called the stochastic realization problem. The concepts of innovations and of innovations representation (IR) will be crucial to the development. In this paper, as in its companion [1], we shall try to present the subject in a tutorial way. A necessary condition for this is that the presentation be (as much as possible) self-contained. Because of this, and of the technical and conceptual dif-

(\*) Laboratoire d'Automatique et d'Analyse des Systèmes, Université Catholique de Louvain, Bâtiment Maxwell, 1348 Louvain-la-Neuve (Belgium)

(\*\*) Department of Mathematics and Systems Engineering, Koninklijke / Shell Laboratorium, Amsterdam (Shell Research B. V.)



ficuity of the subject, we shall not treat the stochastic realization problem in all its generality. First of all we shall limit ourselves to (wide sense) *stationary purely non-deterministic* processes  $\{y_t\}$ . This does simplify the development considerably, and this class of processes is, from a practical point of view, certainly the most interesting. Indeed, a lot of stochastic processes either are (nearly) stationary or are by simple transformations related to stationary processes. Secondly, it turns out that the treatment of the realization problem for a vector observation process requires some background knowledge on the structural properties of multivariable systems which are beyond the scope of this paper. We shall, therefore, restrict ourselves to the discussion of the realization problem for scalar processes.

Again, two types of finite dimensional models will be considered; state-space and ARMA models. In solving the realization problem we are trying to infer structural and parametric properties of lumped models from properties of the covariance function. In order to proceed efficiently, it is useful to study the properties of the covariance function induced by a known finite dimensional model. This is done in sections 2 and 3, first for state-space models, then for ARMA models. In sections 4 and 5 we give the solution of the realization problem, and show how we can obtain both state-space and ARMA models from the process' covariance function. In section 6 we give some additional properties of finite dimensional processes, and in a final section we discuss the results obtained and give some comments about the realization theory for vector processes.

In order to discuss the realization problem we have to touch upon a lot of results related to stochastic processes. Wherever possible we give a proof or a justification for the properties presented. This should increase the value of this paper together with [1] as a study object for someone who wants an introduction to some of the finer points in the theory of stochastic processes.

Most of the results presented are, of course, not new but we try to present the material in a unifying and (hopefully) tutorial way. The stochastic realization problem has been extensively discussed and studied in the last ten years (see e.g. [2] - [10]). The subject is by now well understood, but research is still going on with the objective of designing faster and numerically stable stochastic realization (or spectral factorization) algorithms. The list of references given at the end is by no means exhaustive, but should nevertheless provide a feasible starting point for someone wishing to explore some of the material in more detail.

## 2. STATE-SPACE MODELS

In this section we shall examine wide-sense stationary processes  $\{y_t\}$  that can be described by state-space models. We shall explain what is meant by a finite-

dimensional process and actually define the dimension of a process. We shall also establish the relations between the covariance function of the process  $\{y_t\}$  and the parameters of the state-space model. We shall assume that the  $p$ -vector process  $\{y_t\}$  can be expressed as a linear combination of a related  $n$ -vector process  $\{x_t\}$ , plus additive white noise, i.e.

$$y_t = H x_t + v_t \quad (2.1a)$$

where  $H$  is a constant  $p \times n$  matrix and  $v_t$  is a  $p$ -vector white noise with constant covariance matrix  $M$ .  $M$  is, of course, a symmetric non-negative definite  $p \times p$  matrix.

The process  $\{x_t\}$  introduced in (2.1a) is assumed to be a first order  $n$ -vector Markov process, i.e.

$$x_{t+1} = F x_t + G u_t \quad (2.1b)$$

where  $F$  and  $G$  are constant  $n \times n$  and  $n \times m$  matrices respectively,  $u_t$  is a  $m$ -vector white noise with constant covariance matrix  $\Sigma$ ,  $\Sigma$  being a  $m \times m$  full rank covariance matrix. We shall furthermore assume that the processes  $\{v_t\}$  and  $\{u_t\}$  are mutually correlated, with  $m \times p$  correlation matrix  $S$ , i.e.

$$E \left\{ \begin{bmatrix} u_t \\ v_t \end{bmatrix} \begin{bmatrix} u_\tau^T & v_\tau^T \end{bmatrix} \right\} = \begin{bmatrix} \Sigma & S \\ S^T & M \end{bmatrix} \delta_{t,\tau} \quad (2.2)$$

The  $p$ -vector process  $\{y_t\}$  is thus a linear projection of a  $n$ -vector Markov process,  $\{x_t\}$ .

We shall also assume that  $\{y_t\}$  is a (wide-sense) stationary purely non-deterministic process. This together with an observability condition on  $(H, F)$  implies that the process  $\{x_t\}$  has to be (wide-sense) stationary and purely non-deterministic, and thus for all  $t$ , we have

$$E \{x_t x_t^T\} = \Pi \quad (2.3)$$

where  $\Pi$  is a nonnegative constant symmetric matrix. Stationarity implies that  $\Pi$  is the solution of the following equation

$$\Pi - F \Pi F^T = G \Sigma G^T \quad (2.4)$$

An equation of this type is sometimes called a Lyapunov equation, and a necessary and sufficient condition for the existence of a nonnegative solution is that the matrix  $F$  must have all its eigenvalues strictly inside the unit disc centered at the origin of the complex plane (see e.g. [27]). The matrix  $F$  will then be called an asymptotically stable matrix. Since in this paper we are only interested in purely non-deterministic stationary processes, we shall only consider models of the form (2.1-2.2) where  $F$  is an asymptotically stable matrix. The stochastic process  $\{y_t\}$  as generated by (2.1-2.2) will also be called a *finite dimensional process*. A precise definition will be given a little later.

Next we shall derive an expression for the output covariance function. Because of the stationarity as-

sumption, the covariance function will be a function of the time shift  $(t-\tau)$  only. We shall then adapt our notation accordingly, and define the covariance function as

$$R_y(k) \triangleq E \{ y_{t+k} y_t^T \} \quad (2.5)$$

Because of the assumption that  $\{y_t\}$  is purely non-deterministic, it follows that  $\lim_{k \rightarrow \infty} \|R_y(k)\| = 0$ ,

where  $\|\cdot\|$  is any matrix norm.

The covariance function  $R_y(k)$  can be expressed in terms of the parameters of the state space (Markov) model. An expression is derived as follows: consider first  $k > 0$ . From (2.1) we can write

$$y_{t+k} = H F^k x_t + \sum_{i=0}^{k-1} (H F^i G u_{t+k-(i+1)} + v_{t+k}) \quad (2.6)$$

Post multiplying (2.6) by the transpose of the expression (2.1a) for  $y_t$ , taking expectations and using the whiteness properties of  $\{u_t\}$  and  $\{v_t\}$ , one obtains

$$R_y(k) = H F^k \Pi H^T + H F^{k-1} G S, \quad k > 0 \quad (2.7)$$

For reasons that will become clear further down it will be convenient to define a  $n \times p$  matrix  $L$

$$L \triangleq H \Pi H^T + G S \quad (2.8)$$

so the function  $R_y(k)$  for  $k > 0$  can be written as

$$R_y(k) = H F^{k-1} L \quad (2.9)$$

For  $k < 0$ , we note that  $R_y(k) = E \{ y_{t-|k|} y_t^T \}$ , and because of the stationarity of the process this can be written as

$$R_y(k) = E \{ y_t y_{t+|k|}^T \} = R_y^T(|k|), \quad k < 0 \quad (2.10)$$

Thus it immediately follows that

$$R_y(k) = L^T (F^T)^{|k|-1} H^T, \quad k < 0 \quad (2.11)$$

Finally, for  $k = 0$ , we obtain from (2.1a) and (2.2)

$$R_y(0) = H \Pi H^T + M \quad (2.12)$$

So far in this section, we have developed expressions for the autocorrelation function of a  $p$ -vector process  $\{y_t\}$  given the model parameters  $(F, G, H)$  and the noise covariances  $(\Sigma, S, M)$ . There is one more result we want to formulate now concerning the structure of the autocovariance function.

We define the Hankel matrix of  $R_y$ , as follows

$$\mathcal{H}_{N_1, N_2}(R_y) = \begin{bmatrix} R_y(N_1) & R_y(N_1+1) & \dots & R_y(N_2) \\ R_y(N_1+1) & R_y(N_1+2) & \dots & R_y(N_2+1) \\ \vdots & \vdots & \ddots & \vdots \\ R_y(N_2) & R_y(N_2+1) & \dots & R_y(N_1+N_2-1) \end{bmatrix}$$

$$\text{with } 0 < N_1 \leq N_2 \quad (2.13)$$

Clearly this is a  $p(N_2 - N_1 + 1) \times p(N_2 - N_1 + 1)$  square matrix. Of special interest to us are the matrices  $\mathcal{H}_{1, N}(R_y)$  and  $\mathcal{H}_{1, \infty}(R_y)$ . The first of these two will be called the *finite Hankel matrix of order  $N$* , and the latter the *infinite Hankel matrix*. Since it is clear that we are considering Hankel matrices of  $R_y$ , we shall simplify the notation and use  $\mathcal{H}_{1, N}$  and  $\mathcal{H}_{1, \infty}$ .

We have said earlier that a process  $\{y_t\}$ , generated through a state-space model such as (2.1) is called finite dimensional. We are now in a position to be more specific, and define the dimension of  $\{y_t\}$ .

#### Definition 2.1

The stationary stochastic process  $\{y_t\}$  is a process of dimension  $n'$  if and only if the rank of the infinite Hankel matrix  $\mathcal{H}_{1, \infty}$  is  $n'$ , i.e.

$$\dim(\{y_t\}) = n' \iff \rho(\mathcal{H}_{1, \infty}) = n' \quad \square$$

This is a rather abstract definition, so we shall try to give it somewhat more body by considering further properties of the Hankel matrix. Always assuming that  $\{y_t\}$  is generated by a state-space model (2.1), it follows from the definition of  $\mathcal{H}_{1, N}$  and from the expressions developed for  $R_y(k)$  (equation 2.9) that  $\mathcal{H}_{1, N}$  can be factored into two matrices  $O_N^*$  and  $C_N^*$  as follows

$$\mathcal{H}_{1, N} = O_N^* C_N^* \quad (2.14)$$

where

$$O_N^* \triangleq \begin{bmatrix} H \\ HF \\ HF^2 \\ \vdots \\ HF^{N-1} \end{bmatrix} \quad (2.15)$$

and

$$C_N^* = [L, FL, F^2L, \dots, F^{N-1}L] \quad (2.16)$$

The matrix  $O_N^*$  is an observability matrix, and  $C_N^*$  is a controllability matrix. The infinite Hankel matrix  $\mathcal{H}_{1, \infty}$  can be factored in the same way as  $\mathcal{H}_{1, N}$  into the product of an infinite observability matrix  $O_\infty^*$ , and an infinite controllability matrix  $C^*$  i.e.

$$\mathcal{H}_{1, \infty} = O_\infty^* C^* \quad (2.17)$$

One property is immediately apparent. Recall that the rank of the product of two matrices is bounded by the minimum of the individual ranks of the factors, i.e. here this becomes

$$\rho(\mathcal{H}_{1, \infty}) \leq \min(\rho(O_\infty^*), \rho(C^*)) \quad (2.18)$$

Since  $O_\infty^*$  is a  $(\infty \times n)$  matrix, and  $C^*$  a  $(n \times \infty)$  matrix we have  $\rho(O_\infty^*) \leq n$  and  $\rho(C^*) \leq n$  so it follows that

**Property 2.1.**

The dimension of the stochastic process  $\{y_t\}$  as generated by (2.1) is not greater than the dimension of the state vector,  $x_t$ , i.e.

$$\dim(\{y_t\}) \leq n \quad \square$$

Another property of the Hankel matrix will be formulated now.

**Property 2.2**

If the process  $\{y_t\}$  is of finite dimension  $n'$ , then the Hankel matrix of order  $n'$  has rank  $n'$ , i.e.

$$\rho(\mathcal{H}_{1, \infty}) = n' \Rightarrow \rho(\mathcal{H}_{1, n'}) = n' \quad \square$$

The proof of Property 2.2 will be given in a number of steps. We shall first introduce some notational definitions. Let

$$\mathcal{H}_{1, \infty} = \begin{bmatrix} R_1^\infty \\ R_2^\infty \\ \vdots \end{bmatrix}$$

where  $R_i^\infty \in \mathbb{R}^p \times \infty$  are block rows of  $\mathcal{H}_{1, \infty}$ . If we define the operator  $\sigma_R : \mathbb{R}^p \times \infty \rightarrow \mathbb{R}^p \times \infty$  as the operator which drops the first  $p \times p$  block from an element of  $\mathbb{R}^p \times \infty$  and then shifts the remainder  $p$  positions to the left, it follows from the Hankel structure that

$$R_{i+j}^\infty = \sigma_R^j R_i^\infty \quad j \geq 0, i > 0$$

$$\sigma_R^j C R_i^\infty = C \sigma_R^j R_i^\infty \quad j \geq 0, i > 0$$

where  $C$  is an arbitrary matrix in  $\mathbb{R}^{p \times p}$ .

Also we shall define the numbers  $\rho_k$  as

$$\rho_k = \rho \begin{bmatrix} R_1^\infty \\ \vdots \\ R_k^\infty \end{bmatrix}$$

or  $\rho_k$  is the rank of the  $pk \times \infty$  matrix formed by taking the first  $k$   $R_i^\infty$  elements of  $\mathcal{H}_{1, \infty}$ .

The first result we obtain is

**Lemma 2.1.**

If  $\rho_{k+1} = \rho_k$ , then  $\rho_{k+j} = \rho_k$  for all  $j \geq 0$  and furthermore  $\rho_k = n' = \rho(\mathcal{H}_{1, \infty})$   $\square$

*Proof*

Let  $\rho_{k+1} = \rho_k$ , i.e. there exist matrices  $C_i \in \mathbb{R}^{p \times p}$  such that

$$R_{k+1}^\infty = \sum_{i=1}^k C_i R_i^\infty$$

From the property of the  $\sigma_R$  operator defined above it follows that

$$\sigma_R R_{k+1}^\infty = \sum_{i=1}^k C_i \sigma_R R_i^\infty$$

or

$$\begin{aligned} R_{k+2}^\infty &= C_k R_{k+1}^\infty + \sum_{i=1}^{k-1} C_i R_{i+1}^\infty \\ &= C_k \sum_{i=1}^k C_i R_i^\infty + \sum_{i=1}^{k-1} C_i R_{i+1}^\infty \end{aligned}$$

so  $R_{k+2}^\infty$  is a linear combination of the same elements as  $R_{k+1}^\infty$ , or  $\rho_{k+2} = \rho_k$ . Obviously, we can find expressions for  $R_{k+j}^\infty$  ( $j > 0$ ) as linear combinations of

$$\{R_i^\infty\}_{i=1}^k; \text{ thus } \rho_{k+j} = \rho_k \text{ for all } j \geq 0.$$

$$\text{Now } \rho_k = \lim_{j \rightarrow \infty} \rho_{k+j} = \rho(\mathcal{H}_{1, \infty}) = n' \quad \square \square$$

The next thing we shall show is that the first  $n$  block rows  $R_i^\infty$  are sufficient to span the same space as  $\mathcal{H}_{1, \infty}$ , or in other words

**Lemma 2.2.**

If  $\rho(\mathcal{H}_{1, \infty}) = n'$ , then

$$\rho_{n'} = n' \quad \square$$

*Proof*

We know that there is a finite integer  $k$  such that

$$\rho_1 < \rho_2 < \rho_3 \dots < \rho_{k-1} < \rho_k = \rho_{k+1} = \rho_{k+2} = \dots = n'$$

Because of the strict inequalities for the  $\rho_i$ ,  $i < k$ , we have that

$$n' \geq \rho_i \geq i, \text{ thus } n' \geq \rho_{n'} = n' \text{ or } \rho_{n'} = n' \quad \square \square$$

Rather than considering the block rows  $R_i^\infty \in \mathbb{R}^{p \times \infty}$  of the  $\mathcal{H}_{1, \infty}$  matrix, we can also consider the block columns  $C_i^\infty \in \mathbb{R}^{\infty \times p}$  of  $\mathcal{H}_{1, \infty}$ , i.e.

$$\mathcal{H}_{1, \infty} = [C_1^\infty \ C_2^\infty \ \dots]$$

In a way completely parallel to the way to prove that

$$\rho_{n'} = \rho \begin{bmatrix} R_1^\infty \\ \vdots \\ R_{n'}^\infty \end{bmatrix} = \rho(\mathcal{H}_{1, \infty}) \text{ (Lemma 2), we can prove the following Lemma.}$$

Lemma 2.3.

If  $\rho(\mathcal{H}_{1,\infty}) = n'$ , then

$$\rho[C_1^\infty, C_2^\infty, \dots, C_{n'}^\infty] = n' \quad \square$$

Proof

Left to the amusement of the reader.  $\square \square$

Let us call  $\xi_{n'}^\infty \triangleq [C_1^\infty, \dots, C_{n'}^\infty]$  and

$\mathcal{R}_{n'}^\infty \triangleq [R_1^{\infty T}, \dots, R_{n'}^{\infty T}]^T$ . We know

$$\rho(\xi_{n'}^\infty) = \rho(\mathcal{R}_{n'}^\infty) = n'$$

Consider the following diagram

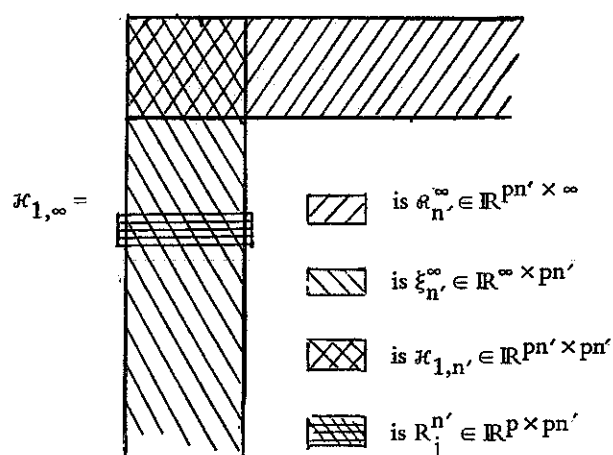


Diagram 2.1.

Also consider

$$R_j^{n'} \triangleq [R_y(j) R_y(j+1) \dots R_y(j+n')] \in \mathbb{R}^{p \times pn'},$$

the  $j$ th block row of the  $\xi_{n'}^\infty$  matrix.

The elements  $R_j^{n'}$  for  $j \in [1, \dots, n']$  are of course the block rows of the  $\mathcal{H}_{1, n'}$  matrix. The proof of Property 2.2 can now be given

Proof of Property 2.2.

It follows from Lemma 2.1, from the fact that the  $R_j^{n'}$  are the truncated  $R_j^\infty$  and from the fact that  $\rho(\mathcal{R}_{n'}^\infty) = n'$ , that there exist coefficients  $C_i \in \mathbb{R}^{p \times p}$  such that

$$R_j^{n'} = \sum_{i=1}^{n'} C_i R_i^{n'} \quad \text{for all } j$$

However, since the  $R_i^{n'}$  are also the block rows of the  $\xi_{n'}^\infty$  matrix, it follows that we can express any  $j$ th block row as a linear combination of the first  $n'$  block rows.

Since  $\rho(\xi_{n'}^\infty) = n'$  we must also have that

$$\rho \begin{bmatrix} R_1^{n'} \\ \vdots \\ R_{n'}^{n'} \end{bmatrix} = n'; \text{ but } \begin{bmatrix} R_1^{n'} \\ \vdots \\ R_{n'}^{n'} \end{bmatrix} = \mathcal{H}_{1, n'} \text{ so } \rho(\mathcal{H}_{1, n'}) = n' \quad \square \square$$

Comment on Property 2.2.

The importance of Property 2.2 is that it holds for arbitrary Hankel matrices. Indeed, in the proof we have not used the fact that the  $\mathcal{H}_{1, \infty}$  matrix is generated through some state space model. Property 2.2 is a very powerful result, and plays an important role in realization theory.

This will be expanded upon in section 4.

One of the questions one might ask is under what conditions will the dimension of the process  $\{y_t\}$  be equal to the dimension  $n$  of the state vector  $x_t$  in the state space model (2.1).

We prove the following property:

Property 2.3.

Let the process  $\{y_t\}$  be generated by the state-space model (2.1-2.2) where  $F$  is an  $n \times n$  matrix. Then  $\dim(\{y_t\}) = n$  if and only if the pair  $(F, H)$  is completely observable and the pair  $(F, L)$  is completely controllable (\*).

Proof

Necessity: Let  $\dim(\{y_t\}) = n$ . Then by Property 2.2  $\rho(\mathcal{H}_{1, n}) = n$ , and since

$$\rho(\mathcal{H}_{1, n}) \leq \min(\rho(O_n^*), \rho(C_n^*))$$

it follows that  $O_n^*$  and  $C_n^*$  must have rank  $n$ .

Sufficiency: If  $O_n^*$  has rank  $n$ , then so does  $O_\infty^*$ , because  $O_n^*$  is a submatrix of  $O_\infty^*$  and they both have  $n$  columns; the same argument holds for  $C_n^*$  and  $C_\infty^*$ . Therefore by the following Sylvester inequality (see [11] p. 66).

$$\rho(O_\infty^*) + \rho(C_\infty^*) - n \leq \rho(O_\infty^* \cdot C_\infty^*) = \rho(\mathcal{H}_{1, \infty}), \quad (2.19)$$

we have  $\rho(\mathcal{H}_{1, \infty}) = n = \dim(\{y_t\})$ .  $\square \square$

Notice that the complete controllability condition depends upon the matrix  $L = F \Pi H^T + S G$ , which depends upon the model's noise covariances  $S$  and  $\Sigma$  (through 2.4) as well as upon the model's dynamic parameters  $F, G, H$ . The controllability of  $(F, L)$  is *not implied* by a controllability condition on  $(F, G)$  as one might naively think. It is easy to show that

(\*) Recall that  $(F, H)$  completely observable is equivalent with  $\rho(O_n^*) = n$  and  $(F, L)$  completely controllable is equivalent with  $\rho(C_n^*) = n$ .

a model of the form (2.1-2.2) where  $(F, H)$  is observable, and  $(F, G)$  is controllable, and where we choose  $S = 0$  and  $F$  singular, will generate a process  $\{y_t\}$  of dimension smaller than  $n$ . Indeed,  $S = 0$  implies that  $L = F \Pi H^T$ , and therefore

$$C_n^* = F [\Pi H^T, L, FL, \dots, F^{n-2}L]$$

Hence  $\rho(C_n^*) \leq \rho(F) < n$  by the assumption that  $F$  is singular.

A model like (2.1-2.2) which is characterized by the elements  $\{F, G, H, \Sigma, S, M\}$  is called a *realization* of the process  $\{y_t\}$ . When  $F \in \mathbb{R}^{n \times n}$  and the dimension of  $\{y_t\}$  is also  $n$ , we say that  $\{F, H, G, \Sigma, S, M\}$  is a *minimal realization* of the process  $\{y_t\}$ .

At this point we want to point out that several state-space models may generate the same process  $\{y_t\}$ , or more precisely processes that have the same covariance function  $R_y(k)$ . These models will be called *covariance equivalent realizations*. A trivial way of generating covariance equivalent realizations is as follows. Let  $\{F, H, G, \Sigma, S, M\}$  be a realization of  $\{y_t\}$ ; then, if  $T \in \mathbb{R}^{n \times n}$  is any non-singular matrix, it follows that  $\{T F T^{-1}, H T^{-1}, T G, \Sigma, S, M\}$  is a realization of  $\{y_t\}$ . This type of transformation does not involve the driving noise covariances  $\Sigma, S, M$ , but involves a coordinate transformation in the state-space and indeed, leads to identical input/output maps. It turns out that there are other equivalent realizations which do not lead to identical input/output maps, but still yield the same covariance function  $R_y(k)$ . For a given  $(F, H)$  which fixes the coordinate system for the state space, any combination of elements  $(G, \Sigma, S, M)$  which yields a given  $L$  (through (2.8)) and a given  $R_y(0)$  (through (2.12)) while satisfying (2.4) will give the same covariance function.

Let  $\{F, H, G, \Sigma, S, M\}$  be a *minimal realization* for  $\{y_t\}$ .

A minimal realization which plays an important role is the *innovations representation* for a stochastic process  $\{y_t\}$ . This realization is defined as follows :

$$s_{t+1} = F s_t + K_p e_t \quad (2.20a)$$

$$y_t = H s_t + e_t \quad (2.20b)$$

where  $e_t$  is a  $p$ -vector white noise (the innovations) with covariance

$$Q = H P H^T + M \quad (2.21)$$

The  $n \times p$  matrix  $K_p$  and the  $n \times n$  matrix  $P \geq 0$  are given by the solution of the following set of coupled algebraic matrix equations :

$$K_p = [F P H^T + G S] [H P H^T + M]^{-1} \quad (2.22a)$$

$$P = F P F^T - K_p [H P H^T + M] K_p^T + G \Sigma G^T \quad (2.22b)$$

These equations link the elements  $\{K_p, Q\}$  of the innovations representation to the elements  $\{G, \Sigma, S, M\}$  of the original model.

In order to show that the model (2.20-2.22) does indeed realize the same process as the model (2.1-2.2), we shall compute the covariance function of  $\{y_t\}$  as generated by the innovations model. It follows by straightforward calculation that we can write

$$R_y(0) = H \hat{\Pi} H^T + Q \quad (2.23a)$$

$$R_y(k) = H F^{k-1} \hat{L} \quad k > 0 \quad (2.23b)$$

where

$$\hat{L} = F \hat{\Pi} H + K_p Q \quad (2.24)$$

and  $\hat{\Pi}$  is the (nonnegative) solution of the Lyapunov equation

$$\hat{\Pi} = F \hat{\Pi} F^T + K_p Q K_p^T \quad (2.25)$$

This follows by exactly the same calculations as the derivation of (2.9). We now show that  $\hat{L} = L$ , and that  $H \hat{\Pi} H^T + Q = H \Pi H^T + M$  (cfr. eq. (2.12)). Let us first establish a relationship between the covariance matrix of the state of the original realization,  $\Pi$ , the covariance matrix of the state of the innovations representation,  $\hat{\Pi} \triangleq E \{s_t s_t^T\}$ , and the matrix  $P$  as defined by equation (2.22b).

*Lemma 2.3.*

$$\Pi = \hat{\Pi} + P \quad (2.26) \square$$

*Proof*

Recall that  $\hat{\Pi} - F \hat{\Pi} F^T = K_p Q K_p^T$  and that

$$P - F P F^T = G \Sigma G^T - K_p Q K_p^T$$

Thus :

$$(\hat{\Pi} + P) - F (\hat{\Pi} + P) F^T = G \Sigma G^T$$

and it follows that  $\Pi = \hat{\Pi} + P$  (see eq. (2.4)).  $\square$

Using now the relationship  $\hat{\Pi} = \Pi - P$  in equations (2.23a-b), we find

$$\begin{aligned} R_y(0) &= H \hat{\Pi} H^T + Q \\ &= H \Pi H^T - H P H^T + Q \\ &= H \Pi H^T + M \end{aligned}$$

and

$$\begin{aligned} R_y(k) &= H F^{k-1} \hat{L} \\ &= H F^{k-1} (F (\Pi - P) H^T + F P H^T + G S) \\ &= H F^{k-1} (F \Pi H^T + G S) \\ &= H F^{k-1} L \end{aligned}$$

where we made use of (2.21)-(2.22). We thus have

established that (2.20)-(2.22) is indeed a covariance equivalent realization of the process  $\{y_t\}$ . This realization will be denoted by  $\{F, H, K_p, Q\}$ .

We shall now state the following properties about  $\{F, H, K_p, Q\}$ .

*Property 2.4.*

Given a minimal state-space realization  $\{F, H, G, \Sigma, S, M\}$  for a  $p$ -vector process  $\{y_t\}$  there always exists a covariance equivalent minimal innovations representation  $\{F, H, K_p, Q\}$  such that

(i) the state noise process  $u_t$  and the measurement noise process  $v_t$  are both replaced by the same  $p$ -vector white noise  $e_t$

(ii) the covariance of the state of the innovations model is always smaller (\*) than the covariance of the state of the original realization

(iii) the innovations representation is invertible, i.e. the filter

$$\sigma_{t+1} = (F - K_p H) \sigma_t + K_p y_t \quad (2.27a)$$

$$\epsilon_t = y_t - H \sigma_t \quad (2.27b)$$

is stable. □

*Proof*

(i) and (ii) follow immediately from the development preceding the formulation of *property 2.4*. Point (iii) is proven in Kalman [12]. □□

*Comment*

Property 2.4 (iii) shows that the innovations representation (2.20) and the filter (2.27) are inverses of one another, i.e. the IR (2.20) is a recursive algorithm that computes the  $\{y_t\}$  sequence from the white noise sequence  $\{e_t\}$ , while the filter (2.27) computes the  $\{\epsilon_t\}$  sequence from the  $\{y_t\}$  sequence. By comparing the present results with section (3.2) of [1], it can be verified that (2.27), together with (2.21)-(2.22), are the equations for the predicted estimate  $\hat{x}_{t+1/t}$  of the state  $x_{t+1}$  of the system (2.1), i.e.  $\sigma_{t+1} = \hat{x}_{t+1/t}$  (\*\*)

We have stated the equations (2.20)-(2.22) of the IR and shown that this IR is covariance equivalent with the given model (2.1)-(2.2). The normal way to derive the equations (2.20)-(2.22) is to compute the Kalman predictor for the system (2.1)-(2.2), and then to invert the predictor, i.e. to rewrite the predictor equations in the form (2.20).

(\*) If  $A$  and  $B$  are square matrices of the same dimension, we say that  $B$  is smaller than  $A$  if  $(A-B)$  is a non-negative definite matrix.

(\*\*) Comparing the expression (2.22a) of the predictor gain  $K_p$  with (3.11) and (3.15c) in [1] shows that we have an additional term + GS here. This difference is due to the fact that we have assumed a correlation between  $\{u_t\}$  and  $\{v_t\}$  here (see (2.2)), while in [1] we assume

$$E \begin{bmatrix} u_t \\ v_t \end{bmatrix} = 0.$$

### 3. ARMA MODELS

In the previous section we discussed stochastic processes that were modeled as linear projections of Markov processes. Properties of such models were discussed and expressions for the covariance function were derived.

In this section we will be interested in another kind of finitely recursive models for the process  $\{y_t\}$ , the so-called mixed autoregressive moving-average (ARMA) models. ARMA models for stochastic processes are widely used in time series analysis [13] and identification (see e.g. [14], [15]).

Let the  $p$ -vector process  $\{y_t\}$  be modeled by the following ARMA model

$$\sum_{i=0}^N A_i y_{t-i} = \sum_{i=0}^M B_i w_{t-i} \quad (3.1)$$

where  $A_i$  and  $B_i$  are  $p \times p$  matrices, and

$$A_0 = B_0 = I, \quad A_N \neq 0 \text{ and } B_M \neq 0.$$

The process  $\{w_t\}$  is a stationary white noise with covariance  $R$ . We shall furthermore assume that  $R$  is full rank. The matrices  $A_i$  are also referred to as the autoregressive (AR) parameters, and the matrices  $B_i$  are called the moving average (MA) parameters.

A more elegant notation can be obtained by introducing the backward shift operator  $q$  defined as follows :

$$q^j s_t \triangleq s_{t-j} \quad (3.2)$$

This allows us to write (3.1) as

$$A(q) y_t = B(q) w_t \quad (3.3)$$

where

$$A(q) \triangleq I + \sum_{i=1}^N A_i q \text{ and } B(q) = I + \sum_{i=1}^M B_i q$$

are matrix polynomials (or polynomial matrices depending on how one looks at it) in the operator variable  $q$ .

We want to give the following definition about polynomial matrices.

*Definition 3.1*

Let  $M(q)$  be a square polynomial matrix in the complex variable  $q$ . When the polynomial  $\det(M(q))$  has all its zeroes strictly outside the unit disc centered at the origin of the complex plane,  $M(q)$  is called a *stable polynomial matrix*. □

In order for  $\{y_t\}$  to be a purely non-deterministic stationary process, we have to satisfy a stability condition on the model (3.3). We quote without proof.

*Property 3.1.*

Let the  $p$ -vector process  $\{y_t\}$  be generated by the model

$$A(q) y_t = B(q) w_t$$



where  $w_t$  is a full rank  $p$ -vector (wide sense) stationary white noise process and  $A(q)$  and  $B(q)$  are  $p \times p$  polynomial matrices in the complex variable  $q$ . Then a necessary and sufficient condition for  $\{y_t\}$  to be (wide sense) stationary and purely non-deterministic is that  $A(q)$  is a stable polynomial matrix.  $\square$

Since in this paper we are only interested in purely non-deterministic stationary processes, we shall always assume that  $A(q)$  is stable. This, together with the fact that  $A(0)$  is non singular, also ensures that  $A^{-1}(q)$  exists, i.e.  $A^{-1}(q)$  is a convergent series.

As in section 2, we shall define the autocovariance function of  $\{y_t\}$  as

$$R_y(k) \triangleq E \{y_{t+k} y_t^T\}$$

Furthermore, we shall denote the cross covariance between  $\{w_t\}$  and  $\{y_t\}$  by

$$R_{wy}(k) \triangleq E \{w_{t+k} y_t^T\}$$

Now, since  $\{w_t\}$  is a white process, and since  $y_t$  does only depend upon past and present  $w_t$ , we have the property

$$R_{wy}(k) = 0 \quad \text{for } k > 0 \quad (3.4)$$

Consider then the autocovariance function  $R_y(k)$  for  $k > M$

$$\begin{aligned} R_y(k) &= E \{y_{t+k} y_t^T\} \\ &= E \left\{ - \sum_{i=1}^N A_i y_{t+k-i} y_t^T + \sum_{i=0}^M B_i w_{t+k-i} y_t^T \right\} \end{aligned}$$

and using (3.4) we obtain

$$R_y(k) + \sum_{i=1}^N A_i R_y(k-i) = 0; \quad \text{for } k > M \quad (3.5)$$

This set of equations for  $k > M$  are sometimes called the *Yule-Walker* equations.

Note that these equations fix a linear relationship between some of the elements of the autocovariance function and the autoregressive (AR) coefficients  $A_i$ . These equations will play an important role in the discussion on realization in section 5. One of the things we would like to do now is to derive an expression for the autocovariance function  $R_y(k)$  as a function of the  $A_i$  and  $B_i$  parameters and the driving noise covariance  $R$ .

In order to do this conveniently, we shall represent the function  $\{R_y(k); k = -\infty, \dots, +\infty\}$  as an infinite matrix series, and call this matrix series the *spectrum of the process*  $\{y_t\}$ .

### Definition 3.2.

The spectrum of the  $p$ -vector process  $\{y_t\}$  is an infinite series defined as follows

$$S_y(q) \triangleq \sum_{i=-\infty}^{\infty} R_y(i) q^{-i} \quad (3.6) \quad \square$$

Similarly we can "transform" the sequences  $y_t$  and  $w_t$  into infinite series  $Y(q)$  and  $W(q)$  as follows :

$$Y(q) \triangleq \sum_{i=-\infty}^{+\infty} y_i q^{-i} \quad (3.7a)$$

$$W(q) \triangleq \sum_{i=-\infty}^{+\infty} w_i q^{-i} \quad (3.7b)$$

By equation (3.3) we have

$$A(q) y_i = B(q) w_i$$

thus

$$A(q) y_i q^{-i} = B(q) w_i q^{-i}$$

and by summing over all  $i \in [-\infty, +\infty]$  we obtain a relationship between  $Y(q)$  and  $W(q)$

$$A(q) Y(q) = B(q) W(q) \quad (3.8)$$

which is just another way of stating that (3.3) holds for all  $t$ .

Remember now that  $R_y(k) = E \{y_k y_0^T\}$  and

$$R_{wy}(k) = E \{w_k y_0^T\}, \quad \text{and that } \{w_t\} \text{ is a white noise,}$$

so  $R_{wy}(k) = R \cdot \delta_{k,0}$  where  $\delta_{k,0}$  is the Kronecker delta,

i.e.  $\delta_{k,0} = 0$  for  $k \neq 0$  and  $\delta_{0,0} = 1$ . Post multiplying equation (3.8) by  $y_0$  and taking expectations yields

$$A(q) S_y(q) = B(q) S_{wy}(q) \quad (3.9)$$

where

$$S_{wy}(q) \triangleq \sum_{i=-\infty}^{+\infty} R_{wy}(i) q^{-i}.$$

The function  $S_{wy}(q)$  is called the *cross spectrum* between the driving noise and the output process. An alternative expression for  $S_{wy}(q)$  may be obtained by noting that, because of the stationarity assumption,

$$\begin{aligned} R_{wy}(k) &= E \{w_0 y_{-k}^T\} \\ &= (E \{y_{-k} w_0^T\})^T \\ &= R_{yw}^T(-k) \end{aligned}$$

so

$$S_{wy}(q) = \sum_{i=-\infty}^{+\infty} R_{yw}^T(-i) q^{-i}$$

or

$$S_{wy}(q) = S_{yw}^T(q^{-1}) \quad (3.10)$$

An expression for  $S_{yw}(q)$  is easily found from

$$\begin{aligned}
S_{yw}(q) &= E \{Y(q) w_0^T\} \\
&= A^{-1}(q) B(q) E \{W(q) w_0^T\} \\
&= A^{-1}(q) B(q) R.
\end{aligned}$$

Therefore

$$S_{wy}(q) = R B^T(q^{-1}) A^{-T}(q^{-1}) \quad (*) \quad (3.11)$$

Combining (3.9) and (3.11) and premultiplying by  $A^{-1}(q)$  yields

$$S_y(q) = A^{-1}(q) B(q) R B^T(q^{-1}) A^{-T}(q^{-1}) \quad (3.12)$$

Equation (3.12) then relates the spectrum of the process  $\{y_t\}$  to the coefficient matrices  $(A_i, B_i)$  and the noise covariance matrix  $R$  of the ARMA model (3.3).

As a shorthand notation for the model (3.1) or (3.3), we shall employ the notation  $[A(q), B(q), R]$ . We shall also say that  $[A(q), B(q), R]$  is a realization of the process  $\{y_t\}$ . Two models  $[A(q), B(q), R]$  and  $[\tilde{A}(q), \tilde{B}(q), \tilde{R}]$  will be called *spectrally equivalent* if the spectra they generate are equal. Since the process  $\{y_t\}$  is, as far as we are concerned, fully specified by its spectrum, any model which is spectrally equivalent to the original model  $[A(q), B(q), R]$  is also called a realization of  $\{y_t\}$ .

Expression (3.12) is a very important relation in connection with the stochastic realization problem. It is clear from the model (3.3) that we can regard

$$H(q) \triangleq A^{-1}(q) B(q) \quad (3.13)$$

as the matrix transfer function between the white noise input  $\{w_t\}$  and the output  $\{y_t\}$ . Equation (3.12) then shows that when a white noise with covariance  $R$  is passed through a stable filter with transfer function matrix  $H(q)$ , then the output  $\{y_t\}$  of this filter is a stationary process whose spectrum is given by

$$S_y(q) = H(q) R H^T(q^{-1}) \quad (3.14)$$

Notice that  $S_y(q)$  is a rational matrix in the variable  $q$ . Conversely it can be shown that a rational spectral matrix can be factored in the product form (3.14), where  $H(q)$  admits a finite dimensional representation as in (3.13). This result is also called the *Spectral Factorization Theorem*. We shall give a precise statement and a proof, for scalar processes, in section 6. References for the vector case will be given there. The problem of how to find  $A(q)$ ,  $B(q)$  and  $R$  from  $S_y(q)$  so as to verify (3.12) is often called the spectral factorization problem. All the methods of solution of the stochastic realization problem for vector ARMA models are based on the equation (3.12).

The further discussion of vector ARMA models becomes

(\*) We use the notation  $A^{-T}$  for  $(A^{-1})^T$

rather involved, and quite outside the scope of an introductory tutorial paper. The special case  $p = 1$ , i.e.  $\{y_t\}$  is a scalar process, is rather interesting however. From here on then, we shall consider scalar ARMA equations. To make this very clear we shall in the sequel replace the matrix coefficients  $A_i$  and  $B_i$  by scalar coefficients  $a_i$  and  $b_i$ .

The stability condition for  $A(q)$  is also simplified, since  $A(q)$  is now a polynomial, so  $A(q)$  is said to be stable if all its zeroes are strictly outside the unit circle. If we now define

$$n = \max(N, M)$$

where  $N$  and  $M$  are the number of AR and MA terms respectively, we can of course write (3.1) as

$$\sum_{i=0}^n a_i y_{t-i} = \sum_{i=0}^n b_i w_{t-i} \quad (3.15)$$

where  $a_0 = b_0 = 1$  and  $a_k = 0$  for  $k \in [N+1, \dots, n]$  and  $b_k = 0$  for  $k \in [M+1, \dots, n]$ .

Clearly, since  $n = \max(N, M)$  it follows that  $a_n$  and  $b_n$  can never be simultaneously zero.

Remember that, in section 2, we defined the dimension of the process  $\{y_t\}$  as the rank of the infinite Hankel matrix formed with the elements of the covariance function  $R_y(k)$ . The same definition, obviously, applies here.

The following property holds :

*Property 3.1.*

If  $\{y_t\}$  is a scalar process generated by the scalar ARMA model (3.15) then  $\dim \{y_t\} \leq n$   $\square$

*Proof*

Consider the semi-infinite Hankel matrix

$$R_{n+1}^\infty = \begin{bmatrix} R_y(1) & R_y(2) & \dots \\ \vdots & \vdots & \\ R_y(n) & R_y(n+1) & \dots \\ R_y(n+1) & R_y(n+2) & \dots \end{bmatrix}$$

By the Yule-Walker equations (3.5) we have

$$R_y(k) = - \sum_{i=1}^n a_i R(k-i) \quad k > n$$

Therefore the last row vector (of infinite dimension) of  $R_{n+1}^\infty$  is a linear combination of the  $n$  previous row vectors, whence

$$\rho(R_{n+1}^\infty) = \rho(R_n^\infty) \leq n. \text{ By the same argument}$$

$$\rho(R_{n+j}^\infty) = \rho(R_n^\infty) \leq n \text{ for all } j \geq 0 \quad \square \square$$

*Definition 3.3.*

An ARMA realization  $[A(q), B(q), R]$  of the process  $\{y_t\}$  is called a *minimal realization* when

$\dim \{y_t\} = n = \max(\delta A(q), \delta B(q))$  where " $\delta A(q)$ " means "the degree of the polynomial  $A(q)$ "  $\square$

**Proposition 3.1.**

A necessary condition for the realization  $[A(q), B(q), R]$  to be a minimal realization of the scalar process  $\{y_t\}$  is that  $A(q)$  and  $B(q)$  are coprime, i.e. they have no common polynomial factors.  $\square$

*Proof*

By contradiction. Consider a stationary process  $\{y_t\}$  of dimension  $n$  and let  $[A(q), B(q), R]$  be a minimal realization of  $\{y_t\}$ , i.e.

$n = \max[\delta A(q), \delta B(q)]$ . Assume that  $A(q)$  and  $B(q)$  have a common factor  $D(q)$  with  $\delta D(q) = n'' > 0$ , so that  $A(q) = D(q) \tilde{A}(q)$  and  $B(q) = D(q) \tilde{B}(q)$ .

Let  $n' = n - n'' = \max[\delta \tilde{A}(q), \delta \tilde{B}(q)]$ .

Then :

$$S_y(q) = A^{-1}(q) B(q) R B^{-1}(q^{-1}) A^{-1}(q^{-1}) \\ = \tilde{A}^{-1}(q) \tilde{B}(q) R \tilde{B}^{-1}(q^{-1}) \tilde{A}^{-1}(q^{-1})$$

The last relation shows that  $[\tilde{A}(q), \tilde{B}(q), R]$  is also a realization of the process  $\{y_t\}$ . By property 3.1 it follows that  $\dim \{y_t\} \leq n' < n$ , which contradicts the assumption.  $\square \square$

As it turns out,  $A(q), B(q)$  coprime is in general not a sufficient condition for the realization  $[A(q), B(q), R]$  to be minimal. If we look at a special class of ARMA models however, namely those where the polynomial  $B(q)$  is invertible, i.e. has all its zeroes outside or on the unit circle, then  $A(q), B(q)$  coprime is indeed necessary and sufficient for  $[A(q), B(q), R]$  to be a minimal realization. The proof of this is based on some results on realization and will be discussed in section 6.

Given now a minimal realization  $[A(q), B(q), R]$  of a scalar process  $\{y_t\}$  we shall show that one can always find a spectrally equivalent realization  $[A(q), C(q), Q]$  which is minimal and invertible, i.e. where  $C(q)$  has all its zeroes on or outside the unit circle. Such a realization will be called an *ARMA innovations representation* (IR) for  $\{y_t\}$ .

**Definition 3.4.**

$[A(q), C(q), Q]$  is an ARMA innovations representation (IR) for  $\{y_t\}$  if it is minimal and invertible.  $\square$

Given now a minimal ARMA model  $[A(q), B(q), R]$  for the (wide sense) stationary process  $\{y_t\}$ , we shall construct an innovations representation for the same process.

Consider the polynomial  $B(q)$ . If  $B(q)$  has all its zeroes outside or on the unit circle we may put  $C(q) = B(q)$  and  $Q = R$  and  $[A(q), C(q), Q]$  will be the required IR.

If  $B(q)$  has some of its zeroes inside the unit circle, we can factor  $B(q)$  as

$$B(q) = B^+(q) B^-(q) \quad (3.16)$$

where  $B^+(q)$  is invertible and  $B^-(q)$  has all its zeroes strictly inside the unit circle.

Let  $\delta B^-(q) = n' \leq n$ , then  $B^-(q)$  can be factored as

$$B^-(q) = \prod_{i=1}^{n'} (1 - \nu_i q) \quad (3.17)$$

where  $\nu_i$  are the inverse zeroes of  $B^-(q)$ , so  $|\nu_i| > 1$ .

We shall now define the polynomial  $\tilde{B}(q)$  as

$$\tilde{B}(q) \triangleq \prod_{i=1}^{n'} (1 - \frac{1}{\nu_i} q) \quad (3.18)$$

It is seen from (3.18) that  $\tilde{B}(q)$  has the numbers  $\{1/\nu_i\} > 1$  as its zeroes, so  $\tilde{B}(q)$  is invertible. The relationship between  $B^-(q)$  and  $\tilde{B}(q)$  is given by

$$B^-(q) = (-1)^{n'} q^{n'} \prod_{i=1}^{n'} \nu_i \tilde{B}(q^{-1}) \quad (3.19)$$

This last relationship follows straightforwardly from (3.17) and (3.18).

Using now (3.16) and (3.19), we see that we can express the spectrum  $S_y(q)$  of  $\{y_t\}$  as

$$S_y(q) = \frac{B^+(q) \tilde{B}(q^{-1}) [(-1)^{n'} q^{n'} \prod_{i=1}^{n'} \nu_i] R [(-1)^{n'} q^{-n'} \prod_{i=1}^{n'} \nu_i] \tilde{B}(q) B^+(q^{-1})}{A(q) A(q^{-1})}$$

By defining

$$Q \triangleq \prod_{i=1}^{n'} \nu_i^2 R \quad (3.20)$$

and

$$C(q) \triangleq B^+(q) \tilde{B}(q) \quad (3.21)$$

we see that  $S_y(q)$  is also realized by the model  $[A(q), C(q), Q]$ , which is clearly an IR, since  $C(q)$  is invertible. It follows from (3.20) and  $|\nu_i| > 1$  that  $Q \geq R$ .

Let us summarize the preceding argument in a proposition :

**Proposition 3.2.**

If the ARMA model  $[A(q), B(q), R]$  is an arbitrary minimal realization of the stationary process  $\{y_t\}$ , then  $[A(q), C(q), Q]$  is the innovations representation of  $\{y_t\}$  where  $C(q)$  is invertible and  $Q \geq R$ .

$C(q)$  and  $R$  are uniquely defined by the set of equations :

$$B(q) = B^+(q) B^-(q) \quad (3.22a)$$

$$B^-(q) = \prod_{i=1}^{n'} (1 - \nu_i q) \quad , \quad |\nu_i| > 1, \quad n' \leq n \quad (3.22b)$$

$$\tilde{B}(q) = \prod_{i=1}^{n'} (1 - \frac{1}{\nu_i} q) \quad (3.22c)$$

$$C(q) = B^+(q) \tilde{B}(q) \quad (3.22d)$$

$$Q = \prod_{i=1}^{n'} \nu_i^2 R \quad (3.22e)$$

Note that in order to obtain the IR from a given model  $[A(q), B(q), R]$ , a factorization of the polynomial  $B(q)$  has to be carried out. Also note that the variance of the driving white noise in the IR is never smaller than the driving white noise in any other spectrally equivalent realization.

#### 4. THE STATE-SPACE REALIZATION PROBLEM

##### 4.1. Introduction

In this section we show how to derive a state-space model for a finite-dimensional stationary process  $\{y_t\}$  described by its covariance function  $R_y(k)$ .

In Section 2 we have established the relations that exist between the parameters of a state-variable model and the covariance function of the output  $y_t$  of that model. We have also shown that to each arbitrary state-variable white-noise driven model we can associate an innovations representation (IR) that has the same output covariance. This IR is the inverse of the Kalman filter associated with the given state-space model. We shall of course heavily rely on the results of Section 2 to solve the inverse problem, namely to derive a state-space model whose output has a specified covariance. We shall give a complete solution to this state-variable covariance realization problem for scalar processes  $\{y_t\}$ .

We have shown in section 2 that the covariance function  $R_y(k)$  of a process that admits a state-variable model has the following structure :

$$R_y(k) = HF^{k-1}L, \quad k > 0 \quad (4.1.a)$$

$$= H \Pi H^T + M, \quad k = 0 \quad (4.1.b)$$

$$= L^T (F^T)^{|k|-1} H^T, \quad k < 0 \quad (4.1.c)$$

where  $H, F$  are respectively the output matrix and the state-transition matrix of the model, and  $L, \Pi$  and  $M$  are related to other model parameters including the covariances of the driving noises.

The state-variable realization problem can now be subdivided into two major steps :

- a. factor  $R_y(k)$ , for  $k > 0$ , into  $HF^{k-1}L$ , i.e. derive  $H, F$  and  $L$  of dimensions  $1 \times n, n \times n$  and  $n \times 1$  respectively from  $R_y(k)$  such that (4.1.a) holds with  $n$  minimal. This step is called, for obvious reasons, the covariance factorization problem.

- b. once  $H, F$  and  $L$  have been computed in step a, obtain a state-variable innovations representation of the form (2.20)-(2.22) with output matrix  $H$  and state-transition matrix  $F$ , where  $K_p$  and  $Q$  are derived from  $H, F, L$  and  $R_y(0)$ .

This second step can be called the problem of state-variable stochastic realization from given covariance factors. We shall treat these two problems separately.

##### 4.2. The covariance factorization problem

We now address ourselves to the problem of factoring a scalar covariance function  $R_y(k)$ ,  $k > 0$ , in the form

$$R_y(k) = HF^{k-1}L, \quad k > 0 \quad (4.1.a)$$

where the dimension  $n$  of  $F$  is minimal. We introduce the following definitions :

*Definition 4.1.*

We shall call the triple  $\{H, F, L\}$  with  $H \in \mathbb{R}^{1 \times n'}$ ,  $F \in \mathbb{R}^{n' \times n'}$ ,  $L \in \mathbb{R}^{n' \times 1}$  a factorization of the scalar covariance function  $R_y(k)$ , if  $R_y(k) = HF^{k-1}L$ ,  $k > 0$ .  $\square$

*Definition 4.2.*

The triple  $\{H, F, L\}$  is a minimal factorization of the scalar covariance function  $R_y(k)$  if it is a factorization, and if  $\dim F = n = \dim(\{y_t\})$ .  $\square$

We first observe that there is no unique minimal factorization of  $R_y(k)$ . Indeed, if  $\{H, F, L\}$  is one triple of minimal factors that satisfies (4.1.a), then

$\{HT^{-1}, TFT^{-1}, TL\}$  is an equivalent triple of minimal factors provided  $T$  is any nonsingular  $n \times n$  matrix. We shall therefore solve the minimal factorization problem by choosing a particular form for  $F$  that makes the solution unique.

Let  $n$  be the rank of the infinite Hankel matrix  $\mathcal{H}_{1,\infty}(R_y)$ , or equivalently the dimension of the  $\{y_t\}$  process (see section 2). Then we know by property 2.2. that

$$\rho(\mathcal{H}_{1,\infty}) = \rho(\mathcal{H}_{1,n}) = n$$

Therefore, the  $(n+1)$ -th row of  $\mathcal{H}_{1,\infty}(R_y)$  can be uniquely expressed as a linear combination of the first  $n$  rows. Let  $a_1, \dots, a_n$  be the coefficients of this linear combination, i.e. let

$$R_y(n+k) = - \sum_{j=1}^n a_j R_y(n+k-j), \quad k \geq 1 \quad (4.2)$$

Notice that equations (4.2) are nothing else but the Yule-Walker equations derived in section 3. Let us show now how the coefficients  $a_j$  can be obtained from the covariance function  $R_y(k)$ .

First, we compute the rank  $n$  of the infinite-dimensional Hankel matrix  $\mathcal{H}_{1,\infty}(R_y)$ .

Then we compute  $a_1, \dots, a_n$  by solving the set of  $n$  linear equations (4.2) for  $k=1, \dots, n$

$$\begin{bmatrix} R_y(1) & R_y(2) & \dots & R_y(n) \\ R_y(2) & R_y(3) & \dots & R_y(n+1) \\ \vdots & \vdots & \ddots & \vdots \\ R_y(n) & R_y(n+1) & \dots & R_y(2n-1) \end{bmatrix} \begin{bmatrix} a_n \\ a_{n-1} \\ \vdots \\ a_1 \end{bmatrix} = \begin{bmatrix} R_y(n+1) \\ R_y(n+2) \\ \vdots \\ R_y(2n) \end{bmatrix} \quad (4.3)$$

The solution is unique because the matrix is invertible since it is  $\mathcal{H}_{1,n}(R_y)$ , which has rank  $n$  (see Property 2.2 in section 2).

#### Comments

1) In practice, the test for the rank of the infinite-dimensional Hankel matrix  $\mathcal{H}_{1,\infty}(R_y)$  is of course not an easy matter. However Property 2.2 suggests the following practical procedure.

Test the rank of submatrices  $\mathcal{H}_{1,j}$  of increasing dimension. Suppose for a given  $j$ ,  $\text{rank}[\mathcal{H}_{1,j}] = j_1 < j$ . Then identify the parameters  $a_1, \dots, a_{j_1}$  by solving the equations (4.8) with  $n = j_1$ , and check whether the Yule-Walker equations (4.3) are satisfied for  $k > j_1$ . If yes, then  $n = j_1$ . If no, then increase the dimension of the Hankel matrix  $\mathcal{H}_{1,j}$  until, for some  $j$ , rank

$[\mathcal{H}_{1,j}] = j_2 < j$ , with  $j_2 > j_1$ . Repeat the procedure with  $j_1$  replaced by  $j_2$ . To test the rank of a finite  $j \times j$  Hankel matrix, one can use a recursive Hankel factorization algorithm such as described in [16], [17], which besides factoring the Hankel matrix into a product  $O^*C^*$  also computes its rank.

2) In many cases one does not have the infinite covariance sequence  $\{R_y(k), k=0, 1, \dots\}$ , but only a finite portion  $\{R_y(k), k=0, 1, \dots, N\}$ . In such case one wants to obtain a model that fits this finite sequence of covariance elements. This is the so-called partial realization problem. In such case, the Hankel matrix that must be factored is of course always finite. We shall not go any further into the partial realization problem; in this paper, we shall assume that we have the infinite covariance sequence and that the rank of  $\mathcal{H}_{1,\infty}(R_y)$  can be computed, or at least estimated, for all practical purposes.

Having thus determined  $n$  coefficients from the Hankel matrix, we can formulate a solution to the covariance factorization problem in the form of a proposition:

#### Proposition 4.1

Let  $n$  be the rank of  $\mathcal{H}_{1,\infty}(R_y)$  and let the elements of the  $(n+1)$ -th row of  $\mathcal{H}_{1,\infty}(R_y)$  be expressed as a function of the elements of the first  $n$  rows as in (4.3). Then a minimal factorization is obtained with the

following choice of  $H, F$  and  $L$ :

$$H = [1 \quad 0 \quad \dots \quad 0] \quad (4.4)$$

$$F = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & 0 \\ -a_n & -a_{n-1} & \dots & \dots & -a_1 \end{bmatrix} \quad (4.5)$$

$$L = [R_y(1) \quad R_y(2) \quad \dots \quad R_y(n)]^T \quad (4.6)$$

□

#### Proof

First we shall prove by induction that

$$F^k L = [R_y(k+1), \dots, R_y(k+n)]^T, \quad k \geq 0 \quad (4.7)$$

Assume that (4.7) holds for  $k$ , then for  $k+1$  we obtain

$$F^{k+1} L = F(F^k L) = F[R_y(k+1), \dots, R_y(k+n)]^T$$

but because of the special structure of  $F$ , this becomes

$$F[R_y(k+1), \dots, R_y(k+n)]^T = \begin{bmatrix} R_y(k+2) \\ \vdots \\ R_y(k+n) \\ -\sum_{j=1}^n a_j R_y(k+n+1-j) \end{bmatrix}$$

From equations (4.3) we see that this last entry can be expressed as  $R_y(k+1+n)$ , thus we have shown that

$$F^{k+1} L = [R_y(k+1+1), \dots, R_y(k+1+n)]^T$$

To conclude the proof of relation (4.7), notice that for  $k=0$

$$L = [R_y(1), \dots, R_y(n)]^T$$

which holds by definition.

Premultiplying  $F^k L$  by  $H$  has the effect of selecting the first entry of  $F^k L$ , thus

$$R_y(k) = H F^{k-1} L \quad \text{for } k \geq 1$$

so we have shown that  $\{H, F, L\}$  is a factorization of  $R_y(k)$ .

It is a minimal factorization because the dimension of  $F$  is  $n$ , the rank of  $\mathcal{H}_{1,\infty}(R_y)$ . If there was a realization with a matrix  $F$  of dimension  $n' < n$ , then by Property 2.1 of section 2 the rank of  $\mathcal{H}_{1,\infty}(R_y)$  would be strictly smaller than  $n$ , which

contradicts the assumption.  $\square$

*Comment :*

The particular choice of H and F in (4.4)-(4.5) yields what is called the observability canonical form.

Notice that the observability matrix  $O^*$  of [H, F] is the identity matrix.

*Corollary 4.1*

The pair (F, L) with F and L defined by equations (4.5) and (4.6) respectively is completely controllable, i.e.

$$\rho[L, FL, \dots, F^{n-1}L] = n \quad \square$$

*Proof*

From the definition of L, and from equation (4.7) one sees that the controllability matrix is

$$[L, FL, \dots, F^{n-1}L] = \mathcal{K}_{1,n}$$

The corollary follows from the fact that  $\rho(\mathcal{K}_{1,n}) = n$ .  $\square$

The factors H, F and L of  $R_Y(k)$  are expressed in terms of the known elements  $R_Y(1), \dots, R_Y(n)$  and in terms of the coefficients  $a_1, \dots, a_n$  of the linear combination (4.3).

The matrix F as defined by equation (4.5) in Proposition 4.1 is a stable matrix, i.e. all its eigenvalues are strictly inside the unit circle. This result is formulated as follows.

*Proposition 4.2*

The matrix F as defined in equation (4.5), where the  $a_i$  are determined from the Yule-Walker equations, i.e. they satisfy

$$R_Y(n+k) + \sum_{i=1}^n a_i R_Y(n+k-i) = 0, \quad k \geq 1,$$

has all its eigenvalues strictly inside the unit circle.  $\square$

*Proof*

The sequence  $R(i+1)$  ( $i=0, 1, \dots$ ) can be considered as the impulse response of the system

$$x(i+1) = Fx(i) + Lu(i)$$

$$R(i+1) = Hx(i)$$

with  $u(0) = 1$ ,  $u(i) = 0$ ,  $i > 0$ .

Since (F, H) is completely observable and (F, L) is completely controllable, and  $\lim_{i \rightarrow \infty} \|R(i)\| = 0$  (which is true because  $\{y_t\}$  is purely non-deterministic) it must follow that F has all its eigenvalues inside the unit circle.  $\square$

#### 4.3. Computation of the gain $K_p$ and the noise covariance Q

Through the covariance factorization method just

described we have obtained H, F, L from  $R_Y(k)$  with a minimal dimensional F matrix (see (4.1.a)). Now we show how to compute  $K_p$  and Q from H, F, L and  $R_Y(0)$  such that the output of the following state-variable model

$$s_{t+1} = F s_t + K_p e_t \quad (4.9.a)$$

$$y_t = H s_t + e_t \quad (4.9.b)$$

with

$$E \{e_t e_{t'}\} = Q \delta_{t,t'} \quad (4.9.c)$$

has the prescribed covariance  $R_Y(k)$ .

We shall solve this problem using the state-variable innovations model described in section 2. We shall actually show that the expression (2.20) and (2.22) we have derived in section 2 for  $K_p$  and Q can be rewritten in terms of quantities that involve only H, F, L and  $R_Y(0)$ .

*Proposition 4.3*

Assuming that  $\{H, F, L\}$  is a minimal factorization of the covariance function  $R_Y(k)$  of a stationary process, then (4.9) is a state-variable model for that process with  $K_p$  and Q computed as follows :

$$K_p = [L - F \hat{\Pi} H^T] [R_Y(0) - H \hat{\Pi} H^T]^{-1} \quad (4.10.a)$$

$$Q = [R_Y(0) - H \hat{\Pi} H^T] \quad (4.10.b)$$

and the  $n \times n$  matrix  $\hat{\Pi}$  is the non-negative definite solution of the algebraic Riccati equation.

$$\hat{\Pi} = F \hat{\Pi} F^T + K_p Q K_p^T \quad (4.11.a)$$

or equivalently

$$\hat{\Pi} = F \hat{\Pi} F^T + [L - F \hat{\Pi} H^T] [R_Y(0) - H \hat{\Pi} H^T]^{-1} [L - F \hat{\Pi} H^T]^T \quad (4.11.b)$$

*Proof*

In section 2 we have shown that for any state-variable realization  $\{F, H, G, \Sigma, S, M\}$

1° the covariance of the output process  $\{y_t\}$  has the form (4.1),

2° there exists a "covariance equivalent" innovations representation (2.20)-(2.22) (whose output process has the same covariance function).

Now we show that this covariance equivalent innovations representation can be expressed entirely in terms of the quantities H, F, L and  $R_Y(0)$  of the given model, with L and  $R_Y(0)$  as defined by (2.8) and (2.12).

Conversely therefore, if we are given the quantities H, F, L and  $R_Y(0)$  derived from the covariance function  $R_Y(k)$  of a process  $\{y_t\}$ , then the innovations representation (2.20)-(2.22), with  $K_p$  and Q expressed as functions of H, F, L and  $R_Y(0)$ , will generate an output

process  $\{y_t\}$  that has the required covariance and is therefore a solution to the stochastic realization problem. Now let us get to the heart of the proof. By replacing in (2.25)  $K_p$  and  $Q$  by their expressions (2.22), we get :

$$\hat{\Pi} = F \hat{\Pi} F^T + [FPH^T + GS][HPH^T + M]^{-1} [FPH^T + GS]^T \quad (4.12)$$

Now we replace  $P$  by  $\Pi - \hat{\Pi}$  following (2.26) :

$$\hat{\Pi} = F \hat{\Pi} F^T + [F \hat{\Pi} H^T - F \hat{\Pi} H^T + GS][H \hat{\Pi} H^T - H \hat{\Pi} H^T + M]^{-1} \cdot [F \hat{\Pi} H^T - F \hat{\Pi} H^T + GS]^T \quad (4.13)$$

Finally, using (2.8) and (2.12), we get :

$$\hat{\Pi} = F \hat{\Pi} F^T + [L - F \hat{\Pi} H^T][R_y(0) - H \hat{\Pi} H^T]^{-1} [L - F \hat{\Pi} H^T]^T \quad (4.14)$$

Similarly :

$$K_p = [L - F \hat{\Pi} H^T][R_y(0) - H \hat{\Pi} H^T]^{-1} \quad (4.15.a)$$

$$Q = [R_y(0) - H \hat{\Pi} H^T] \quad (4.15.b)$$

It is now clear from (4.15) that  $K_p$  and  $Q$  depend only on  $H, F, L, R_y(0)$  and on  $\hat{\Pi}$ , which itself is the non-negative definite solution of the Riccati equation (4.14) that depends only on  $H, L, F$  and  $R_y(0)$ . This completes the proof.  $\square$

For the sake of clarity we recall the major steps of the state-variable stochastic realization procedure :

- 1°) From the covariance function  $R_y(k)$  of the studied finite-dimensional stationary process  $\{y_t\}$ , compute the rank of the Hankel matrix  $\mathcal{H}_{1, \infty}(R_y)$ . Let  $n$  be this rank.
- 2°) Compute a minimal factorization  $H, F, L$  of  $R_y(k)$ ,  $k > 0$ , in observability canonical form (see (4.4)-(4.6)) by solving the normal equations (4.8).
- 3°) Compute the non-negative definite solution  $\hat{\Pi}$  of the algebraic Riccati equation (4.11.b).
- 4°) Compute  $K_p$  and  $Q$  via (4.10).  
Then (4.9) is a state-variable realization of the process  $\{y_t\}$  with the prescribed covariance function  $R_y(k)$ .

#### Comments

1) We have derived the innovations model (4.9)-(4.11) for the process  $\{y_t\}$  with covariance factors  $H, F$  and  $L$  by first showing that the output of the innovations model (2.20)-(2.22) has a covariance  $R_y(k) = HF^{k-1}L$ , and by then making the simple algebraic substitution  $P = \Pi - \hat{\Pi}$  in (2.21)-(2.22) so as to get rid of the unknown quantities  $P, M, G, S$  and  $\Sigma$  which cannot be obtained from the factorization of  $R_y(k)$ . It can also be verified directly that the output of model (4.9)-(4.11) has the prescribed covariance given by (4.1). This direct verification would provide a valid and short

proof of the proposition; however it would not have given the reader a feeling as to where the equations (4.10)-(4.11) originate.

2) Solving the algebraic Riccati equation (4.11.b) for the non-negative definite solution  $\hat{\Pi}$  is by no means an easy matter. One practical way of doing this is to search for the steady-state solution of the iterative Riccati equation

$$\hat{\Pi}_{i+1} = F \hat{\Pi}_i F^T + [L - F \hat{\Pi}_i H^T][R_y(0) - H \hat{\Pi}_i H^T]^{-1} [L - F \hat{\Pi}_i H^T]^T \quad (4.16.a)$$

$$\hat{\Pi}_0 = 0 \quad (4.16.b)$$

It can be shown (see e.g. [18], [10]) that the steady-state solution of (4.16) is the non-negative definite solution of the algebraic Riccati equation (4.11.b).

#### 4.4. Discussion

The development in this section comprised two major steps. First we obtained, from the covariance function, a factorization  $\{H, F, L\}$ . Then, given this factorization and  $R_y(0)$  we could compute the elements  $K_p$  and  $Q$  of the state-space innovations representation  $\{F, H, K_p, Q\}$  for  $\{y_t\}$ . In fact, the development in this second step did not in any way rely on the fact that  $R_y(k)$  is a scalar covariance function, so the results carry over immediately if  $\{y_t\}$  is a  $p$ -vector process. The solution method we described was first published (for the discrete time case) in 1972 [5]. The first step, the factorization of  $R_y(k)$  in  $\{H, F, L\}$  did, in our presentation, hinge on the fact that  $R_y(k)$  was a scalar covariance function. Similar results can be obtained for the  $p$ -vector case, but in order to understand these results a good background knowledge of structural properties of multivariable systems is required. This background can be acquired in e.g. [19]. The covariance factorization for vector processes is equivalent to the deterministic realization problem of vectorial impulse responses namely the factorization of matrix impulse responses  $H(k)$  into  $HF^kG$ . This problem is treated e.g. in [9].

### 5. REALIZATION OF ARMA PROCESSES

#### 5.1. Introduction

In this section we show how to derive an ARMA model for a finite dimensional stationary scalar process  $\{y_t\}$  described by its covariance function  $R_y(k)$ . Of course this ARMA model can be obtained by first deriving a state-variable innovations model from the covariance function using the techniques of section 4, and then computing the transfer function (or, equivalently, an ARMA model) from this state-space model. However, we want to be able to compute an ARMA representation directly from  $R_y(k)$  without going through this indirect route. The results of this section will of course

heavily rely on Section 3, where we have established the relations that exist between an ARMA realization  $[A(q), B(q), R]$  and the spectrum  $S_y(q)$  of the output process  $\{y_t\}$  of this model, namely:

$$S_y(q) = A^{-1}(q) B(q) R B^T(q^{-1}) A^{-T}(q^{-1}) \quad (5.1)$$

We recall that the spectrum is just an alternative description of the covariance function as an infinite power series. One way of solving the ARMA realization problem would be to compute the spectrum of the given covariance function  $R_y(k)$ , to find a closed form expression of  $S_y(q)$  as a rational function of  $q$ , and then to factor  $S_y(q)$  into  $A(q)$ ,  $B(q)$  and  $R$  such that (5.1) holds and that the realization  $[A(q), B(q), R]$  is minimal. This procedure, based on the factorization of  $S_y(q)$ , is called spectral factorization. The spectral factorization problem is of course totally equivalent with the covariance factorization problem, the only difference being that spectral factorization is performed in the spectral domain (where one works with rational functions of the complex variable  $q$ ), while covariance factorization is performed in the time domain (where one works with functions of the time variable  $t$ , or  $k$ ). In this section we shall solve the stochastic realization problem in the time domain, as we have done throughout this paper. We shall therefore use covariance factorization methods rather than spectral factorization methods. The reason for doing so is that we want to exploit the time-domain innovations results of our first paper, rather than resort to algebraic methods. We have given the spectral domain interpretations mainly because they add a lot of insight into the stochastic realization problem, particularly for the ARMA realization problem. Indeed, transforming  $R_y(k)$  to  $S_y(q)$  has enabled us to write the explicit relation (5.1) between the spectrum of  $\{y_t\}$  and an ARMA model for  $\{y_t\}$ . No such explicit relation can be written between  $R_y(k)$  and the parameters  $A_i$ ,  $B_i$  and  $R$ .

Just as in section 4, we shall now restrict our attention to scalar processes  $\{y_t\}$ ; the solution of the vector realization problem requires background knowledge on the structure of multivariable systems. The ARMA realization problem can then be stated as follows: "Given the covariance function  $R_y(k)$ ,  $k = 0, 1, 2, \dots$  (or equivalently the spectrum  $S_y(q)$ ) of a scalar discrete-time zero-mean stationary process  $\{y_t\}$  and assuming  $\dim\{y_t\} = n < \infty$ , find then a minimal realization  $[A(q), B(q), R]$  (see Definition 3.3) where  $A(q)$  and  $B(q)$  are polynomials, and  $R$  is a positive scalar".

We shall see that, just as for the state-variable realization problem, the coefficients  $a_i$  of the polynomial  $A(q)$ , which determine the poles of the model (i.e. the autoregressive structure), will be obtained as the solution of linear Yule-Walker equations, while the constant  $R$  and the coefficients  $b_i$  of  $B(q)$  will be obtained via the asymptotic solution of a set of nonlinear equations.

We have shown in section 3 that every ARMA realization is spectrally equivalent with an invertible ARMA

representation, which was called the ARMA innovations representation. We shall therefore solve the ARMA realization problem by computing the innovations representation, thereby using the results of section 4 of our first paper [1] on predictors for ARMA models.

## 5.2. Computation of the AR parameters

Let  $n$  be the rank of the infinite Hankel matrix  $\mathcal{K}_{1,\infty}(R_y)$ , or equivalently the dimension of the  $\{y_t\}$  process. We assume that  $n$  has been determined from  $\mathcal{K}_{1,\infty}(R_y)$  by the procedure described in section 4. Then, by the same arguments as in section 4 we can uniquely compute the solution  $a_1, \dots, a_n$  of the Yule-Walker equations (4.3). The coefficients  $a_1, \dots, a_n$  will be the AR parameters of our ARMA model. To show this conclusively we have to first construct the complete ARMA model, i.e. also compute the coefficients  $b_i$  and  $R$ . But it is at least plausible that the  $a_i$  are the AR coefficients, since they obey the Yule-Walker equations that the AR parameters of an ARMA model obey (see Section 3).

*Lemma 5.1.*

The polynomial  $A(q) = 1 + \sum_{i=1}^n a_i q^i$  is stable, i.e. all its zeroes are outside the unit circle.  $\square$

*Proof*

We have shown in Proposition 4.2 that the matrix  $F$ , whose last row is  $[-a_n, -a_{n-1}, \dots, -a_1]$  (see (4.5)) has all its eigenvalues strictly inside the unit circle.

where  $\det(\lambda I - F) = \lambda^n + a_1 \lambda^{n-1} + \dots + a_n$ . It turns out that  $A(q) = \det(I - qF)$ . Therefore  $A(q)$  has all its zeroes strictly outside the unit circle.  $\square$

## 5.3. Computation of the MA parameters

To proceed further we now define the auxiliary process

$$r_t \triangleq y_t + \sum_{i=1}^n a_i y_{t-i} = \sum_{i=0}^n a_i y_{t-i} \quad \text{with } a_0 = 1 \quad (5.2)$$

and we show some properties of the  $\{r_t\}$  process.

*Proposition 5.1.*

The process  $\{r_t\}$  is a zero mean stationary finite variance process. Its covariance function is truncated, i.e.  $R_r(t, t-k) = 0$  for  $|k| > n$ .  $\square$

*Proof*

$$1) E\{r_t\} = \sum_{i=0}^n a_i E\{y_{t-i}\} = 0 \quad (5.3)$$

2) By Lemma 5.1 we know that  $A(q) \triangleq \sum_{i=0}^n a_i q^i$  is a

stable polynomial. Therefore  $\{r_t\}$  is a finite variance process because it is obtained by passing the finite



variance process  $\{y_t\}$  through a stable filter.

$$3) R_r(t, t-k) = E \left\{ \left( \sum_{i=0}^n a_i y_{t-i} \right) \left( \sum_{j=0}^n a_j y_{t-k-j} \right) \right\} \\ = \sum_{j=0}^n a_j \left[ \sum_{i=0}^n a_i R_y(k+j-i) \right] \quad (5.4)$$

We see from equation (5.4) that  $R_r(t, t-k)$  is a function of  $k$  only, whence we shall use the notation  $R_r(k)$  for  $R_r(t, t-k)$ .

By section 4 (eq. (4.2)) we have

$$\sum_{i=0}^n a_i R_y(k-i) = 0 \quad k > n; a_0 = 1 \quad (5.5)$$

Therefore it follows from equation (5.4) that

$$R_r(k) = 0 \quad \text{for } |k| > n \quad (5.6a)$$

$$= \sum_{j=0}^n a_j \left[ \sum_{i=0}^n a_i R_y(k+j-i) \right] \quad \text{for } |k| \leq n \quad (5.6b)$$

□ □

Our next step is to obtain a white noise driven model for  $\{r_t\}$ . Since  $\{r_t\}$  is a stationary process with a truncated covariance function of width at most equal to  $n$ , it can be modelled as a purely moving average (MA) process of degree  $n$ .

$$r_t = \sum_{i=0}^n b_i w_{t-i}, \quad b_0 = 1 \quad (5.7)$$

with  $w_t$  a zero mean white noise process with some unknown variance  $R$ . It now remains to show how the coefficients  $b_i$  and the variance  $R$  can be computed. It turns out that there are several ways of doing that, but all of them involve the solution of nonlinear equations as we now show.

The covariance function  $R_r(k)$ ,  $k = 0, 1, \dots, n$  of the  $\{r_t\}$  process can be computed from the covariance function  $R_y(k)$  through (5.6). We can now express  $R_r(k)$ ,  $k = 0, 1, \dots, n$  as a function of the unknown coefficients  $b_1, \dots, b_n$  and  $R$  by using (5.7) :

$$R_r(k) = E \left\{ \left( \sum_{i=0}^n b_i w_{t-i} \right) \left( \sum_{j=0}^n b_j w_{t-k-j} \right) \right\} \\ = R \left( \sum_{i=k}^n b_i b_{i-k} \right), \quad k = 0, 1, \dots, n \quad (5.8)$$

Equation (5.8) constitutes a set of  $n+1$  equations in the  $n+1$  unknowns  $b_1, \dots, b_n$  and  $R$ . However these equations are non-linear. Their solution therefore requires an iterative method; in addition this set of equations, as any set of non-linear equations, admits more than one solution. This last fact is, of course, entirely consistent with our observation in section 3 that there are several minimal ARMA realizations which are spectrally equivalent. We also know that there is a unique minimal invertible realization, the

ARMA innovations representation. Instead of trying to derive an algorithm for the direct solution of equation (5.8), we shall use the general innovations filter (GIF) formulas of our first paper [1] to obtain parameters for a MA model (5.7). We shall thereby insure that the obtained model is an innovations model, i.e.

$$r_t = \epsilon_t + \sum_{i=1}^n c_i \epsilon_{t-i} \\ = C(q) \epsilon_t \quad (5.9)$$

where the polynomial  $C(q)$  is invertible.

In other words, among all solutions of the set of nonlinear equations (5.8), we shall find the unique solution for which the polynomial  $C(q)$  has all its zeroes on or outside the unit circle.

In section 4 of our first paper [1] we have derived finitely recursive expressions for the predictor  $\hat{r}_t/t-1$  of a process  $r_t$  that has the "truncation" property.

The inverse of this predictor can of course be used as a model for the process  $r_t$ , i.e. :

$$r_t = \hat{r}_t/t-1 + \epsilon_t \quad (5.10)$$

The prediction formulas for  $\hat{r}_t/t-1$  assuming observations  $\{r_\tau\}_0^{t-1}$  ( $t > 0$ ), given in [1], are repeated here for convenience

$$\hat{r}_t/t-1 = \sum_{k=1}^n R_{r\epsilon}(t, t-k) R_\epsilon^{-1}(t-k, t-k) \epsilon_{t-k} \quad (5.11a)$$

$$\epsilon_t = r_t - \hat{r}_t/t-1, \quad t = 0, 1, \dots \quad (5.11b)$$

where

$$R_{r\epsilon}(t, \tau) \triangleq R_\epsilon(\tau, \tau) \triangleq 0 \quad \text{for } \tau < 0 \quad (5.11c)$$

We also recall that the coefficients  $R_{r\epsilon}(t, t-k) R_\epsilon^{-1}(t-k, t-k)$  of this predictor can be recursively computed from the covariance function  $R_r(k)$  as follows (see eq. (4.10)-(4.11) of [1]) :

$$\left[ \begin{aligned} R_{r\epsilon}(t, t-k) &= R_r(k) \\ &- \sum_{s=k+1}^n R_{r\epsilon}(t, t-s) R_\epsilon^{-1}(t-s, t-s) R_{r\epsilon}^T(t-k, t-s) \end{aligned} \right] \quad (5.12a)$$

$$R_\epsilon(t, t) = R_\epsilon(t, t) \quad (5.12b)$$

$$t = 0, 1, \dots; \quad k = 0, 1, \dots, n$$

again with the constraints (5.11c) and with initial condition

$$R_{r\epsilon}(0, 0) = R_\epsilon(0, 0) = R_r(0) \quad (5.12c)$$

The prediction filter equation (5.11a) is a non-stationary

filter, and the innovations  $\{\epsilon_t\}$  defined by (5.11b) form a non-stationary process. This non-stationarity is caused by the fact that, although  $\{r_t\}$  is a stationary process, the observation process is not (we only start to observe  $r_t$  at  $t = t_0$ ). However, if we let  $t_0$  tend to  $-\infty$ , the predictor for  $\{r_t\}$  will become time invariant, and  $\{\epsilon_t = r_t - \hat{r}_t/t-1\}$  will be a stationary finite variance process. Reversing the argument, if we start at  $t = 0$ , then for  $t \rightarrow \infty$ , i.e. asymptotically,  $\{\epsilon_t = r_t - \hat{r}_t/t-1\}$  will become stationary, and the limits

$$\lim_{t \rightarrow \infty} R_{r\epsilon}(t, t-i); \quad i = 0, 1, \dots, n$$

do exist. Furthermore  $\lim_{t \rightarrow \infty} R_{r\epsilon}(t, t)$  is the asymptotic variance of the innovations. We may actually note that  $R_{r\epsilon}(t, t)$  (or, equivalently,  $R_\epsilon(t, t)$ ) is always bounded by  $R_r(0)$  since setting  $k = 0$  in (5.12a) shows that  $R_{r\epsilon}(t, t)$  is obtained by subtracting a non-negative definite term from  $R_r(0)$ . Recall also that  $R_r(0)$  is bounded by Proposition 5.1.

Let us then define coefficients  $c_i$  and a constant  $Q$  as

$$Q = \lim_{t \rightarrow \infty} R_{r\epsilon}(t, t) \quad (5.14a)$$

$$c_i = \lim_{t \rightarrow \infty} R_{r\epsilon}(t, t-i) Q^{-1} \quad (5.14b)$$

The model

$$r_t = C(q) \epsilon_t; \quad E\{\epsilon_t \epsilon_\tau\} = Q \delta_{t, \tau}$$

where the coefficients of  $C(q)$  are given by equation (5.14b), is then an innovations realization for the process  $\{r_t\}$ , since it is obtained as the sum of the asymptotic innovations predictor  $\hat{r}_t/t-1$  for  $r_t$  and the innovation  $\epsilon_t$  (see (5.10)). The polynomial  $C(q)$  is therefore invertible.

The preceding argument can be summarized in a Proposition.

**Proposition 5.2.**

Let  $R_r(k)$  be the covariance function of a (wide-sense) stationary process  $\{r_t\}$ , and let  $R_r(k) = 0$  for  $|k| > n$ . Consider the iterative algorithm :

$$\begin{cases} R_{r\epsilon}(i, i-k) = R_r(k) \\ - \sum_{s=k+1}^n R_{r\epsilon}(i, i-s) R_\epsilon^{-1}(i-s, i-s) R_{r\epsilon}^T(i-k, i-s) \end{cases} \quad (5.15a)$$

$$R_\epsilon(i, i) = R_r(i, i) \quad (5.15b)$$

$$i = 0, 1, \dots; \quad k = 0, 1, \dots, n$$

$$\text{with } R_{r\epsilon}(i, i-j) = 0 \text{ and } R_\epsilon(i-j, i-j) = 0 \text{ for } j > i, \quad (5.15c)$$

and with initial condition

$$R_\epsilon(0, 0) = R_r(0) \quad (5.15d)$$

Then

a) the sequences  $R_{r\epsilon}(i, i-k)$ ,  $k = 0, 1, \dots, n$ , converge to a set of constant values  $\bar{c}_k$  as  $i \rightarrow \infty$ .

b) the model  $r_t = C(q) \epsilon_t$  with

$$C(q) = 1 + \sum_{k=1}^n c_k q^k, \quad c_k = \bar{c}_k \cdot \bar{c}_0^{-1} \quad (5.15e)$$

where  $\{\epsilon_t\}$  is a white noise with covariance

$$Q = \bar{c}_0 \quad (5.15f)$$

is an invertible realization of the process  $\{r_t\}$ .  $\square$

**Proof**

The proof is provided by the argument preceding the proposition.  $\square$

We have presently given a complete solution to the stochastic realization problem for scalar ARMA models. Indeed, by putting together the invertible moving average model  $r_t = C(q) \epsilon_t$  and the stable autoregressive model  $A(q) y_t = r_t$ , we obtain the ARMA innovations model

$$A(q) y_t = C(q) \epsilon_t \quad (5.16)$$

We have solved the realization problem by constructing a model piece by piece; it is an easy matter to verify that the process  $\{y_t\}$  generated by (5.16) has indeed the prescribed covariance.

Actually, for  $k > n$ , the covariance function  $R_y(k)$  of the output process  $\{y_t\}$  of (5.16) obeys the Yule-Walker equations (4.2) because the parameters  $a_1, \dots, a_n$  have been obtained through (4.3). In addition it follows from (5.5) and (5.6) that  $R_r(k)$  is entirely determined by the elements  $R_y(0), R_y(1), \dots, R_y(n)$  of the given covariance function. Since the parameters  $c_1, \dots, c_n$  and  $Q$  have been chosen such that (5.6) is satisfied, it follows that the covariance elements  $R_y(0), R_y(1), \dots, R_y(n)$  of the output process  $\{y_t\}$  of the model (5.16) coincide with the first  $n+1$  elements of the given covariance sequence. By the Yule-Walker equations, the complete covariance sequences coincide. We can therefore state this as a Proposition.

**Proposition 5.3.**

Assuming that  $a_1, \dots, a_n$  are the solutions of the Yule-Walker equations (4.3) and that  $c_1, \dots, c_n$  and  $Q$  have been obtained by (5.15), then the model (5.16) is an innovations model for the process  $\{y_t\}$ .  $\square$

For the sake of clarity we recall the major steps of the realization algorithm :

- (i) From the covariance function  $R_y(k)$ , determine the Hankel matrix  $\mathcal{H}_{1, \infty}(R_y(k))$ , and find its rank
- (ii) Solve equations (4.3) to find the  $n$  AR coefficients  $a_i$
- (iii) Compute the (truncated) autocovariance function  $R_r(k)$  from equations (5.5)

- (iv) Using  $R_T(k)$ , iterate algorithm (5.15) to obtain the constants  $\bar{c}_k$  ( $k = 0, \dots, n$ ). The MA parameters are given by  $c_k = \bar{c}_k \bar{c}_0^{-1}$ , and the driving noise covariance by  $Q = \bar{c}_0$ .

#### 5.4. Discussion

It is noteworthy that steps (i) and (ii) are *identical* to the first two steps required in order to obtain a state space model. Actually step (ii) determines the pole structure of the realization (both here and in Section 4) and requires the solution of a set of linear equations. In order to determine the zero structure ( $K_p$  in section 4, and  $C(q)$  here) we have to solve a quadratic equation. Iterative solutions have been given.

Finally we should observe that just as in Section 4, the computation of the moving average part  $r_t = C(q) \epsilon_t$  of our ARMA model did not rely on the fact that  $R_T(k)$  is a scalar covariance function. Actually the algorithm (5.15) is valid for a  $p$ -vector process  $\{r_t\}$ . To the best of our knowledge the use of (5.15) as a solution to the stochastic realization problem has not been published before, even though the prediction formulas (5.11)-(5.12) are well known [20].

### 6. DISCUSSION AND SOME MORE RESULTS

In this section we shall give some comments and further results that can be obtained from the material we have presented in sections 2 through 5. First let us give an extra definition :

#### Definition 6.1

A stationary covariance function  $R_Y(k) \in \mathbb{R}^{P \times P}$  is called factorable if there exists an integer  $n \geq 0$ , and constant matrices  $H \in \mathbb{R}^{P \times n}$ ,  $F \in \mathbb{R}^{n \times n}$  and  $L \in \mathbb{R}^{n \times P}$  such that  $R_Y(k)$  can be expressed as  $R_Y(k) = H F^{k-1} L$  ( $k > 0$ )  $\square$

The following statements about the process  $\{y_t\}$ , its covariance function  $R_Y(k)$  and its spectrum  $S_Y(q)$  are then equivalent :

#### 6.1. Equivalent statements

- (a)  $\{y_t\}$  is generated by passing white noise through a finite dimensional (i.e. lumped) dynamical model.
- (b)  $\{y_t\}$  is a finite dimensional process, i.e.  $\dim(\{y_t\}) = n < \infty$
- (c)  $\{y_t\}$  has a rational spectral density function,  $S_Y(q)$
- (d)  $\{y_t\}$  has a factorable covariance function,  $R_Y(k)$ .

We shall not prove the equivalence between these

statements; they can be demonstrated from the material presented in previous sections (for scalar processes at least). By presenting these equivalent statements we merely wanted to highlight the fact that there are various ways of characterizing finite-dimensional processes.

#### 6.2. The spectral factorization theorem

As promised in Section 3, we shall give a statement and proof of the Spectral Factorization Theorem for scalar processes.

#### Proposition 6.1 (Spectral factorization)

Let  $S_Y(q)$  be a rational spectral density. Then  $S_Y(q)$  can be factored as  $S_Y(q) = H(q) R H(q^{-1})$  where  $H(q)$  is a rational transfer function of the form  $H(q) = B(q) / A(q)$ .  $A(q)$  and  $B(q)$  are finite polynomials, and  $A(q)$  has all its zeroes strictly outside the unit circle.  $\square$

#### Proof

The result follows immediately from Section 5, where we showed that we can always find an innovations representation  $[A(q), C(q), Q]$  for a given rational spectrum  $S_Y(q)$ .  $\square \square$

Proposition 6.1 is an existence theorem. In Section 5 we gave a constructive procedure to compute a factorization of  $S_Y(q)$  if  $S_Y(q)$  is rational. The existence follows thus trivially *a posteriori*.

A similar theorem can be formulated for vector processes, i.e. when  $S_Y(q)$  is a rational matrix function. For results on this see e.g. [21].

#### 6.3. Characterization of a minimal ARMA realization

In section 3, we gave necessary conditions for an ARMA realization  $[A(q), B(q), R]$  to be a minimal realization of a finite dimensional scalar process  $\{y_t\}$ . This was the subject of Proposition 3.1. Using the results from section 5, we shall now be able to give a sufficient condition as follows.

#### Proposition 6.2

Let  $[A(q), B(q), R]$  be a realization for a finite dimensional scalar process  $\{y_t\}$ , and let  $n = \max(\partial A, \partial B)$ . Then  $\dim(\{y_t\}) = n$  if  $A(q)$  and  $B(q)$  are coprime and if  $B(q)$  is invertible.  $\square$

#### Proof

By contradiction. Let  $(A, B)$  be coprime,  $B(q)$  be invertible, and assume  $\dim(\{y_t\}) = n' < n$ . It then follows from the realization theory for ARMA models that there exists a realization  $[A'(q), B'(q), R]$  with  $B'(q)$  invertible and  $n' = \max(\partial A', \partial B')$ . Furthermore, since  $[A, B, R]$  and  $[A', B', R']$  realize the same process we must have

$$\frac{B'(q) R' B'(q^{-1})}{A'(q) A'(q^{-1})} = \frac{B(q) R B(q)}{A(q) A(q^{-1})}$$

$$\begin{aligned} & \text{or } B'(q) R' B'(q^{-1}) A(q) A(q^{-1}) \\ & = B(q) R B(q^{-1}) A'(q) A'(q^{-1}) \end{aligned} \quad (6.1)$$

Let's for convenience define the polynomials  $L(q)$  and  $R(q)$  as

$$L(q) = B'(q) R' B'(q^{-1}) A(q) A(q^{-1}) \quad (6.2a)$$

$$R(q) = B(q) R B(q^{-1}) A'(q) A'(q^{-1}) \quad (6.2b)$$

Now, let

$$B'(q) = \prod_{i=1}^{n'} (1 - \alpha'_i q) \quad B(q) = \prod_{i=1}^n (1 - a_i q) \quad (6.3a)$$

$$A'(q) = \prod_{i=1}^{n'} (1 - \pi'_i q) \quad A(q) = \prod_{i=1}^n (1 - \pi_i q) \quad (6.3b)$$

Since  $\dim \{y_t\} = n'$  it follows from Proposition 3.1 that  $A'(q)$  and  $B'(q)$  are coprime or equivalently  $\forall i, j \in (1, \dots, n'), \alpha'_i \neq \pi'_j$ .

Equation (6.1) expresses an equality between two polynomials, namely  $L(q) = R(q)$ . This must mean that all the zeroes of  $L(q)$  are also zeroes of  $R(q)$ .

From equations (6.2, 6.3) it follows that the zeroes of  $L(q)$  are given by the set  $\mathcal{L}$  which is the union of the two sets  $\mathcal{L}_1$  and  $\mathcal{L}_2$  defined as follows :

$$\mathcal{L} = \mathcal{L}_1 \cup \mathcal{L}_2$$

where

$$\mathcal{L}_1 \triangleq \{\alpha'_i, \frac{1}{\alpha'_i} ; i \in 1, \dots, n'\}$$

$$\mathcal{L}_2 \triangleq \{\pi_i, \frac{1}{\pi_i} ; i \in 1, \dots, n\}$$

The zeroes of  $R(q)$  are given by the set  $\mathcal{R}$  as follows

$$\mathcal{R} \triangleq \mathcal{R}_1 \cup \mathcal{R}_2$$

$$\text{where } \mathcal{R}_1 \triangleq \{\pi'_i, \frac{1}{\pi'_i} ; i \in 1, \dots, n'\}$$

$$\mathcal{R}_2 \triangleq \{a_i, \frac{1}{a_i} ; i \in 1, \dots, n\}.$$

We must have from equation (6.1) that

$$\mathcal{L} = \mathcal{R} \quad (6.4)$$

Since  $\alpha'_i \neq \pi'_j$ , for every zero in  $\mathcal{L}_1$  there must be a corresponding zero in  $\mathcal{R}_2$ , and for every zero in  $\mathcal{R}_1$  there must be a corresponding zero in  $\mathcal{L}_2$ .

At the end there will thus be  $2(n-n') > 0$  zeroes in  $\mathcal{L}_2$  and  $2(n-n') > 0$  zeroes in  $\mathcal{R}_2$  that are not matched by zeroes in  $\mathcal{R}_1$  and  $\mathcal{L}_1$  respectively. These  $2(n-n') > 0$  zeroes in  $\mathcal{L}_2$  have to match the  $2(n-n') > 0$  zeroes in  $\mathcal{R}_2$ . Furthermore, since by assumption  $|\pi_j| > 1$  and  $|\alpha_j| \geq 1$  ( $A(q)$  stable and  $B(q)$  invertible!) there must be a polynomial  $D(q)$  of degree  $n-n' > 0$ , and with all its zeroes strictly outside the unit circle which is common to  $A(q)$  and  $B(q)$ . This contradicts the assump-

tion that  $A(q), B(q)$  were coprime.  $\square \square$

With a simple example we shall illustrate the fact that the invertibility of  $B(q)$  is really required to make Proposition 6.2 work.

### Example 6.1

Take the realization  $A(q) = 1 - 0.5q, B(q) = 1 - 2q, R = r > 0$ ; note that  $B(q)$  is not invertible, because it has a zero at 0.5.

The number  $n$  is defined as  $n = \max(\partial A, \partial B) = 1$ .

In order to determine  $\dim \{y_t\}$ , we look at  $\pi_{1, \infty}$ .

It follows straightforwardly here that

$$R_y(1) = R_y(2) = \dots = 0, \text{ thus } \dim \{y_t\} = 0 < n = 1. \square$$

## 7. CONCLUDING REMARKS

This paper is on stochastic realization theory and algorithms. Basically, we have studied properties of covariance functions. First we have seen what happens if we pass a white noise process through a stable linear filter, i.e., what properties are induced by the filter structure into the covariance function of the output process. This was done in sections 2 and 3 for state space models and ARMA models respectively. Two basic results emerged from this. Firstly the fact that the Hankel matrix of the covariance function has finite rank. Secondly the fact that, if the given realization (we shall call a white noise driven model generating  $\{y_t\}$  a realization for  $\{y_t\}$ ) is minimal, then we can always find a unique invertible minimal realization, or a so-called innovations representation.

In sections 4 and 5 we considered the inverse problem, namely : given a covariance function, how can we infer a realization from it. The link between sections 2, 3 and sections 4, 5 was provided by the Hankel matrix of the covariance function. Indeed, if we find out that the Hankel matrix has a finite rank, then we can infer finite dimensional white noise driven models (i.e. realizations) for the process  $\{y_t\}$ .

The algorithms we have used to construct such models were based on the innovations filters derived in our first paper, and therefore led to innovations realizations (i.e. invertible models).

The construction of minimal invertible realizations, i.e. realization algorithms, involves three steps. First, we have to determine the rank of the infinite Hankel matrix of the covariance function. This is not a trivial problem, and some discussion is given in section 4. Second, we have to solve a linear problem, namely the solution of the Yule-Walker equations. This problem is trivial in the scalar case and more involved in the p-vector case (in this paper we did not discuss the vector case). For state space realizations this linear problem is also called the covariance factorization problem; for ARMA realizations it means that we obtain the AR part of the realization.

In a final (third) step then we had to solve a non-linear problem (actually a quadratic problem). For state space models we had to find a solution for an

Algebraic Riccati Equation (ARE), and for ARMA models we had to solve a (matrix) polynomial spectral factorization problem. For both cases we have given algorithms to compute the solutions but we have not explicitly shown the equivalence of both quadratic problems (see also the discussion in section 4.4 and section 5).

We shall consider now the application of the realization results in estimation theory. In our first paper [1], we presented the GIF (General Innovations Filter) for prediction. As we commented there, the GIF can be recursively computed from the covariance function  $R_y(k)$ , but the recursions are not finite, so a growing memory is required. Here, however, we showed that, if the process  $\{y_t\}$  admits a finite representation, a finite innovations representation (IR) can be obtained from covariance information. Given such an IR, we can obtain, by a trivial transformation, a prediction filter for  $\{y_t\}$ . This is discussed in [6] for state space models and in [22], [23] for ARMA models. We should comment that if our only interest is in prediction, we could get away with just ARMA models since they contain all the necessary information that is required to compute predicted estimates. However  $k$ -step predicted estimates of  $y_t$  (i.e.  $\hat{y}_t + k/t$ ) are more easily computed with state space models.

In addition it may be that we are also interested in the estimation of a process  $\{x_t\}$ , related to the observation process  $\{y_t\}$ . Again from [1], we know that, given cross-covariance information (i.e.  $R_{xy}(k)$ ), the GIF formulas provide a (non-finitely) recursive solution. The construction of finite dimensional estimators based on covariance information has been treated in [6] for the case where the cross covariance  $R_{xy}(k)$  is given in a factored form.

We have in this paper and in [1] restricted ourselves to discrete-time processes. The continuous-time results dealing with the state-space realization problem are given, for instance, in [4], where a factored form of the covariance function is assumed to be given.

As a final comment we want to mention the fact that stochastic realization theory plays an important role in the problem of systems identification. It is by no means clear that realization algorithms (based on an exact knowledge of the covariance function) are also good identification algorithms (based on covariances estimated from a finite data record).

Stochastic realization results have, however, provided important insight in the structural aspects of the identification problem. Some identification schemes are actually directly derived from the stochastic realization theory (see e.g. [24]-[25]). Some of the structure determination methods are also based on these results (see e.g. [26]).

## REFERENCES

1. WOUTERS, W. R. E., and GEVERS, M. : "An innovations approach to the discrete-time linear least-squares estimation problem", *Journal A*, 19,(1), 8-20 (1978).

2. ANDERSON, B. D. O.; MOORE, J. B. and LOO, S. G. : "Spectral factorization of time-varying covariance functions", *IEEE Trans. IT-15*, Sept. 550-557 (1969).
3. BRANDENBURG, L. H. : "Shaping filter models for non-stationary random processes", Ph. D. dissertation, Dept. Elec. Engr., Columbia University, New York, N. Y. (June 1968).
4. KAILATH, T. and GEESEY, R. : "An innovations approach to least-squares estimation - Part IV : Recursive estimation given lumped covariance functions", *IEEE Trans. AC-16*, Dec. 720-727 (1971).
5. GEVERS, M. : "Structural properties of realizations of discrete time Markovian processes", Ph. D. Thesis, T. R. 7050-19, Information Systems Laboratory, Stanford University (May 1972).
6. GEVERS, M. and KAILATH, T. : "An innovations approach to least-squares estimation - Part VI : Discrete-time innovations representations and recursive estimation", *IEEE Trans. AC-18*, Dec. 588-600 (1973).
7. RISSANEN, J. and KAILATH, T. : "Partial realization of random systems", *Automatica*, 8, 389-396 (July 1972).
8. AKAIKE, H. : "Stochastic theory of minimal realization", *IEEE Trans. AC-19*, 667-673 (Dec. 1974).
9. DICKINSON, B. W., KAILATH, T. and MORF, M. : "Canonical matrix fraction and state-space descriptions for deterministic and stochastic linear systems", *IEEE Trans. AC-19*, 656-666 (Dec. 1974).
10. FAURRE, P. : "Stochastic realization algorithms", in "System Identification : Advances and Case Studies", ed. by R. K. Mehra and D. G. Lainiotis, Academic Press, New York, 1-25 (1976).
11. GANTMACHER, F. R. : "The theory of matrices", vol. 1, Chelsea Publishing Cy, (1959).
12. KALMAN, R. E. : "A new approach to linear filtering and prediction problems", *Trans. ASME Ser. D. J. Basic Eng.* 82, 34-35 (1960).
13. BOX, G. E. P. and JENKINS, G. M. : "Time series analysis, forecasting and control", Holden Day (1970).
14. "Special issue on System Identification and Time Series Analysis", *IEEE Trans. AC-19* (Dec. 1974).
15. GUSTAVSSON, I., LJUNG, L. and SÖDERSTRÖM, T. : "Identification of processes in closed loop - Identifiability and accuracy aspects", Survey Paper, *Automatica* 13, 59-75 (Jan. 1977).
16. RISSANEN, J. : "Recursive identification of linear systems", *SIAM J. on Control*, 9 (3), 420-430 (1971).
17. VANBEVEREN, Y. : "Robustesse d'un algorithme de réalisation et son application à l'identification des systèmes faiblement couplés", Doctoral thesis, Université de Paris IX-Dauphine (Dec. 1977).
18. GEVERS, M. and KAILATH, T. : "The discrete-time stochastic realization problem : minimum variance property of the innovations representation", *Proc. 1973 IEEE Conf. on Decision and Control*, San Diego, 168-171 (Dec. 1973).
19. WOLOWICH, W. : "Linear multivariable systems", Springer-Verlag, Berlin (1974).

20. AASNAES, H. B. and KAILATH, T. : "An innovations approach to least-squares estimation, Part VII : Some applications of vector autoregressive moving average models" IEEE Trans. AC-18, Dec., 601-607 (1973).
21. YOULA, D. C. : "On the factorization of rational matrices", IEEE Trans. IT, 7, July, 172-189 (1961).
22. ASTRÖM, K. J. : "Introduction to stochastic control theory", Academic Press, New York, (1970).
23. HANNAN, E. J. : "Multiple time series", J. Wiley, New York, (1970).
24. TSE, E. and WEINERT, H. L. : "Structure determination and parameter identification for multivariable stochastic linear systems", IEEE Trans. AC-20, 603-612 (Oct. 1975).
25. GEVERS, M. and DABLEMONT, S. P. : "Identification and feedback control of an industrial glass furnace", to appear in Automatica.
26. AKAIKE, H. : "Canonical correlation analysis of time series and the use of an information criterion", in "System Identification : Advances and Case Studies", ed. by R. K. Mehra and D. G. Lainiotis, Academic Press, New York, 27-96 (1976).
27. BROCKETT, R. W. : "Finite dimensional linear systems", J. Wiley, New York, (1970).