

A STOCHASTIC APPROACH TO THE MINIMAL  
REALIZATION PROBLEM

Dissertation for the Degree of Ph. D.

MICHIGAN STATE UNIVERSITY

JAMES W. CLARY

1975

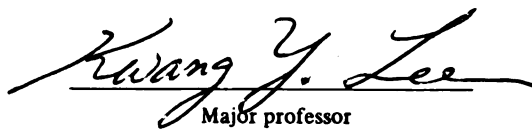
This is to certify that the  
thesis entitled  
A STOCHASTIC APPROACH TO THE  
MINIMAL REALIZATION PROBLEM

presented by

James W. Clary

has been accepted towards fulfillment  
of the requirements for

Ph. D degree in Systems Science

  
Major professor

Date 1/13/75

7195388

## ABSTRACT

### A STOCHASTIC APPROACH TO THE MINIMAL REALIZATION PROBLEM

By

James W. Clary

The goal of this work is to exploit the concept of a predictor space in the continuous time realization problem for systems generating a stationary or analytic impulse response matrix. The predictor space provides a natural representation of the state space and a minimal realization corresponds to selecting a basis of the predictor space.

A representation problem is considered first. Given two second order processes  $y(t)$  and  $z(t)$ , with  $y(t)$  measurable and  $z(t)$  continuous in quadratic mean and purely nondeterministic, the predictor space for  $y(t)$  is formed. Under rather general conditions if a finite dimensional basis for the predictor space exists then  $y(t)$  can be represented as the output of a linear finite dimensional dynamical system driven by the innovation process of  $z(t)$ .

Next the realization problem for stationary systems is considered. Given a stationary impulse response matrix a stochastic system is defined whose input-output covariance

equals it. The output of this system generates a finite dimensional predictor space. The previously developed representation theory then supplies a solution to the realization problem. Two algorithms are discussed in the context of the predictor space approach.

The realization problem for systems generating a time-varying analytic impulse response matrix is considered next. The development is similar to that for the stationary case with the exception of some material on the linear independence of analytic functions. An algorithm by Silverman for time-varying systems is discussed in the context of the theory developed.

A STOCHASTIC APPROACH TO THE MINIMAL REALIZATION PROBLEM

By

James W. Clary

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

DOCTOR OF PHILOSOPHY

Department of Electrical Engineering  
and Systems Science

1975

## ACKNOWLEDGMENTS

The author would like to express his gratitude to his major professor, Dr. Kwang Y. Lee, for introducing him to this topic and for his constant advice and encouragement during the preparation of this thesis.

Gratitude is also expressed to Dr. Robert O. Barr for introducing the author to the subject of Systems Science through his excellent teaching and for providing valuable assistance during his years of graduate study.

Thanks are also due to Dr. Robert A. Schlueter for many helpful discussions and much encouragement. The valuable advice of Dr. V.S. Mandrekar is also gratefully acknowledged.

Finally thanks are due to the author's wife Kay for her love, patience and understanding during the time spent on this work.

## TABLE OF CONTENTS

| Chapter |                                                                               | Page |
|---------|-------------------------------------------------------------------------------|------|
| 1       | INTRODUCTION                                                                  | 1    |
| 2       | MATHEMATICAL BACKGROUND                                                       | 8    |
|         | 2.1 The Realization Problem                                                   | 8    |
|         | 2.2 Hilbert Spaces of Random Vectors                                          | 17   |
| 3       | MARKOVIAN REPRESENTATION OF CONTINUOUS-TIME STOCHASTIC SYSTEMS                | 23   |
|         | 3.1 General Theory                                                            | 24   |
|         | 3.2 Special Cases                                                             | 42   |
| 4       | STOCHASTIC APPROACH TO THE MINIMAL REALIZATION PROBLEM IN THE STATIONARY CASE | 46   |
|         | 4.1 General Theory                                                            | 47   |
|         | 4.2 Silverman's Algorithm                                                     | 68   |
|         | 4.3 The Ho-Kalman Algorithm                                                   | 82   |
|         | 4.4 Example                                                                   | 92   |
| 5       | STOCHASTIC APPROACH TO THE MINIMAL REALIZATION PROBLEM IN THE ANALYTIC CASE   | 101  |
|         | 5.1 General Theory                                                            | 101  |
|         | 5.2 The Silverman Algorithm                                                   | 114  |
|         | 5.3 Example                                                                   | 123  |
| 6       | CONCLUSION                                                                    | 128  |
|         | 6.1 Summary of Results                                                        | 128  |
|         | 6.2 Problems for Future Research                                              | 130  |
|         | BIBLIOGRAPHY                                                                  | 131  |
|         | APPENDIX: GENERALIZED RANDOM PROCESSES                                        | 134  |

## CHAPTER ONE

### INTRODUCTION

Consider an input-output description of an unknown system in the form of an impulse response or weighting pattern matrix. Such a description may arise either through the performance specifications in the case when one designs a new system, or from experimental output obtained by inputting delta functions in the case when one is attempting to identify an unknown system. The problem of obtaining a state space description of the unknown system from the impulse response or weighting pattern matrix is known as the realization problem.

Beginning primarily with the papers of Gilbert [G-2] and Kalman [K-3] in 1963 the realization problem has been the subject of considerable research. Gilbert [G-2] gave a solution to the realization problem for the stationary case with the restriction that each element of the transfer function matrix has distinct poles. The paper by Kalman [K-3] showed that the impulse response matrix determines the controllable and observable portion of a system and gives a solution to the realization problem for stationary systems. His results make clear the importance

of minimal realizations, that is state space representations having the smallest possible dimension for the state vector. The solution to the realization problem by Kalman was presented in the form of an algorithm which reduced the state space of a nonminimal realization until it became minimal.

A paper by Ho and Kalman in 1965 presented a different approach to the realization problem for stationary systems. They showed that the minimal realization problem is equivalent to a certain kind of representation problem involving an infinite series of real matrices  $Y_k$ ,  $k = 1, 2, \dots$ , derived from the impulse response matrix and known as the Markov parameters. Their algorithm, while elegant, is difficult to motivate and explain in the context in which it was originally presented.

A series of papers by Silverman [S-1] and Silverman and Meadows [S-2, S-3] published between 1966 and 1971 presented significant new results on the realization problem and the closely related problem of equivalent system representations. Silverman obtained a new algorithm for the stationary realization problem which generalized to a large class of time varying linear systems, which he called "constant rank systems". This class of systems includes, but is larger than, the class of analytic systems.

The algebraic approach to system theory as presented in Kalman, Falb, and Arbib [K-4] or Padulo and Arbib [P-1] has obtained results relating to the realization problem

for stationary systems and analytic time-varying systems, but at present does not offer an algorithm for time-varying systems of the generality of Silverman's algorithm.

A new approach to the realization problem, called the predictor space approach, has recently been presented by Akaike [A-1] for stationary discrete time systems. Akaike's approach uses a stochastic system driven by white noise to obtain a solution to a given realization problem by the following procedure. He first shows that given two zero-mean real valued second-order stationary Gaussian stochastic processes  $y(n)$  and  $z(n)$  that are mutually stationarily correlated with covariance\*

$$V(\ell) = E\{y(n + \ell)z^T(n)\}$$

$y(n)$  can be obtained as the output of a linear stochastic system driven by the innovation process of  $z(n)$  of the form

$$v(n+1) = Fv(n) + Gw(n+1) \quad (1-1)$$

$$y(n) = Hv(n) + u(n) \quad (1-2)$$

$$w(n+1) = z(n+1) - z(n+1|n) . \quad (1-3)$$

The symbol  $z(n+1|n)$  represents the linear least squares prediction of  $z(n+1)$  given  $z(k)$ ,  $k = n, n-1, n-2, \dots$ .

It also is shown that

---

\* $E\{.\}$  denotes the expected value of the quantity within the brackets. The symbol  $(.)^T$  indicates the transpose operator.

$$V(\ell) = HF^{\ell}E\{v(n)z^T(n)\}, \ell = 0,1,2,\dots \quad (1-4)$$

Then given a stationary impulse response matrix  $W(n)$ , Akaike defines a stochastic system driven by a Gaussian white noise process  $z(n)$  with unit covariance by

$$y(n) = \sum_{m=0}^{\infty} W(m)z(n-m), \quad n = 0,1,2,\dots \quad (1-5)$$

It is then shown that since the input is white noise

$$V(\ell) = W(\ell) \quad (1-6)$$

so that using the previously derived representation (1-1) - (1-3)

$$W(\ell) = HF^{\ell}G, \quad \ell = 0,1,2,\dots \quad (1-7)$$

Thus the linear system

$$x(n+1) = Fx(n) + Gu(n+1) \quad (1-8)$$

$$y(n) = Hx(n) \quad (1-9)$$

is a solution to the realization problem for the given impulse response matrix  $W(\ell)$ . The basic concept in Akaike's approach is that of a predictor space, the space spanned by the projections of  $y(k)$ ,  $k = n, n+1, n+2, \dots$  onto the space spanned by  $z(k)$ ,  $k = n, n-1, n-2, \dots$ . It is the structure of this space that makes possible his representation (1-1) - (1-3).

Using the predictor space approach, Akaike is able to provide new insight into the realization problem and

the structure of the various algorithms for stationary systems. The Silverman and Ho-Kalman algorithms for example are easily analyzed in this context. The predictor space approach provides a powerful tool, as well as a unifying overview, for the realization problem.

It is the purpose of this dissertation to generalize Akaike's predictor space approach to continuous time systems. The outline of the presentation is as follows. Chapter 2 provides an introduction to the realization problem with some special results relating to the stationary case. A section reviewing the necessary background material on Hilbert spaces of random variables and the Hida-Cramer representation for purely nondeterministic stochastic processes is also presented.

In chapter 3 a representation problem for continuous time stochastic processes is considered. It is assumed that two continuous time second-order zero mean stochastic processes  $y(t)$  and  $z(t)$ , with  $y(t)$  measurable and  $z(t)$  purely nondeterministic and continuous in quadratic mean are given. The predictor space, the space spanned by the linear least-squares estimates of  $y(t)$  given  $\{z(\tau): \tau \leq t\}$ , is formed. It is then shown that by assuming a finite dimensional basis for the predictor space and with appropriate continuity and separability hypotheses that  $y(t)$  can be represented as the output of a continuous time linear system driven by the innovation process of  $z(t)$ . The theory is easily

generalized to the case when  $z(t)$  is white noise, and the chapter concludes with some special developments for the stationary case.

Chapter 4 uses the representation theory developed in chapter 3 to solve the realization problem for continuous time stationary systems. Given a stationary impulse response matrix a stochastic system is defined whose input-output cross-covariance is equal to it. It is then shown that the predictor space contains a finite basis and that the representation theory developed in chapter 3 can be applied. It is shown that this representation provides a minimal realization for the given impulse response matrix.

Two algorithms, the Ho-Kalman and the Silverman algorithm, are analyzed for the continuous-time stationary case. Although Akaike has already done this for the discrete-time case, and as shown in chapter 2 there is essentially no difference between the realization problem for the discrete-time and continuous time stationary cases, the justification is considerably more difficult in the context of continuous-time systems. The results demonstrate the power of the predictor space approach. An example illustrating the Silverman and Ho-Kalman algorithms and the predictor space approach to them is also presented.

In chapter 5 the representation theory of chapter 3 is applied to the realization problem for systems generating an analytic impulse response matrix. The procedure

is found to closely parallel the case for stationary systems with the addition of some material on the linear independence of analytic functions. Silverman's algorithm is considered for this case and an example is given.

In chapter 6 a summary, conclusions, and suggestions for further research are presented.

## CHAPTER TWO

### MATHEMATICAL BACKGROUND

This chapter is devoted to basic background material that will be used throughout the thesis. The first section discusses the realization problem for linear systems and some basic results associated with it. Much of the material is based on Silverman [S-1] and Ho and Kalman [H-2]. Relevant discussions of the realization problem may also be found in Kalman [K-3] and Kalman, Falb, and Arbib [K-4].

The second section is devoted to developing the concepts associated with Hilbert spaces of random variables and the Hida-Cramer representation for purely nondeterministic stochastic processes. Basic references here are Wiener and Masani [W-1] and Cramer [C-2]. Material from Lee and Mandrekar [L-1] is also used.

#### 1. The Realization Problem

Since the thesis is concerned only with continuous-time systems, the discussion will be primarily in terms of continuous-time systems. Consider the linear continuous-time system representation

$$\frac{dx(t)}{dt} = F(t)x(t) + G(t)u(t) \quad (2-1)$$

$$y(t) = H(t)x(t) \quad (2-2)$$

where  $x(t)$  is an  $n \times 1$  vector,  $y(t)$  an  $m \times 1$  vector,  $u(t)$  an  $r \times 1$  deterministic vector, and  $F(t)$ ,  $G(t)$  and  $H(t)$  are  $n \times n$ ,  $n \times r$ , and  $m \times n$  matrices respectively. To avoid distracting mathematical complications it will be assumed here that the components of the matrices  $F(t)$ ,  $G(t)$ , and  $H(t)$  are continuous functions of time  $t$ . The system representation given by equations (2-1) and (2-2) will also be denoted by the subscripted triple  $(F(t), G(t), H(t))_n$ .

The solution to equation (2-1) for an initial state  $x_0$  at time  $t_0$  is given by (Weiss and Kalman [W-2])

$$x(t) = \phi(t, t_0)x_0 + \int_{t_0}^t \phi(t, \tau)G(\tau)u(\tau)d\tau \quad (2-3)$$

where the state transition matrix  $\phi(t, \tau)$  is defined by the relationship

$$\phi(t, \tau) = X(t)X^{-1}(\tau) \quad (2-4)$$

and  $X(t)$ , the fundamental matrix for the matrix  $F(t)$  satisfies

$$\frac{dX(t)}{dt} = F(t)X(t) \quad (2-5)$$

From equations (2-2) and (2-3)

$$y(t) = H(t)\phi(t, t_0)x_0 + H(t)\int_{t_0}^t \phi(t, \tau)G(\tau)u(\tau)d\tau. \quad (2-6)$$

By defining

$$K(t, \tau) = H(t)\phi(t, \tau)G(\tau) \quad (2-7)$$

equation (2-6) can be written

$$y(t) = H(t)\phi(t, t_0)x_0 + \int_{t_0}^t K(t, \tau)u(\tau)d\tau. \quad (2-8)$$

The matrix  $K(t, \tau)$  is called the weighting pattern matrix for the system (2-1) and (2-2). For zero initial conditions, i.e.,  $x_0 = 0$ , equation (2-8) shows that the weighting pattern matrix  $K(t, \tau)$  contains all the information needed to describe the input-output relations of the system  $(F(t), G(t), H(t))_n$ .

For a causal system, a system characterized by the property that the behavior of the system at any time  $\bar{t}$  is completely determined by events at times  $t \leq \bar{t}$ , the impulse response matrix  $W(t, \tau)$  is defined by

$$W(t, \tau) = \begin{cases} H(t)\phi(t, \tau)G(\tau) & t \geq \tau \\ 0 & t < \tau \end{cases}. \quad (2-9)$$

The impulse response matrix completely characterizes the input-output properties of  $(F(t), G(t), H(t))_n$ , for zero initial conditions, viewed as a causal system. For the case when  $K(t, \tau)$  is analytic, as when  $F$ ,  $G$ , and  $H$  are constant matrices, the impulse response matrix and

weighting pattern matrix are equivalent representations since  $W(t, \tau)$  will have a unique analytic extension equal to  $K(t, \tau)$  on the region  $\tau > t$ . For the general case, of course,  $W(t, \tau)$  may have many different possible extensions and  $K(t, \tau)$  is not an equivalent representation to  $W(t, \tau)$ . The problem of computing a state space representation  $(F(t), G(t), H(t))_n$  from an impulse response or weighting pattern matrix is known as the realization problem.\* A more formal and precise definition is given below after reviewing the concept of equivalent system representation.

In general there is not a unique solution to a realization problem. Different realizations of the same impulse response or weighting pattern matrix may have quite different characteristics. The concept of equivalent representations is thus necessary to fully understand the realization problem. A system  $(F(t), G(t), H(t))_n$  is algebraically equivalent to a system  $(\bar{F}(t), \bar{G}(t), \bar{H}(t))_n$  if and only if there exists a continuously differentiable matrix  $T(t)$  with  $\det T(t) \neq 0$  for all  $t \in \mathbb{R}$  such that

$$\bar{F}(t) = T(t)F(t)T^{-1}(t) + \frac{dT(t)}{dt} T^{-1}(t) \quad (2-10)$$

$$\bar{G}(t) = T(t)G(t) \quad (2-11)$$

$$\bar{H}(t) = H(t)T^{-1}(t) \quad (2-12)$$

---

\*In the stationary case one could also start with a transfer function matrix. These will play no role in the thesis except in a tangential manner in chapter 4, and will not be discussed here.

The system  $(F(t), G(t), H(t))_n$  is said to be strictly equivalent to  $(F(t), G(t), H(t))_n$  if (2-10) - (2-12) hold for a constant nonsingular matrix  $T$ . In general algebraic equivalence does not preserve the stability properties of a system. For this it is necessary to have topological equivalence (Kalman [K-3]): algebraic equivalence plus the conditions

$$\|T(t)\| \leq C_1 \quad (2-13)$$

$$\|T^{-1}(t)\| \leq C_2 \quad (2-14)$$

for fixed constants  $C_1$  and  $C_2$ , where  $\|\cdot\|$  is the Euclidean norm. It is clear that strict equivalence implies topological equivalence. It may easily be verified that algebraically equivalent systems have the same impulse response or weighting pattern matrix. However systems that are not algebraically equivalent systems can also have the same impulse response or weighting pattern matrix. In general, systems having the same impulse response, or weighting pattern matrix, are called W-equivalent, or K-equivalent, respectively. A formal definition of the realization problem can now be given. A system representation  $(F(t), G(t), H(t))_n$  is said to be an  $n$ th-order impulse response realization if  $W(t, \tau)$  can be expressed in the form of equation (2-9). If for some finite  $n$  such a system representation exists  $W(t, \tau)$  is said to be a realizable impulse response matrix.

The representation  $(F(t), G(t), H(t))_n$  is said to be a minimal impulse response matrix realization if there exists no  $W$ -equivalent system  $(\bar{F}(t), \bar{G}(t), \bar{H}(t))_{\bar{n}}$  with  $\bar{n} < n$ . Weighting pattern realizability is defined analogously.

There is, in general, a difference between the weighting pattern and impulse response realization problems in that even if  $W(t, \tau)$  and  $K(t, \tau)$  agree on the domain  $t \geq \tau$ , they may not give rise to equivalent realizations. Indeed, the respective minimal realizations may even be of different dimension. An example of this case is presented in a paper by Desoer and Varaiya [D-1]. However as already mentioned when the weighting pattern matrix  $K(t, \tau)$  is analytic then the impulse response matrix  $W(t, \tau)$  has a unique analytic extension to the region  $t < \tau$  equal to  $K(t, \tau)$ . It follows that for this case it makes no difference whether the impulse response or weighting pattern matrix is used, the same class of realizations will be obtained. This dissertation will be concerned solely with systems generating analytic impulse response or weighting pattern matrices. The impulse response matrix will be used primarily but the results would be unchanged if the weighting pattern matrix were chosen.

The following result due to Kalman [K-3] is basic in realization theory.

Theorem 2-1:

$W(t, \tau)$  is a realizable impulse response matrix

if and only if it is separable in the form

$$W(t, \tau) = \psi(t)\theta(\tau), \quad t \geq \tau \quad (2-15)$$

where  $\psi(t)$  and  $\theta(\tau)$  are continuous matrices of finite size.

Proof:

From equations (2-4) and (2-9) the impulse response matrix  $W(t, \tau)$  of the system  $(F(t), G(t), H(t))_n$  can be expressed, for  $t \geq \tau$ , as

$$W(t, \tau) = [H(t)X(t)][X^{-1}(\tau)G(\tau)] \quad (2-16)$$

showing the necessity of equation (2-14). The sufficiency of equation (2-15) follows from the obvious realization

$$(0, \theta(t), \psi(t)). \quad (2-17)$$

Q.E.D.

Although theorem 2-1 is of theoretical importance, it may not work well for practical applications. There is first the problem of determining if a function of two variables is separable. Also, even if  $W(t, \tau)$  is time invariant or periodic the realization given by equation (2-17) will not be. Furthermore the representation given by (2-17) will never be asymptotically stable. Silverman [S-1] gives a nice example to illustrate these points. Suppose

$$W(t, \tau) = \begin{cases} e^{-(t-\tau)} & t \geq \tau \\ 0 & t < \tau \end{cases} \quad (2-18)$$

The realization given by equation (2-17) is then

$$(0, e^t, e^{-t}) \quad (2-19)$$

which is time variable, has an unbounded coefficient, and is not asymptotically stable. The realization

$$(-1, 1, 1) \quad (2-20)$$

is much nicer.

Some material from Ho and Kalman [H-2] pertaining to stationary systems will be presented next. Consider the corresponding discrete time and continuous time stationary systems

$$x(t+1) = Fx(t) + Gu(t) \quad (2-21)$$

$$y(t) = Hx(t), \quad t \text{ an integer} \quad (2-22)$$

and

$$\frac{dx(t)}{dt} = Fx(t) + Gu(t) \quad (2-23)$$

$$y(t) = Hx(t), \quad t \text{ a real number,} \quad (2-24)$$

where  $x(t)$  is  $n \times 1$ ,  $y(t)$  is  $m \times 1$ , and  $u(t)$  is an  $r \times 1$  vector. The stationary impulse response matrix for the system in (2-21) and (2-22) will be denoted by  $Y_k$  and is given by

$$Y_k = HF^kG, \quad k = 0, 1, 2, \dots \quad (2-25)$$

For the system given by (2-23) and (2-24) the stationary impulse response matrix  $W(t)$  is given by

$$W(t) = H e^{Ft} G \quad (2-26)$$

$$\begin{aligned}
 &= H \left[ \sum_{k=0}^{\infty} F^k \frac{t^k}{k!} \right] G \\
 &= \sum_{k=0}^{\infty} H F^k G \frac{t^k}{k!} \\
 &= \sum_{k=0}^{\infty} Y_k \frac{t^k}{k!} .
 \end{aligned} \quad (2-27)$$

Hence, in the case of either a stationary discrete time system or stationary continuous time system the realization problem is equivalent to finding a constant system  $(F, G, H)_n$  with the property that

$$Y_k = H F^k G, \quad k = 0, 1, 2, \dots \quad (2-28)$$

The  $Y_k$  are called the Markov parameters of the system. The  $N_1$  by  $N_2$  block matrix built out of the Markov parameters

$$S_{N_1 N_2} = \begin{bmatrix} Y_0 & Y_1 & \cdot & \cdot & \cdot & Y_{N_2-1} \\ Y_1 & Y_2 & \cdot & \cdot & \cdot & Y_{N_2} \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ Y_{N_1-1} & Y_{N_1} & \cdot & \cdot & \cdot & Y_{N_1+N_2-2} \end{bmatrix} \quad (2-29)$$

is called the generalized Hankel matrix. Note that each  $Y_k$  is an  $m \times r$  matrix so  $S_{N_1 N_2}$  is an  $N_1 \cdot m \times N_2 \cdot r$  matrix. The generalized Hankel matrix will play an important role in chapter 4.

Note, finally, that from equation (2-26) a stationary impulse response matrix  $W(t)$  can have a realization of the form (2-23) and (2-24) only if it is analytic.

## 2. Hilbert Spaces of Random Vectors

Let  $(\Omega, A, P)$  be a probability space. The set  $\Omega$  is a set of possible outcomes,  $A$  is a  $\sigma$ -algebra defined on  $\Omega$ , and  $P$  is a probability measure defined on  $A$ . The set  $\Omega$  is called the basic space and its elements are called points. Members of  $A$  are called events.

Let  $\Theta$  be a basic space and  $F$  a  $\sigma$ -algebra of subsets of  $\Theta$ . Then the pair  $(\Theta, F)$  is called a measurable space. If  $\Theta$  is  $n$ -dimensional Euclidean space, denoted  $R^n$ , and  $F$  the smallest  $\sigma$ -algebra of subsets of  $R^n$  containing all rectangles, denoted  $B^n$ , then  $B^n$  is called the  $n$ -dimensional Borel  $\sigma$ -algebra and the sets in  $B^n$  are called Borel sets. A measurable function from a probability space  $(\Omega, A, P)$  into  $(R^n, B^n)$  is called a random variable. For  $n = 1$  such a random variable is called a real random variable. This will be the only case considered in this work and henceforth the term random variable should be interpreted to mean a real random variable.

A second-order random variable  $X$  is one which satisfies

$$E\{X^2\} = \int_{\Omega} |X(\omega)|^2 P(d\omega) < \infty. \quad (2-30)$$

A second-order stochastic process  $\{X(t), t \in T\}$  is a one-parameter family of second-order random variables. The set of all second-order random variables defined on  $(\Omega, A, P)$  is denoted by  $L_2(\Omega, A, P)$ . The set  $L_2(\Omega, A, P)$  is a Hilbert space (Gikhman and Skorokhod [G-3]) with the inner product

$$(x, y)_{L_2} = E\{xy\} \quad (2-31)$$

and norm

$$\|x\|_{L_2} = \sqrt{(x, x)} . \quad (2-32)$$

Let  $L_2^n(\Omega, A, P)$  be the set of all random  $n$  vectors  $x$  with components

$$x_i \in L_2(\Omega, A, P), \quad i = 1, \dots, n .$$

Hence  $x \in L_2^n(\Omega, A, P)$  if and only if

$$\int_{\Omega} |x(\omega)|^2 P(d\omega) < \infty \quad (2-33)$$

where  $|\cdot|$  denotes the Euclidean norm

$$|x|^2 = x^T x = \sum_{i=1}^n |x_i|^2 . \quad (2-34)$$

$L_2^n(\Omega, A, P)$  is a Hilbert space under the usual operations and inner product

$$\begin{aligned} (x, y)_{L_2^n} &= \sum_{i=1}^n (x_i, y_i)_{L_2} \\ &= \int_{\Omega} \sum_{i=1}^n x_i(\omega) y_i(\omega) P(d\omega) . \end{aligned} \quad (2-35)$$

This inner product generates the norm

$$\|x\|_{L_2^n} = \sqrt{(x, x)_{L_2^n}} = \left( \sum_{i=1}^n \|x_i\|_{L_2}^2 \right)^{\frac{1}{2}} \quad (2-36)$$

which in turn induces a topology on  $L_2^n(\Omega, A, P)$ . A sequence  $\{x_k\}$  in  $L_2^n(\Omega, A, P)$  is said to converge to a vector  $x \in L_2^n(\Omega, A, P)$  if and only if

$$\|x_k - x\|_{L_2^n} \rightarrow 0 \quad (2-37)$$

as  $n \rightarrow \infty$  and to be a Cauchy sequence if and only if

$$\|x_k - x_m\|_{L_2^n} \rightarrow 0 \quad (2-38)$$

as  $k, m \rightarrow \infty$ . Two elements  $x$  and  $y$  are regarded as identical if the norm of their difference is zero, that is,  $E\{[x-y]^T[x-y]\} = 0$ , and thus  $P(x = y) = 1$ . Convergence in norm in the Hilbert space  $L_2^n(\Omega, A, P)$  is identical with convergence in quadratic mean in probabilistic terminology.

A linear manifold in  $L_2^n(\Omega, A, P)$  is a nonempty subset  $M$  of  $L_2^n(\Omega, A, P)$  such that if  $x, y \in M$  then  $Ax + By \in M$  for all  $n \times n$  real matrices  $A$  and  $B$ . A subspace of  $L_2^n(\Omega, A, P)$  is a linear manifold closed under the topology induced by the norm  $\|\cdot\|_{L_2^n}$ . A lemma due to Wiener and Masani [W-1] is now given which proves the existence of the projection of an element  $x$  and gives its structure. The space  $L_2^n(\Omega, A, P)$  is abbreviated  $L_2^n$ .

Orthogonal Projection Lemma (Wiener-Masani):

- a)  $M$  is a subspace of  $L_2^n$  if and only if there is a subspace  $M$  of  $L_2$  such that  $M = M^n$ , where  $M^n$  denotes the Cartesian product  $M \times \dots \times M$  with  $n$  factors.  $M$  is the set of all components of all functions in  $M$ .
- b) If  $M$  is a subspace of  $L_2^n$  and  $x \in L_2^n$ , then there exists a unique  $\hat{y}$  such that

$$\hat{y} \in M; \|x - \hat{y}\|_{L_2^n} \leq \|x - y\|_{L_2^n}, \text{ for all } y \in M. \quad (2-39)$$

An element  $\hat{y} \in L_2^n$  satisfies the preceding condition if and only if  $E\{[y - \hat{y}]x^T\} = 0$  for all  $x \in L_2^n$ , i.e.,  $y - \hat{y}$  is orthogonal to  $L_2^n$ , denoted  $y - \hat{y} \perp L_2^n$ . The unique element  $\hat{y}$  of the lemma is called the orthogonal projection of  $y$  on  $L_2^n$  and will be denoted by  $(y|L_2^n)$ . For this  $\hat{y}$ ,  $\hat{y}_i = (y_i|L_2)$ , the orthogonal projection of  $y_i$  on to  $L_2$ .

Consider now the second-order  $n$ -variate stochastic vector  $x(t) \in L_2^n(\Omega, A, P)$ . The present of the process  $x(t)$  is defined as  $L_2^n(x(t))$ , where  $L_2(x(t))$  is the subspace of  $L_2(\Omega, A, P)$  generated by  $\{x_i(t), i = 1, 2, \dots, n\}$ . Similarly the past of the process up to  $t$  is defined by  $L_2^n(x; t)$  where  $L_2(x; t)$  is the subspace of  $L_2(\Omega, A, P)$  generated by  $\{x_i(s), s \leq t, i = 1, 2, \dots, n\}$ .

To conclude this section some aspects of the Hida-Cramer theory on multiplicity and representation will be reviewed (see Cramer [C-2], [C-3], and Lee and Mandrekar

[L-1]). A second-order  $n$ -variate process  $\{x(t), a \leq t \leq b\}$  is called purely nondeterministic if  $\bigcap_{a \leq t \leq b} L_2(x;t)$  contains only the zero element of  $L_2(x;t)$ , i.e., the random variable that is almost everywhere equal to zero. Denote by  $P_t$  the projection operator from  $L_2(x;b)$  onto  $L_2(x;t)$  and assume that  $x(t)$  is zero mean, purely nondeterministic, and continuous in quadratic mean. For any set of mutually orthogonal random variables  $\{z_1, \dots, z_M\} \subset L_2(x;b)$  the  $M$ -variate process  $z(t)$ , considered as a column vector with components  $z_i(t)$  defined by

$$\begin{aligned} z_i(t) &= P_t z_i \\ &= (z_i | L_2(x;t)) \end{aligned} \quad (2-40)$$

will be an orthogonal increment process in that two increments  $\Delta z_i(t)$  and  $\Delta z_j(s)$  are always orthogonal if either  $i \neq j$  or corresponding time intervals are disjoint. The set of random variables  $\{z_1, \dots, z_M\}$  can be chosen [K-5] so that for all  $t \in [a, b]$

$$L_2(x;t) = L_2(z;t). \quad (2-41)$$

The process  $z(t) = (z_1(t), \dots, z_M(t))^T$  is called an innovation process for the  $x(t)$  process. The number  $M$  may be infinite. Define the  $M \times M$  matrix time-spectral function  $F(t)$  by

$$F(t) = E\{z(t)z^T(t)\}. \quad (2-42)$$

For a mean square continuous process  $x(t)$ , with orthogonal increments  $z(t)$ ,  $F(t)$  is continuous and nondecreasing which implies that  $F(t)$  is differentiable almost everywhere. Let  $L_2(F)$  denote the Hilbert space of all row-vector functions  $g(t)$  satisfying

$$\int_a^b g(t) dF(t) g(t)^T < \infty \quad (2-43)$$

with inner product

$$(g, h)_F = \int_a^b g(t) dF(t) h(t)^T. \quad (2-44)$$

Then there occurs the Hida-Cramer representation

$$x(t) = \int_a^t G(t, s) dz(s) \quad (2-45)$$

where  $G(t, s)$  is an  $n \times M$  matrix function with each row  $g_i(t, s) \in L_2(F)$ ,  $i = 1, \dots, n$ . The innovation  $z(t)$  is not uniquely determined but its dimensionality  $M$  is unique and is the smallest number for which (2-45) holds. The number  $M$  is called the multiplicity of  $x(t)$ .

Finally it should be noted that for a zero mean second-order quadratic mean process  $x(t)$  the space  $L_2(x; t)$  is separable. A proof of this may be found in Cramer ([C-3], p. 253).

CHAPTER THREE  
MARKOVIAN REPRESENTATION OF CONTINUOUS-TIME  
STOCHASTIC SYSTEMS

In this chapter the following problem is considered. Given two real vector valued continuous-time zero mean second-order stochastic processes  $y(t)$  and  $z(t)$ , with  $y(t)$  measurable and  $z(t)$  continuous in quadratic mean and purely nondeterministic, find a representation of  $y(t)$  as the output of a linear continuous time system driven by the innovations process of  $z(t)$ . The concept of the predictor space, the Hilbert space spanned by the linear least square predictions of  $y(t)$  given  $z(r)$ ,  $t_0 \leq r \leq t$ , will be used to derive a solution to this problem under appropriate assumptions. Essentially the basic idea is that if the predictor space has a finite basis, this basis can be used as the state vector of a linear system generating  $y(t)$ .

After developing a general theory two special situations will be considered. The first, called the white noise case, will look at the representation when  $z(t)$  is white noise. A minor extension to the general theory will be necessary since white noise is neither second-order nor continuous in quadratic mean. The white noise case will be useful for applications.

The second special case of interest occurs when  $y(t)$  and  $z(t)$  are stationary processes with stationary cross covariance. The development of the properties of the representation for this case will conclude the chapter.

### 1. General Theory

Consider two zero-mean real valued second-order stochastic processes  $y(t)$  and  $z(t)$  with covariance\*

$$V_{yz}(t,s) = E\{y(t)z^T(s)\} \quad (3-1)$$

The vectors  $y$  and  $z$  are  $m \times 1$  and  $r \times 1$  respectively. It is also required that  $y(t)$  be measurable and  $z(t)$  be continuous in quadratic mean and purely nondeterministic.

Denote by  $L_2(z;s)$  the Hilbert space obtained by taking the closure with respect to the mean square norm of the linear manifold spanned by the components of  $z(r)$ ,  $r \leq s$ . The predictors, denoted  $\hat{y}(t|s)$ , are defined for  $t \geq s$  to be the vector formed by projecting the components of  $y(t)$  onto  $L_2(z;s)$ . Letting  $\hat{y}_i(t|s)$  be a component of  $\hat{y}(t|s)$  with  $y_i(t)$  the corresponding component of  $y(t)$

$$\hat{y}_i(t|s) = (y_i(t)|L_2(z;s)), \quad i = 1, \dots, m, \quad s \leq t. \quad (3-2)$$

---

\*The covariance actually plays no role in this section. It will, however, be useful in section 2 and crucial for the applications in later chapters.  $E\{\cdot\}$  indicates the expected value of the quantity in brackets while  $(\cdot)^T$  indicates the transpose.

Equivalently

$$\hat{y}(t|s) = (y(t)|L_2^m(z;s)), \quad s \leq t. \quad (3-3)$$

The predictors are specified by the relationships

$$E\{[\hat{y}(t|s) - y(t)]z^T(r)\} = 0, \quad r \leq s \leq t, \quad (3-4)$$

or

$$E\{\hat{y}(t|s)z^T(r)\} = V_{yz}(t,r). \quad (3-5)$$

The predictor space  $L_2(\hat{y};t|s)$  is defined as the Hilbert space obtained by taking the closure with respect to the mean square norm of the linear manifold spanned by the components of the predictors  $\hat{y}(t|s)$  for  $t \geq s$ . In forming this space  $s$  is to be regarded as fixed while  $t$  varies over all values greater than or equal to  $s$ . It should be noted that  $L_2(\hat{y};t|s)$  is a subspace of  $L_2(z;s)$ .

Following the example of Akaike [A-1] the assumption is made that the predictor space is of finite dimension. Since  $L_2(\hat{y};t|s)$  is formed by projecting components of  $y(t)$  onto  $L_2(z;s)$  with  $s$  fixed while allowing  $t$  to vary over all values greater than  $s$ , a finite dimensional basis, consisting of  $n$  random variables, can be regarded as a function of  $s$ .<sup>\*</sup> The basis, of dimension  $n$ , will be

---

<sup>\*</sup>The basis could also be regarded as being of the form  $\{x_1(t_1), x_2(t_2), \dots, x_n(t_n)\}$  where  $x_i(t_i)$  is an element of  $L_2(\hat{y};t|s)$ ,  $t_i \geq s$ . The point is, for fixed  $s$  the basis can be fixed and changes only as  $s$  varies.

denoted by the vector

$$x(s) = (x_1(s), x_2(s), \dots, x_n(s))^T. \quad (3-6)$$

Letting  $L_2(x(s))$  be the Hilbert space generated by taking the closure with respect to the mean square norm of the linear manifold spanned by the components of  $x(s)$  it is obvious that

$$L_2(\hat{y}; t|s) = L_2(x(s)) . \quad (3-7)$$

Theorem 3-1:

Let  $s$  and  $t$  be fixed with  $s \leq t$ . Then

$$\hat{y}(t|s) = A(t,s)x(s) \quad (3-8)$$

for some  $m \times n$  matrix  $A(t,s)$ .

Proof:

Since  $x(s)$  is a finite dimensional basis for  $L_2(\hat{y}; t|s)$  then

$$\hat{y}(t|s) \in L_2(\hat{y}; t|s) \quad (3-9)$$

implies that

$$\hat{y}(t|s) = Ax(s) \quad (3-10)$$

for some matrix  $A$ . Since  $\hat{y}$  varies with  $t$  and  $s$  then  $A$  must also, so equation (3-8) is justified.

Q.E.D.

Corollary:

Let  $s$  and  $t$  be fixed with  $t \geq s$ . Then

$$y(t) = A(t,s)x(s) + u(t) \quad (3-11)$$

where  $u(t)$  is an  $m \times 1$  vector orthogonal to  $L_2^m(z;s)$ .

Proof:

Set

$$u(t) = y(t) - \hat{y}(t|s). \quad (3-12)$$

Then for  $r \leq s \leq t$

$$\begin{aligned} E\{u(t)z^T(r)\} &= E\{[y(t) - \hat{y}(t|s)]z^T(r)\} \\ &= 0 \end{aligned} \quad (3-13)$$

by equation (3-4). Hence  $u(t)$  is orthogonal to  $L_2^m(z;s)$  and combining equations (3-8) and (3-12) equation (3-11) is obtained.

Q.E.D.

The next goal will be to develop the dynamic behavior of  $x(t)$ . The following results are directed to this end.

Lemma 3-1:

Let  $\{X_n\}$  be a sequence of scalar random variables converging in quadratic mean to the process  $X$ . Then

$$\lim_{n \rightarrow \infty} E\{X_n\} = E\{X\}. \quad (3-14)$$

Proof:

Recall the Schwarz inequality

$$|E\{XY\}|^2 \leq E\{X^2\}E\{Y^2\}. \quad (3-15)$$

Let  $Y = 1$ . Then

$$|E\{[X_n - X]Y\}|^2 \leq E\{[X_n - X]^2\} \cdot E\{Y^2\}. \quad (3-16)$$

That is

$$|E\{X_n - X\}|^2 \leq E\{[X_n - X]^2\}. \quad (3-17)$$

But by hypothesis

$$\lim_{n \rightarrow \infty} E\{[X_n - X]^2\} = 0 \quad (3-18)$$

Hence

$$\lim_{n \rightarrow \infty} |E\{X_n - X\}|^2 = 0 \quad (3-19)$$

which implies

$$\lim_{n \rightarrow \infty} [E\{X_n\} - E\{X\}]^2 = 0. \quad (3-20)$$

Equation (3-20) is equivalent to equation (3-14) so the

lemma is proved.

Q.E.D.

Corollary:

$x(t)$  is a zero mean random variable.

Proof:

By definition each component  $x_i(s)$  of the  $n \times 1$  vector process  $x(s)$  is an element of  $L_2(z;s)$ . Hence  $x_i(t)$  is the limit in quadratic mean of a sequence of finite linear combinations of components of  $z$ . But  $z$  is a zero mean stochastic process so finite linear combinations of components of  $z$  have zero mean. Hence  $x_i(s)$  is a limit in quadratic mean of a sequence of zero mean

processes. Thus by lemma 3-1,  $x_i(s)$  and hence  $x(s)$  have zero means.

Q.E.D.

It is appropriate at this point to make a few remarks concerning measurability since that will be the next property of  $x(t)$  to be considered. Let  $T$  be a Lebesgue measurable parameter set on which is defined a  $\sigma$ -algebra of Lebesgue measurable sets  $L$ . Let  $\Omega$  be a probability space on which is defined a  $\sigma$ -algebra of events  $A$ . The symbol  $L \otimes A$  denotes the smallest  $\sigma$ -algebra defined\* on  $T \times \Omega$  including all sets of the form  $L \times A$  such that  $L \in L$  and  $A \in A$ . Then a process  $\{X(t), t \in T\}$  is said to be a measurable process\*\* if  $X(t, \omega)$  is a  $(t, \omega)$  function measurable with respect to  $L \otimes A$ . The assumption at the beginning of the chapter that  $y(t)$  is measurable is in the sense that its components satisfy this definition.

Lemma 3-2:

$x(t)$  is a measurable second-order random vector.

Proof:

Since a process continuous in quadratic mean has a measurable modification,  $z(t)$  can be taken to be a measurable process. Then  $L_2(z; t)$  is a closed subspace of  $L_2(M)$ , the space of square integrable measurable

---

\*  $A \times B$  denotes the Cartesian product of the sets  $A$  and  $B$ .

\*\* The product measure defined on  $L \otimes A$  is, of course, itself a measure (Rudin [R-1], p. 140).

functions defined on  $M = [t_0, t] \times \Omega$ , where  $\Omega$  is the probability space on which  $\{z(r): t_0 \leq r \leq t\}$  is defined. It is known that  $L_2(M)$  is complete (Rudin [R-1], p. 66), and since a closed subspace of a complete space is complete (Simmons [S-4], p. 73),  $L_2(z;t)$  is complete. Hence every convergent sequence of measurable square integrable functions in  $L_2(z;t)$  converges to a measurable square integrable function in  $L_2(z;t)$ . Since  $x_i(t)$  is the limit in the mean square norm of a sequence of such functions,  $x_i(t)$  is measurable and square integrable. Hence  $x(t)$  is measurable and square integrable. Q.E.D.

It is next necessary to review the concept of weak convergence which will play a key role in the development. A sequence of random vectors  $x_n$  in a separable Hilbert space  $H$  is said to converge weakly to the vector  $x$  in  $H$ , denoted  $x_n \xrightarrow{W} x$ , if for every  $y$  in  $H$

$$\lim_{n \rightarrow \infty} E\{y^T x_n\} = E\{y^T x\}. \quad (3-21)$$

A key lemma for determining weak convergence is the following (Parzen [P-3], p. 272).

Lemma 3-3:

A sequence of random vectors  $x_n$  defined as members of some separable Hilbert space of random vectors converges weakly to a vector  $x$  if and only if the following two conditions hold:

- (i) For some constant  $M$ ,

$$E\{x_n^T x_n\} \leq M, \quad n = 1, 2, \dots \quad (3-22)$$

(ii) For some family of vectors  $\{k(t), t \in T\}$  which spans the Hilbert space spanned by  $\{x_n, n = 1, 2, \dots\}$

$$\lim_{n \rightarrow \infty} E\{x_n^T k(t)\} \quad (3-23)$$

exists for every  $t \in T$ .

The following theorem and corollary due to Halyo and McAlpine [H-1] are useful in developing the dynamic behavior of  $x(t)$ . The proof of the theorem as given by Halyo and McAlpine is presented since it will be necessary in further developments to refer to it. Except for changes in notation, the following theorem, proof, and corollary are directly from [H-1]. The separability of the spaces  $L_2(z;t)$  and its subspaces  $L_2(x;t)$  and  $L_2(x(t))$ , although not explicitly mentioned, is necessary to allow the use of weak convergence.

Theorem 3-2:

Let  $\{x(t): t_0 \leq t\}$  and  $\{z(t): t_0 \leq t\}$  be second order measurable vector random processes with zero mean such that  $L_2(x;t)$  is a subspace of  $L_2(z;t)$  for each  $t$ ,  $\frac{\partial V_{xz}(t,s)}{\partial t}$  is continuous on  $\{(t,s): t_0 \leq s \leq t\}$ , and

$$\{E|([x_i(t) - x_i(s)]|L_2(z;s))|^2\}^{\frac{1}{2}} \leq B|t-s|, \quad s \leq t \quad (3-24)$$

then  $x(t)$  can be written as

$$x(t) = x(t_0) + \int_{t_0}^t w(s)ds + U(t), \quad t_0 \leq t \quad (3-25)$$

where  $U(t)$  is a process with orthogonal increments such that  $U(t) - U(s)$  is uncorrelated to  $z(r)$  whenever  $r \leq s \leq t$ , and  $U(t_0)$  vanishes;  $w(t)$  and  $U(t)$  belong to  $L_2(z; t)$  and

$$V_{wz}(t, s) = \frac{\partial V_{xz}(t, s)}{\partial t}, \quad s \leq t. \quad (3-26)$$

Proof:

Let  $t$  be the smaller of  $t_1$  and  $t_2$  and  $s \leq t$ ; then

$$E \left\{ \left[ \frac{x(t_1) - x(t_2)}{t_1 - t_2} \right] z^T(s) \right\} = E \left\{ \left( \frac{x(t_1) - x(t_2)}{t_1 - t_2} \middle| L_2^n(z; t) \right) z^T(s) \right\} \quad (3-27)$$

$$= \frac{V_{xz}(t_1, s) - V_{xz}(t_2, s)}{t_1 - t_2}. \quad (3-28)$$

The right side of (3-28) converges as  $t_2$  tends to  $t_1$ ;

hence by lemma 3-3,  $\left( \frac{x(t_1) - x(t_2)}{t_1 - t_2} \middle| L_2^n(z; t) \right)$  converges weakly to, say,  $w(t_1)$  in  $L_2(z; t_1)$  since its norm is bounded by  $B$ . Define the process  $U(t)$  as

$$U(t) = x(t) - x(t_0) - \int_{t_0}^t w(s) ds, \quad t_0 \leq t. \quad (3-29)$$

First note that, using the definition of  $z(t)$ , (3-26) follows from (3-28). Now,

$$\int_s^t E\{w(r) z^T(t')\} dr = E\{[x(t) - x(s)] z^T(t')\}, \quad t' \leq s \leq t. \quad (3-30)$$

Interchanging integration and expectation, and using the properties of projections, it can be shown that

$$\left(\int_s^t w(r)dr | L_2^n(z;s)\right) = ([x(t)-x(s)] | L_2^n(z;s)), \quad s \leq t \quad (3-31)$$

$$([U(t) - U(s)] | L_2^n(z;s)) = 0, \quad s \leq t. \quad (3-32)$$

Since  $U(t)$  belongs to  $L_2(z;t)$ ,  $U(t) - U(s)$  is orthogonal to  $L_2(z;s)$ ; and hence to  $U(t') - U(s')$ ,  $s' \leq t' \leq s$ . Therefore  $U(t)$  has orthogonal (i.e., uncorrelated) increments. Also note that since  $w(s)$  belongs to  $L_2(z;s)$ ,  $U(t) - U(s)$  is uncorrelated to  $L_2(w;s)$ ,  $s \leq t$ , and  $x(t_0)$ .

Q.E.D.

Corollary:

Theorem 3-2 remains valid if (3-24) is replaced by the condition that  $E\{|x_i(t)|^2\}$  is bounded on bounded intervals and

$$V_{xz}(t,s) = p(t)q(s) \quad (3-33)$$

where  $p(t)$  is an  $n \times l$  matrix function of  $t$  alone and  $q(s)$  is an  $l \times r$  matrix function of  $s$  alone.

Equation (3-25) will be the basis of the dynamic representation of  $x(t)$ . Note that the continuity of

$\frac{\partial V_{xz}(t,s)}{\partial t}$  and equation (3-24) among the hypotheses of Theorem 3-2 are new assumptions on  $x(t)$  and  $z(s)$ . The continuity of  $\frac{\partial V_{xz}(t,s)}{\partial t}$  is a restriction necessary for the desired representation to exist. Equation (3-24) can be replaced by the separability of  $V_{xz}(t,s)$  and the boundedness on bounded intervals of  $E\{|x_i(t)|^2\}$ , and

these conditions will be satisfied automatically in the applications to be encountered later.

The next task, which extends the work of Halyo and McAlpine in this context and plays the crucial role in the development of the representation theory, will be to show that  $w(t)$  is, in fact, an element of  $L_2^n(x(t))$ . For this the following lemma is necessary.

Lemma 3-4:

Let  $\{x(t): t_0 \leq t\}$  and  $\{z(t): t_0 \leq t\}$  be as in theorem 3-2. Then

$$\left( \frac{x(t_1) - x(t_2)}{t_1 - t_2} \middle| L_2^n(z; t_1) \right) \in L_2^n(x(t_1)), \quad t_0 \leq t_1 \leq t_2. \quad (3-34)$$

Proof:

Clearly

$$\begin{aligned} \left( \frac{x(t_1) - x(t_2)}{t_1 - t_2} \middle| L_2^n(z; t) \right) &= \frac{1}{t_1 - t_2} \left[ (x(t_1) \middle| L_2^n(z; t_1)) \right. \\ &\quad \left. - (x(t_2) \middle| L_2^n(z; t_1)) \right]. \end{aligned} \quad (3-35)$$

By definition  $x(t_1) \in L_2^n(z; t_1)$  so  $(x(t_1) \middle| L_2^n(z; t_1))$  equals  $x(t_1)$  and is an element of  $L_2^n(x(t_1))$ . The problem is to show that  $(x(t_2) \middle| L_2^n(z; t_1)) \in L_2^n(x(t_1))$ . Recall that  $x(t_2)$  is a basis for  $L_2(\mathfrak{Y}; t|t_2)$  and  $L_2(\mathfrak{Y}; t|t_2)$  is an  $n$ -dimensional space generated by linear combinations of components of  $\mathfrak{Y}(t|t_2)$  as  $t$  varies over all values greater than or equal to  $t_2$ . There must exist a set of predictors  $\{\mathfrak{Y}(\tau_1|t_2), \mathfrak{Y}(\tau_2|t_2), \dots, \mathfrak{Y}(\tau_k|t_2)\}$ ,

$\ell \leq n$ ,  $\tau_k \geq t_2$ , the set of whose components contains  $n$  linearly independent elements, that is, a basis for  $L_2(\mathfrak{Y}; t|t_2)$ . This basis can be in general different from  $x(t_2)$ . Also, it must be true that

$$x(t_2) = \sum_{j=1}^{\ell} C_j \mathfrak{Y}(\tau_j | t_2) \quad (3-36)$$

for some set of  $n \times m$  matrices  $C_j$ ,  $j = 1, \dots, \ell$ . This follows from the fact that any two bases of the predictor space are related by a nonsingular transformation. Now

$$(x(t_2) | L_2^n(z; t_1)) = \left( \sum_{j=1}^{\ell} C_j \mathfrak{Y}(\tau_j | t_2) | L_2^n(z; t_1) \right) \quad (3-37)$$

$$= \sum_{j=1}^{\ell} C_j ((y(\tau_j) | L_2^m(z; t_2)) | L_2^m(z; t_1))$$

$$= \sum_{j=1}^{\ell} C_j (y(\tau_j) | L_2^m(z; t_1)). \quad (3-38)$$

The last equality follows from the fact that  $L_2(z; t_1) \subset L_2(z; t_2)$  for  $t_1 \leq t_2$ . Continuing,

$$\begin{aligned} (x(t_2) | L_2^n(z; t_1)) &= \sum_{j=1}^{\ell} C_j \mathfrak{Y}(\tau_j | t_1) \\ &= \sum_{j=1}^{\ell} C_j A(\tau_j, t_1) x(t_1) \end{aligned} \quad (3-39)$$

and the last sum is an element in  $L_2^n(x(t_1))$ . This suffices to prove the lemma. Q.E.D.

Theorem 3-3:

Let  $w(t_1)$  be as defined in theorem 3-2. Then  $w(t_1) \in L_2^n(x(t_1))$ .

Proof:

Note that in the proof of theorem 3-2  $w(t_1)$  is defined as the weak limit of  $\left( \frac{x(t_1) - x(t_2)}{t_1 - t_2} | L_2^n(z; t_1) \right)$  in  $L_2^n(z; t_1)$  as  $t_2 \rightarrow t_1$  by lemma 3-3 since

$E \left\{ \left( \frac{x(t_1) - x(t_2)}{t_1 - t_2} | L_2^n(z; t_1) \right) z^T(s) \right\}$  exists for  $t_0 \leq s \leq t_1 \leq t_2$ , and since

$E \left\{ \left( \frac{x(t_1) - x(t_2)}{t_1 - t_2} | L_2^n(z; t_1) \right)^T \left( \frac{x(t_1) - x(t_2)}{t_1 - t_2} | L_2^n(z; t_1) \right) \right\}$  is bounded by virtue of equation (3-24). That

$E \left\{ \left( \frac{x(t_1) - x(t_2)}{t_1 - t_2} | L_2^n(z; t_1) \right) z^T(s) \right\}$  exists for  $t_0 \leq s \leq t_1 \leq t_2$

implies that  $E \left\{ \left( \frac{x(t_1) - x(t_2)}{t_1 - t_2} | L_2^n(z; t_1) \right) v^T(s) \right\}$  exists for any  $v(s) \in L_2^n(z; s)$ , and in particular for  $v(s) \in L_2^n(x(s))$ .

Setting  $s = t_1$ ,  $E \left\{ \left( \frac{x(t_1) - x(t_2)}{t_1 - t_2} | L_2^n(z; t_1) \right) v^T(t_1) \right\}$  exists for all  $v(t_1) \in L_2^n(x(t_1))$ . Since by lemma 3-4 both

factors in the expected value are elements of  $L_2^n(x(t_1))$

the weak limit must exist as an element of  $L_2^n(x(t_1))$ .

That is,  $w(t_1) \in L_2^n(x(t_1))$ .

Q.E.D.

Corollary:

Let  $x(t)$  and  $w(t)$  be as defined above. Then

$$w(t) = F(t)x(t), \quad t_0 \leq t. \quad (3-40)$$

Proof:

Since  $w(t)$  and  $x(t)$  are both elements of

$L_2(x(t)),$

$$W(t) = \lim_{n \rightarrow \infty} F_n x(t) \quad (3-41)$$

for some sequence of  $n \times n$  matrices  $F_n$ . Since the limit exists, the sequence of real valued matrices converges to a real valued matrix  $F(t)$  and equation (3-40) is justified. Q.E.D.

Combining equations (3-25) and (3-40) yields

$$x(t) = x(t_0) + \int_{t_0}^t F(s)x(s)ds + U(t), \quad t_0 \leq t. \quad (3-42)$$

The next item to be considered is how to relate  $U(t)$  to  $z(t)$ . By the assumption that  $z(t)$  is a quadratic mean continuous purely nondeterministic process, it can be expressed as (Cramer [C-2])

$$z(t) = \int_{t_0}^t K(t,s)dv(s) \quad (3-43)$$

where the  $k \times 1$  vector<sup>\*</sup>  $v(s)$  is the innovation process for  $z(t)$ . The innovation process has the important property that (Kallianpur and Mandrekar [K-5])

$$L_2(z;t) = L_2(v;t). \quad (3-44)$$

Hence  $U(t) \in L_2^n(z;t)$  implies  $U(t) \in L_2^n(v;t)$ . Thus  $U(t)$  can be expressed in terms of the innovation process  $z(t)$  by (Ivkovich and Rozanov [I-1])

---

<sup>\*</sup> $v(s)$  may be of countably infinite dimension.

$$U(t) = \int_{t_0}^t G(t,s)dv(s). \quad (3-45)$$

The orthogonal increment property of  $U(t)$  implies it is a wide-sense Martingale process (Doob [D-2], Lee and Mandrekar [L-1]). Since  $U(t)$  is wide-sense Martingale by theorem 3.3 of Lee and Mandrekar [L-1] it is seen that  $G(t,s)$  is independent of  $t$ . Hence

$$U(t) = \int_{t_0}^t G(s)dv(s). \quad (3-46)$$

Thus, combining equations (3-42) and (3-46)

$$x(t) = x(t_0) + \int_{t_0}^t F(s)x(s)ds + \int_{t_0}^t G(s)dv(s) \quad (3-47)$$

where  $v(s)$  is the innovation process for  $z(t)$ . Equation (3-47) can be written in the form

$$dx(t) = F(t)x(t)dt + G(t)dv(t). \quad (3-48)$$

Equations (3-47) and (3-48) are regarded as equivalent representations and have the solution (Mandrekar and Salehi [M-1])

$$x(t) = \phi(t, t_0)x(t_0) + \int_{t_0}^t \phi(t, r)g(r)dv(r) \quad (3-49)$$

where  $\phi(t, r)$  is the transition matrix defined by  $F(t)$ .

Equations (3-11) and (3-48) now yield the representation

$$dx(t) = F(t)x(t)dt + G(t)dv(t) \quad (3-50)$$

$$y(t) = A(t,s)x(s) + u(t) \quad (3-51)$$

Setting

$$H(t) = A(t, t) \quad (3-52)$$

and letting  $s = t$  in equation (3-51), the desired representation

$$dx(t) = F(t)x(t)dt + G(t)dv(t) \quad (3-53)$$

$$y(t) = H(t)x(t) + u(t) \quad (3-54)$$

is achieved.

The development to this point is summarized by the following theorem.

Theorem 3-4:

Let  $y(t)$  and  $z(t)$  be zero-mean real valued second-order  $m \times 1$  and  $r \times 1$  variate stochastic processes respectively and let  $y(t)$  be measurable and  $z(t)$  be continuous in quadratic mean and purely non-deterministic. Form the predictor space  $L_2(\hat{y}; t|s)$  by projecting the components of  $y(t)$  onto  $L_2(z; s)$  for  $t_0 \leq s \leq t$ . Assume that  $L_2(\hat{y}; t|s)$  has an  $n$  dimensional basis  $x(s)$ . Then if  $\frac{\partial V_{xz}(t, s)}{\partial t}$  is continuous for  $t_0 \leq s \leq t$ , and either  $V_{xz}(t, s)$  is separable and  $E\{|x_i(t)|^2\}$  is bounded on bounded intervals, or  $E\{|(x_i(t) - x_i(s)|_{L_2(z; s)})|^2\}^{\frac{1}{2}} < B|t-s|$  for  $s \leq t$ , the representation given by equation (3-53) and (3-54) can be obtained.

The next point that should be mentioned is the minimality of the dimension of the representation. The

representation is minimal in the sense that since  $x(t)$  is a basis of the predictor space, there does not exist a vector of smaller dimension than  $x(t)$  that contains all the information in the predictor space.

To close this section it seems appropriate to comment briefly on some of the assumptions used and their roles in the derivations. The requirements that  $z(t)$  be continuous in quadratic mean and purely nondeterministic are to ensure the existence of its innovation process. Continuity in quadratic mean may be replaced by the weaker condition that for all  $t$  the limit  $z(t+0)$  exists (Cramer [C-2]). The requirement that  $V_{xz}(t,s)$  be separable, equation (3-25), and the necessity of  $\frac{\partial V_{xz}(t,s)}{\partial t}$  being continuous are necessary that the system structure to be derived will exist. In the applications to the realization problem  $V_{yz}(t,s)$  will be seen to be equal to a deterministic impulse response matrix. It is well known (Kalman [K-3]) that an impulse response matrix has a system realization if and only if it is separable. Since

$$V_{yz}(t,s) = H(t)V_{xz}(t,s) \quad (3-55)$$

the requirement that  $V_{xz}(t,s)$  be separable is a natural one.

Perhaps the most critical assumption was that the predictor space have a finite dimensional basis of constant

dimension  $n$ . Both the finite dimensionality and constant dimensionality are worthy of comment. The finite dimensionality of  $L_2(\hat{y}; t|s)$  played no essential role in deriving equation (3-11). The derivation will go through with  $x(t)$  of countably infinite dimension. After that, mathematical intricacies begin to arise if  $x(t)$  is allowed to have infinite dimension, and since the applications will be for finite dimensional systems only the restrictions to the finite case was felt to be justified.

The question of constant dimension is a tricky one. The eventual concern of the thesis is with the realization problem and, as observed by Kalman (Kalman, Falb, and Arbib [K-4]), to develop a really comprehensive theory of realization the dimension of the state space should be allowed to vary with time. However for the types of systems to be considered in this thesis considerations of this nature need not arise. For more general time-varying systems than the analytic systems considered in this paper, however, the question of constant dimensionality is a significant one.

Finally it should be mentioned that the representation theory developed here is somewhat more general than will be necessary for applications. To be specific  $z(t)$  will always be a Wiener process. The derivation of the representation is not simplified significantly by making this assumption on  $z(t)$  however.

## 2. Special Cases

In using the representation developed in part 1 to solve the realization problem the input process will be an  $r \times 1$  vector zero-mean stationary white noise process. Since white noise is neither measurable nor second-order the theory developed in part 1 is not immediately applicable. The first consideration here will be to resolve this problem. Then some results pertaining to stationary processes will be developed.

Given a zero mean white Gaussian process  $\eta(t)$  set

$$z(t) = \int_{t_0}^t \eta(s) ds. \quad (3-56)$$

Then (3-56) is actually the innovation representation (Kailath [K-1]) for the Wiener process  $z(t)$  which is clearly a second-order purely nondeterministic process continuous in quadratic mean. Hence the theory of part 1 applies. The vector  $dv(s)$  (see equation (3-43)) can be replaced by  $\eta(s)ds$  with the interpretation

$$dz(s) = \eta(s)ds. \quad (3-57)$$

Equations (3-53) and (3-54) become

$$dx(t) = F(t)x(t)dt + G(t)\eta(t)dt \quad (3-58)$$

$$y(t) = H(t)x(t) + u(t) \quad (3-59)$$

and equation (3-49) becomes

$$\mathbf{x}(t) = \Phi(t, t_0)\mathbf{x}(t_0) + \int_{t_0}^t \Phi(t, s)G(s)\eta(s)ds. \quad (3-60)$$

In this manner the theory of part 1 is extended to what will be called the white noise case.

Next consider the situation wherein  $y(t)$  and  $z(t)$  are zero mean real valued second order stationary processes with stationary cross-covariance

$$V_{yz}(t-s) = E\{y(t)z^T(s)\}, \quad (3-61)$$

where  $y(t)$  and  $z(t)$  are, as before,  $m \times 1$  and  $r \times 1$  vectors respectively. Furthermore,  $z(t)$  is a Wiener process generated by stationary white noise  $\eta(t)$  with unit covariance and  $y(t)$  is, as usual, measurable. It will be shown that the representation corresponding to this situation takes the form of a stationary system

$$d\mathbf{x}(t) = F\mathbf{x}(t)dt + G\eta(t)dt \quad (3-62)$$

$$y(t) = H\mathbf{x}(t) + u(t) \quad (3-63)$$

where  $F$ ,  $G$ , and  $H$  are constant matrices. Equations (3-58) and (3-59) are still valid and the derivation starts from them. Let  $t \geq s$ . Then

$$V_{yz}(t-s) = E\left\{\left[H(t)\Phi(t, t_0)\mathbf{x}(t_0) + \int_{t_0}^t W(t, \tau)\eta(\tau)d\tau\right]z^T(s)\right\} \quad (3-64)$$

where

$$W(t, \tau) = H(t)\Phi(t, \tau)G(\tau). \quad (3-65)$$

The usual assumption that  $\mathbf{x}(t_0)$  is independent of  $z(s)$  is made and since both are zero mean processes (3-64)

reduces to

$$V_{yz}(t-s) = \int_{t_0}^t W(t, \tau) E\{\eta(\tau) z^T(s)\} d\tau. \quad (3-66)$$

By assumption

$$z(s) = \int_{t_0}^s \eta(r) dr. \quad (3-67)$$

Hence

$$\begin{aligned} V_{yz}(t-s) &= \int_{t_0}^s \int_{t_0}^t W(t, \tau) \delta(\tau-r) d\tau dr \\ &= \int_{t_0}^s W(t, r) dr. \end{aligned} \quad (3-68)$$

Now from (3-68)

$$\begin{aligned} \frac{\partial V_{yz}(t-s)}{\partial s} &= -\dot{V}_{yz}(t-s) \\ &= -W(t, s). \end{aligned} \quad (3-69)$$

Hence

$$W(t, s) = W(t-s). \quad (3-70)$$

Recall also the assumption that  $V_{yz}(t, s)$  is separable, that is

$$V_{yz}(t-s) = \alpha(t) \beta(s). \quad (3-71)$$

This implies  $W(t-s)$  is also separable,

$$W(t-s) = -\alpha(t) \dot{\beta}(s). \quad (3-72)$$

Since  $W(t-s)$  is a realizable stationary impulse response matrix it is analytic in both  $t$  and  $s$  for  $t \geq s$ .

Hence both  $\alpha(t)$  and  $\beta(s)$  must be analytic functions.

Since  $\alpha(t)$  and  $\beta(t)$  are continuously differentiable then by lemma 5 of Youla [Y-1]  $W(t-s)$  has a constant coefficient realization. That is,  $W(t-s)$  is of the form

$$W(t-s) = He^{F(t-s)}G.$$

Thus the representation

$$dx(t) = Fx(t)dt + G\eta(t)dt \quad (3-74)$$

$$y(t) = Hx(t) + u(t) \quad (3-75)$$

is achieved. This concludes the discussion of the representation problem.

## CHAPTER FOUR

### STOCHASTIC APPROACH TO THE MINIMAL REALIZATION PROBLEM IN THE STATIONARY CASE

This chapter will be concerned with the realization problem for stationary systems. The stationary realization problem may be stated as follows. There is given a stationary impulse response or weighting pattern matrix  $W(t-s)$  and the objective is to find matrices  $F$ ,  $G$ , and  $H$  such that

$$W(t-s) = He^{F(t-s)}G \quad (4-1)$$

where  $F$  is of minimal dimension. The idea used in the predictor space approach is to define a stochastic system having input-output covariance  $V_{yz}(t-s)$  equal to  $W(t-s)$ . Then, if a basis of the predictor space can be found, the results of the previous chapter can be used to write down a Markovian representation determining a minimal realization for the impulse response matrix  $W(t-s)$ .

The theory developed in the first part of the chapter will be used in the rest of the chapter to examine two important realization algorithms. The power of the predictor space approach is demonstrated by the ease with which these algorithms are explained in this context.

## 1. General Theory

Consider the  $m \times r$  matrix  $W(t, \tau)$  on which are made the following assumptions:

A-1)  $W(t, \tau)$  is stationary in the sense that

$$W(t+\sigma, \tau+\sigma) = W(t, \tau) \quad \text{for all } t, \tau, \text{ and } \sigma.$$

A-2)  $W(t, \tau)$  is an impulse response matrix, implying that  $W(t, \tau)$  is separable and equals 0 for  $t < \tau$ .

$$\text{A-3)} \quad \int_{t_0}^t W(t, \tau)^T W(t, \tau) d\tau < \infty, \quad t_0 \leq t < \infty.$$

The usual abuse of notation will prevail and  $W(t, \tau)$  will be written as  $W(t-\tau)$  with  $t \geq \tau$ .

Let  $\eta(t)$  be an  $r$ -dimensional stationary Gaussian white noise with

$$E\{\eta(t)\} = 0, \quad E\{\eta(t)\eta^T(s)\} = I \cdot \delta(t-s). \quad (4-1)$$

Denote by  $y(t)$  the  $m$ -dimensional output of the stochastic system defined by

$$y(t) = \int_{t_0}^t W(t-\tau) \eta(\tau) d\tau \quad (4-2)$$

when the stochastic integral exists. Define the Wiener process  $z(t)$  by

$$z(t) = \int_{t_0}^t \eta(s) ds. \quad (4-3)$$

Then  $\eta(t)$  is the innovation process for  $z(t)$  and equation (4-2) can be written

$$y(t) = \int_{t_0}^t W(t-\tau) dz(\tau). \quad (4-4)$$

The goal here is to justify using the representation theory developed in chapter 3. To do that it will be necessary to verify that the hypotheses of theorem 3-4 are satisfied.

That  $y(t)$  and  $z(t)$  are both zero mean and second order is clear. Furthermore, as noted in chapter 3,  $z(t)$  is continuous in quadratic mean as well as purely nondeterministic. Since  $W(t-\tau)$  satisfies assumption A-3 the stochastic integral in equation (4-2), or equivalently equation (4-4), exists and  $y(t)$  is quadratic mean continuous and measurable (Wong [W-3], pp. 145-146).

First form the predictors  $\hat{y}(t|s)$  by

$$\begin{aligned} \hat{y}(t|s) &= (y(t) | L_2^m(z;s)) \\ &= \left( \int_{t_0}^t W(t-\tau) dz(\tau) | L_2^r(z;s) \right) \\ &= \int_{t_0}^s W(t-\tau) dz(\tau). \end{aligned} \quad (4-5)$$

The last equality follows from the fact that for  $\tau > s$

$$(dz(\tau) | L_2^r(z;s)) = 0 \quad (4-6)$$

by virtue of the orthogonal increment property of  $z(t)$ . Forming the predictor space  $L_2(\hat{y};t|s)$  suppose that it has a finite dimensional basis  $x(s)$ . Then from the corollary to theorem 3-1

$$y(t) = A(t,s)x(s) + u(t). \quad (4-7)$$

Since by equations (4-4) and (4-5)

$$\hat{y}(t|t) = y(t) \quad (4-8)$$

the term  $u(t)$  defined by  $\hat{y}(t|t) - y(t)$  is not present when  $t = s$  and equation (4-7) becomes

$$y(t) = H(t)x(t). \quad (4-9)$$

Equation (4-9) will be useful in showing that the cross covariance  $V_{xz}(t,s)$  must be separable. First it is necessary to compute  $V_{yz}(t,s)$ . Using equation (4-4) and the properties of orthogonal increment processes (Wong [W-3], p. 144) for  $t_0 \leq s \leq t$

$$\begin{aligned} V_{yz}(t,s) &= E\left\{\int_{t_0}^t W(t-\tau)dz(\tau)z^T(s)\right\} \\ &= E\left\{\int_{t_0}^t W(t-\tau)dz(\tau)\int_{t_0}^s dz^T(\tau)\right\} \\ &= \int_{t_0}^s W(t-\tau)d\tau. \end{aligned} \quad (4-10)$$

By hypothesis  $W(t-\tau)$  is separable in the form

$$W(t-\tau) = p(t)q(\tau) \quad (4-11)$$

for some  $m \times \ell$  matrix function  $p(t)$  and  $\ell \times r$  matrix function  $q(\tau)$ . Hence

$$V_{yz}(t,s) = p(t)\int_{t_0}^s q(\tau)d\tau \quad (4-12)$$

so  $V_{yz}(t,s)$  is separable. Since

$$V_{yz}(t,s) = H(t)V_{xz}(t,s) \quad (4-13)$$

it is clear that  $V_{xz}(t,s)$  must also be separable.

It is next necessary to show that  $\frac{\partial V_{xz}(t,s)}{\partial t}$  must be continuous for  $t_0 \leq s \leq t$ . Combining equations (4-10) and (4-13)

$$H(t)V_{xz}(t,s) = \int_{t_0}^s W(t-\tau)d\tau \quad (4-14)$$

and the right hand side of the equation is analytic for  $t \geq \tau$ . From the discussion of the stationary case at the end of chapter 3 the matrix output coefficient matrix  $H(t)$  is time invariant and can be denoted by simply  $H$ . Thus

$$HV_{xz}(t,s) = \int_{t_0}^s W(t-\tau)d\tau \quad (4-15)$$

implies  $V_{xz}(t,s)$  is analytic and hence possesses continuous derivatives of any order for  $t_0 \leq s \leq t$ .

Finally to show that  $E\{x_i(t)^2\}$  for each component  $x_i(t)$  of  $x(t)$  is bounded on bounded intervals the fact that in the case of a stationary system a minimal realization is observable is used (Silverman [S-1]). Hence the matrix  $H$  must have full rank which implies that the diagonal elements of  $H^TH$  cannot contain a zero. Since

$$E\{x^T(t)H^THx(t)\} = E\{y^T(t)y(t)\} < \infty \quad (4-16)$$

it follows that for each  $i$ ,  $i = 1, 2, \dots, n$ ,

$$E\{x_i^2(t)\} < \infty. \quad (4-17)$$

Hence the desired property that  $E\{|x_i^2(t)|\}$  is bounded on bounded intervals must hold.

Thus the necessary hypotheses of theorem 3-4 will all be satisfied if a finite basis  $x(s)$  of the predictor space  $L_2(\hat{y}; t|s)$  can be found. The search for such a basis is now begun.

Set  $t - \tau = \lambda$ . Since  $W(\lambda)$  is analytic it has a Taylor series expansion

$$W(\lambda) = \sum_{k=0}^{\infty} Y_k \frac{\lambda^k}{k!}, \quad \lambda \geq 0. \quad (4-18)$$

The coefficients  $Y_k$ ,  $k = 0, 1, 2, \dots$  are called the Markov parameters (Ho and Kalman [H-2]) and are given by

$$Y_k = \left. \frac{d^k W(\lambda)}{d\lambda^k} \right|_{\lambda=0}, \quad k = 0, 1, 2, \dots \quad (4-19)$$

Using equation (4-18) the predictors can be expressed as

$$\hat{y}(t|s) = \int_{t_0}^s \left( \sum_{k=0}^{\infty} Y_k \frac{(t-\tau)^k}{k!} \right) dz(\tau), \quad t_0 \leq s \leq t. \quad (4-20)$$

Lemma 4-1:

$$\hat{y}(t|s) = \sum_{k=0}^{\infty} Y_k \int_{t_0}^s \frac{(t-\tau)^k}{k!} dz(\tau), \quad t_0 \leq s \leq t. \quad (4-21)$$

Proof:

The equality in equation (4-21) is in the sense of mean square equality. Define

$$S_n = \int_{t_0}^s \left[ \sum_{k=0}^n Y_k \frac{(t-\tau)^k}{k!} \right] dz(\tau). \quad (4-22)$$

From a criterion by Doob ([D-2], p. 429) the sequence of stochastic integrals  $S_n$  converges in the mean to  $\hat{y}(t|s)$  if and only if the integrands converge in the mean to  $W(t-\tau)$ . This criterion is clearly satisfied in this case and since

$$S_n = \sum_{k=0}^n Y_k \int_{t_0}^s \frac{(t-\tau)^k}{k!} dz(\tau)$$

the lemma is proved.

Theorem 4-1:

The  $p$ th quadratic mean derivative of  $\hat{y}(t|s)$  with respect to  $t$  exists and is given by

$$\frac{\partial^p \hat{y}(t|s)}{\partial t^p} = \sum_{k=p}^{\infty} \frac{Y_k}{(k-p)!} \int_{t_0}^s (t-\tau)^{k-p} dz(\tau) \quad (4-23)$$

for  $t_0 \leq s \leq t$ .

Proof:

The  $p$ th derivative  $\frac{\partial^p \hat{y}(t|s)}{\partial t^p}$  exists in a mean square sense if  $\frac{\partial^{2p} V_{\hat{y}}(t_1|s, t_2|s)}{\partial t_1^p \partial t_2^p}$  exists at the point  $(t, t)$  (Papoulis [P-2], p. 317) where

$$V_{\hat{y}}(t_1|s, t_2|s) = E\{\hat{y}(t_1|s)\hat{y}(t_2|s)^T\}. \quad (4-24)$$

Using the representation given by equation (4-5) and the orthogonal increment property of  $z(t)$

$$V_{\hat{y}}(t_1|s, t_2|s) = E\left\{\left[\int_{t_0}^s W(t_1-\tau)dz(\tau)\right]\left[\int_{t_0}^s W(t_2-\tau)dz(\tau)\right]^T\right\} \quad (4-25)$$

$$= \int_{t_0}^s W(t_1-\tau)W^T(t_2-\tau)d\tau, \quad t_0 \leq s \leq t. \quad (4-26)$$

Since  $W(\lambda)$  is analytic it is clear from (4-26) that mixed partials of  $V_{\hat{y}}(t_1|s, t_2|s)$  of any order exist for all  $t_1$

and  $t_2$ . Thus  $\frac{\partial^p \hat{y}(t|s)}{\partial t^p}$  exists in a mean square sense for all positive integers  $p$ . Since the only portion of  $\int_{t_0}^s W(t-\tau)dz(\tau)$  that depends on  $t$  is the deterministic portion of the integrand  $W(t-\tau)$ , it is clear that for  $t_0 \leq s \leq t$

$$\begin{aligned} \frac{\partial^p \hat{y}(t|s)}{\partial t^p} &= \int_{t_0}^s \frac{\partial^p W(t-\tau)}{\partial t^p} dz(\tau) \\ &= \int_{t_0}^s \left( \sum_{k=p}^{\infty} Y_k \frac{(t-\tau)^{k-p}}{(k-p)!} \right) dz(\tau) \\ &= \sum_{k=p}^{\infty} Y_k \int_{t_0}^s \frac{(t-\tau)^{k-p}}{(k-p)!} dz(\tau), \end{aligned} \quad (4-27)$$

where the last equality follows by a proof similar to that for lemma 4.1.\*

Q.E.D.

---

\*To prove rigorously that the process  $y(t|s)$  has the mean square derivative  $\frac{\partial \hat{y}(t|s)}{\partial t}$  it is necessary to show that  $\lim_{\epsilon \rightarrow 0} \left\| \frac{\hat{y}(t+\epsilon|s) - \hat{y}(t|s)}{\epsilon} - \frac{\partial \hat{y}(t|s)}{\partial t} \right\| = 0$ . Doing that for  $\hat{y}(t|s)$  results in an easy calculation justifying equation (4-27). Also, for the case when  $s = t$ ,  $\hat{y}(t|t) = \int_{t_0}^t W(t-\tau)dz(\tau)$  the  $t$  occurring as a limit of integration is independent of the  $t$  occurring in the integrand and  $\frac{\partial \hat{y}(t|t)}{\partial t}$  is to be interpreted as  $\left. \frac{\partial \hat{y}(t|s)}{\partial t} \right|_{s=t}$ .

Stating the result of the theorem in a different way, for  
 $t_0 \leq s \leq t$

$$\begin{bmatrix} Y_0 & Y_1 & Y_2 & \cdot & \cdot & \cdot \\ Y_1 & Y_2 & Y_3 & \cdot & \cdot & \cdot \\ Y_2 & Y_3 & Y_4 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \begin{bmatrix} \int_{t_0}^s \frac{(t-\tau)^0}{0!} dz(\tau) \\ \int_{t_0}^s \frac{(t-\tau)^1}{1!} dz(\tau) \\ \int_{t_0}^s \frac{(t-\tau)^2}{2!} dz(\tau) \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} = \begin{bmatrix} \hat{y}(t|s) \\ \frac{\partial \hat{y}(t|s)}{\partial t} \\ \frac{\partial^2 \hat{y}(t|s)}{\partial t^2} \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \quad (4-28)$$

The matrix consisting of  $Y_i$ 's on the left is called the generalized Hankel matrix (Ho and Kalman [H-2]) and will be denoted  $S_\infty$ . Equation (4-28) will be of considerable importance in this section.

Lemma 4-2:

Let  $V_{\hat{y}}(t_1|s, t_2|s)$  be as defined in theorem 4-1. Then  $V_{\hat{y}}(t_1|s, t_2|s)$  is analytic in  $t_1$  and  $t_2$  for  $t_1, t_2 \geq t_0$  and

$$\frac{\partial^{k+m} V_{\hat{y}}(t_1|s, t_2|s)}{\partial t_1^k \partial t_2^m} = E \left\{ \frac{\partial^k \hat{y}(t_1|s)}{\partial t_1^k} \cdot \frac{\partial^m \hat{y}^T(t_2|s)}{\partial t_2^m} \right\}. \quad (4-29)$$

Proof:

The analyticity of  $V_{\hat{y}}(t_1|s, t_2|s)$  is obvious from the analyticity of  $W(\lambda)$  and equation (4-24). Equation (4-29) is a standard result (see Papoulis [P-2], p. 317).

Q.E.D.

For notational convenience  $\frac{\partial^{k+m} V_{\hat{y}}(t_1|s, t_2|s)}{\partial t_1^k \partial t_2^m}$  will be denoted by  $V_{\hat{y}}^{(k,m)}(t_1|s, t_2|s)$ .

Theorem 4-2:

For  $t_0 \leq s \leq t$  and  $t_0 \leq s \leq t + \tau$ ,

$$\hat{y}(t + \tau | s) = \sum_{k=0}^{\infty} \frac{\partial^k \hat{y}(t | s)}{\partial t^k} \frac{\tau^k}{k!} . \quad (4-30)$$

Proof:

Since the existence of the derivatives of  $\hat{y}(t | s)$  has already been shown, the first step will be to show the convergence of the right hand side of (4-30) in the mean square sense. Define

$$S_p = \sum_{k=0}^p \frac{\partial^k \hat{y}(t | s)}{\partial t^k} \frac{\tau^k}{k!} \quad (4-31)$$

For  $\ell < p$

$$\begin{aligned} E \left\{ \left[ S_p - S_\ell \right] \left[ S_p - S_\ell \right]^T \right\} &= E \left\{ \left[ \sum_{k=\ell+1}^p \frac{\partial^k \hat{y}(t | s)}{\partial t^k} \frac{\tau^k}{k!} \right] \left[ \sum_{j=\ell+1}^p \frac{\partial^j \hat{y}(t | s)}{\partial t^j} \frac{\tau^j}{j!} \right]^T \right\} \\ &= E \left\{ \sum_{k,j=\ell+1}^p \frac{\partial^k \hat{y}(t | s)}{\partial t^k} \frac{\partial^j \hat{y}(t | s)^T}{\partial t^j} \frac{\tau^{k+j}}{k! j!} \right\} \quad (4-32) \\ &= \sum_{k,j=\ell+1}^p V_{\hat{y}}^{(k,j)}(t | s, t | s) \frac{\tau^{j+k}}{j! k!} \\ &= \sum_{k=\ell+1}^p \left( \sum_{j=\ell+1}^p V_{\hat{y}}^{(k,j)}(t | s, t | s) \frac{\tau^j}{j!} \right) \frac{\tau^k}{k!} \\ &\rightarrow 0 \end{aligned}$$

as  $\ell$  tends to  $\infty$  since  $\sum_{j=0}^{\infty} V_{\hat{y}}^{(k,j)}(t | s, t | s) \frac{\tau^j}{j!}$  converges.

Therefore,\*

---

\*tr is the trace operator.

$$\|S_p - S_\ell\| = \text{tr } E\{[S_p - S_\ell][S_p - S_\ell]^T\} \quad (4-33)$$

$$\rightarrow 0$$

as  $\ell$  tends to  $\infty$ . Thus the sequence is Cauchy and hence converges in  $L_2^m(z;s)$  for  $t_0 \leq s \leq t$ ,  $t_0 \leq s \leq t + \tau$ .

To prove (4-30) it will be necessary to show that

$$\|\hat{y}(t + \tau | s) - \sum_{k=0}^{\infty} \frac{\partial^k \hat{y}(t | s)}{\partial t^k} \frac{\tau^k}{k!}\| = 0. \quad (4-34)$$

Now

$$E \left\{ \left[ \hat{y}(t + \tau | s) - \sum_{k=0}^{\infty} \frac{\partial^k \hat{y}(t | s)}{\partial t^k} \frac{\tau^k}{k!} \right] \frac{\partial^\ell \hat{y}(t | s)^T}{\partial t^\ell} \right\}$$

$$= V_{\hat{y}}^{(0, \ell)}(t + \tau | s, t | s) - \sum_{k=0}^{\infty} V_{\hat{y}}^{(k, \ell)}(t | s, t | s) \frac{\tau^k}{k!}$$

$$= 0 \quad (4-35)$$

since  $\sum_{k=0}^{\infty} V_{\hat{y}}^{(\ell, k)}(t | s, t | s) \frac{\tau^k}{k!}$  is just the Taylor series for  $V_{\hat{y}}^{(0, \ell)}(t + \tau | s, t | s)$ .

Thus  $\hat{y}(t + \tau | s) - \sum_{k=0}^{\infty} \frac{\partial^k \hat{y}(t | s)}{\partial t^k} \frac{\tau^k}{k!}$  is orthogonal to  $\frac{\partial^\ell \hat{y}(t | s)}{\partial t^\ell}$  for all  $\ell$  and so it is orthogonal to  $\sum_{k=0}^{\infty} \frac{\partial^k \hat{y}(t | s)}{\partial t^k} \frac{\tau^k}{k!}$ . Therefore

$$\begin{aligned}
& E \left\{ \left[ \hat{y}(t+\tau | s) - \sum_{k=0}^{\infty} \frac{\partial^k \hat{y}(t | s)}{\partial t^k} \frac{\tau^k}{k!} \right] \left[ \hat{y}(t+\tau | s) - \sum_{j=0}^{\infty} \frac{\partial^j \hat{y}(t | s)}{\partial t^j} \frac{\tau^j}{j!} \right]^T \right\} \\
&= E \left\{ \left[ \hat{y}(t+\tau | s) - \sum_{k=0}^{\infty} \frac{\partial^k \hat{y}(t | s)}{\partial t^k} \frac{\tau^k}{k!} \right] \hat{y}(t+\tau | s)^T \right\} \\
&= V_{\hat{y}}(t+\tau | s, t+\tau | s) - \sum_{k=0}^{\infty} V_{\hat{y}}^{(k,0)}(t | s, t+\tau | s) \frac{\tau^k}{k!} \\
&= 0,
\end{aligned} \tag{4-36}$$

since the second term is again just the Taylor series expansion of the first. Thus

$$\begin{aligned}
\| \hat{y}(t+\tau | s) - \sum_{k=0}^{\infty} \frac{\partial^k \hat{y}(t | s)}{\partial t^k} \frac{\tau^k}{k!} \| &= \text{tr } E \left\{ \left[ \hat{y}(t+\tau | s) - \sum_{k=0}^{\infty} \frac{\partial^k \hat{y}(t | s)}{\partial t^k} \frac{\tau^k}{k!} \right] \left[ \hat{y}(t+\tau | s) - \sum_{j=0}^{\infty} \frac{\partial^j \hat{y}(t | s)}{\partial t^j} \frac{\tau^j}{j!} \right]^T \right\} = 0,
\end{aligned} \tag{4-37}$$

and the theorem is proved.

Q.E.D.

Corollary:

Let  $\hat{y}_i(t+\tau | s)$  be the  $i$ th component of  $\hat{y}(t+\tau | s)$ . Then

$$\hat{y}_i(t+\tau | s) = \sum_{k=0}^{\infty} \frac{\partial^k \hat{y}_i(t | s)}{\partial t^k} \frac{\tau^k}{k!}, \quad t_0 \leq s \leq t. \tag{4-38}$$

Proof:

The corollary is only a restatement of equation (4-30).

Q.E.D.

For notational convenience define the symbol  $\Gamma(t | s)$  by

$$\Gamma(t|s) = \left[ \hat{y}(t|s)^T, \frac{\partial \hat{y}(t|s)^T}{\partial t}, \dots, \frac{\partial^k \hat{y}(t|s)^T}{\partial t^k}, \dots \right]^T. \quad (4-39)$$

It is thus the right hand side of equation (4-28). The vector  $\Gamma(t|s)$  will be regarded as a vector whose components are the scalar components of

$\hat{y}(t|s), \frac{\partial \hat{y}(t|s)}{\partial t}, \frac{\partial^2 \hat{y}(t|s)}{\partial t^2}, \dots$ . The linear manifold spanned by  $\Gamma(t|s)$ , denoted by  $S\{\Gamma(t|s)\}$ , consists of all random variables which can be written in the form

$$a = \sum_{i=1}^N C_i \frac{\frac{\partial^{k_i} \hat{y}_{m_i}(t|s)}{\partial t^{k_i}}}{\partial t^{k_i}} \quad (4-40)$$

for some integer  $N$  and real constants  $C_1, \dots, C_N$ . The

symbol  $\frac{\frac{\partial^{k_i} \hat{y}_{m_i}(t|s)}{\partial t^{k_i}}}{\partial t^{k_i}}$  is the  $m_i$ th component of  $\frac{\partial^{k_i} \hat{y}(t|s)}{\partial t^{k_i}}$ .

The Hilbert space spanned by  $\Gamma(t|s)$ , denoted  $L_2(\Gamma(t|s))$ , is the closure of  $S\{\Gamma(t|s)\}$  with respect to the mean square norm. The variables  $t$  and  $s$  are regarded as fixed when considering  $L_2(\Gamma(t|s))$ .

Theorem 4-3:

$$L_2(\Gamma(t|s)) = L_2(\hat{y}; t|s), \quad t_0 \leq s \leq t. \quad (4-41)$$

Proof:

Recall that when considering the space  $L_2(\hat{y}; t|s)$ ,  $s$  is regarded as fixed while  $t$  varies over all values greater than or equal to  $s$ . This is in contrast to the

space  $L_2(\Gamma(t|s))$  where both  $t$  and  $s$  are regarded as fixed. Thus the  $s$  in  $L_2(\Gamma(t|s))$  is the same as that in  $L_2(\hat{y}; t|s)$ , but the use of the symbol  $t$  must be correctly interpreted.

That  $L_2(\Gamma(t|s)) \subset L_2(\hat{y}; t|s)$  is trivial since for any stochastic process  $a(t) \in L_2(\hat{y}; t|s)$  it follows automatically that  $\frac{da(t)}{dt} \in L_2(\hat{y}; t|s)$  whenever the derivative exists in the mean square sense. (Wong [W-3], p. 79). Hence

$$S\{\Gamma(t|s)\} \subset L_2(\hat{y}; t|s). \quad (4-42)$$

Since  $L_2(\hat{y}; t|s)$  is closed, taking the closure of the left hand side yields

$$L_2(\Gamma(t|s)) \subset L_2(\hat{y}; t|s). \quad (4-43)$$

Now let  $a \in L_2(\hat{y}; t|s)$ . Then by definition

$$a = \lim_{p \rightarrow \infty} a_p \quad (4-44)$$

where, for each  $p$ ,  $a_p$  is a linear combination of components of  $\hat{y}(t_{p_i}|s)$  with  $t_{p_i} \geq s$ . This can be expressed as

$$a_p = \sum_{i=1}^{N_p} C_{p_i} \hat{y}_{p_{ji}}(t_{p_i}|s) \quad (4-45)$$

where  $C_{p_1}, \dots, C_{p_{N_p}}$  are real constants,  $t_{p_i} \geq s$  for all  $p_i$ , and  $\hat{y}_{p_{ji}}(t_{p_i}|s)$  is the  $p_{ji}$ th component of  $\hat{y}(t_{p_i}|s)$ . By the corollary to theorem 4-2, for  $s < t$  and  $s < t_{p_i}$ ,

$$\hat{y}_{p_{ji}}(t_{p_i}|s) = \sum_{k=0}^{\infty} \frac{\partial^k \hat{y}_{p_{jk}}(t|s)}{\partial t^k} \frac{(t_{p_i} - t)^k}{k!} \quad (4-46)$$

Hence

$$a_p = \sum_{i=0}^{N_p} C_{p_i} \left( \sum_{k=0}^{\infty} \frac{\partial^k \hat{y}_{p_{jk}}(t|s)}{\partial t^k} \frac{(t_{p_i} - t)^k}{k!} \right). \quad (4-47)$$

Let

$$S_N = \sum_{k=0}^N \frac{\partial^k \hat{y}_{p_{jk}}(t|s)}{\partial t^k} \frac{(t_{p_i} - t)^k}{k!} \quad (4-48)$$

Since  $S_N$  is a linear combination of elements of  $\Gamma(t|s)$ ,

$S_N \in L_2(\Gamma(t|s))$ . Since  $L_2(\Gamma(t|s))$  is closed,

$\lim_{N \rightarrow \infty} S_N \in L_2(\Gamma(t|s))$ . Therefore,

$$\sum_{k=0}^{\infty} \frac{\partial^k \hat{y}_{p_{jk}}(t|s)}{\partial t^k} \frac{(t_{p_i} - t)^k}{k!} \in L_2(\Gamma(t|s)). \quad (4-49)$$

Thus  $a_p \in L_2(\Gamma(t|s))$  since  $a_p$  is just a finite linear combination of such elements. Invoking once again the fact that  $L_2(\Gamma(t|s))$  is closed, l.i.m.  $a_p$  is an element of  $L_2(\Gamma(t|s))$ . Thus

$$L_2(\hat{y}; t|s) \subset L_2(\Gamma(t|s)). \quad (4-50)$$

Q.E.D.

Corollary:

$\Gamma(t|s)$  spans  $L_2(\hat{y}; t|s)$  for  $t_0 \leq s \leq t$ .

Proof:

The corollary is immediate from the equality

$$L_2(\hat{y}; t|s) = L_2(\Gamma(t|s)). \quad (4-51)$$

Q.E.D.

For convenient reference equation (4-28) is reproduced again here:

$$\begin{bmatrix} Y_0 & Y_1 & Y_2 & \dots \\ Y_1 & Y_2 & Y_3 & \dots \\ Y_2 & Y_3 & Y_4 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} \int_{t_0}^s \frac{(t-\tau)^0}{0!} dz(\tau) \\ \int_{t_0}^s \frac{(t-\tau)^1}{1!} dz(\tau) \\ \int_{t_0}^s \frac{(t-\tau)^2}{2!} dz(\tau) \\ \vdots \end{bmatrix} = \begin{bmatrix} \hat{y}(t|s) \\ \frac{\partial \hat{y}(t|s)}{\partial t} \\ \frac{\partial^2 \hat{y}(t|s)}{\partial t^2} \\ \vdots \end{bmatrix} = \Gamma(t|s).$$

It has been established by the corollary to theorem 4-3 that  $\Gamma(t|s)$  contains among its components a basis for the space  $L_2(\hat{y}; t|s)$ . The basis, though, is as yet unidentified. However, if a finite dimensional basis for the linear space spanned by the rows of the Hankel matrix is found, define the vector  $x(t|s)$  as the corresponding elements of the predictor space where the correspondence is realized by taking the inner product of the row vectors\* of the Hankel matrix forming the basis and the vector  $Z(t,s)$  defined by

$$Z(t,s) = \left[ \int_{t_0}^s \frac{(t-\tau)^0}{0!} dz^T(\tau), \int_{t_0}^s \frac{(t-\tau)^1}{1!} dz^T(\tau), \dots \right]^T. \quad (4-52)$$

---

\*The row vectors must be distinguished from the block row vectors such as  $[Y_k, Y_{k+1}, Y_{k+2}, \dots]$ .

Then  $x(t|s)$  is the basis for  $L_2(\hat{y};t|s)$ . Essentially the idea is that if the predictor space has a finite dimensional basis, then it can be identified by examining the Hankel matrix.

Theorem 4-4:

Suppose the linear space spanned by the rows of the Hankel matrix has a basis consisting of a finite number of (scalar) row vectors of the matrix. Define the vector  $x(t|s)$  as that vector whose components are given by taking the inner product of the rows of the Hankel matrix forming the basis and the vector  $Z(t,s)$ . Then  $x(t|s)$  is a basis for the predictor space.

Proof:

By equation (4-28) a set of components of  $\Gamma(t|s)$  is linearly independent if and only if corresponding rows of the Hankel matrix are. Thus by the hypothesis on the Hankel matrix the components of the vector  $x(t|s)$  are a maximal linearly independent subset of the components of  $\Gamma(t|s)$  and hence span  $L_2(\Gamma(t|s))$ . Then by the corollary to theorem 4-3 the components of  $x(t|s)$  are a linearly independent set of random variables spanning  $L_2(\hat{y};t|s)$ . Therefore, by definition,  $x(t|s)$  is a basis of the predictor space.

Q.E.D.

The next thing to show is that a finite dimensional basis of the linear space spanned by the rows of the  $S_\infty$  can be found. It has been shown by Silverman

[S-1] that for a stationary system the rank of the Hankel matrix is equal to the dimension of a minimal realization of the system if the transfer function of the impulse response matrix of the system is a strictly proper rational matrix.\* He also showed that for a stationary system, this is a necessary and sufficient condition for a realization to exist. Since the assumption of separability made at the beginning of the section is also sufficient to guarantee a realization for the stationary impulse response matrix  $W(t-s)$ , this implies that the Hankel matrix has finite rank. Hence a basis for the linear space spanned by the row vectors can be found and the  $x(t|s)$  so defined is of the dimension of a minimal realization of  $W(t-s)$ . These results are summarized by the following theorem.

Theorem 4-5:

Let  $W(t-s)$  be a stationary separable impulse response matrix satisfying equation (4-2). Then the Hankel matrix defined by  $W(t-s)$  has the property that a finite dimensional basis of the space spanned by its row vectors can be found. Furthermore, the vector  $x(t|s)$  defined by taking the inner product of these row vectors and  $Z(t,s)$  is the same dimension as a minimal realization of  $W(t-s)$ .

---

\*  $G(s)$  is a strictly proper rational matrix if the numerator of every element of  $G(s)$  is of lower degree than its denominator.

Hence, a basis for the predictor space  $L_2(\mathcal{Y}; t|s)$  can be found by examining the rows of the Hankel matrix  $S_N$  for some  $N$ . The basis, however, is a function of two variables  $s$  and  $t$  in apparent contrast to chapter 3 where the basis of  $L_2(\mathcal{Y}; t|s)$  was considered as a function of  $s$  alone. The vector  $x(t|s)$ , moreover, is a basis of  $L_2(\mathcal{Y}; t|s)$  for any value of  $t$  greater than or equal to  $s$ . In this sense  $x(t|s)$  is independent of  $t$  in its role as a basis. The problem arises though that the derivatives in the representations of chapter 3, for instance in the representation given by equations (3-74) and (3-75), are total derivatives. That is, it is necessary to consider

$$\frac{dx(t|s)}{ds} = \frac{\partial x(t|s)}{\partial s} + \frac{\partial x(t|s)}{\partial t} \frac{dt}{ds} . \quad (4-53)$$

This total derivative is not defined because the term  $\frac{dt}{ds}$  is not defined. In general there is no relationship between  $t$  and  $s$ . If, however, consideration is explicitly restricted to the case  $t = s$  then

$$\frac{dt}{ds} = 1 \quad (4-54)$$

and the total derivative is defined. Furthermore by restricting consideration to the case  $t = s$  a basis depending only on  $s$  has been found. Define

$$x(s) = x(s|s) = x(t|s)|_{t=s} . \quad (4-55)$$

As discussed at the beginning of the section the criteria of theorem 3-4 are satisfied since now it has been shown that a finite basis  $x(s)$  of the predictor space  $L_2(\hat{y}; t|s)$  can be found. The representation given by equations (3-74) and (3-75) is justified,

$$dx(s) = Fx(s)ds + Gdz(s) \quad (4-56)$$

$$y(s) = Hx(s). \quad (4-57)$$

No noise term  $u(t)$  appears in equation (4-57) since

$$\hat{y}(s|s) = y(s) \quad (4-58)$$

and  $u(t)$  was defined by equation (3-12) as

$$u(t) = y(t) - \hat{y}(t|s).$$

The solution to equation (4-56) is

$$x(s) = \int_{t_0}^s e^{F(s-r)} Gdz(r) \quad (4-59)$$

and

$$y(s) = H \int_{t_0}^s e^{F(s-r)} Gdz(r). \quad (4-60)$$

#### Theorem 4-6:

The matrices  $F$ ,  $G$ , and  $H$  occurring in equations (4-53) and (4-54) determine a minimal realization for the impulse response matrix  $W(t-s)$ .

#### Proof:

Let  $t_0 < \tau < t$ . Then using equation (4-57) and the representation

$$\eta(t)dt = dz(t), \quad (4-61)$$

it follows that

$$V_{y\eta}(t, \tau) = HE \left\{ \int_{t_0}^t e^{F(t-r)} G \eta(r) dr \eta^T(\tau) \right\} \quad (4-62)$$

$$\begin{aligned} &= H \int_{t_0}^t e^{F(t-r)} G \delta(r-\tau) dr \\ &= H e^{F(t-\tau)} G. \end{aligned} \quad (4-63)$$

Furthermore, for a symmetric delta function,\* by equation (4-2)

$$\begin{aligned} V_{y\eta}(t, \tau) &= E \left\{ \int_{t_0}^t W(t-r) \eta(r) dr \eta^T(\tau) \right\} \\ &= \int_{t_0}^t W(t-r) E\{\eta(r) \eta^T(\tau)\} dr \\ &= \int_{t_0}^t W(t-r) \delta(r-\tau) dr \\ &= \begin{cases} \frac{W(t-\tau)}{2}, & \tau = t_0 \text{ or } \tau = t \\ W(t-\tau), & t_0 < \tau < t \\ 0, & \tau \notin [t_0, t] \end{cases} \end{aligned} \quad (4-64)$$

Hence, for  $t_0 < \tau < t$

$$W(t-\tau) = H e^{F(t-\tau)} G. \quad (4-65)$$

---

\*The symmetric delta function is defined by

$$\int_a^b f(t) \delta(t-s) dt = \begin{cases} \frac{1}{2} f(s-0), & s = b \\ \frac{1}{2} f(s+0), & s = a \\ \frac{1}{2} [f(s-0) + f(s+0)], & a < s < b \\ 0, & s \notin [a, b] \end{cases}$$

That equation (4-65) also holds at  $\tau = t_0$  and  $\tau = t$  follows from the analyticity of the functions involved. Since analytic functions defined on an interval of the real line have a unique analytic extension to the real line, and analytic functions equal on an interval of the real line are equal on their whole domain of analyticity, (Olmstead [0-1], p. 268), equation (4-65) can be extended to hold for all  $t$  and  $\tau$ . In particular it holds for  $\tau = t$  and  $\tau = t_0$ . The minimality of the realization follows from theorem 4-5.

Q.E.D.

Thus, the Markovian representation developed in chapter 3 yields a solution to the deterministic realization problem for the stationary case. The following theorem is interesting in that it shows explicitly that the components of  $x(s)$  generate the predictor space.

Theorem 4-7:

$$\hat{y}(t|s) = H e^{F(t-s)} x(s), \quad t_0 \leq s \leq t. \quad (4-66)$$

Proof:

Using equations (4-5) and (4-59), and theorem 4-6,

$$\begin{aligned} \hat{y}(t|s) &= \int_{t_0}^s W(t-r) dz(r) \\ &= H \int_{t_0}^s e^{F(t-r)} G dz(r) \\ &= H e^{F(t-s)} \int_{t_0}^s e^{F(s-r)} G dz(r) \\ &= H e^{F(t-s)} x(s). \end{aligned} \quad \text{Q.E.D.}$$

This completes the results for this section. It is to be emphasized that the material developed here should not be regarded as an algorithm, but as a theoretical setting which is useful in analyzing and understanding the realization problem and associated algorithms, as will be demonstrated in the next several sections.

## 2. Silverman's Algorithm

In this section the predictor space approach will be used to analyze Silverman's realization algorithm for stationary systems (Silverman [S-2]). Akaike [A-1] has previously done this for the discrete-time case. The analysis for continuous time systems is considerably more complex.

The analysis starts from the assumption that the dimension of the system is somehow known.

### Theorem 4-8:

Suppose it is known that the impulse response matrix  $W(t-\tau)$  can be realized by a system of dimension  $n$ . Then the search for a basis of the predictor space can be limited to within the space spanned by the components of

$$\left[ \mathbf{y}(t|s)^T, \frac{\partial \mathbf{y}(t|s)^T}{\partial t}, \dots, \frac{\partial^{n-1} \mathbf{y}(t|s)^T}{\partial t^{n-1}} \right]^T.$$

### Proof:

Since  $W(t-\tau)$  can be realized by a system of dimension  $n$ , theorem 4-6 can be used to represent the predictors as

$$\hat{y}(t|s) = \int_{t_0}^s H e^{F(t-r)} G dz(r) \quad (4-67)$$

where  $F$  is an  $n \times n$  matrix. Then

$$\frac{\partial^p \hat{y}(t|s)}{\partial t^p} = \int_{t_0}^s H F^p e^{F(t-r)} G dz(r). \quad (4-68)$$

If  $p = n$ , then the characteristic equation of the  $n \times n$  matrix  $F$ ,

$$F^n = \sum_{j=0}^{n-1} \lambda_j F^j \quad (4-69)$$

can be used to write

$$\begin{aligned} \frac{\partial^n \hat{y}(t|s)}{\partial t^n} &= \sum_{j=0}^{n-1} \lambda_j \int_{t_0}^s H F^j e^{F(t-r)} G dz(r) \\ &= \sum_{j=0}^{n-1} \lambda_j \frac{\partial^j \hat{y}(t|s)}{\partial t^j}. \end{aligned} \quad (4-70)$$

Then

$$\begin{aligned} \frac{\partial^{n+1} \hat{y}(t|s)}{\partial t^{n+1}} &= \sum_{j=0}^{n-1} \lambda_j \frac{\partial^{j+1} \hat{y}(t|s)}{\partial t^{j+1}} \\ &= \sum_{j=0}^{n-2} \lambda_j \frac{\partial^{j+1} \hat{y}(t|s)}{\partial t^{j+1}} \\ &\quad + \lambda_{n-1} \sum_{j=0}^{n-1} \lambda_j \frac{\partial^j \hat{y}(t|s)}{\partial t^j} \\ &= \sum_{j=0}^{n-1} \gamma_j \frac{\partial^j \hat{y}(t|s)}{\partial t^j}. \end{aligned} \quad (4-71)$$

Continuing by induction it is clear that all elements in the vector  $[\hat{y}(t|s)^T, \frac{\partial \hat{y}(t|s)}{\partial t}^T, \dots]^T$  are linearly

dependent on the components of the first  $n$  block elements  $[\hat{y}(t|s)^T, \dots, \frac{\partial^{n-1} \hat{y}(t|s)^T}{\partial t^{n-1}}]^T$ .

Q.E.D.

Corollary:

The search for a basis of the linear space spanned by the row vectors of the Hankel matrix may be restricted to the first  $n$  block rows. That is, it is only necessary to look at

$$S_{n,\infty} = \begin{bmatrix} Y_0 & Y_1 & \dots & Y_k & \dots \\ Y_1 & Y_2 & \dots & Y_{k+1} & \dots \\ \vdots & \vdots & & \vdots & \\ \vdots & \vdots & & \vdots & \\ Y_{n-1} & Y_n & \dots & Y_{n+k-1} & \dots \end{bmatrix}. \quad (4-72)$$

Proof:

The corollary is immediate from the theorem.

Q.E.D.

It is equally easy to see that attention may be restricted to the first  $n$  block columns of  $S_{n,\infty}$ .

Indeed, since

$$Y_k = HF^k G \quad (4-73)$$

then again using the characteristic polynomial of  $F$

$$\begin{aligned} Y_n &= HF^n G \\ &= H \sum_{j=0}^{n-1} \lambda_j F^j G \\ &= \sum_{j=0}^{n-1} \lambda_j Y_j. \end{aligned} \quad (4-74)$$

Clearly, by induction, for  $p \geq n$  the block entry  $Y_p$  can be written as a linear combination of  $Y_0, \dots, Y_{n-1}$ . Thus  $S_{n,\infty}$  has only  $n$  linearly independent block columns and it is only necessary to consider

$$S_{n,n} = \begin{bmatrix} Y_0 & Y_1 & \cdot & \cdot & \cdot & Y_{n-1} \\ Y_1 & Y_2 & \cdot & \cdot & \cdot & Y_n \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ Y_{n-1} & Y_n & \cdot & \cdot & \cdot & Y_{2n-2} \end{bmatrix}. \quad (4-75)$$

Lemma 4-3:

$$E \left\{ \frac{\partial^k \hat{y}(t|t)}{\partial t^k} \frac{d^j \eta^T(t)}{dt^j} \right\} = (-1)^j E \left\{ \frac{\partial^{k+j} \hat{y}(t|t)}{\partial t^{k+j}} \eta^T(t) \right\},$$

$$j, k = 0, 1, 2, \dots \quad (4-76)$$

Proof:

The quantities  $\frac{d^j \eta^T(t)}{dt^j}$ ,  $j = 0, 1, 2, \dots$  have a meaning only in the sense of generalized random processes. A discussion of generalized random processes and proof of the lemma will be found in the appendix.

Q.E.D.

Theorem 4-9:

$$E \left\{ \frac{\partial^k \hat{y}(t|t)}{\partial t^k} \frac{d^j \eta^T(t)}{dt^j} \right\} = \frac{(-1)^j}{2} Y_{k+j}, \quad j, k = 0, 1, 2, \dots \quad (4-77)$$

Proof:

From equation (4-27) for any nonnegative integer  $p$

$$E \left\{ \frac{\partial^P \hat{y}(t|t)}{\partial t^P} \eta^T(t) \right\} = E \left\{ \int_{t_0}^t \frac{\partial^P W(t-\tau)}{\partial t^P} \eta^T(\tau) d\tau \eta^T(t) \right\} \quad (4-78)$$

$$= \int_{t_0}^t \frac{\partial^P W(t-\tau)}{\partial t^P} \delta(\tau-t) d\tau$$

$$= \frac{1}{2} \frac{\partial^P W(t-\tau)}{\partial t^P} \Big|_{\tau=t} \quad (4-79)$$

Thus from equation (4-19)

$$E \left\{ \frac{\partial^P \hat{y}(t|t)}{\partial t^P} \eta^T(t) \right\} = \frac{1}{2} Y_P. \quad (4-80)$$

Combining equations (4-76) and (4-80) yields equation (4-77).

Q.E.D.

Corollary:

For all positive integers  $j$  and  $k$

$$E \left\{ \begin{bmatrix} \hat{y}(t|t) \\ \frac{\partial \hat{y}(t|t)}{\partial t} \\ \vdots \\ \frac{\partial^{j-1} \hat{y}(t|t)}{\partial t^{j-1}} \end{bmatrix} \begin{bmatrix} \eta^T(t), -\frac{d\eta^T(t)}{dt}, \dots, (-1)^{k-1} \frac{d^{k-1} \eta^T(t)}{dt^{k-1}} \end{bmatrix} \right\}$$

$$= \frac{1}{2} \begin{bmatrix} Y_0 & Y_1 & \dots & Y_{k-1} \\ Y_1 & Y_2 & \dots & Y_k \\ \vdots & \vdots & \ddots & \vdots \\ Y_{j-1} & Y_j & \dots & Y_{j+k-2} \end{bmatrix} \quad (4-81)$$

Proof:

The corollary is a direct consequence of theorem 4-9.

Q.E.D.

At this point it is appropriate to present Silverman's algorithm for stationary systems (Silverman [S-1]). The algorithm assumes that integers  $j$  and  $k$  can be found such that

$$\text{rank } S_{jk} = \text{rank } S_{j+1, k+\ell} = n, \ell = 0, 1, 2, \dots \quad (4-82)$$

The integers  $j$  and  $k$  are usually taken to be the smallest values such that equation (4-82) holds, but this is not necessary for the algorithm. The integer  $n$  defined by equation (4-82) is the dimension of the minimal realization of the impulse response matrix generating the Hankel matrix.

Given  $S_{jk}$ , let  $M$  denote the  $n \times k \cdot r$  submatrix formed from the first  $n$  independent rows of  $S_{jk}$ . Define the  $n \times k \cdot r$  submatrix  $\bar{M}$  of  $S_{j+1, k}$  in the following way. Denote the scalar rows of the Hankel matrix by  $Y_\alpha$ ,  $\alpha = 1, 2, 3, \dots$ . If  $Y_\ell$  is one of the rows selected for the matrix  $M$ , then the row  $Y_{\ell+m}$  ( $m$  is dimension of output) is selected to be the corresponding row in the matrix  $\bar{M}$ . The following four submatrices of  $S_{j+1, k}$  are then defined.

- A -- The nonsingular  $n \times n$  matrix formed from the first  $n$  independent columns of  $M$ .
- $\bar{A}$  -- The  $n \times n$  matrix occupying the same column positions in  $\bar{M}$  as does  $A$  in  $M$ .
- $A_1$  -- The  $m \times n$  matrix occupying the same column positions in  $S_{1k}$  as does  $A$  in  $M$ .
- $A_2$  -- The  $n \times r$  matrix occupying the first  $r$  (dimension of input) columns of  $M$ .

Then setting

$$F = \bar{A}A^{-1}, G = A_2, H = A_2A^{-1} \quad (4-83)$$

yields a minimal realization of the impulse response matrix

$$W(t-s) = He^{F(t-s)}G. \quad (4-84)$$

After a preliminary lemma, equation (4-81) will be employed in examining the Silverman algorithm in the context of predictor space theory.

Lemma 4-4:

If equation (4-82) is satisfied by the Hankel matrix  $S_{jk}$  for some  $j, k$ , and  $n$ , then the matrix  $S_{nn}$  satisfies equation (4-82).

Proof:

Since the dimension of the realization is  $n$ , from the proof of theorem 4-8 and the discussion following its corollary it is clear that the  $n + 1$ st block row of  $S_{n+1,n}$  is but a linear combination of the  $n$  block rows of  $S_{nn}$ . Similarly the  $n+1, \dots, n+l$  block columns of  $S_{n+1,n+l}$  are but linear combinations of the  $n$  block columns of  $S_{n+1,n}$ . Hence

$$\text{rank } S_{nn} = \text{rank } S_{n+1,n} = \text{rank } S_{n+1,n+l}^{\ell} = 0, 1, 2, \dots \quad (4-85)$$

Q.E.D.

The only purpose of this lemma is to justify working with the specific matrix  $S_{nn}$  in the following analysis. This

is desirable since, as it is known, the realization being sought is of dimension  $n$  or less. Theorem 4-8 guarantees the existence of a finite dimensional basis within the vector\*  $[\hat{y}(t|t)^T, \frac{\partial \hat{y}(t|t)^T}{\partial t}, \dots, \frac{\partial^{n-1} \hat{y}(t|t)^T}{\partial t^{n-1}}]^T$ .

Now from equation (4-81) picking the matrix  $M$  of the first  $n$  linearly independent rows of  $S_{nn}$  is clearly equivalent to picking the first  $n$  linearly independent elements of  $[\hat{y}(t|t)^T, \frac{\partial \hat{y}(t|t)^T}{\partial t}, \dots, \frac{\partial^{n-1} \hat{y}(t|t)^T}{\partial t^{n-1}}]^T$ , which amounts to picking a basis  $x(t|t) = x(t)$  for the predictor space. This is formalized in the following theorem.

Theorem 4-10:

If a Hankel matrix  $S_{nn}$  satisfies equation (4-82) with rank  $S_{nn} = n$ , then the corresponding vector  $[\hat{y}(t|t)^T, \frac{\partial \hat{y}(t|t)^T}{\partial t}, \dots, \frac{\partial^{n-1} \hat{y}(t|t)^T}{\partial t^{n-1}}]^T$  contains exactly  $n$  linearly independent elements and these form a basis for the predictor space  $L_2(\hat{y}; t|t)$ .

Proof:

Since  $S_{nn}$  satisfies equation (4-82) the dimension of the realization is  $n$ , theorem 4-8 applies, and the vector  $[\hat{y}(t|t)^T, \frac{\partial \hat{y}(t|t)^T}{\partial t}, \dots, \frac{\partial^{n-1} \hat{y}(t|t)^T}{\partial t^{n-1}}]^T$  contains

---

\*The point is that although Silverman's algorithm may work with the Hankel matrix  $S_{jk}$ ,  $j < n$ ,  $k < n$ , and the analysis to be carried out here can still be done, it would be unjustified from the viewpoint of the theory developed in this work.

among its components a basis of the predictor space  $L_2(\hat{y}; t|t)$ . Hence a maximal set of linearly independent components from this vector will constitute a basis.

From equation (4-81) the vector

$$\left[ \hat{y}(t|t)^T, \frac{\partial \hat{y}(t|t)^T}{\partial t}, \dots, \frac{\partial^{n-1} \hat{y}(t|t)^T}{\partial t^{n-1}} \right]^T \text{ must have exactly}$$

$n$  linearly independent elements when the rank of  $S_{nn}$

is  $n$ . Hence the  $n$  linearly independent elements from

$$\left[ \hat{y}(t|t)^T, \frac{\partial \hat{y}(t|t)^T}{\partial t}, \dots, \frac{\partial^{n-1} \hat{y}(t|t)^T}{\partial t^{n-1}} \right]^T \text{ form a maximal linear}$$

independent set and thus a basis for the predictor space.

Q.E.D.

Denoting the basis formed picking the first  $n$  linearly independent elements of

$$\left[ \hat{y}(t|t)^T, \frac{\partial \hat{y}(t|t)^T}{\partial t}, \dots, \frac{\partial^{n-1} \hat{y}(t|t)^T}{\partial t^{n-1}} \right]^T \text{ by } x(t|t) = x(t),$$

it is obvious from equation (4-81) that

$$M = 2E \left\{ x(t) \left[ \eta^T(t), -\frac{d\eta^T(t)}{dt}, \dots, (-1)^{n-1} \frac{d^{n-1} \eta^T(t)}{dt^{n-1}} \right] \right\}. \quad (4-86)$$

If  $x_i(t)$  is a component of  $x(t)$ , then the element  $m$  rows beneath it in the vector

$$\left[ \hat{y}(t|t)^T, \frac{\partial \hat{y}(t|t)^T}{\partial t}, \dots, \frac{\partial^{n-1} \hat{y}(t|t)^T}{\partial t^{n-1}} \right]^T \text{ is just }^* \frac{\partial x_i(t|t)}{\partial t}.$$

Thus the rows of the matrix  $\bar{M}$  correspond to the vector

$$\frac{\partial x(t|t)}{\partial t} \text{ and so}$$

---

\*Note that once again the partial is understood as  $\frac{\partial x(t|t)}{\partial t} = \frac{\partial x(t|s)}{\partial t} \Big|_{s=t}$ . A careful distinction must be made between  $\frac{\partial x(t|t)}{\partial t}$  and  $\frac{dx(t|s)}{dt}$ .

$$\bar{M} = 2E \left\{ \frac{\partial \mathbf{x}(t|t)}{\partial t} \left[ \eta^T(t), \frac{-d\eta^T(t)}{dt}, \dots, (-1)^{n-1} \frac{d^{n-1} \eta^T(t)}{dt^{n-1}} \right] \right\}. \quad (4-87)$$

Example: Writing out the matrices and vectors of equation (4-76) explicitly for the case  $j = k = n$  yields

$$2E \left\{ \begin{array}{l} \left[ \begin{array}{l} \hat{y}_1(t|t) \\ \hat{y}_2(t|t) \\ \vdots \\ \hat{y}_m(t|t) \\ \frac{\partial \hat{y}_1(t|t)}{\partial t} \\ \vdots \\ \frac{\partial \hat{y}_m(t|t)}{\partial t} \\ \vdots \\ \frac{\partial^{n-1} \hat{y}_1(t|t)}{\partial t^{n-1}} \\ \vdots \\ \frac{\partial^{n-1} \hat{y}_m(t|t)}{\partial t^{n-1}} \end{array} \right] \begin{array}{l} * \\ * \\ \vdots \\ * \\ \checkmark \\ \checkmark \end{array} \left\{ \begin{array}{l} \mathbf{x}(t|t) \\ \frac{\partial \mathbf{x}(t|t)}{\partial t} \end{array} \right\} \hat{\mathbf{y}}(t|t) \\ \left[ \begin{array}{l} \eta_1(t), \eta_2(t), \dots, \eta_r(t), \frac{-d\eta_1(t)}{dt}, \dots, \frac{-d\eta_r(t)}{dt}, \dots, (-1)^{n-1} \frac{d^{n-1} \eta_1(t)}{dt^{n-1}}, \dots, (-1)^{n-1} \frac{d^{n-1} \eta_r(t)}{dt^{n-1}} \end{array} \right] \end{array} \right\}$$

$$= \begin{array}{l} * \\ * \\ \vdots \\ * \\ \checkmark \\ \checkmark \end{array} \left[ \begin{array}{cccc} y_{11} & y_{12} & \cdot & \cdot & y_{1, (n-1) \cdot r} \\ y_{21} & y_{22} & \cdot & \cdot & y_{2, (n-1) \cdot r} \\ \vdots & \vdots & & & \vdots \\ y_{m1} & y_{m2} & \cdot & \cdot & y_{m, (n-1) \cdot r} \\ y_{m+1, 1} & y_{m+1, 2} & \cdot & \cdot & y_{m+1, (n-1) \cdot r} \\ y_{m+2, 1} & y_{m+2, 2} & \cdot & \cdot & y_{m+2, (n-1) \cdot r} \\ \vdots & \vdots & & & \vdots \\ y_{n \cdot m, 1} & y_{n \cdot m, 2} & \cdot & \cdot & y_{n \cdot m, (n-1) \cdot r} \end{array} \right] \left\{ \begin{array}{l} Y_0, \dots, Y_{n-1} \end{array} \right\}$$

The  $y_{ij}$  are the actual elements of the Hankel matrix, not the Markov parameters. Suppose the two starred rows of the Hankel matrix are the rows of  $M$ . Then the corresponding starred elements of the vector of  $\hat{y}$ 's are the elements of the basis of  $L_2(\hat{y}; t|t)$ , i.e.,  $x(t|t)$ . The two checkmarked rows of the Hankel matrix are then the rows of  $\bar{M}$  and the two corresponding checkmarked elements in the column of  $\hat{y}$ 's are just  $\frac{\partial x(t|t)}{\partial t}$ . Picking the  $n \times n$  matrix  $A$  corresponds to selecting the first  $n$  independent elements from the vector  $\left[ \eta^T(t), \frac{-d\eta^T(t)}{dt}, \dots, (-1)^{n-1} \frac{d^{n-1}\eta^T(t)}{dt^{n-1}} \right]$ . The resulting vector will be called  $v(t)$ . Thus

$$A = 2E\{x(t|t)v^T(t)\}. \quad (4-88)$$

Similarly the  $n \times n$  matrix  $\bar{A}$  is just

$$\bar{A} = 2E\left\{\frac{\partial x(t|t)}{\partial t} v^T(t)\right\}. \quad (4-89)$$

The  $m \times n$  matrix  $A_1$  is given by

$$A_1 = 2E\{\hat{y}(t|t)v^T(t)\}. \quad (4-90)$$

Finally, the  $n \times r$  matrix  $A_2$  is given by

$$A_2 = 2E\{x(t|t)\eta^T(t)\}. \quad (4-91)$$

Now using equation (4-54) and the fact that  $\hat{y}(t|t) = y(t)$

$$A_1 A^{-1} = 2E\{y(t)v^T(t)\} [2E\{x(t|t)v^T(t)\}]^{-1} \quad (4-92)$$

$$= HE\{x(t|t)v^T(t)\} [E\{x(t|t)v^T(t)\}]^{-1} \quad (4-93)$$

$$= H . \quad (4-94)$$

Next, using equation (4-56)

$$\begin{aligned} A_2 &= 2E\{x(t|t)\eta^T(t)\} \\ &= 2E\left\{\int_{t_0}^t e^{F(t-r)} G \eta(r) dr \eta^T(t)\right\} \end{aligned} \quad (4-95)$$

$$\begin{aligned} &= 2 \int_{t_0}^t e^{F(t-r)} G \delta(r-t) dr \\ &= 2 \cdot \frac{1}{2} e^{F(t-t)} G \\ &= G . \end{aligned} \quad (4-96)$$

Before computing the final product,  $\bar{A} A^{-1}$ , note that  $\bar{A}$  involves the term  $\frac{\partial x(t|t)}{\partial t}$ . It will be necessary to interpret, and find an expression for, this term. Let  $x_i(t|t)$  be a component of  $x(t|t)$ . Then from equation (4-81), for some integer  $p \leq n$

$$x_i(t|s) = \frac{\partial^p \hat{y}_j(t|s)}{\partial t^p} \quad (4-97)$$

where  $\hat{y}_j(t|s)$  is a component of  $\hat{y}(t|s)$ . Since from equation (4-27)

$$\frac{\partial^p \hat{y}(t|s)}{\partial t^p} = \int_{t_0}^s \frac{\partial^p W(t-r)}{\partial t^p} dz(r), \quad (4-98)$$

denoting the  $j$ th row of  $\frac{\partial^p W(t-r)}{\partial t^p}$  by  $\frac{\partial^p W_j(t-r)}{\partial t^p}$

$$x_i(t|s) = \int_{t_0}^s \frac{\partial^p W_j(t-r)}{\partial t^p} dz(r) \quad (4-99)$$

The following theorem can now be proven.

Theorem 4-11:

$$\frac{\partial x(t|t)}{\partial t} = Fx(t|t) \quad (4-100)$$

where  $\frac{\partial x(t|t)}{\partial t}$  is interpreted as  $\frac{\partial x(t|s)}{\partial t} \Big|_{s=t}$ .

Proof:

From equation (4-99)

$$x_i(t|s) = \int_{t_0}^s \frac{\partial^p W_j(t-r)}{\partial t^j} dz(r).$$

Then the vector  $x(t|s)$  can be represented as

$$x(t|s) = \int_{t_0}^s M(t-r) dz(r) \quad (4-101)$$

where  $M(t-r)$  is an  $n \times r$  matrix whose  $i$ th row is of

the form  $\frac{\partial^p W_j(t-r)}{\partial t^p}$  for some  $p \leq n$  and some  $j$  between 1 and  $r$ . Hence,

$$\frac{\partial x(t|s)}{\partial t} = \int_{t_0}^s \frac{\partial M(t-r)}{\partial t} dz(r), \quad (4-102)$$

and

$$\begin{aligned} \frac{\partial x(t|t)}{\partial t} &= \frac{\partial x(t|s)}{\partial t} \Big|_{s=t} \\ &= \int_{t_0}^t \frac{\partial M(t-r)}{\partial t} dz(r). \end{aligned} \quad (4-103)$$

But from equation (4-59)

$$x(t|t) = \int_{t_0}^t e^{F(t-r)} G dz(r).$$

Then identifying the integrand in equation (4-59) with that in equation (4-101)

$$\begin{aligned} \frac{\partial x(t|t)}{\partial t} &= \int_{t_0}^t \frac{\partial e^{F(t-r)}}{\partial t} G dz(r) & (4-104) \\ &= F \int_{t_0}^t e^{F(t-r)} G dz(r) \\ &= F x(t|t) . \end{aligned}$$

Q.E.D.

Now computing the final product

$$\begin{aligned} \bar{A} A^{-1} &= 2E \left\{ \frac{\partial x(t|t)}{\partial t} v^T(t) \right\} \left[ 2E \{ x(t|t) v^T(t) \} \right]^{-1} & (4-105) \\ &= E \left\{ F x(t|t) v^T(t) \right\} \left[ E \{ x(t|t) v^T(t) \} \right]^{-1} \\ &= FE \left\{ x(t|t) v^T(t) \right\} \left[ E \{ x(t|t) v^T(t) \} \right]^{-1} \\ &= F. \end{aligned}$$

Thus, in the context of the predictor space approach to realization, Silverman's algorithm turns out to be a rather straightforward result. The same machinery that was used here will next be used to consider the Ho-Kalman algorithm.

### 3. The Ho-Kalman Algorithm

In this section the predictor space approach is used to analyze the Ho-Kalman algorithm (Ho and Kalman [H-2]). A summary of the algorithm is provided first. A neat and concise summary has been provided by Tether [T-1].

As with the Silverman algorithm it is necessary that equation (4-82) be satisfied. This will be assumed to be the case. Let  $\sigma^K$  denote the shift operator defined by

$$\sigma^K \begin{bmatrix} Y_0 & Y_1 & \cdot & \cdot & \cdot & Y_{j-1} \\ Y_1 & Y_2 & \cdot & \cdot & \cdot & Y_j \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ Y_{i-1} & Y_i & \cdot & \cdot & \cdot & Y_{i+j-2} \end{bmatrix} = \begin{bmatrix} Y_K & Y_{K+1} & \cdot & \cdot & \cdot & Y_{K+j-1} \\ Y_{K+1} & Y_{K+2} & \cdot & \cdot & \cdot & Y_{K+j} \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ Y_{K+i-1} & Y_{K+i} & \cdot & \cdot & \cdot & Y_{K+i+j-2} \end{bmatrix}. \quad (4-106)$$

Then the algorithm proceeds as follows

1) Pick  $N_1$  and  $N_2$  such that  $S_{N_1 N_2}$  satisfies equation (4-82).

2) Find nonsingular square matrices  $P$  and  $Q$  such that

$$P S_{N_1 N_2} Q = \begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix} = J = U_n \tilde{U}_n, \quad (4-107)$$

where  $I_n$  is an  $n \times n$  unit matrix and  $n$  is the constant defined by equation (4-82). The matrices  $U_n$  and  $\tilde{U}_n$  are given by

$$U_n = \begin{bmatrix} I_n \\ 0 \end{bmatrix} (N_1 m \times n), \quad \tilde{U} = [I_n \quad 0] (n \times N_2 \cdot r) \quad (4-108)$$

3) Let  $E_m$  be the block matrix

$$E_m = [I_m \quad 0_m \quad \dots \quad 0_m] \quad (m \times mN_1) \quad (4-109)$$

and  $E_r$  be

$$E_r = [I_r \quad 0_r \quad \dots \quad 0_r] \quad (r \times N_2 r) \quad (4-110)$$

4) Then

$$F = \tilde{U}_n [J^T P \quad \sigma \quad S_{N_1 N_2} Q \quad J^T] U_n \quad (4-111)$$

$$G = \tilde{U}_n [J^T P \quad S_{N_1 N_2} E_r^T] \quad (4-112)$$

$$H = E_m [S_{N_1 N_2} Q \quad J^T] U_n \quad (4-113)$$

yields a minimal realization of the impulse response matrix generating the Hankel matrix  $S_{N_1 N_2}$ .

For the analysis the specific Hankel matrix  $S_{nn}$  will be used. As shown in the previous section  $S_{nn}$  satisfies equation (4-82). Using equation (4-81), equation (4-107) from step 2 can be written

$$2PE \left\{ \begin{bmatrix} y(t|t) \\ \frac{\partial y(t|t)}{\partial t} \\ \vdots \\ \frac{\partial^{n-1} y(t|t)}{\partial t^{n-1}} \end{bmatrix} \begin{bmatrix} n^T(t), \frac{-d n^T(t)}{dt}, \dots, (-1)^{n-1} \frac{d^{n-1} n^T(t)}{dt^{n-1}} \end{bmatrix} \right\} Q = \begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix}. \quad (4-114)$$

Partition the matrices  $P$  and  $Q$  into the form

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdot & \cdot & \cdot & P_{1n} \\ P_{21} & P_{22} & \cdot & \cdot & \cdot & P_{2n} \\ \vdots & \vdots & & & & \vdots \\ \vdots & \vdots & & & & \vdots \\ P_{n1} & P_{n2} & \cdot & \cdot & \cdot & P_{nn} \end{bmatrix}_{nm \times nm}, \quad Q = \begin{bmatrix} Q_{11} & Q_{12} & \cdot & \cdot & \cdot & Q_{1n} \\ Q_{21} & Q_{22} & \cdot & \cdot & \cdot & Q_{2n} \\ \vdots & \vdots & & & & \vdots \\ \vdots & \vdots & & & & \vdots \\ Q_{n1} & Q_{n2} & \cdot & \cdot & \cdot & Q_{nn} \end{bmatrix}_{nr \times nr}$$

(4-115)

where  $P_{ij}$  is an  $m \times m$  matrix and  $Q_{ij}$  is an  $r \times r$  matrix. Then (4-114) becomes

$$2E \left\{ \begin{bmatrix} \sum_{i=1}^n P_{1i} \frac{\partial^{i-1} \hat{y}(t|t)}{\partial t^{i-1}} \\ \sum_{i=1}^n P_{2i} \frac{\partial^{i-1} \hat{y}(t|t)}{\partial t^{i-1}} \\ \vdots \\ \sum_{i=1}^n P_{ni} \frac{\partial^{i-1} \hat{y}(t|t)}{\partial t^{i-1}} \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n (-1)^{i-1} Q_{i1}^T \frac{d^{i-1} \eta(t)}{dt^{i-1}} \\ \sum_{i=1}^n (-1)^{i-1} Q_{i2}^T \frac{d^{i-1} \eta(t)}{dt^{i-1}} \\ \vdots \\ \sum_{i=1}^n (-1)^{i-1} Q_{in}^T \frac{d^{i-1} \eta(t)}{dt^{i-1}} \end{bmatrix}^T \right\} = \begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix}$$

(4-116)

Equation (4-116) will be shown to imply that the first  $n$

elements of the vector

$$\begin{bmatrix} \sum_{i=1}^n P_{1i} \frac{\partial^{i-1} \hat{y}(t|t)}{\partial t^{i-1}} \\ \sum_{i=1}^n P_{2i} \frac{\partial^{i-1} \hat{y}(t|t)}{\partial t^{i-1}} \\ \vdots \\ \sum_{i=1}^n P_{ni} \frac{\partial^{i-1} \hat{y}(t|t)}{\partial t^{i-1}} \end{bmatrix}$$

are linearly

independent and hence form a basis for the predictor space.

Theorem 4-12:

Equation (4-116) holds only if the first  $n$  ele-

ments of the vector

$$\begin{bmatrix} \sum_{i=1}^n P_{1i} \frac{\partial^{i-1} \hat{y}(t|t)}{\partial t^{i-1}} \\ \sum_{i=1}^n P_{2i} \frac{\partial^{i-1} \hat{y}(t|t)}{\partial t^{i-1}} \\ \vdots \\ \sum_{i=1}^n P_{ni} \frac{\partial^{i-1} \hat{y}(t|t)}{\partial t^{i-1}} \end{bmatrix}$$

are linearly

independent.

Proof: Denote the scalar components of  
by  $u_i$ ,  $i = 1, \dots, n \cdot m$ .

$$\begin{bmatrix} \sum_{i=1}^n P_{1i} \frac{\partial^{i-1} \hat{y}(t|t)}{\partial t^{i-1}} \\ \sum_{i=1}^n P_{2i} \frac{\partial^{i-1} \hat{y}(t|t)}{\partial t^{i-1}} \\ \vdots \\ \sum_{i=1}^n P_{ni} \frac{\partial^{i-1} \hat{y}(t|t)}{\partial t^{i-1}} \end{bmatrix}$$

Also denote the scalar components of

$$\left[ \sum_{i=1}^n (-1)^{i-1} \frac{d^{i-1} \eta^T(t)}{dt^{i-1}} Q_{i1}, \dots, \sum_{i=1}^n (-1)^{i-1} \frac{d^{i-1} \eta^T(t)}{dt^{i-1}} Q_{in} \right]$$

by  $v_i$ ,  $i = 1, \dots, nr$ .

Then by equation (4-112)

$$2E\{u_j v_k\} = \begin{cases} \delta_{jk}; & j, k = 1, \dots, n \\ 0; & j = n+1, \dots, nm \end{cases} \quad (4-117)$$

Suppose the first  $n$  components of the vector

$[u_1, u_2, \dots, u_{nm}]^T$  are not linearly independent. Then there

exists  $\alpha_1, \dots, \alpha_n$ , not all equal to zero, such that

$$\sum_{i=1}^n \alpha_i u_i = 0. \quad (4-118)$$

Suppose  $\alpha_j$  is nonzero. Then

$$u_j = - \sum_{\substack{i=1 \\ i \neq j}}^n \frac{\alpha_i}{\alpha_j} u_i. \quad (4-119)$$

Thus for  $k = 1, 2, \dots, n$

$$\begin{aligned} 2E\{u_j v_k\} &= -2 \sum_{\substack{i=1 \\ i \neq j}}^n \frac{\alpha_i}{\alpha_j} E\{u_i v_k\} \\ &= \sum_{\substack{i=1 \\ i \neq j}}^n \frac{\alpha_i}{\alpha_j} \delta_{ik} \\ &= -2 \frac{\alpha_k}{\alpha_j}. \end{aligned} \quad (4-120)$$

Hence, for equation (4-117) to hold  $\alpha_k$  must be zero for  $k = 1, 2, \dots, n, k \neq j$ . Then by equation (4-118)

$$\alpha_j u_j = 0. \quad (4-121)$$

Since

$$E\{u_j \eta_j\} = 1 \quad (4-122)$$

it must be that  $\alpha_j$  equals zero, contradicting the assumption of linear dependence. Thus the theorem is proved.

Q.E.D.

Corollary:

The first  $n$  components of the vector

$$\begin{bmatrix} \sum_{i=1}^n P_{1i} \frac{\partial^{i-1} \hat{y}(t|t)}{\partial t^{i-1}} \\ \sum_{i=1}^n P_{2i} \frac{\partial^{i-1} \hat{y}(t|t)}{\partial t^{i-1}} \\ \vdots \\ \sum_{i=1}^n P_{ni} \frac{\partial^{i-1} \hat{y}(t|t)}{\partial t^{i-1}} \end{bmatrix}$$

form a basis for the predictor space.

Proof:

From theorem 4-10 the vector  $\left[ \hat{y}(t|t)^T, \frac{\partial \hat{y}(t|t)^T}{\partial t}, \dots, \frac{\partial^{n-1} \hat{y}(t|t)^T}{\partial t^{n-1}} \right]^T$  contains exactly  $n$  linearly independent elements which form a basis of the predictor space. Since the first  $n$  elements of the vector

$$\begin{bmatrix} \sum_{i=1}^n P_{1i} \frac{\partial^{i-1} \hat{y}(t|t)}{\partial t^{i-1}} \\ \sum_{i=1}^n P_{2i} \frac{\partial^{i-1} \hat{y}(t|t)}{\partial t^{i-1}} \\ \vdots \\ \sum_{i=1}^n P_{ni} \frac{\partial^{i-1} \hat{y}(t|t)}{\partial t^{i-1}} \end{bmatrix}$$

are linearly independent they are

simply a nonsingular linear transformation of the first basis and hence another basis.

Q.E.D.

The basis defined by the above corollary will, as usual, be denoted by  $x(t|t)$ .

Next consider the quantity  $^* P(\sigma S_{nn})Q$ . From

---

\*The symbol  $\sigma$  is the shift operator defined above.

equation (4-106)

$$\sigma S_{nn} = \begin{bmatrix} Y_1 & Y_2 & \cdot & \cdot & \cdot & Y_n \\ Y_2 & Y_3 & \cdot & \cdot & \cdot & Y_{n+1} \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ Y_n & Y_{n+1} & \cdot & \cdot & \cdot & Y_{2n-1} \end{bmatrix} . \quad (4-123)$$

From equation (4-81)

$$2E \left\{ \begin{bmatrix} \frac{\partial \hat{y}(t|t)}{\partial t} \\ \frac{\partial^2 \hat{y}(t|t)}{\partial t^2} \\ \cdot \\ \cdot \\ \frac{\partial^n \hat{y}(t|t)}{\partial t^n} \end{bmatrix} \left[ \eta^T(t), \frac{-d\eta^T(t)}{dt}, \dots, (-1)^{n-1} \frac{d^{n-1} \eta^T(t)}{dt^{n-1}} \right] \right\} \\ = \begin{bmatrix} Y_1 & Y_2 & \cdot & \cdot & \cdot & Y_n \\ Y_2 & Y_3 & \cdot & \cdot & \cdot & Y_n \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ Y_n & Y_{n+1} & \cdot & \cdot & \cdot & Y_{2n-1} \end{bmatrix} . \quad (4-124)$$

Again partitioning the matrices P and Q as in equation (4-115)

$$P\sigma S_{nn}Q = 2E \left\{ \begin{bmatrix} \sum_{i=1}^n P_{1i} \frac{\partial^i \hat{y}(t|t)}{\partial t^i} \\ \sum_{i=1}^n P_{2i} \frac{\partial^i \hat{y}(t|t)}{\partial t^i} \\ \vdots \\ \sum_{i=1}^n P_{ni} \frac{\partial^i \hat{y}(t|t)}{\partial t^i} \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n (-1)^{i-1} Q_{i1}^T \frac{d^{i-1} \eta(t)}{dt^{i-1}} \\ \sum_{i=1}^n (-1)^{i-1} Q_{i2}^T \frac{d^{i-1} \eta(t)}{dt^{i-1}} \\ \vdots \\ \sum_{i=1}^n (-1)^{i-1} Q_{in}^T \frac{d^{i-1} \eta(t)}{dt^{i-1}} \end{bmatrix}^T \right\} \quad (4-125)$$

Since

$$\begin{bmatrix} \sum_{i=1}^n P_{1i} \frac{\partial^i \hat{y}(t|t)}{\partial t^i} \\ \vdots \\ \sum_{i=1}^n P_{ni} \frac{\partial^i \hat{y}(t|t)}{\partial t^i} \end{bmatrix} = \frac{\partial}{\partial t} \begin{bmatrix} \sum_{i=1}^n P_{1i} \frac{\partial^{i-1} \hat{y}(t|t)}{\partial t^{i-1}} \\ \vdots \\ \sum_{i=1}^n P_{ni} \frac{\partial^{i-1} \hat{y}(t|t)}{\partial t^{i-1}} \end{bmatrix} \quad (4-126)$$

it is clear that the first  $n$  elements of the vector

$$\begin{bmatrix} \sum_{i=1}^n P_{1i} \frac{\partial^i \hat{y}(t|t)}{\partial t^i} \\ \vdots \\ \sum_{i=1}^n P_{ni} \frac{\partial^i \hat{y}(t|t)}{\partial t^i} \end{bmatrix}$$

are simply  $\frac{\partial x(t|t)}{\partial t} = \frac{\partial x(t|s)}{\partial t} \Big|_{s=t}$ . Now consider the right hand side of equation (4-111).

$$\begin{aligned}
\tilde{U}_n [J^T P_{\sigma} S_{nn} Q J^T] U_n &= \\
&= \begin{bmatrix} I_n & 0 \end{bmatrix} \begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix} 2E \left\{ \begin{bmatrix} \sum_{i=1}^n P_{1i} \frac{\partial^i \hat{y}(t|t)}{\partial t^i} \\ \vdots \\ \sum_{i=1}^n P_{ni} \frac{\partial^i \hat{y}(t|t)}{\partial t^i} \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n (-1)^{i-1} \frac{d^{i-1} \eta^T(t)}{dt^{i-1}} Q_{i1} \\ \vdots \\ \sum_{i=1}^n (-1)^{i-1} \frac{d^{i-1} \eta^T(t)}{dt^{i-1}} Q_{in} \end{bmatrix} \right\} \begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} I_n \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} I_n & 0 \end{bmatrix} 2E \left\{ \begin{bmatrix} \frac{\partial x(t|t)}{\partial t} \\ 0 \end{bmatrix} \begin{bmatrix} v^T(t) & 0 \end{bmatrix} \right\} \begin{bmatrix} I_n \\ 0 \end{bmatrix} \quad (4-127)
\end{aligned}$$

where  $v^T(t)$  consists of the first  $n$  scalar components of the  $1 \times r \cdot n$  vector

$$\left[ \sum_{i=1}^n (-1)^{i-1} \frac{d^{i-1} \eta^T(t)}{dt^{i-1}} Q_{i1}, \dots, \sum_{i=1}^n (-1)^{i-1} \frac{d^{i-1} \eta^T(t)}{dt^{i-1}} Q_{in} \right].$$

From theorem 4-11  $\frac{\partial x(t|t)}{\partial t}$  can be replaced by  $Fx(t|t)$ .

Hence

$$\begin{aligned}
\tilde{U}_n [J^T P_{\sigma} S_{nn} Q J^T] U_n &= \begin{bmatrix} I_n & 0 \end{bmatrix} 2E \left\{ \begin{bmatrix} Fx(t|t) \\ 0 \end{bmatrix} \begin{bmatrix} v^T(t) & 0 \end{bmatrix} \right\} \begin{bmatrix} I_n \\ 0 \end{bmatrix} \quad (4-128) \\
&= 2 \begin{bmatrix} I_n & 0 \end{bmatrix} E \left\{ \begin{bmatrix} Fx(t|t) v^T(t) & 0 \\ 0 & 0 \end{bmatrix} \right\} \begin{bmatrix} I_n \\ 0 \end{bmatrix}
\end{aligned}$$

$$= 2F E\{x(t|t) v^T(t)\} \quad (4-129)$$

$$= F \cdot I \quad (4-130)$$

from equation (4-116). This proves equation (4-107).

Next consider  $\tilde{U}_n [J P S_{nn} E_r^T]$ . Omitting some steps which would be redundant

$$\tilde{U}_n [J^T P \ S_{nn} E_r^T] = [I_n \ 0] 2E \left\{ \begin{bmatrix} x(t|t) \\ 0 \end{bmatrix} \begin{bmatrix} \eta^T(t), \frac{-d\eta^T(t)}{dt}, \dots, \\ (-1)^{n-1} \frac{d^{n-1}\eta^T(t)}{dt^{n-1}} \end{bmatrix} \right\} \begin{bmatrix} I_r \\ 0_r \\ \vdots \\ 0_r \end{bmatrix} \quad (4-131)$$

$$= 2E\{x(t|t)\eta^T(t)\} \\ = 2E\left\{\int_{t_0}^t e^{F(t-r)} G_{\eta}(r) dr \eta^T(t)\right\} \quad (4-132)$$

$$= 2 \int_{t_0}^t e^{F(t-r)} G_{\delta}(r-t) dr \\ = 2 \cdot \frac{1}{2} G \\ = G, \quad (4-133)$$

where the representation in equation (4-56) was used to obtain the equality (4-132).

Finally consider  $E_m[S_{nn} Q J^T] U_n$ . Again some redundant steps are omitted

$$E_m[S_{nn} Q J^T] U_n = \\ [I_m \ 0_m \ \dots \ 0_m] 2E \left\{ \begin{bmatrix} \hat{y}(t|t) \\ \frac{\partial \hat{y}(t|t)}{\partial t} \\ \vdots \\ \frac{\partial^{n-1} \hat{y}(t|t)}{\partial t^{n-1}} \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n (-1)^{i-1} \frac{d^{i-1}\eta^T(t)}{dt^{i-1}} Q_{i1}, \\ \dots, \sum_{i=1}^n (-1)^{i-1} \frac{d^{i-1}\eta^T(t)}{dt^{i-1}} Q_{in} \end{bmatrix} \right\} \begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} I_n \\ 0 \end{bmatrix} \quad (4-134)$$

$$= 2E\{\hat{y}(t|t)v^T(t)\} \quad (4-135)$$

where  $v^T(t)$  is again the first  $n$  components of the vector defined in (4-127). Continuing,

$$\begin{aligned} E_m[S_{nn}Q J^T]U_n &= H \cdot 2E\{x(t|t)v^T(t)\} \\ &= H \cdot I \end{aligned} \quad (4-136)$$

from equation (4-110).

Thus the Ho-Kalman algorithm follows by a straightforward analysis using the predictor space approach. As noted by Akaike [A-1] the basis of the algorithm is that it provides a linear transformation of the vector  $\left[ \hat{y}(t|t)^T, \frac{\partial \hat{y}(t|t)^T}{\partial t}, \dots, \frac{\partial^{n-1} \hat{y}(t|t)^T}{\partial t^{n-1}} \right]^T$  such that the first  $n$  components of the transformed vector are a basis of the predictor space.

#### 4. Example

An example is presented here to explicitly demonstrate the Silverman algorithm, the Ho-Kalman algorithm, and several aspects of the predictor space approach. The example starts with the system

$$\begin{aligned} \begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} &= \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix}; \\ \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \end{aligned} \quad (4-137)$$

Let

$$\bar{F} = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}, \quad \bar{G} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \quad \bar{H} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}. \quad (4-138)$$

Then it may easily be verified that

$$e^{\bar{F}t} = \begin{bmatrix} e^{2t} & e^{2t-t} \\ 0 & e^t \end{bmatrix}; \quad W(t-\tau) = \begin{bmatrix} e^{2(t-\tau)} & 2e^{2(t-\tau)} \\ 0 & 2e^{t-\tau} \end{bmatrix}. \quad (4-139)$$

Also, the Markov parameters  $Y_K$  are given by

$$Y_K = \bar{H}\bar{F}^K\bar{G} = \begin{bmatrix} 2^k & 2^{2k+1} \\ 0 & 2 \end{bmatrix}, \quad k = 0, 1, 2, \dots \quad (4-140)$$

Part I: Silverman's algorithm.

The Hankel matrix for the system (4-137) is

$$S_\infty = \begin{array}{c|cccccccccccc|cccccccc|cccccccc|cccccccc|} \begin{array}{c} 1 \\ 0 \\ 2 \\ 0 \\ 4 \\ 0 \\ 8 \\ 0 \\ 16 \\ 0 \\ 32 \\ \vdots \\ \vdots \\ \vdots \end{array} & \begin{array}{c} 2 \\ 2 \\ 4 \\ 2 \\ 8 \\ 2 \\ 16 \\ 2 \\ 32 \\ 2 \\ 64 \\ \vdots \\ \vdots \\ \vdots \end{array} & \begin{array}{c} 2 \\ 0 \\ 4 \\ 0 \\ 8 \\ 0 \\ 16 \\ 0 \\ 32 \\ 0 \\ 64 \\ \vdots \\ \vdots \\ \vdots \end{array} & \begin{array}{c} 4 \\ 2 \\ 8 \\ 2 \\ 16 \\ 2 \\ 32 \\ 2 \\ 64 \\ 2 \\ 128 \\ \vdots \\ \vdots \\ \vdots \end{array} & \begin{array}{c} 4 \\ 0 \\ 8 \\ 0 \\ 16 \\ 0 \\ 32 \\ 0 \\ 64 \\ 0 \\ 128 \\ \vdots \\ \vdots \\ \vdots \end{array} & \begin{array}{c} 8 \\ 2 \\ 16 \\ 2 \\ 32 \\ 2 \\ 64 \\ 2 \\ 128 \\ 2 \\ 256 \\ \vdots \\ \vdots \\ \vdots \end{array} & \begin{array}{c} 8 \\ 0 \\ 16 \\ 0 \\ 32 \\ 0 \\ 64 \\ 0 \\ 128 \\ 0 \\ 256 \\ \vdots \\ \vdots \\ \vdots \end{array} & \begin{array}{c} 16 \\ 2 \\ 32 \\ 2 \\ 64 \\ 2 \\ 128 \\ 2 \\ 256 \\ 2 \\ 512 \\ \vdots \\ \vdots \\ \vdots \end{array} & \begin{array}{c} 16 \\ 0 \\ 64 \\ 0 \\ 128 \\ 0 \\ 256 \\ 0 \\ 512 \\ 0 \\ 1024 \\ \vdots \\ \vdots \\ \vdots \end{array} & \begin{array}{c} 32 \\ 2 \\ 128 \\ 2 \\ 256 \\ 2 \\ 512 \\ 2 \\ 1024 \\ 2 \\ 2048 \\ \vdots \\ \vdots \\ \vdots \end{array} & \begin{array}{c} 32 \\ 0 \\ 64 \\ 0 \\ 128 \\ 0 \\ 256 \\ 0 \\ 512 \\ 0 \\ 1024 \\ \vdots \\ \vdots \\ \vdots \end{array} & \dots & \end{array} \quad (4-141)$$

where the dashed lines partition the matrix into its Markov parameters  $Y_k$ . It can be determined by inspection that

$$\text{rank } S_{1,1} = \text{rank } S_{2,1+\ell} = 2, \ell = 0, 1, 2, \dots \quad (4-142)$$

Hence, by equation (4-82) it can be determined that the dimension of a minimal realization of this system is two.

The matrix  $M$  is just  $S_{11}$ , that is

$$M = \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix} \quad (4-143)$$

and

$$\bar{M} = \begin{bmatrix} 2 & 4 \\ 0 & 2 \end{bmatrix}. \quad (4-144)$$

Clearly  $A$ ,  $A_1$  and  $A_2$  all equal  $M$  and  $\bar{A}$  equals  $\bar{M}$ .

Then, using equation (4-83),

$$\begin{aligned} F &= \bar{A} A^{-1} \\ &= \begin{bmatrix} 2 & 4 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}^{-1} \\ &= \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \end{aligned} \quad (4-145)$$

$$G = A_2 = \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}, \quad (4-146)$$

and

$$H = A_1 A^{-1} = I. \quad (4-147)$$

The nonsingular transformation

$$T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

defines a strict equivalence between the system

$(F, G, H)_2$  and the system given by equations (4-137) and (4-138),  $(\bar{F}, \bar{G}, \bar{H})_2$ .

Part II: The Ho-Kalman algorithm.

Once again the algorithm starts with

$$S_{11} = \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}$$

since  $S_{11}$  satisfies equation (4-82). It is first necessary to find square matrices  $P$  and  $Q$  such that

$$PS_{11}Q = I_2. \quad (4-150)$$

It is easily seen that

$$Q = \begin{bmatrix} 1 & -1 \\ 0 & 1/2 \end{bmatrix}, \quad P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (4-151)$$

satisfy this condition. In this particular example  $U_n$ ,  $\tilde{U}_n$ ,  $J$ ,  $E_m$ , and  $E_r$  are all  $I_2$ . From equations (4-111) - (4-113)

$$\begin{aligned} F &= P\sigma S_{11}Q \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1/2 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \end{aligned} \quad (4-152)$$

$$\begin{aligned} G &= PS_{11} \\ &= \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix} \end{aligned} \quad (4-153)$$

and

$$\begin{aligned}
H &= S_{11}Q \\
&= \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1/2 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.
\end{aligned} \tag{4-154}$$

This is the same realization as was arrived at via the Silverman algorithm. In general these algorithms will yield equivalent, but not the same, realizations.

Part III: The predictor space approach.

This part of the example illustrates the Silverman algorithm in the context of the predictor space approach.

Equation (4-28) becomes for this example

$$\begin{bmatrix}
1 & 2 & 2 & 4 & 4 & 8 & 8 & 16 & 16 & \dots \\
0 & 2 & 0 & 2 & 0 & 2 & 0 & 2 & 0 & \dots \\
\hline
2 & 4 & 4 & 8 & 8 & 16 & 16 & 32 & 32 & \dots \\
0 & 2 & 0 & 2 & 0 & 2 & 0 & 2 & 0 & \dots \\
\hline
4 & 8 & 8 & 16 & 16 & 32 & 32 & 64 & 64 & \dots \\
0 & 2 & 0 & 2 & 0 & 2 & 0 & 2 & 0 & \dots \\
\hline
8 & 16 & 16 & 32 & 32 & 64 & 64 & 128 & 128 & \dots \\
0 & 2 & 0 & 2 & 0 & 2 & 0 & 2 & 0 & \dots \\
\hline
\vdots & \vdots & & & & & \vdots & \vdots & & \dots
\end{bmatrix}
\begin{bmatrix}
\int_{t_0}^s \eta_1(\tau) d\tau \\
\int_{t_0}^s \eta_2(\tau) d\tau \\
\int_{t_0}^s (t-\tau) \eta_1(\tau) d\tau \\
\int_{t_0}^s (t-\tau) \eta_2(\tau) d\tau \\
\int_{t_0}^s \frac{(t-\tau)^2}{2!} \eta_1(\tau) d\tau \\
\int_{t_0}^s \frac{(t-\tau)^2}{2!} \eta_2(\tau) d\tau \\
\vdots \\
\vdots
\end{bmatrix}
=
\begin{bmatrix}
\hat{y}_1(t|s) \\
\hat{y}_2(t|s) \\
\frac{\partial \hat{y}_1(t|s)}{\partial t} \\
\frac{\partial \hat{y}_2(t|s)}{\partial t} \\
\vdots
\end{bmatrix}.
\tag{4-155}$$

By inspection it can be determined that

$$\begin{bmatrix} \hat{y}_1(t|s) \\ \hat{y}_2(t|s) \\ \frac{\partial \hat{y}_1(t|s)}{\partial t} \\ \frac{\partial \hat{y}_2(t|s)}{\partial t} \\ \vdots \\ \vdots \end{bmatrix} = \begin{bmatrix} \sum_{k=0}^{\infty} 2^k \int_{t_0}^s \frac{(t-\tau)^k}{k!} \eta_1(\tau) d\tau + \sum_{k=0}^{\infty} 2^{k+1} \int_{t_0}^s \frac{(t-\tau)^k}{k!} \eta_2(\tau) d\tau \\ 2 \sum_{k=0}^{\infty} \int_{t_0}^s \frac{(t-\tau)^k}{k!} \eta_2(\tau) d\tau \\ 2\hat{y}_1(t|s) \\ \hat{y}_2(t|s) \\ \vdots \\ \vdots \end{bmatrix} \quad (4-156)$$

Clearly  $\hat{y}_1(t|s)$  and  $\hat{y}_2(t|s)$  form a maximal linearly independent subset of  $\Gamma(t|s)$  and hence are the basis  $x(t|s)$ . Then from equation (4-156)

$$\begin{aligned} \frac{\partial x(t|s)}{\partial t} &= \begin{bmatrix} \frac{\partial \hat{y}_1(t|s)}{\partial t} \\ \frac{\partial \hat{y}_2(t|s)}{\partial t} \end{bmatrix} \\ &= \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{y}_1(t|s) \\ \hat{y}_2(t|s) \end{bmatrix} \\ &= \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} x(t|s). \end{aligned} \quad (4-157)$$

From theorem 4-11 this shows that

$$F = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}. \quad (4-158)$$

Assume now that the system input vector  $\eta(t)$  has independent white noise components  $\eta_1(t)$  and  $\eta_2(t)$ , each with unit covariance. Then the vector  $v^T(t)$ , used in analyzing Silverman's algorithm, consisting of the first  $n$  independent components from the vector

$$\left[ \eta^T(t), \frac{-d\eta^T(t)}{dt}, \dots, (-1)^{n-1} \frac{d^{n-1}\eta^T(t)}{dt^{n-1}} \right] \text{ is just } \eta^T(t).$$

Hence

$$2E\{x(t|t)v^T(t)\} =$$

$$2E \left\{ \begin{bmatrix} \sum_{k=0}^{\infty} 2^k \int_{t_0}^t \frac{(t-\tau)^k}{k!} \eta_1(\tau) d\tau + \sum_{k=0}^{\infty} 2^{k+1} \int_{t_0}^t \frac{(t-\tau)^k}{k!} \eta_2(\tau) d\tau \\ 2 \sum_{k=0}^{\infty} \int_{t_0}^t \frac{(t-\tau)^k}{k!} \eta_2(\tau) d\tau \end{bmatrix} \begin{bmatrix} \eta_1(t) \\ \eta_2(t) \end{bmatrix}^T \right\}$$

$$= 2 \begin{bmatrix} \sum_{k=0}^{\infty} 2^k \int_{t_0}^t \frac{(t-\tau)^k}{k!} \delta(\tau-t) d\tau & \sum_{k=0}^{\infty} 2^{k+1} \int_{t_0}^t \frac{(t-\tau)^k}{k!} \delta(\tau-t) d\tau \\ 0 & 2 \sum_{k=0}^{\infty} \int_{t_0}^t \frac{(t-\tau)^k}{k!} \delta(\tau-t) d\tau \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}$$

$$= G.$$

(4-159)

Similar computations can be used to show that

$$2E \left\{ \frac{\partial x(t|t)}{\partial t} v^T(t) \right\} \left[ 2E \{ x(t|t) v^T(t) \} \right]^{-1} = F$$

and

$$2E\{\hat{y}(t|t)v^T(t)\}[2E\{x(t|t)v^T(t)\}]^{-1} = H$$

but it is not necessary to do so for this example. The matrix  $H$  is  $I$  follows from the fact that  $\hat{y}(t|t) = x(t|t)$  and  $F$  was determined by inspection from equation (4-156).

Part IV:

An alternate approach to this example which also illustrates some of the predictor space concepts is as follows. Using the impulse response matrix given by equation (4-139) the output of the stochastic system defined by equation (4-2) is given by

$$y(t) = \int_{t_0}^t \begin{bmatrix} e^{2(t-\tau)} & 2e^{2(t-\tau)} \\ 0 & 2e^{t-\tau} \end{bmatrix} \begin{bmatrix} \eta_1(\tau) \\ \eta_2(\tau) \end{bmatrix} d\tau \quad (4-160)$$

where  $\eta_1(\tau)$  and  $\eta_2(\tau)$  are again assumed to be independent white noise processes with unit covariance. Then

$$\hat{y}(t|s) = \int_{t_0}^s \begin{bmatrix} e^{2(t-\tau)} & 2e^{2(t-\tau)} \\ 0 & 2e^{t-\tau} \end{bmatrix} \begin{bmatrix} \eta_1(\tau) \\ \eta_2(\tau) \end{bmatrix} d\tau. \quad (4-161)$$

It can be easily seen by actually carrying out the differentiations that for  $k = 0, 1, 2, \dots$ ,

$$\frac{\partial^k \hat{y}(t|s)}{\partial t^k} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}^k \hat{y}(t|s). \quad (4-162)$$

Hence  $\hat{y}(t|s)$  is seen to form a maximal linearly independent subset for the vector  $\Gamma(t|s)$  and hence is a basis  $x(t|s)$  for the predictor space  $L_2(\hat{y};t|s)$ .

Again noting that  $v^T(t) = \eta^T(t)$  the realization given by Silverman's algorithm is easily derived. Since  $\hat{y}(t|t) = x(t|t)$  and  $v^T(t) = \eta^T(t)$ , using equation (4-161)

$$\begin{aligned}
 2E\{x(t|t)v^T(t)\} &= \\
 2E\left\{\int_{t_0}^t \begin{bmatrix} e^{2(t-\tau)} & 2e^{2(t-\tau)} \\ 0 & 2e^{t-\tau} \end{bmatrix} \begin{bmatrix} \eta_1(\tau) \\ \eta_2(\tau) \end{bmatrix} d\tau [\eta_1(t), \eta_2(t)]\right\} \\
 &= 2 \int_{t_0}^t \begin{bmatrix} e^{2(t-\tau)} & 2e^{2(t-\tau)} \\ 0 & 2e^{t-\tau} \end{bmatrix} \begin{bmatrix} \delta(t-\tau) & 0 \\ 0 & \delta(t-\tau) \end{bmatrix} d\tau \\
 &= \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix} \\
 &= G.
 \end{aligned} \tag{4-163}$$

The matrix  $\bar{F}$  may be similarly computed from

$$2E\left\{\frac{\partial x(t|t)}{\partial t} v^T(t)\right\} [2E\{x(t|t)v^T(t)\}]^{-1}$$

or simply read off from equation (4-162). The matrix  $H$  is the identity matrix since  $y(t|t) = x(t|t)$  implies

$$2E\{\hat{y}(t|t)v^T(t)\} [2E\{x(t|t)v^T(t)\}]^{-1} = I. \tag{4-164}$$

This example concludes the discussion of the realization problem for stationary impulse response matrices.

# CHAPTER FIVE

## STOCHASTIC APPROACH TO THE MINIMAL REALIZATION PROBLEM IN THE ANALYTIC CASE

In this chapter the minimal realization problem for analytic impulse response matrices is considered. The development is similar to that for the stationary case.

### 1. General Theory

Consider the  $m \times r$  matrix  $W(t, \tau)$  on which are made the following assumptions:

A-1)  $W(t, \tau)$  is analytic in both  $t$  and  $\tau$   
for  $t \geq \tau$ ,

A-2)  $W(t, \tau)$  is a realizable impulse response matrix,

A-3)  $\int_{t_0}^t W(t, \tau)^T W(t, \tau) d\tau < \infty, t_0 \leq t < \infty.$

Following the procedure in chapter 4, denote by  $y(t)$  the  $m$ -dimensional output of a stochastic system driven by the  $r$ -dimensional input  $\eta(t)$  where  $\eta(t)$  is a white noise process with

$$E\{\eta(t)\} = 0, \quad E\{\eta(t)\eta^T(s)\} = I \cdot \delta(t-s). \quad (5-1)$$

The output of the stochastic system is defined by

$$y(t) = \int_{t_0}^t W(t, \tau) \eta(\tau) d\tau \quad (5-2)$$

when the integral exists. Define the Wiener process  $z(t)$  by

$$z(t) = \int_{t_0}^t \eta(s) ds \quad (5-3)$$

so that  $\eta(t)$  is the innovation process for  $z(t)$  and (5-2) can be written

$$y(t) = \int_{t_0}^t W(t, \tau) dz(\tau). \quad (5-4)$$

The goal here again is to justify using the representation theory developed in chapter 3. To do this it is again necessary to verify that the hypotheses of theorem 3-4 are satisfied.

That  $y(t)$  and  $z(t)$  are both zero mean real valued second-order processes is clear. Since  $z(t)$  is quadratic mean continuous and hence measurable and since  $W(t, \tau)$  satisfies A-3 the integral in equation (5-2) exists and  $y(t)$  is also quadratic mean continuous and measurable (Wong [W-3], pp. 145-146).

Next form the predictors  $\hat{y}(t|s)$  by

$$\begin{aligned} \hat{y}(t|s) &= (y(t) | L_2^m(z; s)) \\ &= \int_{t_0}^t W(t, \tau) (dz(\tau) | L_2^r(z; s)) \\ &= \int_{t_0}^s W(t, \tau) dz(\tau), \end{aligned} \quad (5-5)$$

where the last equality follows from the fact that

$$(dz(\tau)|L_2^r(z;s)) = 0, \quad \tau > s \quad (5-6)$$

by virtue of the orthogonal increment property of  $z(t)$ .  
Forming the predictor space  $L_2(\hat{y};t|s)$  suppose that a finite dimensional basis  $x(s)$  can be found for it. Then from theorem 3-1

$$\hat{y}(t|s) = A(t,s)x(s). \quad (5-7)$$

From equations (5-4), (5-5), and

$$\hat{y}(t|t) = H(t)x(t) \quad (5-8)$$

it is the case that

$$y(t) = H(t)x(t). \quad (5-9)$$

Equation (5-9) will be useful in showing that  $V_{xz}(t,s)$  must be separable. First it is necessary to compute  $V_{yz}(t,s)$ . Using equation (5-4) and the orthogonal increment property of  $z(t)$  for  $t_0 \leq s \leq t$

$$\begin{aligned} V_{yz}(t,s) &= E \left\{ \int_{t_0}^t W(t,\tau) dz(\tau) z^T(s) \right\} \\ &= E \left\{ \int_{t_0}^t W(t,\tau) dz(\tau) \int_{t_0}^s dz^T(\tau) \right\} \\ &= \int_{t_0}^s W(t,\tau) d\tau. \end{aligned} \quad (5-10)$$

By hypothesis A-2 and theorem 2-1  $W(t,\tau)$  is separable in the form

$$W(t,\tau) = p(t)q(\tau). \quad (5-11)$$

Hence

$$V_{yz}(t,s) = p(t) \int_{t_0}^s q(\tau) d\tau \quad (5-12)$$

so  $V_{yz}(t,s)$  is separable. Since

$$V_{yz}(t,s) = H(t)V_{xz}(t,s) \quad (5-13)$$

it is clear that  $V_{xz}(t,s)$  is also separable.

It is next necessary to show that  $\frac{\partial V_{xz}(t,s)}{\partial t}$  must be continuous for  $t_0 \leq s \leq t$ . Combining equations (5-10) and (5-13)

$$H(t)V_{xz}(t,s) = \int_{t_0}^s W(t,\tau) d\tau \quad (5-14)$$

and the right hand side of the equation is analytic for  $t \geq s \geq \tau \geq t_0$ . Hence  $H(t)$  and  $V_{xz}(t,s)$  must be analytic also in this region and thus possess continuous derivatives of any order in this region.

Finally it is necessary to show that the components of  $E\{x(t)^T x(t)\}$  must be bounded on bounded intervals. Since the components of  $H(t)$  are continuous and  $y(t)$  is a second-order process, from equation (5-9)  $x(t)$  must also be second-order, which automatically yields the desired property.

All the prerequisites of theorem 3-4 have been established except for the existence of a finite basis for the predictor space. The search for that basis for the most part parallels the stationary case. Since  $W(t,\tau)$

is analytic it has the Taylor series expansion in  $\tau$ , for  $t \geq \tau$ ,

$$W(t, \tau) = \sum_{k=0}^{\infty} \frac{\partial^k W(t, \tau)}{\partial \tau^k} \Big|_{\tau=\bar{\tau}} \cdot \frac{(\tau-\bar{\tau})^k}{k!}, \quad t_0 \leq \bar{\tau} \leq t. \quad (5-15)$$

There is no problem about the derivatives of  $W(t, \tau)$  existing at  $\tau = t$  since  $W(t, \tau)$  has a unique analytic extension to the region  $t < \tau$  and this extension can be used to ensure that the derivative is well defined.

Using equation (5-15) the predictors can be expressed as

$$\hat{y}(t|s) = \int_{t_0}^s \left( \sum_{k=0}^{\infty} \frac{\partial^k W(t, \tau)}{\partial \tau^k} \Big|_{\tau=\bar{\tau}} \cdot \frac{(\tau-\bar{\tau})^k}{k!} \right) dz(\tau). \quad (5-16)$$

Lemma 5-1:

$$\hat{y}(t|s) = \sum_{k=0}^{\infty} \frac{\partial^k W(t, \tau)}{\partial \tau^k} \Big|_{\tau=\bar{\tau}} \cdot \int_{t_0}^s \frac{(\tau-\bar{\tau})^k}{k!} dz(\tau), \quad (5-17)$$

$$t_0 \leq s \leq t.$$

$$\bar{\tau} \leq t$$

Proof:

The proof is by the same procedure as for lemma 4-1.

Q.E.D.

Theorem 5-1:

The  $p$ th quadratic mean derivative of  $\hat{y}(t|s)$  with respect to  $t$  exists for  $t_0 \leq s \leq t$  and is given by

$$\frac{\partial^p \hat{y}(t|s)}{\partial t^p} = \sum_{k=0}^{\infty} \frac{\partial^{p+k} W(t, \tau)}{\partial t^p \partial \tau^k} \Big|_{\tau=\bar{\tau}} \cdot \int_{t_0}^s \frac{(\tau-\bar{\tau})^k}{k!} dz(\tau), \quad \bar{\tau} \leq t. \quad (5-18)$$

Proof:

The proof of this theorem is virtually identical to the proof of theorem 4-1.

Q.E.D.

For ease of notation  $\left. \frac{\partial^{p+k} W(t, \tau)}{\partial t^p \partial \tau^k} \right|_{\tau=\bar{\tau}}$  will be denoted by  $W_{p,k}(t, \bar{\tau})$ . Stating the result of the theorem in a different way, for  $t_0 \leq s \leq t$  and  $\bar{\tau} \leq t$

$$\begin{bmatrix} W_{0,0}(t, \bar{\tau}) & W_{0,1}(t, \bar{\tau}) & W_{0,2}(t, \bar{\tau}) & \dots \\ W_{1,0}(t, \bar{\tau}) & W_{1,1}(t, \bar{\tau}) & W_{1,2}(t, \bar{\tau}) & \dots \\ W_{2,0}(t, \bar{\tau}) & W_{2,1}(t, \bar{\tau}) & W_{2,2}(t, \bar{\tau}) & \dots \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} \int_{t_0}^s \frac{(\tau - \bar{\tau})^0}{0!} dz(\tau) \\ \int_{t_0}^s \frac{(\tau - \bar{\tau})^1}{1!} dz(\tau) \\ \int_{t_0}^s \frac{(\tau - \bar{\tau})^2}{2!} dz(\tau) \\ \vdots \\ \vdots \end{bmatrix}$$

$$= \begin{bmatrix} \hat{y}(t|s) \\ \frac{\partial \hat{y}(t|s)}{\partial t} \\ \frac{\partial^2 \hat{y}(t|s)}{\partial t^2} \\ \vdots \\ \vdots \end{bmatrix} \cdot \quad (5-19)$$

Equation (5-19) will be abbreviated

$$S(t, \bar{\tau})Z(\bar{\tau}, s) = \Gamma(t|s)$$

with the corresponding definitions for  $S(t, \bar{\tau}), Z(\bar{\tau}, s)$ , and  $\Gamma(t|s)$ .

Lemma 5-2:

Let  $V_{\hat{y}}(t_1|s, t_2|s)$  be defined by

$$V_{\hat{y}}(t_1|s, t_2|s) = E\{\hat{y}(t_1|s)\hat{y}^T(t_2|s)\}. \quad (5-20)$$

Then  $V_{\hat{y}}(t_1|s, t_2|s)$  is analytic in  $t_1$  and  $t_2$  for  $t_1, t_2 \geq s \geq t_0$  and

$$\frac{\partial^{k+p} V_{\hat{y}}(t_1|s, t_2|s)}{\partial t_1^k \partial t_2^p} = E \left\{ \frac{\partial^k \hat{y}(t_1|s)}{\partial t_1^k} \frac{\partial^p \hat{y}^T(t_2|s)}{\partial t_2^p} \right\}. \quad (5-21)$$

Proof:

From equation (5-5)

$$\begin{aligned} V_{\hat{y}}(t_1|s, t_2|s) &= E \left\{ \left[ \int_{t_0}^s W(t_1, \tau) dz(\tau) \right] \left[ \int_{t_0}^s W(t_2, \tau) dz(\tau) \right]^T \right\} \\ &= \int_{t_0}^s W(t_1, \tau) W^T(t_2, \tau) d\tau, \quad t_0 \leq s \leq t. \end{aligned} \quad (5-22)$$

Hence the analyticity of  $V_{\hat{y}}(t_1|s, t_2|s)$  with respect to  $t_1$  and  $t_2$  follows from the analyticity of  $W(t, \tau)$ .

Equation (5-21) is a standard result (see Papoulis [P-2], p. 317).

Q.E.D.

Theorem 5-2:

For  $t_0 \leq s \leq t$  and  $t_0 \leq s \leq t + \tau$

$$\hat{y}(t + \tau|s) = \sum_{k=0}^{\infty} \frac{\partial^k \hat{y}(t|s)}{\partial t^k} \frac{\tau^k}{k!}. \quad (5-23)$$

Proof:

The proof is the same as for theorem 4-2.

Q.E.D.

Forming the space  $L_2(\Gamma(t|s))$  as in chapter 4 the following analogue to theorem 4-3 occurs.

Theorem 5-3:

$$L_2(\Gamma(t|s)) = L_2(\hat{y}; t|s), \quad t_0 \leq s \leq t. \quad (5-24)$$

Proof:

The proof is identical to that for theorem 4-3.

Q.E.D.

As in chapter 4 it has been established that  $\Gamma(t|s)$  contains among its components a basis for the predictor space  $L_2(\hat{y}; t|s)$ . The basis, though, is as yet unidentified. However if a finite dimensional basis for the linear space spanned by the scalar rows of the matrix  $S(t, \bar{t})$  is found\*, define the vector  $x(t|s)$  as the corresponding elements of the predictor space where the correspondence is realized by taking the inner product of the row vectors of  $S(t, \bar{t})$  forming the basis and the vector  $Z(t, s)$ . Then, as in chapter 4,  $x(t|s)$  is a basis for the predictor space.

Theorem 5-4:

Suppose that the linear space spanned by the rows of  $S(t, \bar{t})$  has a basis consisting of a finite number of (scalar) row vectors of the matrix. Define the vector  $x(t|s)$  as that vector whose components are given by taking the inner product of the rows of  $S(t, \bar{t})$  forming the basis and the vector  $Z(t, s)$ . Then  $x(t|s)$  is a finite dimensional basis for the predictor space.

---

\*Recall that when considering the space  $L_2(\Gamma(t|s))$ , both  $s$  and  $t$  are regarded as fixed. Hence  $t$  in  $S(t, \bar{t})$  is regarded as fixed also.

Proof:

The proof is completely analogous to that of theorem 4-4.

Q.E.D.

The next thing to show is that a finite basis of the linear space spanned by the rows of  $S(t, \bar{\tau})$  can be found. By assumption A-2 and theorem 2-1 it follows that  $W(t, \tau)$  is separable in the form

$$W(t, \tau) = p(t)q(\tau). \quad (5-25)$$

Since  $W(t, \tau)$  is analytic in both variables so are  $p(t)$  and  $q(\tau)$  (Kalman [K-3]). Furthermore the columns of  $p(t)$  and rows of  $q(\tau)$  may be taken to be linearly independent and the number of linearly independent columns in  $p(t)$  and rows in  $q(\tau)$  is an integer  $n$  uniquely determined by  $W(t, \tau)$  (Youla [Y-1]).\* Consider the matrix

$$\begin{bmatrix} W_{0,p}^T(t, \bar{\tau}) \mid W_{1,p}^T(t, \bar{\tau}) \mid W_{2,p}^T(t, \bar{\tau}) \mid \dots \end{bmatrix} = \frac{d^p q^T(\bar{\tau})}{d\tau^p} \left[ p^T(t) \mid \frac{dp^T(t)}{dt} \mid \frac{d^2 p^T(t)}{dt^2} \mid \dots \right]. \quad (5-26)$$

By a theorem by Chen [C-1] from the fact that  $p^T(t)$  is an  $n \times m$  analytic matrix with linearly independent rows

---

\*In this case the decomposition (5-25) is said to be globally reduced. It can be shown (Youla Y-1) that  $n$  is the dimension of a minimal realization of  $W(t, \tau)$ .

$$\text{rank} \left[ p^T(t) \mid \frac{dp^T(t)}{dt} \mid \frac{d^2 p^T(t)}{dt^2} \mid \dots \right] = n. \quad (5-27)$$

This immediately implies that

$$\text{rank} \left[ \frac{d^k p^T(t)}{dt^k} \mid \frac{d^{k+1} p^T(t)}{dt^{k+1}} \mid \dots \right] \leq n, \quad k = 0, 1, 2, \dots, \quad (5-28)$$

since the matrix in (5-28) is a submatrix of the one in (5-27). Equation (5-27) implies that

$$\text{rank} \left[ \frac{d^p q^T(\bar{t})}{d\tau^p} p^T(t) \mid \frac{d^p q^T(\bar{t})}{d\tau^p} \cdot \frac{dp^T(t)}{dt} \mid \frac{d^p q^T(\bar{t})}{d\tau^p} \frac{d^2 p^T(t)}{dt^2} \mid \dots \right] \leq n \quad (5-29)$$

That is,

$$\text{rank} \left[ w_{0,p}^T(t, \bar{t}) \mid w_{1,p}^T(t, \bar{t}) \mid w_{2,p}^T(t, \bar{t}) \mid \dots \right] \leq n. \quad (5-30)$$

Next define

$$m_q^p(t) = \text{rank} \left[ w_{0,p}^T(t, \bar{t}) \mid w_{1,p}^T(t, \bar{t}) \mid \dots \mid w_{q-1,p}^T(t, \bar{t}) \right], \quad (5-31)$$

and

$$m_q^p = \sup_t m_q^p(t). \quad (5-32)$$

Then by another theorem by Chen [C-1] if  $\alpha$  is the least integer such that  $m_\alpha^p = m_{\alpha+1}^p = v$ , then

$$m_\alpha^p = m_\beta^p = v, \quad \beta = \alpha+1, \alpha+2, \dots, \quad (5-33)$$

and

$$\alpha \leq n - m_1^p + 1, \quad \alpha \leq v \leq \alpha \cdot m. \quad (5-34)$$

Since  $1 \leq m_1^p \leq n$ , by equation (5-34)  $\alpha \leq n$ . Hence it has

been shown that

$$\begin{aligned} & \text{rank} \begin{bmatrix} W_{0,0}(t, \bar{\tau}) & W_{0,1}(t, \bar{\tau}) & . & . & . \\ W_{1,0}(t, \bar{\tau}) & W_{1,1}(t, \bar{\tau}) & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \end{bmatrix} \\ & = \text{rank} \begin{bmatrix} W_{0,0}(t, \bar{\tau}) & W_{0,1}(t, \bar{\tau}) & . & . & . \\ W_{1,0}(t, \bar{\tau}) & W_{1,1}(t, \bar{\tau}) & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ W_{n-1,0}(t, \bar{\tau}) & W_{n-1,1}(t, \bar{\tau}) & . & . & . \end{bmatrix} \leq n. \end{aligned} \quad (5-35)$$

Thus a basis for the linear space spanned by the rows of  $S(t, \bar{\tau})$  can be found among the rows of the matrix

$$S_{n,\infty}(t, \bar{\tau}) = \begin{bmatrix} W_{0,0}(t, \bar{\tau}) & W_{0,1}(t, \bar{\tau}) & W_{0,2}(t, \bar{\tau}) & . & . & . \\ W_{1,0}(t, \bar{\tau}) & W_{1,1}(t, \bar{\tau}) & W_{1,2}(t, \bar{\tau}) & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ W_{n-1,0}(t, \bar{\tau}) & W_{n-1,1}(t, \bar{\tau}) & W_{n-1,2}(t, \bar{\tau}) & . & . & . \end{bmatrix} \quad (5-36)$$

The following theorem is now justified.

Theorem 5-5:

Let  $W(t, \tau)$  be an analytic realizable impulse response matrix. Then the matrix  $S(t, \bar{\tau})$  defined by  $W(t, \tau)$  has the property that a finite dimensional basis of the space spanned by its row vectors can be found. Furthermore the vector  $x(t|s)$  defined by taking the inner product of

these row vectors and  $Z(t,s)$  is a basis for the predictor space  $L_2(\hat{y};t|s)$ .

Hence a finite dimensional basis for the predictor space  $L_2(\hat{y};t|s)$  can be found. As in chapter 4 it is necessary to restrict the basis to the case  $x(s|s) = x(t|s)|_{t=s}$  in order that  $\frac{dx(t|s)}{ds}$  be defined. All of the criteria of theorem 3-4 have now been shown to be satisfied and the representation

$$dx(s|s) = F(s)x(s|s) + G(s)dz(s) \quad (5-37)$$

$$y(s) = H(s)x(s|s) \quad (5-38)$$

is justified. The solution to (5-37) and (5-38) can be represented as

$$x(s|s) = \int_{t_0}^s \phi(s,\tau)G(\tau)dz(\tau) \quad (5-39)$$

$$y(s) = H(s) \int_{t_0}^s \phi(s,\tau)G(\tau)dz(\tau) \quad (5-40)$$

where  $\phi(s,\tau)$  is the transition matrix for equation (5-37) satisfying the property

$$\frac{\partial \phi(s,\tau)}{\partial s} = F(s)\phi(s,\tau)$$

Theorem 5-6:

The matrices  $F(s)$ ,  $G(s)$ , and  $\phi(s,\tau)$  occurring in equations (5-39) and (5-40) determine a minimal realization for the impulse response matrix  $W(t,\tau)$ .

Proof:

Let  $t_0 < \tau < t$ . Using equation (5-40) with  $dz(\tau)$  replaced by  $\eta(\tau)d\tau$

$$\begin{aligned} V_{y\eta}(t, \tau) &= H(t) E \left\{ \int_{t_0}^t \phi(t, r) G(r) \eta(r) dr \eta^T(\tau) \right\} \\ &= H(t) \int_{t_0}^t \phi(t, r) G(r) \delta(r - \tau) dr \\ &= H(t) \phi(t, \tau) G(\tau). \end{aligned} \quad (5-42)$$

But using equation (5-2) for  $t_0 < \tau < t$

$$\begin{aligned} V_{y\eta}(t, \tau) &= E \left\{ \int_{t_0}^t W(t, r) \eta(r) dr \eta^T(\tau) \right\} \\ &= W(t, \tau). \end{aligned} \quad (5-43)$$

Hence for  $t_0 < \tau < t$

$$W(t, \tau) = H(t) \phi(t, \tau) G(\tau). \quad (5-44)$$

By the same argument as for theorem 4-6 equation (5-44) can be extended to hold for  $t_0 \leq \tau \leq t$ . Hence the representation (5-37) and (5-38) determines a realization for  $W(t, \tau)$ . The realization must be minimal since by the discussion preceeding theorem 5-5 the number of components in  $x(t|s)$  is less than or equal to  $n$ , the dimension of a minimal realization of  $W(t, \tau)$ . Since  $x(t|s)$  does determine a realization, it has in fact  $n$  components.

Q.E.D.

Corollary:

Let  $n$  be the dimension of a finite dimensional basis of  $L_2(\hat{y}; t|s)$ . Then  $\text{rank } S_{n,\infty}(t, \bar{\tau}) = n$ .

Proof:

The corollary is proved by the discussion at the end of the proof of theorem 5-6 and equation (5-19).

Q.E.D.

This completes the results for this section. In the next section the algorithm by Silverman is considered for the case of a time-varying analytic impulse response matrix.

## 2. Silverman's Algorithm for Time-Varying Systems

This section uses the predictor space approach to analyze Silverman's algorithm for time-varying systems in the case where the system impulse response matrix is analytic. Silverman's algorithm for the time-varying case is actually stated for a class of systems larger than just those having an analytic impulse matrix. Hence it must be realized that the analysis presented here does not reflect the full generality of the algorithm.

Define the matrix

$$S_{N_1 N_2}(t, \tau) = \begin{bmatrix} W_{0,0}(t, \tau) & W_{0,1}(t, \tau) & \dots & W_{0, N_2-1}(t, \tau) \\ W_{1,0}(t, \tau) & W_{1,1}(t, \tau) & \dots & W_{1, N_2-1}(t, \tau) \\ \vdots & \vdots & \ddots & \vdots \\ W_{N_1-1,0}(t, \tau) & W_{N_1-1,1}(t, \tau) & \dots & W_{N_1-1, N_2-1}(t, \tau) \end{bmatrix} \quad (5-46)$$

for  $N_1, N_2 = 1, 2, \dots$ . Then the following theorem by Silverman [S-2] may be stated.

Theorem 5-7 (Silverman):

Suppose that for some finite positive integers  $\ell, k$  and  $n$  ( $\ell \leq n, k \leq n$ ) the matrices  $S_{\ell k}(t, \tau)$  and  $S_{\ell+1, k+1}(t, \tau)$  have rank  $n$  for all  $t \geq \tau$  and a fixed  $n \times n$  submatrix  $A(t, \tau)$  of  $S_{\ell k}(t, \tau)$  also has rank  $n$  for all  $t \geq \tau$ . Then  $W(t, \tau)$  is realizable as an  $n$ th order uniformly observable and uniformly controllable system.

The realization is given by the following construction which parallels that for the stationary case. Let  $M(t, \tau)$  denote the  $n \times k : r$  submatrix formed by the first  $n$  linearly independent rows of  $S_{\ell k}(t, \tau)$ . Define the  $n \times k : r$  matrix  $\bar{M}(t, \tau)$  as a submatrix of  $S_{\ell+1, k}(t, \tau)$  in the following manner. If  $s_j(t, \tau)$  is one of the scalar rows of  $S_{\ell k}(t, \tau)$  selected for the matrix  $M(t, \tau)$ , then the row  $s_{j+m}(t, \tau)$  is selected as the corresponding row in the matrix  $\bar{M}(t, \tau)$ . The following four submatrices of  $S_{\ell+1, k}(t, \tau)$  are then defined.

$A(t, \tau)$  -- The  $n \times n$  submatrix of  $M(t, \tau)$  having rank  $n$  for all  $t \geq \tau$ .

$\bar{A}(t, \tau)$  -- The  $n \times n$  matrix occupying the same column positions in  $\bar{M}(t, \tau)$  as does  $A$  in  $M(t, \tau)$ .

$A_1(t, \tau)$  -- The  $m \times n$  matrix occupying the same column positions in  $S_{1k}(t, \tau)$  as does

$A(t, \tau)$  in  $M(t, \tau)$ .

$A_2(t, \tau)$  -- The  $n \times r$  matrix occupying the first  $r$  columns of  $M(t, \tau)$ .

Then setting

$$\begin{aligned} F(t) &= \bar{A}(t, t)A^{-1}(t, t), \quad G(t) = A_2(t, t), \\ H(t) &= A_1(t, t)A^{-1}(t, t) \end{aligned} \quad (5-47)$$

yields a minimal realization of the impulse response matrix  $W(t, \tau)$ .

It should be noted that theorem 5-7 gives a sufficient condition, but not a necessary condition for a realization to exist. Furthermore the minimality of the realization is equivalent to controllability and observability ([S-1], Theorem 24). The properties of uniform controllability and uniform observability are much stronger than controllability and observability however [S-4].

The necessary machinery for the analysis of the algorithm is developed next.

Theorem 5-8:

Suppose it is known that the analytic impulse response matrix  $W(t, \tau)$  can be realized by a system of dimension  $n$ . Then the search for a basis of the predictor space can be limited to the components of the vector  $\Gamma_n(t|s)$  defined by

$$\Gamma_n(t|s) = \left[ \hat{y}(t|s)^T, \frac{\partial \hat{y}(t|s)^T}{\partial t}, \dots, \frac{\partial^{n-1} \hat{y}(t|s)^T}{\partial t^{n-1}} \right]. \quad (5-48)$$

Proof:

By the corollary to theorem 5-6

$$\text{rank } S_{n,\infty}(t, \bar{\tau}) = n$$

where  $n$  is the dimension of a minimal realization of  $W(t, \tau)$ . Furthermore, by equation (5-35)

$$\text{rank } S(t, \bar{\tau}) \leq n .$$

Hence by equation (5-19)  $\Gamma(t|s)$  can have no more than  $n$  linearly independent elements. By the corollary to theorem 5-6 and equation (5-19)  $\Gamma_n(t|s)$  must contain  $n$  linearly independent components. Hence  $\Gamma_n(t|s)$  contains a maximal linearly independent subset of the components of  $\Gamma(t|s)$  and thus by theorem 5-3 a basis for the predictor space.

Q.E.D.

That the search for a basis of the linear space spanned by the row vectors of  $S(t, \bar{\tau})$  may be restricted to the first  $n$  block rows is already established by equation (5-35) and the corollary to theorem 5-6. It can also be shown that it is only necessary to consider the first  $n$  block columns. By an argument entirely analogous to that preceeding equation (5-35) it can be shown that

$$\begin{aligned}
& \text{rank} \begin{bmatrix} W_{0,0}(t, \bar{t}) & W_{0,1}(t, \bar{t}) & W_{0,2}(t, \bar{t}) & \dots & \\ W_{1,0}(t, \bar{t}) & W_{1,1}(t, \bar{t}) & W_{1,2}(t, \bar{t}) & \dots & \\ W_{2,0}(t, \bar{t}) & W_{2,1}(t, \bar{t}) & W_{2,2}(t, \bar{t}) & \dots & \\ \vdots & \vdots & \vdots & \ddots & \\ \vdots & \vdots & \vdots & \ddots & \end{bmatrix} \\
& = \text{rank} \begin{bmatrix} W_{0,0}(t, \bar{t}) & W_{0,1}(t, \bar{t}) & \dots & W_{0,n-1}(t, \bar{t}) \\ W_{1,0}(t, \bar{t}) & W_{1,1}(t, \bar{t}) & \dots & W_{1,n-1}(t, \bar{t}) \\ W_{2,0}(t, \bar{t}) & W_{2,1}(t, \bar{t}) & \dots & W_{2,n-1}(t, \bar{t}) \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \end{bmatrix} \leq n \quad (5-49)
\end{aligned}$$

Equation (5-49) implies that in searching for the basis of the linear space spanned by  $S_{n,\infty}(t, \bar{t})$  that it is not necessary to consider elements in the columns occurring after the first  $n$  block columns in testing for linear independence. The linear independence will show up in the first  $n$  block columns and attention may be restricted to  $S_{nn}(t, \bar{t})$ .

An analogue to equation (4-76) will now be developed.

Theorem 5-9:

$$E \left\{ \frac{\partial^k \hat{y}(t|t)}{\partial t^k} \eta^{(j)}(\bar{t})^T \right\} = \begin{cases} \frac{1}{2} W_{k,j}(t, \bar{t}), & \bar{t} = t_0 \text{ or } \bar{t} = t \\ W_{k,j}(t, \bar{t}), & t_0 < \bar{t} < t \end{cases} \quad (5-50)$$

Proof:

From equation (A-15) of the appendix for  $j, k = 0, 1, 2, \dots$ ,

$$\begin{aligned}
E \left\{ \frac{\partial^k \hat{y}(t|t)}{\partial t^k} \eta^{(j)}(\bar{\tau}) \right\} &= E \left\{ \int_{t_0}^t \frac{\partial^{k+j} W(t, \tau)}{\partial t^k \partial \tau^j} dz(\tau) \eta^T(\bar{\tau}) \right\} \\
&= E \left\{ \int_{t_0}^t \frac{\partial^{k+j} W(t, \tau)}{\partial t^k \partial \tau^j} \eta(\tau) d\tau \eta^T(\bar{\tau}) \right\} \\
&= \int_{t_0}^t \frac{\partial^{k+j} W(t, \tau)}{\partial t^k \partial \tau^j} \delta(\tau - \bar{\tau}) d\tau \\
&= \begin{cases} \frac{1}{2} \cdot \frac{\partial^{k+j} W(t, \bar{\tau})}{\partial t^k \partial \tau^j} & \bar{\tau} = t \text{ or } \bar{\tau} = t_0 \\ \frac{\partial^{k+j} W(t, \bar{\tau})}{\partial t^k \partial \tau^j} & t_0 < \bar{\tau} < t \\ 0 & \bar{\tau} \notin [t_0, t] \end{cases} \quad (5-51)
\end{aligned}$$

Q.E.D.

Corollary 1:For all positive integers  $n$  and  $t_0 < \bar{\tau} < t$ 

$$\begin{aligned}
&E \left\{ \begin{bmatrix} \hat{y}(t|t) \\ \frac{\partial \hat{y}(t|t)}{\partial t} \\ \vdots \\ \frac{\partial^{n-1} \hat{y}(t|t)}{\partial t^{n-1}} \end{bmatrix} \begin{bmatrix} \eta(\bar{\tau})^T, \eta^1(\bar{\tau})^T, \dots, \eta^{(n-1)}(\bar{\tau})^T \end{bmatrix} \right\} \\
&= \begin{bmatrix} W_{0,0}(t, \bar{\tau}) & W_{0,1}(t, \bar{\tau}) & \dots & W_{0,n-1}(t, \bar{\tau}) \\ W_{1,0}(t, \bar{\tau}) & W_{1,1}(t, \bar{\tau}) & \dots & W_{1,n-1}(t, \bar{\tau}) \\ \vdots & \vdots & & \vdots \\ W_{n-1,0}(t, \bar{\tau}) & W_{n-1,1}(t, \bar{\tau}) & \dots & W_{n-1,n-1}(t, \bar{\tau}) \end{bmatrix} \quad (5-52)
\end{aligned}$$

Proof:

The corollary is a direct consequence of the lemma.

Q.E.D.

Equation (5-52) will be represented by  $E\{\Gamma_n(t, t) \Upsilon_n(\bar{t})\} = S_{nn}(t, \bar{t})$ . The symbols  $\Gamma_n(t|t)$  and  $S_{nn}(t, \bar{t})$  are already defined and

$$\Upsilon_n(\bar{t}) = \left[ \eta(\bar{t})^T, \eta^1(\bar{t})^T, \dots, \eta^{(n-1)}(\bar{t})^T \right]. \quad (5-53)$$

Corollary 2:

For all positive integers  $n$  and  $\bar{t} = t_0$  or  $\bar{t} = t$

$$E \left\{ \Gamma_n(t|t) \Upsilon_n(\bar{t}) \right\} = \frac{1}{2} S_{nn}(t, \bar{t}). \quad (5-54)$$

Proof:

The corollary is a direct consequence of the theorem.

Q.E.D.

Suppose now that the hypotheses of theorem 5-7 are satisfied. Then from theorem 5-8 the vector  $\Gamma_n(t|t)$  contains an  $n$  dimensional basis  $x(t|t)$  for the predictor space  $L_2(\hat{y}; t|s)$ . It is clear from equation (5-54) that

$$M(t, t) = 2E\{x(t|t) \Upsilon_n(t)\} \quad (5-55)$$

and by the same reasoning as in the stationary case

$$\bar{M}(t, t) = 2E\left\{ \frac{\partial x(t|t)}{\partial t} \Upsilon_n(t) \right\}. \quad (5-56)$$

The  $n \times n$  matrix  $A(t, t)$  is a submatrix of  $M(t, t)$  selected by taking the linearly independent columns of  $M(t, t)$ . Denoting the linearly independent elements of  $\eta_n(t)$  corresponding to the linearly independent columns of  $M(t, t)$  by  $v^T(t)$ ,

$$A(t, t) = 2E\{x(t|t)v^T(t)\}. \quad (5-57)$$

The  $n \times n$  matrix  $\bar{A}(t, t)$  occupying the same column positions in  $\bar{M}(t, t)$  as does  $A$  in  $M(t, t)$  is then given by

$$\bar{A}(t, t) = 2E\left\{\frac{\partial x(t|t)}{\partial t} v^T(t)\right\}. \quad (5-58)$$

The  $m \times n$  matrix  $A_1(t, t)$  occupying the same column positions in  $S_{1n}(t, t)$  as does  $A(t, t)$  in  $M(t, t)$  is given by

$$A_1(t, t) = 2E\{\hat{y}(t|t)v^T(t)\}. \quad (5-59)$$

Finally, the  $n \times r$  matrix  $A_2(t, t)$  occupying the first  $r$  columns of  $M(t, t)$  is given by

$$A_2(t, t) = 2E\{x(t|t)\eta^T(t)\}. \quad (5-60)$$

Now using equation (5-38) and the fact that  $\hat{y}(t|t) = y(t)$

$$\begin{aligned} A_1(t, t)A^{-1}(t, t) &= 2E\{H(t)x(t|t)v^T(t)\}[2E\{x(t|t)v^T(t)\}]^{-1} \\ &= H(t). \end{aligned} \quad (5-61)$$

Next, using equation (5-39)

$$\begin{aligned}
A_2(t, t) &= 2E\{x(t|t)\eta^T(t)\} \\
&= 2E\left\{\int_{t_0}^t \phi(t, \tau)G(\tau)\eta(\tau)d\tau \eta^T(t)\right\} \\
&= 2 \int_{t_0}^t \phi(t, \tau)G(\tau)\delta(\tau-t)d\tau \\
&= 2 \cdot \frac{1}{2}\phi(t, t)G(t) \\
&= G(t).
\end{aligned} \tag{5-62}$$

By a proof virtually identical to that for theorem 4-11 it can be shown that

$$\frac{\partial x(t|t)}{\partial t} = F(t)x(t|t). \tag{5-63}$$

Hence

$$\begin{aligned}
\bar{A}(t, t)A^{-1}(t, t) &= 2E\{F(t)x(t|t)v^T(t)\}[2E\{x(t|t)v^T(t)\}]^{-1} \\
&= F(t).
\end{aligned} \tag{5-64}$$

Thus the verification of Silverman's algorithm for the case of an analytic impulse response matrix turns out to be nearly identical to that for the stationary case. As noted in the discussion after theorem 5-7 the realization must be observable and controllable. The properties of uniform observability and controllability will not be discussed here since consideration of them would involve developing the properties of generalized controllability and observability matrices, (see Silverman [S-2] or Silverman and Meadows [S-3]) a topic outside of the scope of this

dissertation. A simple example is given next to illustrate the concepts used here.

### 3. Example

Consider the linear time-varying system

$$\frac{d}{dt} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} t & t^2 \\ 0 & t \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix} \quad (5-65)$$

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}. \quad (5-66)$$

The transition matrix for this system,  $\Phi(t, \tau)$ , is given by

$$\Phi(t, \tau) = \begin{bmatrix} e^{\frac{t^2 - \tau^2}{2}} & e^{\frac{t^2 - \tau^2}{2}} \frac{(t^3 - \tau^3)}{3} \\ 0 & e^{\frac{t^2 - \tau^2}{2}} \end{bmatrix}. \quad (5-67)$$

The impulse response matrix  $W(t, \tau)$  for  $t \geq \tau$

is

$$W(t, \tau) = \begin{bmatrix} e^{\frac{t^2 - \tau^2}{2}} & 2e^{\frac{t^2 - \tau^2}{2}} \frac{(t^3 - \tau^3)}{3} \\ 0 & 2e^{\frac{t^2 - \tau^2}{2}} \end{bmatrix}. \quad (5-68)$$

For  $t \geq \tau$

$$\begin{bmatrix} \frac{1}{n!} \frac{\partial^n}{\partial t^n} \left[ e^{\frac{t^2 - \tau^2}{2}} \right] & \frac{1}{n!} \frac{\partial^n}{\partial t^n} \left[ e^{\frac{t^2 - \tau^2}{2}} \frac{(t^3 - \tau^3)}{3} \right] \\ \frac{1}{n!} \frac{\partial^n}{\partial t^n} \left[ e^{\frac{t^2 - \tau^2}{2}} \frac{(t^3 - \tau^3)}{3} \right] & \frac{1}{n!} \frac{\partial^n}{\partial t^n} \left[ e^{\frac{t^2 - \tau^2}{2}} \right] \\ \frac{1}{n!} \frac{\partial^n}{\partial t^n} \left[ e^{\frac{t^2 - \tau^2}{2}} \right] & \frac{1}{n!} \frac{\partial^n}{\partial t^n} \left[ e^{\frac{t^2 - \tau^2}{2}} \frac{(t^3 - \tau^3)}{3} \right] \\ \frac{1}{n!} \frac{\partial^n}{\partial t^n} \left[ e^{\frac{t^2 - \tau^2}{2}} \frac{(t^3 - \tau^3)}{3} \right] & \frac{1}{n!} \frac{\partial^n}{\partial t^n} \left[ e^{\frac{t^2 - \tau^2}{2}} \right] \end{bmatrix} \quad (5-69)$$

The matrix has been partitioned into the submatrices

$W_{j,k}(t,\tau)$ . For  $t = \tau$

$$S_{33}(t,t) = \left[ \begin{array}{cc|cc|cc} 1 & 0 & -t & -2t^2 & t^2-1 & 4t(t^2-1) \\ 0 & 2 & 0 & -2t & 0 & 2(t^2-1) \\ t & 2t^2 & -t^2 & -4t^3 & t(t^2-1) & 6t^2(t^2-1) \\ 0 & 2t & 0 & -2t^2 & 0 & 2t(t^2-1) \\ t^2+1 & 4t(t^2+1) & -t(t^2+1) & -6t^2(t^2+1) & (t^2-1)(t^2+1) & 8t(t^2+1)(t^2-1) \\ 0 & 2(t^2+1) & 0 & -2t(t^2+1) & 0 & 2t^2(t^2-1) \end{array} \right] \quad (5-70)$$

The conditions of theorem 5-7 are satisfied by  $S_{11}(t,\tau)$  and  $S_{22}(t,\tau)$ . It can easily be verified that  $\text{rank } S_{11}(t,\tau) = 2$  and  $\text{rank } S_{22}(t,\tau) = 2$  for all  $\tau \geq t$ . The fixed  $n \times n$  submatrix  $A(t,\tau)$  of  $S_{kk}(t,\tau) = S_{11}(t,\tau)$  mentioned in theorem 5-7 is  $S_{11}(t,\tau)$ . That is

$$A(t,\tau) = \left[ \begin{array}{cc} e^{\frac{t^2-\tau^2}{2}} & 2/3e^{\frac{t^2-\tau^2}{2}}(t^3-\tau^3) \\ 0 & 2e^{\frac{t^2-\tau^2}{2}} \end{array} \right]. \quad (5-71)$$

A realization is now constructed from  $S_{33}(t,t)$  by using Silverman's algorithm. The matrix  $M(t,\tau)$  is just  $A(t,\tau)$ .

The matrix  $\bar{M}(t,\tau)$  is given by

$$\bar{M}(t,\tau) = \left[ \begin{array}{cc} te^{\frac{t^2-\tau^2}{2}} & 2e^{\frac{t^2-\tau^2}{2}} \left[ t^2 + \frac{1}{3}(t^3-\tau^3)t \right] \\ 0 & 2te^{\frac{t^2-\tau^2}{2}} \end{array} \right]. \quad (5-72)$$

The matrix  $A(t, \tau)$  is already defined and  $\bar{A}(t, \tau)$  is  $\bar{M}(t, \tau)$ . The matrices  $A_1(t, \tau)$  and  $A_2(t, \tau)$  both equal  $A(t, \tau)$ . Thus

$$A_2(t, t) = A(t, t) = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \quad (5-73)$$

and

$$A_1(t, t)A^{-1}(t, t) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (5-74)$$

yield the input and output coefficient matrices as expected. Also

$$\begin{aligned} \bar{A}(t, t)A^{-1}(t, t) &= \begin{bmatrix} t & 2t^2 \\ 0 & 2t \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \\ &= \begin{bmatrix} t & t^2 \\ 0 & t \end{bmatrix} \end{aligned} \quad (5-75)$$

as expected.

Next the predictor space approach is illustrated.

Denote the two components of  $\hat{y}(t|s)$  by  $\hat{y}_1(t|s)$  and  $\hat{y}_2(t|s)$ . Also denote the  $\alpha\beta$ th component of the  $2 \times 2$  matrix  $W_{jk}(t, \tau)$  by  $W_{jk}^{\alpha\beta}(t, \tau)$ . Then by equation (5-19)

$$\begin{aligned} \hat{y}_1(t|s) &= \sum_{k=0}^{\infty} W_{ok}^{11}(t, \bar{\tau}) \int_{t_0}^s \frac{(\tau - \bar{\tau})^k}{k!} \eta_1(\tau) d\tau \\ &+ \sum_{k=0}^{\infty} W_{ok}^{12}(t, \bar{\tau}) \int_{t_0}^s \frac{(\tau - \bar{\tau})^k}{k!} \eta_2(\tau) d\tau \end{aligned} \quad (5-76)$$

and

$$\hat{y}_2(t|s) = \sum_{k=0}^{\infty} W_{ok}^{22}(t, \bar{\tau}) \int_{t_0}^s \frac{(\tau - \bar{\tau})^k}{k!} \eta_2(\tau) d\tau. \quad (5-77)$$

Also

$$\begin{aligned} \frac{\partial \hat{y}_1(t|s)}{\partial t} &= \sum_{k=0}^{\infty} W_{1k}^{11}(t, \bar{\tau}) \int_{t_0}^s \frac{(\tau - \bar{\tau})^k}{k!} \eta_1(\tau) d\tau \\ &+ \sum_{k=0}^{\infty} W_{1k}^{12}(t, \bar{\tau}) \int_{t_0}^s \frac{(\tau - \bar{\tau})^k}{k!} \eta_2(\tau) d\tau \end{aligned} \quad (5-78)$$

and

$$\frac{\partial \hat{y}_2(t|s)}{\partial t} = \sum_{k=0}^{\infty} W_{2k}^{22}(t, \bar{\tau}) \int_{t_0}^s \frac{(\tau - \bar{\tau})^k}{k!} \eta_2(\tau) d\tau. \quad (5-79)$$

The basis of the predictor space  $x(t|t)$  is just the vector  $\hat{y}(t|t)$ . Consider equation (5-57). It is assumed that the white noise input vector to the system,  $\eta(t) = [\eta_1(t), \eta_2(t)]^T$ , consists of independent white noises  $\eta_1(t)$  and  $\eta_2(t)$  each with unit covariance. Hence  $v(t) = \eta(t)$  and

$$\begin{aligned} A(t, t) &= 2E\{x(t|t)v^T(t)\} \\ &= 2E\{\hat{y}(t|t)\eta^T(t)\} \end{aligned}$$

$$\begin{aligned} &= 2E \left[ \begin{array}{c} \sum_{k=0}^{\infty} W_{0k}^{11}(t, t) \int_{t_0}^t \frac{(\tau - t)^k}{k!} \eta_1(\tau) d\tau + \sum_{k=0}^{\infty} W_{0k}^{12}(t, t) \int_{t_0}^t \frac{(\tau - t)^k}{k!} \eta_2(\tau) d\tau \\ \sum_{k=0}^{\infty} W_{0k}^{22}(t, t) \int_{t_0}^t \frac{(\tau - t)^k}{k!} \eta_2(\tau) d\tau \end{array} \right] \begin{bmatrix} \eta_1(t) \\ \eta_2(t) \end{bmatrix} \\ &= 2E \left[ \begin{array}{cc} \sum_{k=0}^{\infty} W_{0k}^{11}(t, t) \int_{t_0}^t \frac{(\tau - t)^k}{k!} \delta(\tau - t) d\tau & \sum_{k=0}^{\infty} W_{0k}^{12}(t, t) \int_{t_0}^t \frac{(\tau - t)^k}{k!} \delta(\tau - t) d\tau \\ 0 & \sum_{k=0}^{\infty} W_{0k}^{22}(t, t) \int_{t_0}^t \frac{(\tau - t)^k}{k!} \delta(\tau - t) d\tau \end{array} \right] \end{aligned} \quad (5-80)$$

$$\begin{aligned} &= \begin{bmatrix} W_{00}^{11}(t, t) & W_{00}^{12}(t, t) \\ 0 & W_{00}^{22}(t, t) \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \end{aligned} \quad (5-81)$$

Similarly

$$\begin{aligned}
 \bar{A}(t, t) &= 2E \left\{ \frac{\partial x(t|t)}{\partial t} v^T(t) \right\} \\
 &= 2E \left\{ \frac{\partial \hat{y}(t|t)}{\partial t} \eta^T(t) \right\} \\
 &= \begin{bmatrix} w_{10}^{11}(t, t) & w_{10}^{12}(t, t) \\ 0 & w_{10}^{22}(t, t) \end{bmatrix} \\
 &= \begin{bmatrix} t & 2t^2 \\ 0 & 2t \end{bmatrix}.
 \end{aligned} \tag{5-82}$$

The matrices  $A_1(t, t)$  and  $A_2(t, t)$  both equal  $A(t, t)$  and so all four matrices necessary for the implementation of the Silverman algorithm have been computed from the predictor space approach.

With this example the discussion of the realization problem for analytic impulse response matrices is concluded.

## CHAPTER SIX

### CONCLUSION

In this thesis a new approach to the realization problem, the predictor space approach, has been extended to continuous time finite dimensional systems having analytic impulse response matrices. This approach provides a unifying overview to the algorithms for the solution to the realization problem.

#### 1. Summary of Results

A representation problem which provides the foundation of the predictor space approach is considered first. There is given two second order zero mean stochastic processes  $y(t)$  and  $z(t)$ , with  $y(t)$  measurable and  $z(t)$  purely nondeterministic and continuous in quadratic mean. The predictor space is defined as the Hilbert space of random variables spanned by the projections of  $y(t)$  onto the space spanned by  $z(s)$  for  $t \geq s$ . A finite basis for the predictor space is assumed and then it is shown that  $y(t)$  can be expressed in terms of this basis and an error term.

The predictor space forms the state space for the stochastic system representation. The dynamic behavior of

the basis is developed using a theorem by Halyo and McAlpine, the theory of weak convergence in Hilbert spaces, and the Hida-Cramer representation for purely nondeterministic processes that are continuous in quadratic mean.

To conclude the chapter two special cases are considered. It is shown that if the given purely nondeterministic process is white noise, then by considering it to be the innovation process for a measurable Wiener process the theory remains valid. Finally it is established that if the two given processes are stationary with stationary cross covariance, the system representation obtained has stationary coefficient matrices.

Next the realization problem for stationary systems is considered. Given an impulse response matrix known to be realizable, a stochastic system driven by white noise is defined whose input-output covariance equals the given impulse response matrix. It is shown that the input and output of the stochastic system satisfy the prerequisites for the application of the previously developed representation theory with the exception of the existence of a finite basis for the predictor space of the system output. The proof that such a basis can be found is then carried out. Having established that the representation theory developed in chapter 3 applies, it is shown that the coefficient matrices of the stochastic representation provide a solution to the realization problem for the given impulse response matrix.

Two of the key algorithms in realization theory, an algorithm due to Silverman and the Ho-Kalman algorithm are analyzed. It is shown that both can be interpreted in terms of various manipulations of a basis of the predictor space. An example is given which illustrates many of the concepts used.

The realization problem for time-varying analytic impulse response matrices is considered next. The development parallels that for the stationary case for the most part, although some theorems on the rank of analytic matrices are needed. The Silverman algorithm is analyzed for this case. Another example is also given to illustrate the theory.

## 2. Problems for Future Research

The most pressing problem for future research is to consider developing the predictor space approach for more general time varying systems. If this could be done it would furnish insight into an area where algorithms at present are sorely lacking. The algorithm by Silverman can be used for "constant rank systems", a somewhat more general class of systems than considered here. It would be of interest to extend the predictor space approach to at least this class of systems.

It might also be of interest to attempt an extension of the predictor space approach to some classes of non-linear systems. This is an area with few results at present and any insight shed by a new approach would be of interest.

## BIBLIOGRAPHY

## BIBLIOGRAPHY

- [A-1] Akaike, H., "Stochastic theory of minimal representation," IEEE Trans. Automat. Contr., AC-19 (1974), pp. 667-674.
- [C-1] Chen, C.T., "Linear Independence of Analytic Functions," SIAM J. Appl. Math., 15 (1967), pp. 1272-1274.
- [C-2] Cramér, H., "Stochastic processes as curves in Hilbert space," Theory Prob. Appl., 9 (1964), pp. 169-179.
- [C-3] Cramér, H., "On the structure of purely non-deterministic stochastic processes," Arkiv Mat., 4 (1961), pp. 249-266.
- [D-1] Desoer, C.A. and Varaiya, P., "The minimal realization of a nonanticipative impulse response matrix," SIAM J. Appl. Math., 15 (1967), pp. 754-764.
- [D-2] Doob, J.L., Stochastic Processes, John Wiley, New York, 1953.
- [G-1] Gel'fand, I.M. and Vilenkin, N. Ya., Generalized Functions, Vol. 4, Applications of Harmonic Analysis, Academic Press, New York, 1964.
- [G-2] Gilbert, E.G., "Controllability and observability in multivariable control systems," SIAM J. Control, 1 (1963), pp. 128-151.
- [G-3] Gikhman, I.J. and Skorokhod, A.V., Introduction to the Theory of Random Processes, W.B. Saunders Co., Philadelphia (1969).
- [H-1] Halyo, N. and McAlpine, G., "On the spectral factorization of nonstationary vector random processes," IEEE Trans. Automat. Contr., AC-19 (1974) pp. 674-679.
- [H-2] Ho, B.L. and Kalman, R.E., "Effective construction of linear state-variable models from input/output functions," in Proc. 3rd Allerton Conf. Circuits and Systems Theory (1965), pp. 449-459.

- [I-1] Ivkovich, Z. and Rozanov, Y.A., "On the canonical Hida-Cramér representation for random processes," Theory Prob. Appl. 16 (1971), pp. 351-356.
- [K-1] Kailath, T., "The innovations approach to detection and estimation theory," Proc. IEEE, 58 (1970), pp. 680-695.
- [K-2] Kailath, T., "RKHS approach to detection and estimation problems - part I: deterministic signals in Gaussian noise," IEEE Trans. Inform. Theory, IT-17 (1971), pp. 530-549.
- [K-3] Kalman, R.E., "Mathematical description of linear dynamical systems," SIAM J. Control, 1 (1963), pp. 152-192.
- [K-4] Kalman, K.E., Falb, P.L., and Arbib, M.A., Topics in Mathematical System Theory, McGraw-Hill, New York (1969).
- [K-5] Kallianpur, G. and Mandrekar, V., "Multiplicity and representation theory of purely non-deterministic stochastic processes," Theory Prob. Appl., 10 (1965), pp. 553-581.
- [L-1] Lee, K.Y. and Mandrekar, V., "Multiplicity and martingale approach to linear state estimation," Proc. Princeton Conference on Information Sciences and Systems, 1973.
- [M-1] Mandrekar, V. and Salehi, H., "Operator valued wide-sense Markov processes and solutions of infinite-dimensional linear differential systems driven by white noise," Mathematical System Theory, 4 (1970), pp. 340-356.
- [O-1] Olmstead, J.M.H., Real Variables, An Introduction to the Theory of Functions, Appleton-Century-Crofts, New York, 1959.
- [P-1] Padulo, L. and Arbib, M.A., System Theory, W.B. Saunders Co., Philadelphia, 1974.
- [P-2] Papoulis, A., Probability, Random Variables, and Stochastic Processes, McGraw-Hill, New York, 1965.
- [P-3] Parzen, E., Time Series Analysis Papers, Holden Day, San Francisco, 1967.
- [P-4] Prohorov, Yu, V. and Rozanov, Yu. A., Probability Theory, Springer-Verlag, New York, 1969.

- [R-1] Rudin, W., Real and Complex Analysis, McGraw-Hill, New York, 1966.
- [S-1] Silverman, L.M., "Realization of linear dynamical systems," IEEE Trans. Automat. Contr., AC-16 (1971), pp. 554-567.
- [S-2] Silverman, L.M., "Representation and realization of time-variable linear systems," Dept. Elec. Eng., Columbia Univ., New York, N.Y., Tech. Report 94, June 1966.
- [S-3] Silverman, L.M. and Meadows, H.E., "Equivalent realizations of linear systems," SIAM J. Appl. Math., 17 (1969), pp. 393-408.
- [S-4] Silverman, L.M. and Meadows, H.E., "Controllability and observability in time-variable linear systems," SIAM J. Control, 5 (1968), pp. 64-73.
- [S-5] Simmons, G.F., Introduction to Topology and Modern Analysis, McGraw-Hill, New York, 1963.
- [T-1] Tether, A., "Construction of minimal linear state-variable models from finite input-output data," IEEE Trans. Automat. Contr., AC-15 (1970), pp. 427-436.
- [W-1] Wiener, N. and Masani, P., "The prediction theory of multivariate stochastic processes," Acta. Math., 98 (1957), pp. 111-150.
- [W-2] Weiss, L. and Kalman, R.E., "Contributions to linear systems theory," Int. J. Eng. Sci., 3 (1965), pp. 141-171.
- [W-3] Wong, E., Stochastic Processes in Information and Dynamical Systems, McGraw-Hill, New York, 1971.
- [Y-1] Youla, D.C., "The synthesis of linear dynamical systems from prescribed weighting patterns," J. SIAM Appl. Math., 14 (1966), pp. 527-549.

## APPENDIX

## APPENDIX

### GENERALIZED RANDOM PROCESSES

Let  $(\Omega, A, P)$  be a probability space and consider the linear space  $L_2(\Omega, A, P)$  of all zero-mean finite variance random variables with norm equal to the variance. A second order random process, say  $\{x(t, \omega) : t \in T\}$ , is then a mapping associating with every  $t \in T$  a random variable  $x(t, \omega)$  in  $L_2(\Omega, A, P)$ . White noise is not such a second-order random process. To obtain a rigorous theory for white noise it is necessary to go to the theory of generalized random processes. An important reference in this area is Gel'fand and Vilenkin [G-1]. The book by Prohorov and Rozanov [P-4] is also helpful and an intuitive discussion may be found in Kailath [K-2].

Let  $U$  be a real or complex linear space of elements  $u$ . A function  $x = x(u)$  of the parameter  $u \in U$  whose values  $x(u)$  are real or complex random variables on a probability space  $(\Omega, A, P)$  is called a random functional. The random functional  $x(u)$  is called linear if for any  $u, v \in U$  and numbers  $\alpha, \beta$

$$x(\alpha u + \beta v) = \alpha x(u) + \beta x(v) \quad (A-1)$$

with probability 1. The finite dimensional probability

distribution functions of a generalized process are defined, in an analogous manner to the usual case, by

$$F(a_1, a_2, \dots, a_n) =$$

$$P\{\omega: x(u_1, \omega) \leq a_1, x(u_2, \omega) \leq a_2, \dots, x(u_n, \omega) \leq a_n\} \quad (A-2)$$

for all finite collections  $\{u_1, \dots, u_n\}$  of elements of  $U$ .

The linear space  $U$  is usually taken to be the set  $D$  of all real or complex functions  $\varphi$  defined on the real line which are infinitely differentiable and have compact support. The concept of convergence used to define a topology on  $D$  is that  $\varphi_n$  converges to  $\varphi$  as  $n$  goes to infinity, denoted  $\varphi_n \rightarrow \varphi$ , if all functions  $\varphi_n = \varphi_n(t)$  vanish outside a fixed finite interval which is the same for each function and  $\frac{d^k \varphi_n(t)}{dt^k}$  converges uniformly to  $\frac{d^k \varphi(t)}{dt^k}$ , as  $n$  goes to infinity, for each derivative of order  $k = 0, 1, 2, \dots$ . The space  $D$  equipped with this topology is called the space of testing functions.

The random functional  $x(\varphi)$  defined on  $D$  is called continuous if the convergence in  $D$  of the functions  $\varphi_{kj}(t)$  to  $\varphi_j(t)$ ,  $j = 1, \dots, n$ , implies

$$\lim_{k \rightarrow \infty} (x(\varphi_{k1}), \dots, x(\varphi_{kn})) = (x(\varphi_1), \dots, x(\varphi_n)). \quad (A-3)$$

A continuous linear random functional on  $D$  is called a generalized random function. In the case that  $D$  consists of functions of one variable, the corresponding random

function will be called a generalized random process. This is the case of interest here and for the rest of this appendix D will mean the space of testing functions restricted to functions of one variable.

A number of operations are defined on generalized random processes in a manner analogous to that by which they are defined for generalized functions. By a linear combination  $\alpha x_1 + \beta x_2$  of the generalized random processes  $x_1$  and  $x_2$  is meant the random process  $x$  associating with every function  $\varphi(t) \in D$  the generalized random variable  $\alpha x_1(\varphi) + \beta x_2(\varphi)$ . The product  $f(t) \cdot x$  of an infinitely differentiable function  $f(t)$  and a generalized random process  $x$  is defined as the process for which there corresponds to the function  $\varphi(t) \in D$  the generalized random process  $x(f\varphi)$ . The derivative of a generalized random process  $x(\varphi)$  for  $\varphi \in D$  is defined by

$$x'(\varphi) = -x(\varphi'). \quad (\text{A-4})$$

It is important to note that the derivative of a generalized random process always exists and is a generalized random process. This is in contrast to ordinary random processes which may not have derivatives that are ordinary random processes. In particular, the derivative of the Wiener process does not exist as an ordinary random process, but it does exist as a generalized random process.

The material necessary to establish lemma 4-3 and theorem 5-9 is now developed. It will first be shown that

$$E \left\{ \frac{\partial^k \hat{y}(t|t)}{\partial t^k} \eta^{(j)}(\bar{\tau}) \right\} = E \left\{ \int_{t_0}^t \frac{\partial^{k+j} W(t, \tau)}{\partial t^k \partial \tau^j} \eta(\tau) d\tau \eta^T(\bar{\tau}) \right\} \quad (A-5)$$

where

$$\begin{aligned} \frac{\partial^k \hat{y}(t|t)}{\partial t^k} &= \left. \frac{\partial^k \hat{y}(t|s)}{\partial t^k} \right|_{s=t} \\ &= \int_{t_0}^t \frac{\partial^k W(t, \tau)}{\partial t^k} dz(\tau) \end{aligned} \quad (A-6)$$

and  $\eta(t)$  is a stationary Gaussian white noise interpreted via equation (4-3) as the derivative of the Wiener process  $z(t)$ . Define the generalized Wiener process  $z(\varphi)$  by

$$z(\varphi) = \int_{-\infty}^{\infty} \varphi(t) dz(t) \quad (A-7)$$

for any  $r \times 1$  vector  $\varphi(t)$  whose components  $\varphi_i(t)$  are testing functions. Then for  $k = 0, 1, 2, \dots$ ,

$$E \left\{ \frac{\partial^k \hat{y}(t|t)}{\partial t^k} z(\varphi)^T \right\} = E \left\{ \int_{t_0}^t \frac{\partial^k W(t, \tau)}{\partial t^k} dz(\tau) \left[ \int_{-\infty}^{\infty} \varphi(\tau) dz(\tau) \right]^T \right\}. \quad (A-8)$$

Since  $W(t, \tau)$  is an impulse response matrix it is identically

zero for  $t < \tau$  so  $\int_{t_0}^t \frac{\partial^k W(t, \tau)}{\partial t^k} dz(\tau)$  may be replaced by

$\int_{t_0}^{\infty} \frac{\partial^k W(t, \tau)}{\partial t^k} dz(\tau)$ . Furthermore, assuming the input  $z(\tau)$

to be zero for  $\tau < t_0$  the lower limit of integration to may be replaced by  $-\infty$ . Hence for  $k = 0, 1, 2, \dots$ ,

$$\begin{aligned}
E \left\{ \frac{\partial^k \hat{y}(t|t)}{\partial t^k} z(\varphi)^T \right\} &= E \left\{ \int_{-\infty}^{\infty} \frac{\partial^k W(t, \tau)}{\partial t^k} dz(\tau) \left[ \int_{-\infty}^{\infty} \varphi(\tau) dz(\tau) \right]^T \right\} \\
&= \int_{-\infty}^{\infty} \frac{\partial^k W(t, \tau)}{\partial t^k} \varphi^T(\tau) d\tau
\end{aligned} \tag{A-9}$$

by the properties of the orthogonal increment process  $z(\tau)$ .

Next, by definition, the white noise process  $\eta(\varphi)$  is given by

$$\begin{aligned}
\eta(\varphi) &= z'(\varphi) \\
&= -z(\varphi').
\end{aligned} \tag{A-10}$$

Hence

$$\begin{aligned}
E \left\{ \frac{\partial^k \hat{y}(t|t)}{\partial t^k} \eta(\varphi)^T \right\} &= -E \left\{ \frac{\partial^k \hat{y}(t|t)}{\partial t^k} z(\varphi')^T \right\} \\
&= - \int_{-\infty}^{\infty} \frac{\partial^k W(t, \tau)}{\partial t^k} \varphi'(\tau)^T d\tau.
\end{aligned} \tag{A-11}$$

Consider  $\eta^{(j)}(\varphi)$  for  $j = 0, 1, 2, \dots$ . By definition

$$\eta^{(j)}(\varphi) = (-1)^j \eta(\varphi^{(j)}). \tag{A-12}$$

Then for  $j = 0, 1, 2, \dots$

$$\begin{aligned}
E \left\{ \frac{\partial^k \hat{y}(t|t)}{\partial t^k} \eta^{(j)}(\varphi)^T \right\} &= (-1)^j E \left\{ \frac{\partial^k \hat{y}(t|t)}{\partial t^k} \eta(\varphi^{(j)})^T \right\} \\
&= (-1)^{j+1} \int_{-\infty}^{\infty} \frac{\partial^k W(t, \tau)}{\partial t^k} \varphi^{(j+1)}(\tau)^T d\tau
\end{aligned} \tag{A-13}$$

where the last equality is by the use of equation (A-11).

Integrating by parts  $j$  times and using the fact that  $\varphi(\tau)$  has compact support

$$\begin{aligned}
& (-1)^{j+1} \int_{-\infty}^{\infty} \frac{\partial^k W(t, \tau)}{\partial t^k} \varphi^{(j+1)}(\tau)^T d\tau \\
&= (-1)^{2j+1} \int_{-\infty}^{\infty} \frac{\partial^{k+j} W(t, \tau)}{\partial t^k \partial \tau^j} \varphi'(\tau)^T d\tau \\
&= - \int_{-\infty}^{\infty} \frac{\partial^{k+j} W(t, \tau)}{\partial t^k \partial \tau^j} \varphi'(\tau)^T d\tau. \tag{A-14}
\end{aligned}$$

Hence

$$\begin{aligned}
E \left\{ \frac{\partial^k y(t|t)}{\partial t^k} \eta^{(j)}(\varphi)^T \right\} &= - \int_{-\infty}^{\infty} \frac{\partial^{k+j} W(t, \tau)}{\partial t^k \partial \tau^j} \varphi'(\tau)^T d\tau \\
&= -E \left\{ \int_{-\infty}^{\infty} \frac{\partial^{k+j} W(t, \tau)}{\partial t^k \partial \tau^j} dz(\tau) \left[ \int_{-\infty}^{\infty} \varphi'(\tau) dz(\tau) \right]^T \right\} \\
&= -E \left\{ \int_{-\infty}^{\infty} \frac{\partial^{k+j} W(t, \tau)}{\partial t^k \partial \tau^j} dz(\tau) z(\varphi')^T \right\} \\
&= E \left\{ \int_{t_0}^t \frac{\partial^{k+j} W(t, \tau)}{\partial t^k \partial \tau^j} dz(\tau) \eta(\varphi)^T \right\} \tag{A-15}
\end{aligned}$$

Next suppose that  $W(t, \tau)$  is stationary so that

$$W(t, \tau) = W(t - \tau). \tag{A-16}$$

Then

$$\begin{aligned}
\frac{\partial^{k+j} W(t, \tau)}{\partial t^k \partial \tau^j} &= \frac{\partial^{k+j} W(t - \tau)}{\partial t^k \partial \tau^j} \\
&= (-1)^j \frac{\partial^{k+j} W(t - \tau)}{\partial t^{k+j}}.
\end{aligned} \tag{A-17}$$

Substituting this into equation (A-15) yields

$$\begin{aligned}
E \left\{ \frac{\partial^k \varphi(t|t)}{\partial t^k} \eta^{(j)}(\varphi) \right\} &= (-1)^j E \left\{ \int_{t_0}^t \frac{\partial^{k+j} W(t-\tau)}{\partial t^{k+j}} dz(\tau) \eta(\varphi)^T \right\} \\
&= (-1)^j E \left\{ \frac{\partial^{k+j} \varphi(t|t)}{\partial t^{k+j}} \eta(\varphi)^T \right\}. \quad (A-16)
\end{aligned}$$

Interpreting equation (4-76) in this sense establishes lemma 4-3.

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 03046 5227