RL-TR-93-146
Final Technical Report
July 1993

AD-A270 608

STATE-SPACE MODELS FOR MULTICHANNEL DETECTION

Scientific Studies Corporation

J.R. Roman and D.W. Davis

DTIC
ELECTE
OCT 06 1993
S E D

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

93-23193

Rome Laboratory
Air Force Materiel Command
Griffiss Air Force Base, New York

93 10 4 014

This report has been reviewed by the Rome Laboratory Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RL-TR-93-146 has been reviewed and is approved for publication.

APPROVED: James H. Michels

JAMES H. MICHELS
Project Engineer

FOR THE COMMANDER: Donald W. Hanson

DONALD W. HANSON
Director
Surveillance and Photonics Directorate

If your address has changed or if you wish to be removed from the Rome Laboratory mailing list, or if the addressee is no longer employed by your organization, please notify RL (OCTM) Griffiss AFB NY 13441. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 1993	Final Apr 92 - Jan 93

4. TITLE AND SUBTITLE
STATE-SPACE MODELS FOR MULTICHANNEL DETECTION

5. FUNDING NUMBERS
C - F30602-92-C-0091
PE - 61102F
PR - 2304
TA - E8
WU - P6

6. AUTHOR(S)
J.R. Roman, D.W. Davis

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Scientific Studies Corporation
2250 Quail Ridge
Palm Beach Gardens FL 33418

8. PERFORMING ORGANIZATION REPORT NUMBER
SSC-TR-93-02

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Rome Laboratory (OCTM)
26 Electronic Parkway
Griffiss AFB NY 13441-4514

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
RL-TR-93-146

11. SUPPLEMENTARY NOTES
Rome Laboratory Project Engineer: Dr. James Michels/OCTM/(315)330-4431

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

In multichannel identification and detection (or model-based multichannel detection) problems the parameters of a model are identified from the observed channel process, and the identified model is used to facilitate the detection of a signal in the observed process. A model-based multichannel detection algorithm was developed in the context of an innovations-based detection algorithm (IBDA) formulation for surveillance radar system applications. The state space model class was adopted to model the vector channel process because it is more general than the time series model class used in most analyses to date. An IBDA methodology was developed based on the canonical correlations algorithm which for state-space model identification offers performance advantages over alternative techniques. A computer simulation was developed to validate the methodology and the algorithm, and to carry out performance assessments. Simulation results indicate that the algorithm is capable of discriminating between the null hypothesis (clutter plus noise) and the alternative hypothesis (signal plus clutter plus noise). In summary, the applicability of the approach to radar system problems has been established.

14. SUBJECT TERMS
Multichannel systems, Model-based detection, Identification, Canonical correlations

15. NUMBER OF PAGES
142

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	U/L

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18
298-102

TABLE OF CONTENTS

Accesion For

NTIS CRA&I ☒
DTIC TAB ☐
Unannounced ☐
Justification .....

By .....
Distribution/

Availability Codes

Dist	Avail and/or Special
A-1	

DTIC QUALITY INSPECTED 3

LIST OF FIGURES

LIST OF TABLES

1.0 INTRODUCTION

In multichannel identification problems the outputs of multiple channels (or sensors) are available, and it is desired to identify the parameters of an analytical model to represent the phenomena being observed via the channel outputs. Similarly, in multichannel detection problems the outputs of multiple channels are available, and it is desired to determine the presence (or absence) of a desired signal component in the channel data. In the combined problem of multichannel identification and detection a model is estimated for the phenomena being observed via the channel outputs, and the identified model is used to facilitate the detection of a desired signal in the channel output data. Multichannel identification and detection is thus referred to also as model-based multichannel detection. In all of these problems the channel data is available simultaneously over many channels of the same type, or over many distinct channels (each channel corresponding to a different sensor type).

This report is a summary of the work carried out in this program. Specifically, the development of state space algorithms for model-based multichannel detection in the context of surveillance radar system applications is addressed. In surveillance radar systems (radar arrays) the channels correspond to separate antenna apertures (or elements of a single aperture array). The desired signal may or may not be present in the channel output data at any given time. The data in each channel generally includes noise (broadband interference) as well as "clutter" (narrowband interference), with low signal-to-clutter ratio and, possibly, low signal-to-noise ratio also. Model-based detection methods must discriminate between the condition of target embedded in clutter and noise, and the condition of clutter and noise only.

Figure 1-1 illustrates a radar array system consisting of multiple subarrays or array elements. The output of each subarray (or each individual array element) is a complex-valued, scalar, digital sequence, denoted as $\{x_i(n)\}$. The collection of the J scalar sequences is arranged into a J-dimensional vector, $\{\underline{x}(n)\}$, which is input to a multichannel processor (not shown in the figure).

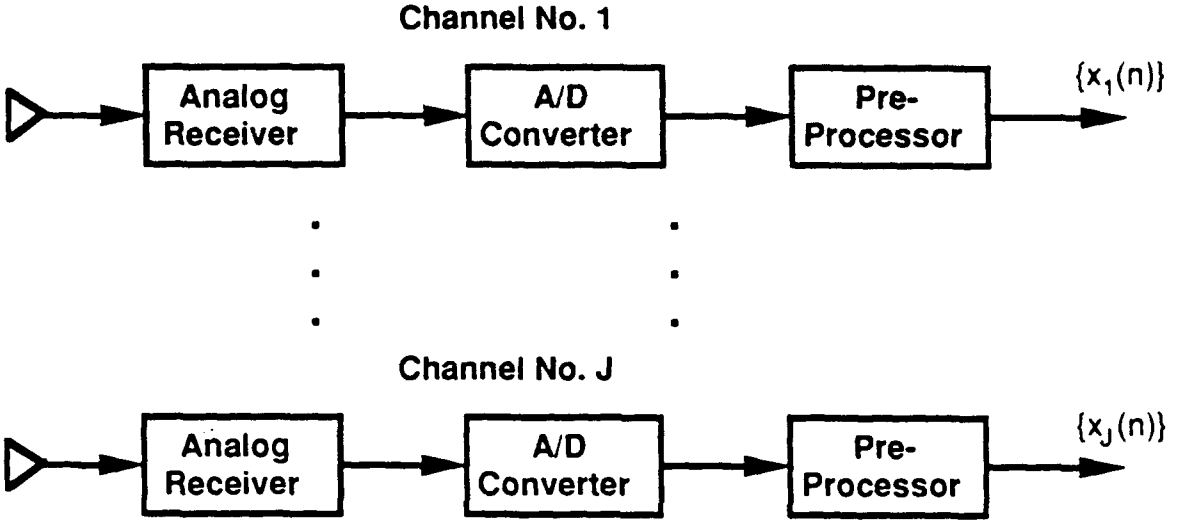


Figure 1-1. . Radar array with J subarrays or individual elements.

In this study the multivariate (multiple input, multiple output) state space model class was adopted to represent the multichannel radar data, and advanced system identification techniques were applied to estimate the model parameters. The modeling of the complex-valued pre-processed radar signals for multichannel detection using the state space model class is one of the contributions of this work. State space models have been used in the context of target tracking (where the detected radar signal is processed further to estimate a trajectory) and for the determination of weights in antenna array sidelobe canceling and related problems, but not for multichannel detection. Model-based detection has been carried out using the more-restricted time

series models (Michels, 1991; Metford and Haykin, 1985), which are included within the class of state space models and can be represented as such.

The methodology formulated in this study is based on a state space identification algorithm developed by Desai et al. (1985), which in turn is based on the stochastic realization concepts formulated by Akaike (1974; 1975) and Faurre (1976). This identification algorithm has several unique features. Foremost among these, the algorithm identifies the model parameters in the innovations representation. As a result, a steady-state Kalman filter design is obtained as an inherent by-product of the algorithm, without having to solve a nonlinear Riccati equation. Implementation of the algorithm is carried out using the singular value decomposition (SVD), which is a stable numerical technique.

An important distinction in the context of radar system applications is that the vector random processes which represent the channel data are complex-valued processes in most cases. Most time series techniques and models have been formulated for complex as well as real processes. The same, however, cannot be said about state-space techniques; state-space methods and results available in the literature have been defined almost exclusively for the case of real-valued processes, including the stochastic realization algorithms adopted herein. In this study the stochastic realization formulations and algorithms were extended to the case of complex-valued processes, which is the formulation presented in this report.

A computer simulation was generated as part of this program to validate the methodology and the algorithms, and to carry out limited simulation-based analyses. This software was exercised with simulated multichannel data generated at RL, and the modeling

and identification results compare favorably with the results obtained at RL using auto-regressive models.

In summary, the analytical and simulation results obtained in this program indicate that the SSC algorithm and methodology for model-based multichannel detection has the potential to result in significant advances for radar system applications.

1.1 Notation

Vector variables are denoted by underscored lower-case letters (including Greek letters). Matrices are denoted by upper-case letters (including Greek letters). Some scalars (such as the order of the state variable model) are denoted also by upper-case letters. Vector spaces are denoted by upper-case script letters, such as $\mathcal{V}$. The expectation operator is denoted as $E[\cdot]$; superscript T and H are used to denote the matrix and vector transpose and the Hermitian transpose operators, respectively; and an asterisk (*) denotes the complex conjugate operator. I_M denotes an M-dimensional identity matrix, $O_{N,J}$ denotes an NxJ null (zero) matrix, O_M denotes an M-dimensional (square) null matrix, and $\underline{0}_M$ denotes an M-dimensional zero vector. $|A|$ denotes the determinant of matrix A; A^{-1} denotes the inverse of matrix A; $A^{\dagger}$ denotes the pseudoinverse of A; rank(A) denotes the rank of A; A(i,j) and a_{ij} are both used to denote the (i,j)th element of matrix A; and $\mathcal{V}_1 | \mathcal{V}_2$ denotes the orthogonal projection of $\mathcal{V}_1$ onto $\mathcal{V}_2$. A caret (^) over a variable denotes an estimate of the variable, a bar (−) over a variable is used to represent the mean of the variable, and ln(a) denotes the natural logarithm of a. The symbol $\perp$ denotes "is orthogonal to;" $\oplus$ denotes the direct sum of vector spaces; $\forall$ denotes "for all;" and $\in$ denotes "is an element of."

Where possible, the symbols used herein to represent variables match the symbols used by Michels (1991) to facilitate enhancing the software available at Rome Laboratory (RL) with the techniques developed in this program. This philosophy forces the use of non-standard symbols to represent the parameters of a state variable model. Of course, notational convention should not be a major issue provided all symbols are defined appropriately. However, it is important to mention this point in order to avoid possible confusion on the part of the reader.

1.2 Report Overview

An introduction to the model-based multichannel detection problem is presented in Section 2.0. This section includes also the definition of the state space model class and several related concepts, including the backward model associated with a forward model, and the innovations representation for a random process. The parameter identification algorithm is presented in Section 3.0, using an approach which differs from the approach of Desai et al. (1985). This alternative approach given here is simple, and enhances intuition. As mentioned earlier, this algorithm is the backbone of the Scientific Studies Corporation (SSC) multichannel detection methodology presented herein. Kalman filtering of the channel data to generate the innovations sequence is discussed in Section 4.0. The innovations sequence is fed to a likelihood ratio detector which generates the detection decision, as described in Section 5.0. A discussion of the software generated in the program is presented in Section 6.0, along with several simulation results which demonstrate the signal discrimination capability of the algorithm. Section 7.0 includes the main conclusions and recommendations borne out of this study.

The three appendices provide background material in a form which is not readily available elsewhere. Appendix A presents a methodology for generating the state space representation of three conventional time series models (moving-average, auto-regressive, and auto-regressive moving-average). Appendix B presents a summary of relevant aspects of deterministic realization theory and algorithms. The extension of canonical correlations to complex-valued variables is presented in Appendix C. This is an important result for Section 3.0.

2.0 MODEL-BASED MULTICHANNEL DETECTION

The model-based approach to multichannel detection involves processing the channel data to identify the parameters of a model for the multichannel system, and determination of a detection decision utilizing the identified parameters to filter the channel data. Model parameters can be identified on-line, as the channel data is received and processed. Alternatively, the model parameters can be identified off-line for various conditions and stored in the processor memory to be accessed in real-time as required.

There are two general classes of linear parametric models for vector random processes: time series models and state space models. Time series models include moving-average (MA) models, auto-regressive (AR) models, and auto-regressive moving-average (ARMA) models. State space models are more general than time series models; in fact, MA, AR, and ARMA models can be represented by state space models (Appendix A). In the state space literature, the determination of the model parameters based on output data (and, sometimes, input data also) is referred to as a stochastic identification or a stochastic realization problem.

Time series models have been applied to the multichannel detection problem, and the performance results obtained provide encouragement for further research (see, for example, Michels, 1991, and the references therein). The results obtained by Michels (1991) assume that the multichannel output process can be modeled as a vector AR process. Given the generality of state-space models and the wealth of results available in the state-space literature, the state space model class was selected in this program to represent the multichannel signals in the model-based multichannel detection problem for radar systems.

In the case of time series models, two types of model parameter estimation algorithms have been established in the literature: (a) algorithms which operate on channel output correlation matrices, such as the extended Levinson algorithm (Anderson and Moore, 1979), and (b) algorithms which operate on the channel output data directly (without the need to compute channel output correlation matrices), such as the Levinson-Wiggins-Robinson algorithm (Wiggins and Robinson, 1965) and the Strand-Nuttall algorithm (Strand, 1977; Nuttall, 1976).

The state-space parameter identification algorithm adopted for this study operates on channel output correlation matrices. The algorithm formulation is due to Desai et al. (1985), and is based on the stochastic realization concepts developed by Akaike (1974, 1975) and Faurre (1976). Implementation of the algorithm is carried out via the singular value decomposition. This algorithm has several attractive features, including direct estimation of the parameters for a Kalman filter, without the requirement to solve a nonlinear Riccati equation.

2.1 Multichannel Detection

Detection problems in the context of radar systems can be postulated as hypothesis testing problems, where a choice has to be made among two or more hypotheses. The detection problems addressed in this report involve the following two hypotheses:

H_0: Desired signal is absent

H_1: Desired signal is present

H_0 is referred to as the null hypothesis, and H_0 is the alternative hypothesis. The model-based approach to the multichannel detection problem is couched on the assumption that the vector random process at the output of the channels can be represented as the output of a linear system (filter) under each of the two hypotheses, and that a unique parametric model corresponds to each hypothesis. Furthermore, the two parametric models (one for each of the two hypotheses) must be sufficiently different to allow selection of the correct hypothesis by the evaluation of measures that are sensitive to those differences.

A particular measure that has produced robust experimental results in the model-based detection context (Metford and Haykin, 1985) is the log-likelihood ratio (LLR) test. This test is the result of solving the hypothesis testing problem using the Neyman-Pearson criterion. The LLR test in the context of model-based detection is calculated using the innovations sequence at the output of each of the two linear filters. This presents practical and implementation advantages.

Figure 2-1 illustrates the architecture of an on-line innovations-based multichannel detector. In the case of a radar array system, each of J radar receiver channels collects the electromagnetic energy arriving at its aperture, and processes it to generate a discrete-time random sequence, denoted as $\{x_i(n)\}$, which contains the desired information. The J random sequences $\{x_i(n)\}$ are represented in vector form as $\{\underline{x}(n)\}$. Michels (1991) has formulated the binary detection problem for multichannel systems. Specifically, the null hypothesis, H_0, corresponds to the case of clutter and noise present in the observation process $\{\underline{x}(n)\}$, and the alternative hypothesis, H_1, corresponds to the case of signal, clutter, and noise present in the observation process $\{\underline{x}(n)\}$. That

is, the detection decision must be made between the following two models,

(2-1a) H_0: $\underline{x}(n) = \underline{c}(n) + \underline{w}(n)$ $n \geq n_o$

(2-1b) H_1: $\underline{x}(n) = \underline{s}(n) + \underline{c}(n) + \underline{w}(n)$ $n \geq n_o$

where n_o denotes the initial observation time, $\{\underline{c}(n)\}$ denotes the clutter process, $\{\underline{w}(n)\}$ denotes all the array channel noise processes, and $\{\underline{s}(n)\}$ denotes the desired signal (target) process.

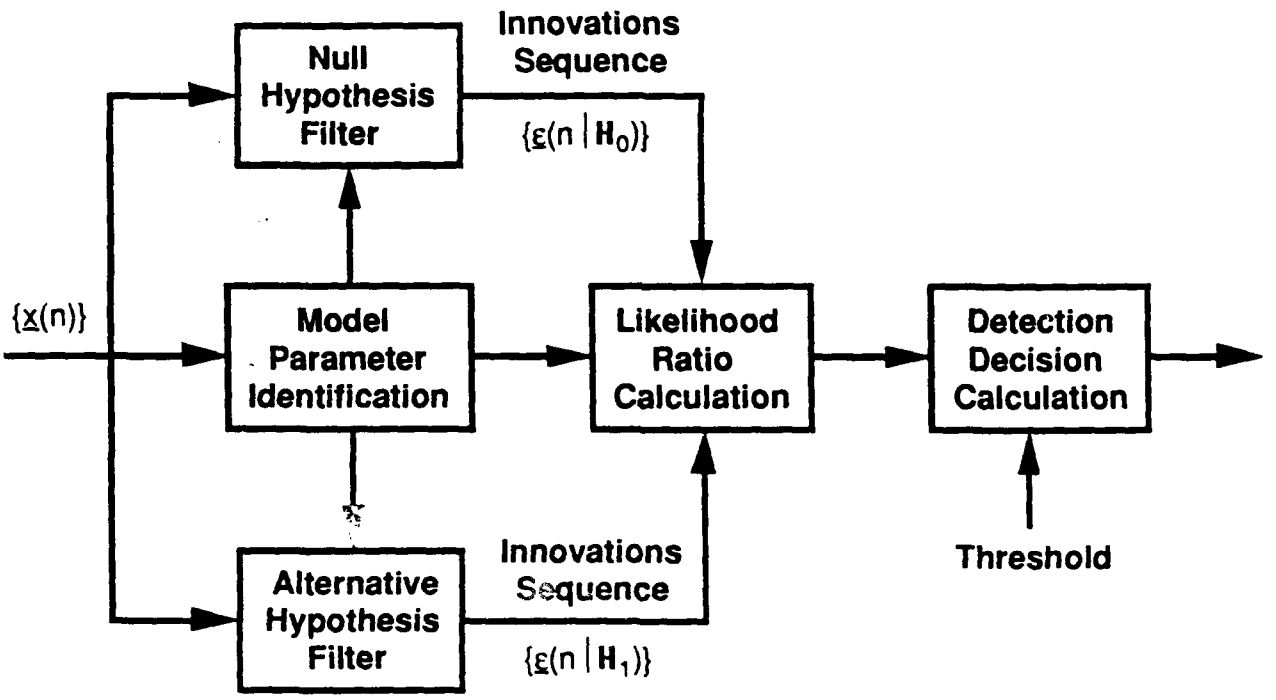


Figure 2-1. Innovations-based multichannel detector with on-line parameter identification.

In the model-based approach pursued herein, a distinct state variable model is associated with each of the two hypotheses, and a Kalman filter is designed for each model. Each Kalman filter processes the observation sequence $\{\underline{x}(n)\}$ to generate a vector

innovations sequence: $\{\underline{\varepsilon}(n|H_0)\}$ denotes the innovations sequence at the output of the null hypothesis filter, and $\{\underline{\varepsilon}(n|H_1)\}$ denotes the innovations sequence at the output of the alternative hypothesis filter. These innovations sequences are used in a likelihood ratio test with a pre-stored threshold to carry out the detection decision.

As indicated in the detection configuration of Figure 2-1, the two filters can be determined in real-time by processing the observation sequence for a prescribed time interval. This approach provides the most adaptability, but may present a large computational burden for some applications. It also presents conceptual challenges, such as real-time determination of model order for each of the two filters. Alternatively, the filter design can be carried out off-line for each of the two hypotheses, and the resulting filter design implemented in the real-time configuration. This alternative approach is less robust to changes in the operational environment, but requires a simpler processor architecture, which is important in many real-time applications. Careful design of the filters off-line using adequate simulated and real data can lead to acceptable performance. Also, many pairs of fixed filters may be designed to cover distinct operational conditions. The filter for the alternative hypothesis will be of higher order than the filter for the null hypothesis because the observation process for the alternative hypothesis has more information (the signal component).

Michels (1991) has developed a likelihood ratio calculation and detection decision model which are compatible with the formulation adopted herein. Both of these capabilities are available at RL, and, where appropriate, the methodology presented in this report is compatible with these capabilities.

2.2 State Space Model

The class of multiple-input, multiple-output state variable models can represent effectively the channel output process for radar applications. Consider a discrete-time, stationary, complex-valued, zero-mean, Gaussian random process $\{\underline{x}(n)\}$ defined as the output of the following state space model representation for the system giving rise to the observed process:

$$\underline{y}(n+1) = F\underline{y}(n) + G\underline{u}(n) \qquad n \geq n_o \tag{2-2a}$$

$$\underline{x}(n) = H^H\underline{y}(n) + D^H\underline{w}(n) \qquad n \geq n_o \tag{2-2b}$$

$$E[\underline{y}(n_o)] = \underline{0}_N \tag{2-2c}$$

$$E[\underline{y}(n_o)\underline{y}^H(n_o)] = P_o \tag{2-2d}$$

Here $n = n_o$ denotes the initial time (which can be adopted as 0 since the system is stationary). Also, $\underline{y}(n)$ is the N-dimensional state of the system with $\underline{y}(n_o)$ a Gaussian random vector; $\underline{u}(n)$ is the J-dimensional, zero-mean, stationary, Gaussian, white input noise process; and $\underline{w}(n)$ is the J-dimensional, zero-mean, stationary, Gaussian, white measurement noise process. The output (or measurement) process $\{\underline{x}(n)\}$ is also a J-dimensional vector process. Matrix F is the NxN system matrix, G is NxJ input noise distribution matrix, H^H is the JxN output distribution matrix, D^H is the JxJ output noise distribution matrix, and P_o is the correlation matrix of the initial state. All these matrices are time-invariant. Matrix P_o is Hermitian ($P_o^H = P_o$, and all its eigenvalues are real-valued) and positive definite (all its eigenvalues are positive).

System (2-2) is assumed to be asymptotically stable, which means that all the eigenvalues of matrix $\mathbf{F}$ are inside the unit circle. Also, system (2-2) is assumed to be reachable and observable, which implies that the dimension N of the state vector (also the order of the system) is minimal (Anderson and Moore, 1979). That is, there is no system of lesser order which has identical input/output behaviour. Lastly, system (2-2) is assumed to be minimum-phase (its zeros are also inside the unit circle). This last assumption implies that the system is defined uniquely by second-order statistics. The output distribution matrices are defined with the conjugate operator in order to have notation consistent with that of the single-output system case, where both $\mathbf{H}$ and $\mathbf{D}$ become vectors, and nominally vectors are defined as column vectors.

The input noise process correlation matrix is given as (all matrices defined hereafter have appropriate dimensions)

(2-3a) $$E[\underline{u}(k)\underline{u}^H(k)] = R_{uu}(0) = Q \qquad k \geq n_o$$

(2-3b) $$E[\underline{u}(k)\underline{u}^H(k-n)] = R_{uu}(n) = [0] \qquad k \geq n_o \text{ and } n \neq 0$$

and the output noise process correlation matrix is given as

(2-4a) $$E[\underline{w}(k)\underline{w}^H(k)] = R_{ww}(0) = C \qquad k \geq n_o$$

(2-4b) $$E[\underline{w}(k)\underline{w}^H(k-n)] = R_{ww}(n) = [0] \qquad k \geq n_o \text{ and } n \neq 0$$

Notice that matrices Q and C are Hermitian. Matrix Q is at least a positive semidefinite matrix since it is an auto-correlation matrix (all the eigenvalues of a positive semidefinite matrix are non-negative), and matrix C is assumed to be positive definite (this can be relaxed to positive semi-definite, but positive

definiteness is more realistic since in the radar problem $\underline{w}(n)$ represents channel noise and other such noise processes which are independent from channel to channel).

In the most general form for this model the input and output noise processes are correlated, with a cross-correlation matrix defined as

(2-5a) $$E[\underline{u}(k)\underline{w}^H(k)] = R_{uw}(0) = S \qquad k \geq n_o$$

(2-5b) $$E[\underline{u}(k)\underline{w}^H(k-n)] = R_{uw}(n) = [0] \qquad k \geq n_o \text{ and } n \neq 0$$

In general, matrix S is not Hermitian. Both the input and output noise processes are uncorrelated with the present and past values of the state process, and this is expressed in terms of cross-correlation matrices as

(2-6a) $$E[\underline{y}(k)\underline{u}^H(k-n)] = R_{yu}(n) = [0] \qquad k \geq n_o \text{ and } n \geq 0$$

(2-6b) $$E[\underline{y}(k)\underline{w}^H(k-n)] = R_{yw}(n) = [0] \qquad k \geq n_o \text{ and } n \geq 0$$

The correlation matrix of the state is defined as

(2-7) $$E[\underline{y}(n)\underline{y}^H(n)] = R_{yy}(n) = P(n) \qquad k \geq n_o \text{ and } n \geq 0$$

It follows from (2-2a) and the above definitions that the state correlation matrix satisfies the following recurrence relation,

(2-8) $$P(n+1) = FP(n)F^H + GQG^H \qquad n \geq n_o$$

In general, matrix $P(n)$ is Hermitian and positive definite. Since system (2-2) is stationary and asymptotically stable, and since

matrix Q is positive definite, then the following steady-state (large n) value exists for the recursion (2-8):

$$P(n+1) = P(n) = P \tag{2-9}$$

Under steady-state conditions Equation (2-8) becomes a Lyapunov equation for the steady-state correlation matrix,

$$P = FPF^H + GQG^H \tag{2-10}$$

The conditions for steady-state also insure that the solution to Equation (2-10) exists, is unique (for the selected state space basis), and is positive definite (Anderson and Moore, 1979). Matrix P is unique for a given state space basis. However, if the basis of the input noise vector and/or the basis of the state vector are changed by a similarity and/or an input transformation, then a different state correlation matrix results from Equation (2-10).

The correlation matrix sequence of the output process $\{\underline{x}(n)\}$ is defined as

$$E[\underline{x}(k)\underline{x}^H(k-n)] = R_{xx}(n) = \Lambda_n \qquad \forall\, k \text{ and } n \geq 0 \tag{2-11a}$$

$$R_{xx}(-n) = R_{xx}^H(n) \qquad \forall\, n \tag{2-11b}$$

For system (2-2) the correlation matrix $R_{xx}(n)$ can be expressed in factored form, with the system parameter matrices as factors:

$$\Lambda_n = R_{xx}(n) = H^H F^{n-1} \Gamma \qquad n > 0 \tag{2-12a}$$

$$\Lambda_n = R_{xx}(n) = \Gamma^H [F^{n-1}]^H H = \Gamma^H [F^H]^{n-1} H \qquad n < 0 \tag{2-12b}$$

Here F^{n-1} denotes F raised to the $(n-1)$th power, and Γ denotes the following cross-correlation matrix

$$\Gamma = E[\underline{y}(n)\underline{x}^H(n-1)] = R_{yx}(1) = FP(n)H + GSD \qquad \forall\, n > 0 \tag{2-13}$$

The correlation matrix sequence factorization in Equation (2-12) is the key to most correlation-based stochastic realization algorithms. The zero-lag ($n = 0$) output correlation matrix is

$$R_{xx}(0) = H^H P(n) H + D^H C D = \Lambda_o \tag{2-14}$$

and matrix $R_{xx}(0)$ is Hermitian and at least positive semidefinite. In steady-state, P replaces $P(n)$ in Equations (2-13) and (2-14).

As can be inferred from the above relations, the system parameters $\{F, G, H, D, Q, C, S, P, \Gamma\}$ completely define the second-order statistics (the correlation matrix sequence $\{R_{xx}(n)\}$) of the output process, and it is said that system (2-2) realizes the output correlation matrix sequence. Conversely, the second-order statistics of the output process provide sufficient information to identify the system parameters, although not uniquely. Since the output process has mean equal to zero and is Gaussian-distributed, the second-order statistics define the process completely.

From the system identification (stochastic realization) point of view, the problem addressed herein can be stated as follows: given the output data sequence $\{\underline{x}(n)\}$ of system (2-2), estimate a set of system parameters $\{F, G, H, D, Q, C, S, P, \Gamma\}$ which generates the same output correlation matrix sequence as system (2-2). Furthermore, the identified parameter set must correspond to a system realization of minimal order (with state vector $\underline{y}$ of minimal dimension). The solution to this problem is pursued herein via a two-step approach: first an estimate of the output correlation

matrix sequence is calculated, and then the estimated correlation sequence is used to determine the system parameters.

It is well known (Anderson and Moore, 1979) that there can be an infinity of systems (2-2) with the same output correlation matrix sequence. The set of all systems that have the same output correlation matrix sequence is an equivalence class, and any two systems belonging to the set are said to be correlation equivalent (Candy, 1976). For example, the output correlation matrix sequence remains invariant to a similarity transformation applied to the state vector. Similarly, the output correlation matrix sequence remains invariant also to a non-singular transformation applied to the input noise and/or to the output noise. As shown by Candy (1976), the equivalence class of correlation equivalent systems is defined including other operations besides a change of basis.

As inferred from these comments, the solution to the system identification problem is not unique. It is also true that most of the possible system parameter solutions do not possess desirable properties. There is, however, a solution which has several features of importance. This solution is referred to as the innovations representation for system (2-2), and is discussed in Section 2.3. The identification algorithm of Section 3.0 generates system parameter matrix estimates for the innovations representation.

In general, the system matrix parameters resulting from the identification algorithm will be represented in a different basis, and should be denoted with a different symbol (say, F_1 instead of F, etc.); nevertheless, the same symbol will be used in this report in order to simplify notation.

Several definitions and notation associated with the input /output behaviour of system (2-2) are important. Consider first the L-term (finite) controllability matrix of system (2-2), $\mathcal{C}_L$; this matrix is defined as an NxJL partitioned matrix of the form

$$\mathcal{C}_L = [\,G \;\; FG \;\; \cdots \;\; F^{L-1}G\,] \tag{2-15}$$

As is well-known, for a minimal-order system matrix $\mathcal{C}_L$ has rank N for $L \geq N$. The controllability matrix maps the input space onto the state space. Analogously, the L-term observability matrix of system (2-2) is the following JLxN partitioned matrix,

$$\mathcal{O}_L = \begin{bmatrix} H^H \\ H^H F \\ \vdots \\ H^H F^{L-1} \end{bmatrix} \tag{2-16}$$

and for a minimal-order system the rank of matrix $\mathcal{O}_L$ is equal to N for $L \geq N$. The observability matrix maps the state space onto the output space. Classical realization theory for the deterministic case (see Appendix B) is based on the fact that a deterministic system block Hankel matrix can be represented as the product of the observability and controllability matrices. Let $H_{L,L}$ denote the JLxJL deterministic block Hankel matrix with the impulse response matrix A(i+j-1) as its (i,j)th block element (a block Hankel matrix is a matrix in which the (i,j)th block element is a function of i+j). That is,

$$(2\text{-}17) \qquad H_{L,L} = \mathcal{O}_L \mathcal{C}_L = \begin{bmatrix} A(1) & A(2) & \cdots & A(L) \\ A(2) & A(3) & \cdots & A(L+1) \\ \vdots & \vdots & \ddots & \vdots \\ A(L) & A(L+1) & \cdots & A(2L-1) \end{bmatrix}$$

The form of Equation (2-17) follows from the definition of the impulse response matrix sequence $\{A(n)\}$ for a deterministic system,

$$(2\text{-}18) \qquad A(n) = H^H F^{n-1} G \qquad\qquad n \geq 1$$

As shown in Appendix B, for $L \geq N$ the rank of the deterministic block Hankel matrix, $H_{L,L}$, is equal to the system order, N. In fact, it is true also that $\text{rank}(H_{N+k,N+k}) = N$ for $k \geq 1$, and that the elements of the impulse response matrix sequence $\{A(n)\}$ satisfy a set of recursion relations (Equation (B-7)) of order equal to the minimal polynomial of matrix F. The block columns (rows) of $H_{L,L}$ satisfy the same recursion relations due to the sequential arrangement of the matrices $\{A(n)\}$ as block elements of $H_{L,L}$.

Notice that the representation (2-18) of the impulse response matrix sequence is of the same form as the representation of the correlation matrix sequence in Equation (2-12). Due to this similarity the matrix elements of the correlation matrix sequence $\{\Lambda_n\}$ satisfy the same set of recursion relations as the matrix elements of the impulse response matrix sequence $\{A(n)\}$, and the above-discussed properties of the deterministic Hankel matrix are also properties of the stochastic Hankel matrix.

Associated with system (2-2) is a backward time model which is defined from the system model (2-2). Backward time models play a role in the formulation of a large class of stochastic realization algorithms. The backward time model for system (2-2)

is defined as a discrete-time, stationary, complex-valued, zero-mean, Gaussian random process with a state space representation of the form (Faurre, 1976)

$$\underline{s}(n) = F^H \underline{s}(n+1) + \underline{v}_i(n) \tag{2-19a}$$

$$\underline{x}(n) = \Gamma^H \underline{s}(n) + \underline{v}_o(n) \tag{2-19b}$$

where $\underline{s}(n)$ is the N-dimensional state vector, $\underline{v}_i(n)$ is the N-dimensional input noise vector, and $\underline{v}_o(n)$ is the J-dimensional output noise vector. Both noise vectors are uncorrelated in time (white), have mean equal to zero, and are Gaussian-distributed. It is important to note that matrix Γ in Equation (2-19b) is the same matrix which appears in the factorization of the output correlation matrices in Equation (2-12), and is defined in Equation (2-13).

The L-term observability matrix for the backward system (2-19) is the following JLxN partitioned matrix,

$$\mathcal{D}_L = \begin{bmatrix} \Gamma^H \\ \Gamma^H F^H \\ \vdots \\ \Gamma^H (F^H)^{L-1} \end{bmatrix} \tag{2-20}$$

The backward system is completely observable also, which implies that $\text{rank}(\mathcal{D}_L) = N$. Also of interest is the conjugate transpose of $\mathcal{D}_L$, which is,

$$\mathcal{D}_L^H = \begin{bmatrix} \Gamma & F\Gamma & \cdots & F^{L-1}\Gamma \end{bmatrix} \tag{2-21}$$

Matrix $\mathcal{D}_L^H$ is a controllability matrix for the matrix pair (F, Γ), and as such, it can be viewed as the controllability matrix for the dual system corresponding to the backward model (if system B is a dual for system A, then the input of system B is the output of system A, and the output of system B is the input of system A; that is, the roles of input and output are interchanged). However, in this report it is preferable to refer to $\mathcal{D}_L$ as the backward observability matrix.

In the context of stochastic realization theory, the significance of the backward model follows from Equation (2-20) and the Hankel matrix of output correlation matrices, as shown next. Define a stochastic Hankel matrix $\mathcal{H}_{L,L}$ as the following JLxJL block matrix,

$$\mathcal{H}_{L,L} = \begin{bmatrix} \Lambda_1 & \Lambda_2 & \cdots & \Lambda_L \\ \Lambda_2 & \Lambda_3 & \cdots & \Lambda_{L+1} \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda_L & \Lambda_{L+1} & \cdots & \Lambda_{2L-1} \end{bmatrix} \tag{2-22}$$

where the block elements $\{\Lambda_n\}$ are the elements of the output correlation matrix sequence, Equation (2-12). It follows from Equations (2-12), (2-16), (2-21), and (2-22) that

$$\mathcal{H}_{L,L} = \mathcal{O}_L \mathcal{D}_L^H \tag{2-23}$$

This equation is fundamental to stochastic realization theory from conceptual as well as algorithmic viewpoints. From a conceptual viewpoint, Equation (2-23) is a factorization of the Hankel matrix into the observability matrices of the forward and backward systems, and thus hints at the underlying structure of the

correlations between the past and future output vectors (as discussed below and in Section 3.0).

From an algorithmic viewpoint, the similarities between the deterministic and stochastic block Hankel matrices and their respective factorizations implies that the properties which are true for the deterministic case are true also for the stochastic case. Specifically, the most important of these properties are:

i) $\text{rank}(\mathcal{H}_{L,L}) = N$ for $L \geq N$,

ii) $\text{rank}(\mathcal{H}_{N+k,N+k}) = N$ for $k \geq 1$, and

iii) the block columns (rows) of $\mathcal{H}_{L,L}$ satisfy the same recursion relations as the block columns (rows) of $H_{L,L}$.

Furthermore, the similarities between Equation (2-17) and Equations (2-22) and (2-23) allow the application of classical deterministic realization concepts, insight, and algorithms (see Appendix B) to the stochastic realization problem formulated with output correlation matrices.

Other important matrices in stochastic realization theory include the JLxJL "future" and "past" block correlation matrices. These matrices are the correlation matrices of future and past output block vectors defined as

$$\text{(2-24)} \qquad \underline{x}_p = \underline{x}(n-1;n-L) = \begin{bmatrix} \underline{x}(n-1) \\ \underline{x}(n-2) \\ \vdots \\ \underline{x}(n-L) \end{bmatrix}$$

$$(2\text{-}25) \qquad \underline{x}_F = \underline{x}(n;n+L-1) = \begin{bmatrix} \underline{x}(n) \\ \underline{x}(n+1) \\ \vdots \\ \underline{x}(n+L-1) \end{bmatrix}$$

With respect to the time instant n, vector $\underline{x}_P$ represents the past of the process $\{\underline{x}(n)\}$, and vector $\underline{x}_F$ represents the future of the process $\{\underline{x}(n)\}$. Given these definitions, the following matrices can be introduced:

$$(2\text{-}26) \qquad \mathcal{R}_{P:L,L} = E[\underline{x}_P \underline{x}_P^H] = \begin{bmatrix} \Lambda_0 & \Lambda_1 & \cdots & \Lambda_{L-1} \\ \Lambda_{-1} & \Lambda_0 & \cdots & \Lambda_{L-2} \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda_{1-L} & \Lambda_{2-L} & \cdots & \Lambda_0 \end{bmatrix}$$

$$(2\text{-}27) \qquad \mathcal{R}_{F:L,L} = E[\underline{x}_F \underline{x}_F^H] = \begin{bmatrix} \Lambda_0 & \Lambda_{-1} & \cdots & \Lambda_{1-L} \\ \Lambda_1 & \Lambda_0 & \cdots & \Lambda_{2-L} \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda_{L-1} & \Lambda_{L-2} & \cdots & \Lambda_0 \end{bmatrix}$$

where $\mathcal{R}_{F:L,L}$ and $\mathcal{R}_{P:L,L}$ are the $JL \times JL$ future and past block correlation matrices, respectively. Both of these matrices are Hermitian (as well as block Hermitian), and they exhibit a block Toeplitz structure (a block Toeplitz matrix is a matrix in which the (i,j)th block element is a function of i-j). It is important to note that, in general, the conjugate transpose of $\mathcal{R}_{F:L,L}$ is not equal to $\mathcal{R}_{P:L,L}$, even though these matrices are the block Hermitian of each other; that is, matrices $\mathcal{R}_{F:L,L}$ and $\mathcal{R}_{P:L,L}$ are not the element-by-element Hermitian of each other.

Another matrix of interest is the block cross-correlation matrix between the future and the past, which is defined as

$$(2\text{-}28) \qquad \mathcal{R}_{F:L,P:L} = E[\underline{x}_F \underline{x}_P^H] = \begin{bmatrix} \Lambda_1 & \Lambda_2 & \cdots & \Lambda_L \\ \Lambda_2 & \Lambda_3 & \cdots & \Lambda_{L+1} \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda_L & \Lambda_{L+1} & \cdots & \Lambda_{2L-1} \end{bmatrix} = \mathcal{H}_{L,L} = \mathcal{O}_L \mathcal{D}_L^H$$

Equations (2-26)-(2-28) are valid for all n because the process $\{\underline{x}(n)\}$ is stationary. The stochastic realization approach of Akaike (1974, 1975) is based on these block correlation matrices.

For $L \geq N$, Equations (2-26)-(2-28) define the correlation structure of system (2-2). As indicated in Equation (2-28), the block cross-correlation matrix $\mathcal{R}_{F:L,P:L}$ is equal to the stochastic block Hankel matrix, Equation (2-22). Thus, as hinted earlier, the cross-correlation between the past and future outputs admits a factorization in terms of the forward system and backward system observability matrices.

2.3 Innovations Representation

The innovations representation is a very powerful concept in the theory of linear stochastic systems due to its simplicity and its characteristics. Several texts and papers discuss this concept in detail. The discussion herein is adapted mostly from Anderson and Moore (1979), which provide a lucid presentation.

The innovations representation for a system (2-2) is a discrete-time, stationary, complex-valued, system of the form

$$(2\text{-}29a) \qquad \underline{\alpha}(n+1) = F\underline{\alpha}(n) + K\underline{\varepsilon}(n) \qquad n \geq n_o$$

(2-29b) $\underline{\chi}(n) = H^H \underline{\alpha}(n) + \underline{\varepsilon}(n)$ $\quad n \geq n_o$

(2-29c) $\underline{\alpha}(n_o) = \underline{0}_N$

(2-29d) $E[\underline{\alpha}(n_o)\underline{\alpha}^H(n_o)] = \Pi(n_o) = \Pi_o = [0]$

(2-29e) $E[\underline{\alpha}(n)\underline{\alpha}^H(n)] = \Pi(n)$ $\quad n > n_o$

(2-29f) $\Pi(n) = \Pi \quad \text{as} \quad n \rightarrow \infty$

(2-29g) $R_{\chi\chi}(n) = R_{xx}(n)$ $\quad \forall\, n$

here $\underline{\alpha}(n)$ is the N-dimensional state, $\underline{\chi}(n)$ is the J-dimensional output, and the input process $\{\underline{\varepsilon}(n)\}$ is the innovations process for system (2-2). That is, $\{\underline{\varepsilon}(n)\}$ is a J-dimensional, zero-mean, white Gaussian process with correlation matrix structure given as

(2-30a) $\Omega = E[\underline{\varepsilon}(k)\underline{\varepsilon}^H(k)] = R_{xx}(0) - H^H \Pi H = \Lambda_o - H^H \Pi H$ $\quad k \geq n_o$

(2-30b) $E[\underline{\varepsilon}(k)\underline{\varepsilon}^H(k-n)] = [0]$ $\quad k \geq n_o$ and $n \neq 0$

The state correlation matrix $\Pi(n)$ has a steady-state value because the system is asymptotically stable (stationary), and the steady-state value, Π, is obtained as the limiting solution to the following recursion

(2-31a) $\Pi(n+1) = F\Pi(n)F^H + [F\Pi(n)H - \Gamma]\,[\Lambda_o - H^H\Pi(n)H]^{-1}\,[F\Pi(n)H - \Gamma]^H$ $\quad n \geq n_o$

(2-31b) $\Pi(n_o) = \Pi_o = [0]$

Matrix K in Equation (2-29a) is given as

(2-32a) $K = [\Gamma - F\Pi H]\,\Omega^{-1} = [\Gamma - F\Pi H]\,[\Lambda_o - H^H\Pi H]^{-1}$

(2-32b) $$K = GSD\,\Omega^{-1} = GSD\,[\Lambda_o - H^H \Pi H]^{-1}$$

where the second relation follows from the definition of Γ in Equation (2-13) and of Ω in Equation (2-30a). In the cases where the inverse of the correlation matrix Ω does not exist, its pseudoinverse is used instead in Equations (2-31) and (2-32).

Matrices F, H, Λ_o, and Γ are as defined for system (2-2). That is, system (2-29) is related to system (2-2). In fact, system (2-29) as defined above is the steady-state *innovations* representation for system (2-2). This representation has the following important features.

(a) First and foremost, the correlation matrix sequence of $\{\chi(n)\}$ is equal to the correlation matrix sequence of $\{\underline{x}(n)\}$, as indicated in Equation (2-29g). That is, the processes $\{\chi(n)\}$ and $\{\underline{x}(n)\}$ are correlation equivalent. This means that the innovations representation is a valid solution to the system identification problem defined herein.

(b) Of all the correlation equivalent representations for a given output correlation sequence, the innovations representation has the smallest state correlation matrix, Π (smallest is meant in the sense of positive definiteness; that is, Π_1 is smaller than Π_2 if $\Pi_2 - \Pi_1$ is a positive definite matrix). This property of the innovations model is significant because the state correlation matrix is a measure of the uncertainty in the state.

(c) The innovations representation is directly related to the steady-state Kalman filter (in the one-step predictor formulation) for system (2-2). In fact, the steady-state Kalman filter for system (2-2) is available immediately upon definition of the steady-state innovations representation, and viceversa. Specifically, matrix K of Equations (2-29a) and (2-31) is the steady-state Kalman gain of the optimal one-step predictor for system (2-2). This is true provided that the eigenvalues of $F-KH^H$ are stable. Thus, the innovations model is defined as above for all processes of the form (2-2), but the steady-state Kalman filter is defined only if $F-KH^H$ is stable.

(d) The process $\{\underline{\varepsilon}(n)\}$ in Equations (2-29) and (2-30) is correlation equivalent to the innovations sequence of system (2-2). This is the reason for referring to system (2-29) as the "innovations representation" for system (2-2).

(e) The innovations model (2-29) is causally invertible. This means that the present and past of the process $\{\underline{\varepsilon}(n)\}$ can be constructed from the present and past values of the output process $\{\underline{\chi}(n)\}$. The converse statement is true also; that is, any causally invertible model is an innovations representation for some system. Causal invertibility of system (2-29) can be demonstrated easily. From Equation (2-29b),

$$\underline{\varepsilon}(n) = -H^H\underline{\alpha}(n) + \underline{\chi}(n) \tag{2-33}$$

Substituting this expression for $\underline{\varepsilon}(n)$ into Equation (2-29a) results in

(2-34) $$\underline{\alpha}(n+1) = [F - KH^H]\underline{\alpha}(n) + K\chi(n)$$

These relations demonstrate the causal invertibility of the innovations model (the input and output variables have traded places).

(f) Matrix $F - KH^H$ in the inverted innovations model is a stable matrix. This follows from the fact that the matrix pair (F, H) is observable, and implies that the Kalman filter for system (2-2) is stable also.

(g) The transfer function of the innovations model (2-29) is minimum phase. This is related to the fact that the innovations model is correlation equivalent to system (2-2), and second-order moment information (the output correlation matrix sequence) does not contain any phase information.

(h) The innovations representation for a system of the form (2-2) is unique. Given that the innovations representation has the same output covariance sequence as system (2-2), the fact that it is unique eliminates searching for other representations for system (2-2) with the properties listed herein.

(i) The innovations model (2-29) can be computed from the output correlation matrix sequence of system (2-2). This fact simplifies the parameter identification problem because the set of parameter matrices that must be estimated from the data is reduced to just five: $\{F, H, \Gamma, \Pi, \Lambda_0\}$ (given these parameter matrices, the innovations covariance, Ω, and the

Kalman gain, K, are obtained using Equations (2-30a) and (2-32a), respectively).

All the features listed above are of relevance to the identification approach presented in Section 3.0 because the selected parameter identification algorithm generates the innovations representation for the given output correlation matrix sequence, following feature (i).

The backward model has an associated backward innovations model with parameter matrices F, Γ, and the backward Kalman gain. Most of the features (a)-(i) that describe the forward innovations model are valid also for the backward innovations model, with a notable exception of feature (b), which needs to be replaced by the following statement: For each valid correlation equivalent representation for a given output correlation sequence, the state correlation matrix is smaller than the inverse of the state correlation matrix for the backward innovations model. More specifically, let Π_b denote the state correlation matrix for the backward innovations model in steady-state conditions, and let Σ denote the state correlation matrix for any valid correlation equivalent representation of an output correlation sequence. Then, $\Pi_b^{-1} - \Sigma$ is a positive definite matrix. This result provides an upper bound for the state correlation matrix of a correlation equivalent representation. Combining this with the lower bound of property (b) of the forward innovations model gives

$$\Pi \leq \Sigma \leq \Pi_b^{-1} \tag{2-35}$$

As before, inequality between two matrices is intended in the sense of positive semi-definiteness of the matrix difference.

3.0 MULTICHANNEL SYSTEM IDENTIFICATION

Identification of the model parameter matrices $\{F, H, \Gamma, \Pi, \Lambda_0\}$ for the innovations representation is carried out based on the predictor space concept and the canonical correlations methodology formulated by Akaike (1974; 1975), and using the specific algorithmic development of Desai et al. (1985), extended to the case of complex-valued data. Their results, in turn, are built upon the correlation equivalence results obtained by Faurre (1976), and the deterministic realization theory and algorithm of Ho (Kalman et al., 1969). The identification algorithm requires the output correlation matrix sequence; since the true output correlation sequence is not available, an estimate must be obtained.

3.1 Covariance Sequence Estimation

The first step in the modeling/identification procedure is the estimation of the output correlation matrix sequence $\{R_{xx}(n)\} = \{\Lambda_n\}$ for $n \geq 0$ (for notational simplicity, $n_o = 0$ is assumed in this section) given a finite-length realization of the output process, $\{\underline{x}(n) \mid n = 0, 1, \ldots, N_T - 1\}$. There are two nominal estimators for correlation matrices. The first estimator is of the form

$$\hat{\Lambda}_n = \hat{R}_{xx}(n) = \frac{1}{N_T - n} \sum_{k=n}^{N_T-1} \underline{x}(k)\underline{x}^H(k-n) \qquad n \leq N_T - 1 \tag{3-1}$$

Estimator (3-1) provides an unbiased estimate of the output correlation matrix sequence (that is, the expected value of (3-1) is equal to the true correlation matrix sequence), but there have been cases where the use of this estimator has led to computational difficulties. In particular, sometimes when estimator (3-1) is used to form a Toeplitz block correlation

matrix, the Toeplitz matrix is not positive definite (or at least positive semi-definite) as it should be.

The second estimator is of the form

$$\hat{\Lambda}_n = \hat{R}_{xx}(n) = \frac{1}{N_T} \sum_{k=0}^{N_T-1} \underline{x}(k)\underline{x}^H(k-n) \qquad n \le N_T - 1 \tag{3-2}$$

with zeros used in the place of missing data $\{\underline{x}(-1), \underline{x}(-2), \ldots\}$. Estimator (3-2) provides a biased estimate as a result of padding the data record with leading zeros. However, division by N_T in (3-2) for all lags drives the correlation estimates to exhibit an enhanced monotonically decreasing behaviour as a function of n (the enhancement is with respect to the correlation estimates resulting from Equation (3-1)). Such a feature is desirable because the output correlation sequence of a stationary system (with matrix F stable) is monotonically decreasing. This feature of estimator (3-2) has provided improved performance (in relation to estimator (3-1)) in algorithms such as the scalar Yule-Walker method for spectrum estimation by insuring that the Toeplitz correlation matrix which arises in that problem be at least positive semidefinite. It is possible that this feature of estimator (3-2) be of similar relevance with the Hankel matrix and the Toeplitz matrices that arise in the stochastic realization problem considered in Section 3.0. For large values of N_T and small values of the maximum lag considered, N_{max}, estimator (3-2) approximates the behaviour of estimator (3-1), and any differences become insignificant. However, for small values of N_T and/or for values of N_{max} close to N_T, each estimator may offer specific advantages in the context of distinct problems. Which estimator is preferable in the context of the problem of interest herein is a topic for future investigation.

3.2 Canonical Correlations Algorithm

In comparison with other alternative stochastic realization techniques, the canonical correlations algorithm used herein has several advantages for multichannel detection applications, as listed next (Desai et al., 1985).

- Identifies the parameters for a model in the state-space class, which is more general than the time series class.

- An approximately balanced (in the stochastic sense) state space realization is generated, thus providing a built-in and robust mechanism for model order selection.

- Identifies the innovations representation of the system and generates the state correlation matrix and the Kalman gain directly. Thus, the Kalman filter is obtained without having to solve a nonlinear discrete matrix Riccati equation.

- Implementation of the algorithm involves the singular value decomposition (SVD), which is a stable numerical method.

These features offer enhanced model-based detection performance in relation to algorithms such as those based on time series models. A discussion of the canonical correlations algorithm is provided in the remainder of this section. This discussion complements and extends the material presented in Appendices B and C, as it is applied to the stochastic realization problem.

The canonical correlations identification algorithm is based on the concept of the correlation structure between the past and

future of the output process $\{\underline{x}(n)\}$, in the context of Hilbert spaces of random variables (Akaike, 1974, 1975; Faurre, 1976). Consider the stochastic process $\{\underline{x}(n)\}$ and define infinite-dimensional block vectors $\underline{x}_p$ and $\underline{x}_f$ as

$$\underline{x}_p = \begin{bmatrix} \underline{x}(n-1) \\ \underline{x}(n-2) \\ \underline{x}(n-3) \\ \vdots \end{bmatrix} \tag{3-3}$$

$$\underline{x}_f = \begin{bmatrix} \underline{x}(n) \\ \underline{x}(n+1) \\ \underline{x}(n+2) \\ \vdots \end{bmatrix} \tag{3-4}$$

These vectors represent the past and future of the process with respect to time n, as in the case of the finite-length vectors in Equations (2-24) and (2-25). Note that the time n can be any instant of time because the process is stationary.

Vector $\underline{x}_p$ indexed at time n (as defined in Equation (3-4)) spans a vector space denoted as $\mathcal{X}^-(n)$, which represents the set of all possible linear combinations of the elements of $\underline{x}_p$, and is referred to as the "past" of the process $\{\underline{x}(n)\}$. Analogously, vector $\underline{x}_f$ indexed at time n spans a vector space denoted as $\mathcal{X}^+(n)$, representing the set of all possible linear combinations of the elements of $\underline{x}_f$, and is referred to as the "future" of the process $\{\underline{x}(n)\}$. The time index is relevant in the stochastic realization problem considered in this section because the process $\{\underline{x}(n)\}$ is a dynamic time series. In contrast, Appendix C presents the canonical correlations formulation for the static multivariate

case. Now let $\mathcal{A}(n)$ be the space generated by the orthogonal projection of $\mathcal{X}^+(n)$ onto $\mathcal{X}^-(n)$; that is,

$$\mathcal{A}(n) = \mathcal{X}^+(n) \mid \mathcal{X}^-(n) \tag{3-5}$$

$\mathcal{A}(n)$ is referred to as the state space of the process $\{\underline{x}(n)\}$ because it is spanned by the state of the innovations representation (Equations (2-29)). For a system of the form (2-2), $\mathcal{A}(n)$ is finite-dimensional, with dimension equal to N. The space $\mathcal{X}^+(n)$ can be represented as the direct sum of two orthogonal subspaces,

$$\mathcal{X}^+(n) = \mathcal{X}^+(n) \mid \mathcal{X}^-(n) \oplus \mathcal{E}(n) = \mathcal{A}(n) \oplus \mathcal{E}(n) \tag{3-6}$$

where $\mathcal{A}(n) \perp \mathcal{E}(n)$, and $\mathcal{E}(n)$ is the space spanned by the innovations process, $\{\underline{\varepsilon}(n)\}$. Equation (3-6) defines the geometric structure of the space $\mathcal{X}^+(n)$. This structure is true for all n because the process is stationary.

Since all the random variables involved are zero-mean and Gaussian-distributed, an orthogonal projection in these vector spaces is equivalent to conditional expectation (Faurre, 1976). Specifically, $\mathcal{A}(n)$ is the space spanned by the elements of

$$\hat{\underline{x}}_{\mathcal{F}} = E[\underline{x}_{\mathcal{F}} \mid \underline{x}_{\mathcal{P}}] = E[\underline{x}_{\mathcal{F}} \underline{x}_{\mathcal{P}}^H] \left(E[\underline{x}_{\mathcal{P}} \underline{x}_{\mathcal{P}}^H]\right)^{-1} \underline{x}_{\mathcal{P}} = \mathcal{H} \mathcal{R}_{\mathcal{P}}^{-1} \underline{x}_{\mathcal{P}} = \mathcal{O} \mathcal{D}^H \mathcal{R}_{\mathcal{P}}^{-1} \underline{x}_{\mathcal{P}} \tag{3-7}$$

Here the caret (^) denotes the conditional expectation (which is also an optimal estimate given the underlying conditions); also, $\mathcal{H}$, $\mathcal{O}$, $\mathcal{D}$, and $\mathcal{R}_{\mathcal{P}}$ are the infinite-dimensional versions of Equations (2-22), (2-16), (2-20), and (2-26), respectively. Now the algebraic representation of the geometric expression (3-6) is obtained as

$$\underline{x}_{\mathcal{F}} = \hat{\underline{x}}_{\mathcal{F}} + \underline{\varepsilon}_n = E[\underline{x}_{\mathcal{F}} \mid \underline{x}_{\mathcal{P}}] + \underline{\varepsilon}_n \tag{3-8}$$

where $\underline{\varepsilon}_n$ is an infinite-dimensional block vector having the innovations sequence as block elements,

$$(3\text{-}9) \qquad \underline{\varepsilon}_n = \begin{bmatrix} \underline{\varepsilon}(n) \\ \underline{\varepsilon}(n+1) \\ \underline{\varepsilon}(n+2) \\ \vdots \end{bmatrix}$$

Of particular interest is the output vector at time n, which is the first block element of Equation (3-8),

$$(3\text{-}10a) \qquad \underline{x}(n) = \hat{\underline{x}}(n|n-1) + \underline{\varepsilon}(n) = E[\underline{x}(n) \,|\, \underline{x}_p] + \underline{\varepsilon}(n)$$

$$(3\text{-}10b) \qquad \underline{x}(n) = E[\underline{x}(n)\,\underline{x}_p^H](E[\underline{x}_p \underline{x}_p^H])^{-1}\,\underline{x}_p + \underline{\varepsilon}(n) = E[\underline{x}(n)\,\underline{x}_p^H]\,\mathcal{R}_p^{-1}\,\underline{x}_p + \underline{\varepsilon}(n)$$

where $\hat{\underline{x}}(n|n-1)$ denotes the minimum variance estimate of the output process at time n based on output measurements up to time $n-1$. This last expression is suggestive of the output equation of the innovations representation. Indeed, it does correspond to Equation (2-29b), as shown next.

Let $\underline{\alpha}(n)$ denote the following N-dimensional vector (since matrix $\mathcal{D}^H$ has N rows),

$$(3\text{-}11) \qquad \underline{\alpha}(n) = \mathcal{D}^H \mathcal{R}_p^{\dagger}\,\underline{x}_p$$

The elements of $\underline{\alpha}(n)$ span the space $\mathcal{A}(n)$. This is true because the elements of $\hat{\underline{x}}_f = E[\underline{x}_f \,|\, \underline{x}_p]$ span $\mathcal{A}(n)$, and because the observability matrix has full rank. In fact, $\underline{\alpha}(n)$ is the <u>state of the innovations representation</u> at time n. This provides the final piece of information needed to complete the innovations

representation output equation. Recall that matrix H^H occupies the first J rows of the observability matrix. Then, from Equations (3-7)-(3-11), it follows that

(3-12) $$\underline{x}_{\mathcal{J}} = O\underline{\alpha}(n) + \underline{\varepsilon}_n$$

(3-13) $$E[\underline{x}(n)\underline{x}_{\mathcal{P}}^H] = H^H \mathcal{D}^H$$

(3-14) $$\underline{x}(n) = H^H \mathcal{D}^H \mathcal{R}_{\mathcal{P}}^{-1} \underline{x}_{\mathcal{P}} + \underline{\varepsilon}(n) = H^H \underline{\alpha}(n) + \underline{\varepsilon}(n)$$

Equation (3-12) is an analytic representation of the statement that the observability matrix maps the state space onto the output space. And Equation (3-14) is the output equation for the innovations representation.

The system identification (stochastic realization) problem can be stated now as follows: determine the factors O and $\mathcal{D}$ of the stochastic Hankel matrix,

(3-15) $$\mathcal{H} = O\mathcal{D}^H$$

in the basis of the innovations representation. In that basis, the state vector is defined as in Equation (3-11), and its correlation matrix is

(3-16) $$\Pi = E[\underline{\alpha}(n)\underline{\alpha}^H(n)] = \mathcal{D}^H \mathcal{R}_{\mathcal{P}}^{-1} E[\underline{x}_{\mathcal{P}}\underline{x}_{\mathcal{P}}^H] \mathcal{R}_{\mathcal{P}}^{-1} \mathcal{D} = \mathcal{D}^H \mathcal{R}_{\mathcal{P}}^{-1} \mathcal{D}$$

Canonical correlations constitute an effective approach for carrying out the factorization of the block Hankel matrix $\mathcal{H}$ in the basis of the innovations representation.

In practical applications, data is available for finite time and the formulation presented above is approximated using block

vectors of finite dimension (JL). With that constraint, the canonical variables approach presented in Appendix C can be applied directly. Consider the JL-dimensional vectors $\underline{x}_P$ and $\underline{x}_F$ of Equations (2-24) and (2-25), with a sufficiently large number (L) of block elements. With reference to Appendix C, let $\underline{x}_P$ replace $\underline{z}$ and let $\underline{x}_F$ replace $\underline{v}$ in the formulation. More specifically, the JL-dimensional canonical variables $\underline{\mu}(n)$ and $\underline{\beta}(n)$ are defined by the following transformations on the past and future vectors,

$$\underline{\mu}(n) = T_P \underline{x}_P \tag{3-17}$$

$$\underline{\beta}(n) = T_F \underline{x}_F \tag{3-18}$$

Then, it is desired to determine the canonical variables $\underline{\mu}(n)$ and $\underline{\beta}(n)$, the canonical correlations $\{\rho_i\}$, and the JLxJL transformation matrices T_P and T_F such that the following conditions are satisfied (Appendix C):

$$E[\underline{\mu}(n)\underline{\mu}^H(n)] = T_P E[\underline{x}_P \underline{x}_P^H] T_P^H = T_P \mathcal{R}_{P:L,L} T_P^H = I_{JL} \tag{3-19}$$

$$E[\underline{\beta}(n)\underline{\beta}^H(n)] = T_F E[\underline{x}_F \underline{x}_F^H] T_F^H = T_F \mathcal{R}_{F:L,L} T_F^H = I_{JL} \tag{3-20}$$

$$E[\underline{\beta}(n)\underline{\mu}^H(n)] = T_F E[\underline{x}_F \underline{x}_P^H] T_P^H = T_F \mathcal{H}_{L,L} T_P^H = R_{\beta\mu} \tag{3-21}$$

$$R_{\beta\mu} = \begin{bmatrix} \rho_1 & 0 & \cdots & 0 \\ 0 & \rho_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \rho_{JL} \end{bmatrix} \tag{3-22a}$$

$$1 \geq \rho_1 \geq \rho_2 \geq \ldots \geq \rho_{JL} \geq 0 \tag{3-22b}$$

Determination of the model order, N, is desired also as part of the procedure. The canonical correlations approach is also well-suited to the determination of model order because the model order is equal to the rank of the block Hankel matrix $\mathcal{H}_{L,L}$. Model order determination is discussed further in Section 3.3.

Following the development in Appendix C, the matrix square root is determined for the past and future correlation matrices using the SVD. This gives

$$\mathcal{R}_{P:L,L} = U_P S_P U_P^H = \left(U_P S_P^{1/2} U_P^H\right)\left(U_P S_P^{1/2} U_P^H\right) = \mathcal{R}_{P:L,L}^{1/2}\,\mathcal{R}_{P:L,L}^{1/2} \tag{3-23}$$

$$\mathcal{R}_{F:L,L} = U_F S_F U_F^H = \left(U_F S_F^{1/2} U_F^H\right)\left(U_F S_F^{1/2} U_F^H\right) = \mathcal{R}_{F:L,L}^{1/2}\,\mathcal{R}_{F:L,L}^{1/2} \tag{3-24}$$

Now transform the past and future vectors, $\underline{x}_P$ and $\underline{x}_F$, into two JL-dimensional random vectors defined as

$$\underline{\theta} = \mathcal{R}_{P:L,L}^{-1/2}\underline{x}_P = U_P S_P^{-1/2} U_P^H\,\underline{x}_P \tag{3-25}$$

$$\underline{\gamma} = \mathcal{R}_{F:L,L}^{-1/2}\underline{x}_F = U_F S_F^{-1/2} U_F^H\,\underline{x}_F \tag{3-26}$$

Given these definitions, it is easy to show that

$$E[\underline{\theta}\,\underline{\theta}^H] = I_{JL} \tag{3-27}$$

$$E[\underline{\gamma}\,\underline{\gamma}^H] = I_{JL} \tag{3-28}$$

$$E[\underline{\gamma}\,\underline{\theta}^H] = R_{\gamma\theta} = \mathcal{R}_{F:L,L}^{1/2}\,\mathcal{H}_{L,L}\,\mathcal{R}_{P:L,L}^{-1/2} \tag{3-29}$$

Notice that the random vectors $\underline{\theta}$ and $\underline{\gamma}$ are correlation-normalized, but their cross-correlation matrix, $R_{\gamma\theta}$, is not diagonal. Thus, $\underline{\theta}$

and $\underline{\gamma}$ are not the desired canonical variables. However, the correlation coefficient of any element θ_i of $\underline{\theta}$ and any element γ_i of $\underline{\gamma}$ is bounded between unity and zero because these variables are correlation-normalized (Equations (3-27) and (3-28)). Therefore, the diagonalization of the cross-correlation matrix $R_{\gamma\theta}$ must be carried out using only unitary operations in order to maintain the correlation-normalized property.

Diagonalization of matrix $R_{\gamma\theta}$ of Equation (3-29) is carried out using the SVD, which results in

$$R_{\gamma\theta} = U_R \Delta_R V_R^H \tag{3-30}$$

Here U_R and V_R are unitary JLxJL matrices, and Δ_R is a JLxJL diagonal matrix with non-negative elements along the diagonal. The diagonal elements of Δ_R are bounded by unity and zero, and are arranged in order of decreasing magnitude, with the largest at the (1,1) location:

$$\Delta_R = \begin{bmatrix} \delta_1 & 0 & \cdots & 0 & 0 \\ 0 & \delta_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \delta_{JL-1} & 0 \\ 0 & 0 & \cdots & 0 & \delta_{JL} \end{bmatrix} \tag{3-31a}$$

$$1 \geq \delta_1 \geq \delta_2 \geq \ldots \geq \delta_{JL} \geq 0 \tag{3-31b}$$

The transformations which diagonalize the cross-correlation matrix $R_{\gamma\theta}$ are identified by inspection of Equation (3-30),

$$U_R^H R_{\gamma\theta} V_R = \Delta_R \tag{3-32}$$

Therefore, the desired canonical variables are defined by the following transformations on the vectors $\underline{\theta}$ and $\underline{\gamma}$ (and on vectors $\underline{X}_P$ and $\underline{X}_F$, from Equations (3-25) and (3-26)),

$$\underline{\mu}(n) = V_R^H \underline{\theta} = V_R^H \mathcal{R}_{P:L,L}^{1/2} \underline{X}_P = V_R^H U_P S_P^{-1/2} U_P^H \underline{X}_P \tag{3-33}$$

$$\underline{\beta}(n) = U_R^H \underline{\gamma} = U_R^H \mathcal{R}_{F:L,L}^{1/2} \underline{X}_F = U_R^H U_F S_F^{-1/2} U_F^H \underline{X}_F \tag{3-34}$$

and the desired canonical correlations are obtained from Equations (3-31) and (3-32) as

$$E[\underline{\beta}(n)\underline{\mu}^H(n)] = R_{\beta\mu} = U_R^H R_{\gamma\theta} V_R = \Delta_R \tag{3-35a}$$

$$\rho_i = \delta_i \qquad i = 1, 2, \ldots, JL \tag{3-35b}$$

Since matrices U_R and V_R are unitary, the norms of vectors $\underline{\mu}(n)$ and $\underline{\beta}(n)$ are equal to the norms of vectors $\underline{\theta}$ and $\underline{\gamma}$, respectively, and the auto-correlation matrix of each of the vectors $\underline{\mu}(n)$ and $\underline{\beta}(n)$ is an identity matrix, as required. The transformation matrices T_P and T_F of Equations (3-17) and (3-18) are obtained from Equations (3-33) and (3-34) as

$$T_P = V_R^H \mathcal{R}_{P:L,L}^{1/2} = V_R^H U_P S_P^{-1/2} U_P^H \tag{3-36}$$

$$T_F = U_R^H \mathcal{R}_{F:L,L}^{1/2} = U_R^H U_F S_F^{-1/2} U_F^H \tag{3-37}$$

This completes the generation of the canonical correlations and associated parameters.

Consider Equation (3-21) now that all the matrices in that equation are known. It is thus possible to solve for the block Hankel matrix as

$$\mathcal{H}_{L,L} = T_F^{-1} R_{\beta\mu} \left(T_P^H\right)^{-1} = \mathcal{R}_{F:L,L}^{1/2} U_R R_{\beta\mu} V_R^H \mathcal{R}_{P:L,L}^{1/2} \tag{3-38}$$

This expression can be factored as follows:

$$\mathcal{H}_{L,L} = \mathcal{O}_L \mathcal{D}_L^H = \left(\mathcal{R}_{F:L,L}^{1/2} U_R R_{\beta\mu}^{1/2}\right)\left(R_{\beta\mu}^{1/2} V_R^H \mathcal{R}_{P:L,L}^{1/2}\right) \tag{3-39}$$

from which the forward and backward innovations observability matrices are determined by inspection. However, a more representative expression for the forward and backward innovations observability matrices is obtained by recognizing that for an Nth-order system the last $JL - N$ canonical correlations are equal to zero. That is, the JLxJL canonical correlation matrix $R_{\beta\mu}$ is partitioned as

$$R_{\beta\mu} = \begin{bmatrix} R_{R1} & [0] \\ [0] & R_{R2} \end{bmatrix} = \begin{bmatrix} R_{R1} & [0] \\ [0] & [0] \end{bmatrix} \tag{3-40}$$

with the NxN diagonal submatrix R_{R1} as,

$$R_{R1} = \begin{bmatrix} \delta_1 & 0 & \cdots & 0 \\ 0 & \delta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \delta_N \end{bmatrix} \tag{3-41a}$$

$$1 \geq \delta_1 \geq \delta_2 \geq \ldots \geq \delta_N > 0 \tag{3-41b}$$

The unitary matrices U_R and V_R are partitioned also into submatrices of dimensions corresponding to the partitions of matrix $R_{\beta\mu}$. Specifically,

$$U_R = [\, U_{R1} \;\; U_{R2} \,] \tag{3-42}$$

$$V_R^H = \begin{bmatrix} V_{R1}^H \\ V_{R2}^H \end{bmatrix} \tag{3-43}$$

where U_{R1} is JLxN, U_{R2} is JLx(JL-N), V_{R1} is JLxN, and V_{R2} is JLx(JL-N). Given these partitions, the forward and backward innovations observability matrices are obtained as

$$\mathcal{O}_L = \mathcal{R}_{F:L,L}^{1/2} U_R R_{\beta\mu}^{1/2} = \mathcal{R}_{F:L,L}^{1/2} U_{R1} R_{R1}^{1/2} \tag{3-44}$$

$$\mathcal{D}_L^H = R_{\beta\mu}^{1/2} V_R^H \mathcal{R}_{P:L,L}^{1/2} = R_{R1}^{1/2} V_{R1}^H \mathcal{R}_{P:L,L}^{1/2} \tag{3-45}$$

This expression for $\mathcal{D}_L$ allows determination of the state of the innovations representation via the finite-data approximation to Equation (3-11). That is,

$$\underline{\alpha}(n) = \mathcal{D}_L^H \mathcal{R}_{P:L,L}^{-1} \underline{x}_P = R_{R1}^{1/2} V_{R1}^H \mathcal{R}_{P:L,L}^{-1/2} \underline{x}_P = R_{R1}^{1/2} \underline{\mu}_1(n) \tag{3-46}$$

where $\underline{\mu}_1(n)$ denotes the first N elements of $\underline{\mu}(n)$. Similarly, the innovations state correlation matrix is determined via the finite-dimensional approximation to Equation (3-16), or directly from Equation (3-46),

$$\Pi = E[\, \underline{\alpha}(n) \underline{\alpha}^H(n) \,] = \mathcal{D}_L^H \mathcal{R}_{P:L,L}^{-1} \mathcal{D}_L = R_{R1} \tag{3-47}$$

Thus, the **NxN** innovations state correlation matrix is diagonal, with its diagonal elements equal to the non-zero canonical correlations. As stated in Section 2.3, the state of the innovations representation has the smallest correlation matrix (in the sense of positive definiteness) of all the admissible correlation-equivalent representations for system (2-2).

The system parameter matrices can be identified using Equation (3-39) and the procedure of Appendix B.3. However, an approach based on Equation (3-21) and the procedure in Appendix B.2 requires less computations, and is the approach preferred herein. The key to the approach is to recognize that Equation (3-21), with $R_{\beta\mu}$ as in Equations (3-40) and (3-41), can be factored into the following two factors:

$$T_2 O_L = \begin{bmatrix} R_{R1}^{1/2} \\ O_{JL-N,N} \end{bmatrix} \tag{3-48}$$

$$\mathcal{D}_L^H T_1^H = \begin{bmatrix} R_{R1}^{1/2} & O_{N,JL-N} \end{bmatrix} \tag{3-49}$$

Given this factorization, proceed as follows. First, operate on the block Hankel matrix $\mathcal{H}_{L,L}$ only on the left with matrix T_2 to obtain

$$T_2 \mathcal{H}_{L,L} = (T_2 O_L)\, \mathcal{D}_L^H = \begin{bmatrix} R_{R1}^{1/2} \\ O_{JL-N,N} \end{bmatrix} \begin{bmatrix} \Gamma & F\Gamma & \cdots & F^{L-1}\Gamma \end{bmatrix} \tag{3-50}$$

where Equation (3-48) has been applied. Now let Z_Γ denote the **NxJ** upper-left-hand submatrix in Equation (3-50),

$$Z_\Gamma = R_{R1}^{1/2} \Gamma \tag{3-51}$$

Then matrix Γ is obtained as

$$\Gamma = R_{R1}^{-1/2} Z_\Gamma \tag{3-52}$$

An analogous procedure is followed to obtain H^H. That is, operate on the block Hankel matrix $\mathcal{H}_{L,L}$ only on the right with the Hermitian of matrix T_1 to obtain

$$\mathcal{H}_{L,L} T_1^H = \mathcal{O}_L \left(\mathcal{D}_L^H T_1^H \right) = \begin{bmatrix} H^H \\ H^H F \\ \vdots \\ H^H F^{L-1} \end{bmatrix} \begin{bmatrix} R_{R1}^{1/2} & O_{N,JL-N} \end{bmatrix} \tag{3-53}$$

where Equation (3-49) has been applied. Now let Z_H denote the JxN upper-left-hand submatrix in Equation (3-53); that is,

$$Z_H = H^H R_{R1}^{1/2} \tag{3-54}$$

Finally, matrix H^H is obtained as

$$H^H = Z_H R_{R1}^{-1/2} \tag{3-55}$$

To determine the system matrix, F, it is necessary to define first a column-shifted (row-shifted) block Hankel matrix as

$$\overleftarrow{\mathcal{H}}_{L,L} = \begin{bmatrix} \Lambda_2 & \Lambda_3 & \cdots & \Lambda_{L+1} \\ \Lambda_3 & \Lambda_4 & \cdots & \Lambda_{L+2} \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda_{L+1} & \Lambda_{L+2} & \cdots & \Lambda_{2L} \end{bmatrix} = \mathcal{O}_L F \mathcal{D}_L^H \tag{3-56}$$

The key to the determination of matrix F is the factorization of the column-shifted Hankel matrix indicated in Equation (3-56). It follows from Equations (3-48), (3-49), and (3-56) that pre-multiplication of $\vec{\mathcal{H}}_{L,L}$ by T_2 and post-multiplication of $\vec{\mathcal{H}}_{L,L}$ by the Hermitian of T_1 results in the following:

$$(3\text{-}57a) \qquad T_2 \vec{\mathcal{H}}_{L,L} T_1^H = (T_2 O_L) F \left(\mathcal{D}_L^H T_1^H\right) = \begin{bmatrix} R_{R1}^{1/2} \\ O_{JL-N,N} \end{bmatrix} [F] \begin{bmatrix} R_{R1}^{1/2} & O_{N,JL-N} \end{bmatrix}$$

$$(3\text{-}57b) \qquad T_2 \vec{\mathcal{H}}_{L,L} T_1^H = \begin{bmatrix} R_{R1}^{1/2} F R_{R1}^{1/2} & O_{N,JL-N} \\ O_{JL-N,N} & O_{JL-N,JL-N} \end{bmatrix} = \begin{bmatrix} Z_F & O_{N,JL-N} \\ O_{JL-N,N} & O_{JL-N,JL-N} \end{bmatrix}$$

In this equation the NxN matrix Z_F is defined implicitly as

$$(3\text{-}58) \qquad Z_F = R_{R1}^{1/2} F R_{R1}^{1/2}$$

The NxN matrix F is obtained easily as

$$(3\text{-}59) \qquad F = R_{R1}^{-1/2} Z_F R_{R1}^{-1/2}$$

This completes the factorization of the output correlation matrix sequence, $\{\Lambda_n\}$.

Determination of the remaining matrix parameters for the innovations model (2-29) is described next. The zero-lag output correlation matrix is estimated directly from the output sequence as

$$(3\text{-}60) \qquad \Lambda_o = \frac{1}{N_T} \sum_{k=0}^{N_T-1} \underline{x}(k) \underline{x}^H(k)$$

where N_T is the total number of output data vectors (length of the output sequence) used in the algorithm. Output correlation lag estimation accuracy depends on this number; thus, N_T should be selected to be sufficiently large. The innovations correlation matrix is obtained as in Equation (2-30a),

$$\Omega = \Lambda_0 - H^H \Pi H \tag{3-61}$$

and the one-step prediction filter (Kalman) gain follows from Equation (2-32a) as

$$K = [\Gamma - F\Pi H]\ \Omega^{-1} = [\Gamma - F\Pi H]\ [\Lambda_0 - H^H \Pi H]^{-1} \tag{3-62}$$

This completes the canonical correlations model parameter identification algorithm.

The canonical correlations approach leads to several alternative solutions to the system identification problem based on the selection of the basis for the factorization of $\mathcal{H}$, and each alternative solution has distinct properties and features. Of interest herein is the solution that corresponds to the backward innovations representation, because it provides additional insight into the canonical correlations formulation. The backward innovations representation solution is obtained in a manner analogous to the development completed above.

Consider the orthogonal projection of $\mathcal{X}^-(n)$ onto $\mathcal{X}^+(n)$ (recall that the preceding development is based on the orthogonal projection of $\mathcal{X}^+(n)$ onto $\mathcal{X}^-(n)$), and let $\mathcal{B}(n-1)$ denote the space generated by the orthogonal projection of $\mathcal{X}^-(n)$ onto $\mathcal{X}^+(n)$. That is,

$$\mathcal{B}(n-1) = \mathcal{X}^-(n) \mid \mathcal{X}^+(n) \tag{3-63}$$

$\mathcal{B}(n-1)$ is the backward state space of the process $\{\underline{x}(n)\}$ because it is spanned by the state of the backward innovations representation. $\mathcal{B}(n-1)$ is finite-dimensional, with dimension equal to N. The space $\mathcal{X}^-(n)$ can be represented as the direct sum of two orthogonal subspaces,

$$\mathcal{X}^-(n) = \mathcal{X}^-(n) \mid \mathcal{X}^+(n) \oplus \mathcal{W}(n-1) = \mathcal{B}(n-1) \oplus \mathcal{W}(n-1) \tag{3-64}$$

where $\mathcal{B}(n-1) \perp \mathcal{W}(n-1)$, and $\mathcal{W}(n-1)$ is the space spanned by the backward innovations process, denoted herein as $\{\underline{\omega}(n)\}$. As before, the geometric structure of the space $\mathcal{X}^-(n)$ defined by Equation (3-64) is valid for all n because the process is stationary.

The space $\mathcal{B}(n-1)$ is spanned by the elements of the conditional expectation of the past given the future,

$$\hat{\underline{x}}_{\mathcal{P}} = E[\underline{x}_{\mathcal{P}} \mid \underline{x}_{\mathcal{F}}] = E[\underline{x}_{\mathcal{P}} \underline{x}_{\mathcal{F}}^H] \left(E[\underline{x}_{\mathcal{F}} \underline{x}_{\mathcal{F}}^H]\right)^{-1} \underline{x}_{\mathcal{F}} = \mathcal{H}^H \mathcal{R}_{\mathcal{F}}^{-1} \underline{x}_{\mathcal{F}} = \mathcal{D}\mathcal{O}^H \mathcal{R}_{\mathcal{F}}^{-1} \underline{x}_{\mathcal{F}} \tag{3-65}$$

which is also the minimum variance estimate of the past for the case of a zero-mean, Gaussian-distributed process. This leads to the algebraic representation of the geometric expression (3-64),

$$\underline{x}_{\mathcal{P}} = \hat{\underline{x}}_{\mathcal{P}} + \underline{\omega}_{n-1} = \mathcal{D}\mathcal{O}^H \mathcal{R}_{\mathcal{F}}^{-1} \underline{x}_{\mathcal{F}} + \underline{\omega}_{n-1} \tag{3-66}$$

where $\underline{\omega}_{n-1}$ is an infinite-dimensional block vector having the backward innovations sequence as block elements; that is,

$$\underline{\omega}_{n-1} = \begin{bmatrix} \underline{\omega}(n-1) \\ \underline{\omega}(n-2) \\ \underline{\omega}(n-3) \\ \vdots \end{bmatrix} \tag{3-67}$$

Now define an N-dimensional vector as

$$\underline{\phi}(n-1) = \mathcal{O}^H \mathcal{R}_f^{-1} \underline{x}_f \tag{3-68}$$

The elements of $\underline{\phi}(n-1)$ span the space $\mathcal{B}(n-1)$, and $\underline{\phi}(n-1)$ is the state of the backward innovations representation at time $n-1$. Using this definition, Equation (3-66) is re-written as

$$\underline{x}_p = \mathcal{D}\underline{\phi}(n-1) + \underline{\omega}_{n-1} \tag{3-69}$$

and this equation is an analytic representation of the statement that the backward system observability matrix maps the backward state space onto the output space. Of particular interest is the first block row of Equation (3-69). Specifically (recall that Γ^H occupies the first J rows of the backward observability matrix),

$$\underline{x}(n-1) = \Gamma^H \underline{\phi}(n-1) + \underline{\omega}(n-1) \tag{3-70a}$$

$$\underline{x}(n) = \Gamma^H \underline{\phi}(n) + \underline{\omega}(n) \tag{3-70b}$$

In Equation (3-70b), and in the remainder of this section, the time argument $n-1$ is replaced by n for notational simplicity (this is permissible because the system is time-invariant). Equation (3-70) is the output equation for the backward innovations representation.

The finite-time approximation and the canonical correlations approach to the parameter identification problem apply also to the backward formulation, and lead to a solution analogous to the forward case. In particular, the state vector is obtained from Equations (3-34), (3-44), and (3-68) as

(3-71) $$\underline{\phi}(n) = \mathcal{O}_L^H \mathcal{R}_{F.L.L}^{-1} \underline{x}_F = R_{R1}^{1/2} \underline{\beta}(n)$$

And the steady-state correlation matrix of the backward innovations representation, Π_b, follows simply as

(3-72) $$\Pi_b = E\left[\underline{\phi}(n)\underline{\phi}^H(n)\right] = \mathcal{O}_L^H \mathcal{R}_{F.L.L}^{-1} \mathcal{O}_L = R_{R1}$$

Notice that the forward and backward innovations representation state correlation matrices are equal to each other (Equations (3-47) and (3-48)),

(3-73) $$\Pi_b = R_{R1} = \Pi$$

A system representation in which the forward and backward state correlation matrices are both diagonal and equal to each other is said to be in balanced coordinates in the stochastic sense (Desai et al., 1985). Balanced coordinates allow effective model order selection and/or model order reduction (Moore, 1981).

3.3 Model Order Determination

Model order determination is a necessary decision for any identification algorithm in applications where the true order of the system generating the channel output data is unknown, or where the true process generating the data may not be a member of the model class adopted to represent the data. In the second case the model generated by the algorithm is a "representation model," as opposed to a "physical model" (a model based on accurate analyses of the underlying physical processes). Determination of the model order is always a difficult problem, and the solution is rarely clear-cut. The canonical correlations identification algorithm

adopted herein does have several strong features that lead to robust and straightforward criteria for model order estimation. Principally, the algorithm identifies the model parameters of the innovations representation for the multichannel process in stochastic balanced coordinates.

Model order selection in the algorithm is based, in one form or another, on the canonical correlations $\{\rho_i\}$, which are the diagonal values of matrix $R_{\beta\mu}$. Thus, it is important to recall that the canonical correlations are real-valued, non-negative, bounded by unity and zero, and are arranged along the diagonal of matrix $R_{\beta\mu}$ in order of decreasing magnitude. Furthermore, the steady-state correlation matrix of the state of the forward (Π) and of the backward (Π_b) innovations models in balanced coordinates are diagonal, with the diagonal elements equal to the canonical correlations. In a balanced representation the position of a state in the state vector is indicative of the importance of the contribution of that state to the output correlation sequence (the first state is equal in importance or more important than the second state; etc.), and the magnitude of the corresponding correlation matrix element is representative of the relative contribution of that state (Moore, 1981). Thus, a simple model order selection approach is to identify the negligible canonical correlations, and select the model order equal to the number of non-negligible canonical correlations.

In most situations involving a finite amount of data, all the canonical correlations are different from zero. This is due to the fact that the singular value decomposition of the Hankel matrix is imperfect for finite data cases because the measurement noise corrupts the estimation of the output correlation matrices. In such cases, model order can be estimated by identifying jump discontinuities in the magnitude of the canonical correlations,

and/or by identifying the correlations at which diminishing returns occur (when the criterion value changes by a negligible amount after increasing the number of states by one).

In the absence of one or more jump discontinuities, external information may be required, such as prior knowledge of the system being modeled. Alternatively, a reasonable model order can be selected, and various analyses can be carried out to reduce the order of the model taking advantage of the features of a state space realization in balanced coordinates.

Model order can be determined also by inspecting the normalized running sum of the canonical correlations. The ith canonical correlation normalized running sum is defined as

$$\text{(3-74)} \qquad NRS_i = \frac{\sum_{k=1}^{i} \rho_k}{\sum_{k=1}^{JL} \rho_k} \qquad i = 1, 2, \ldots, JL$$

Notice that the JLth normalized running sum is equal to unity. Notice also that the parameter NRS_i is the fraction of the past-to-future correlations covered by retaining the ith largest canonical correlations.

Other criteria can be applied for model order determination. Squaring the canonical correlations emphasizes discontinuities, and thus provides a good criterion. The normalized running sum of the squared canonical correlations, which is defined as

$$(3\text{-}75) \qquad NRSS_i = \frac{\sum_{k=1}^{i} \rho_k^2}{\sum_{k=1}^{JL} \rho_k^2} \qquad i = 1, 2, \ldots, JL$$

is still another useful criterion. These last two criteria are heuristic, since there is no significance to the square value of a correlation coefficient, nor to its normalized running sum. However, these two criteria generally perform better model order determination than the canonical correlations and their running sums.

The mutual information between the past and future vectors is the basis for the definition of two other model order determination criteria. Mutual information does have statistical significance, and generally provides effective model order determination. Consider first a set of variables $\{\kappa_i\}$ defined as the following nonlinear function of the canonical correlations:

$$(3\text{-}76) \qquad \kappa_i = -\ln(1 - \rho_i^2) \qquad i = 1, 2, \ldots, JL$$

This set of variables, referred to herein as <u>log parameters</u>, are part of the definition of mutual information, and can be used for model order determination by detection of jump discontinuities or other such behaviour in the sequence. Gelfand and Yaglom (1959) have defined the <u>mutual information</u> between the past and future as the following parameter,

$$(3\text{-}77) \qquad \eta = \frac{1}{2} \sum_{m=1}^{JL} \kappa_m$$

Given this definition, the normalized mutual information parameter for an ith-order model (with $i < N$) is then defined as

$$(3\text{-}78) \qquad \eta_i = \frac{\frac{1}{2}\sum_{m=1}^{i} \kappa_m}{\eta} = \frac{\sum_{m=1}^{i} \kappa_m}{\sum_{m=1}^{JL} \kappa_m} \qquad i = 1, 2, \ldots, JL$$

The value of this parameter represents the fraction of the mutual information in the past about the future that is retained by the state in an ith-order model representation of the output process.

Table 3-1 lists the model order determination criteria presented herein. In an off-line model order determination mode, the procedure to follow with each of the criteria is to examine the sequence of criteria parameter values for discontinuities, diminishing returns, etc., and to select the model order for which a maximum of information is retained. In an on-line mode, one procedure to follow with each of the criteria is to select the model order which corresponds to the criterion value that meets or exceeds a pre-selected threshold. As an example consider the NRS parameters. For this criterion, the model order which corresponds to the parameter value which meets or exceeds a threshold such as 0.95 is selected. Another procedure is to define a threshold which is applied to the increase in value that occurs between two consecutive values of the criterion parameter. A change of a few percent is a reasonable threshold value in many cases.

CRITERION DESCRIPTION	SYMBOL
Canonical correlations	$\{\rho_i\}$
Normalized running sum of canonical correlations	$\{NRS_i\}$
Squared canonical correlations	$\{\rho_i^2\}$
Normalized running sum of squared canonical correlations	$\{NRSS_i\}$
Log parameters	$\{\kappa_i\}$
Normalized mutual information parameters	$\{\eta_i\}$

Table 3-1. List of candidate model order determination criteria.

An important issue related to model order is the selection of the number of block columns (rows) in the block Hankel matrix, L. Based on the rank properties of the block Hankel matrix, the value for L should be selected to satisfy

$$JL > N_E \tag{3-79}$$

where N_E is the expected (or true) model order. If such a value is not available, the best guess at an upper bound for the true model order should be used.

4.0 INNOVATIONS SEQUENCE GENERATION

In the approach pursued in this program, an unknown system of the form (2-2) is modeled as an innovations representation (2-29). Thus, once the innovations model parameters have been identified, an optimal Kalman filter can be configured to generate the innovations sequence, $\{\underline{\varepsilon}(n)\}$, for the likelihood ratio calculations. The approach described in this section is applied to the observation data under each of the two hypotheses.

Any one of several equivalent Kalman filter formulations can be applied to generate the innovations sequence. However, the one-step predictor formulation offers significant advantages in the context of the intended application (Anderson and Moore, 1979). Specifically, the one-step predictor formulation generates the innovations sequence and the filter state update with a simple structure in the case where the input and output noises are correlated ($S \neq [0]$ in Equation (2-5a)), and thus imposes less real-time computational requirements than other formulations. Also, the model identification algorithm generates the parameters for the innovations model. Thus, the one-step predictor formulation is adopted in this work. Strictly speaking, the terminology "one-step predictor" should be used hereafter, but use of the term "Kalman filter" is accepted universally. Both terms are used herein.

The steady-state one-step predictor formulation for the innovations model (2-29) is a linear, time-invariant system described by the following equations:

(4-1a) $$\hat{\underline{\alpha}}(n+1|n) = F\hat{\underline{\alpha}}(n|n-1) + K\underline{\varepsilon}(n) \qquad n \geq n_o$$

(4-1b) $$\underline{\varepsilon}(n) = \underline{x}(n) - \hat{\underline{x}}(n|n-1) = \underline{x}(n) - H^H\hat{\underline{\alpha}}(n|n-1) \qquad n \geq n_o$$

(4-1c) $\hat{\underline{\alpha}}(n_o|n_o\text{-}1) = \underline{0}$

Here $\hat{\underline{\alpha}}(n+1|n)$ is the estimate of the innovations model state vector at time $n+1$ based on observation data up to time n, $\hat{\underline{x}}(n|n\text{-}1)$ is the estimate of the observation vector at time n based on observation data up to time $n\text{-}1$, and $\underline{\varepsilon}(n)$ is the innovations associated with the observation $\underline{x}(n)$. Matrix K is the steady-state filter gain matrix. The filter initial condition is set equal to zero because the innovations model initial condition is zero, Equation (2-29c). A block diagram of the Kalman filter is presented in Figure 4-1, displaying the channel output vector as input, and the innovations sequence vector as output.

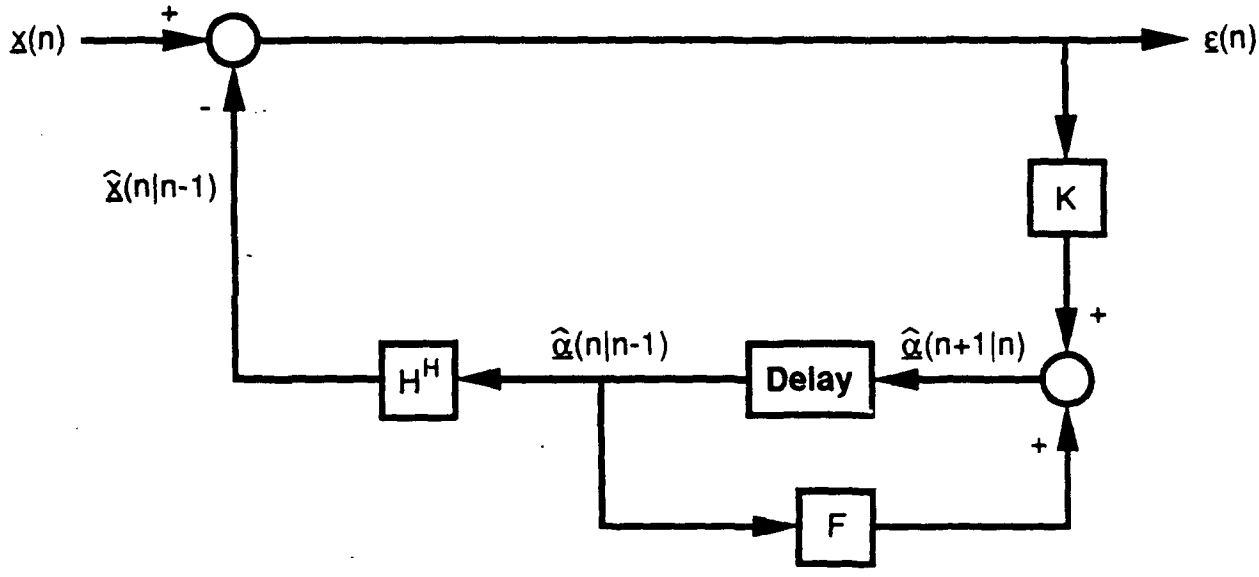


Figure 4-1. Kalman filter block diagram, emphasizing the innovations sequence generation filter function.

The steady-state filter is an approximation to the optimal time-varying filter. If the channel output process is in steady-state, this approximation provides acceptable performance. Additionally, the steady-state filter provides a significant

reduction in the real-time computational requirements over the time-varying filter. In the cases where the channel output process is not in steady-state, filter performance is suboptimal, and the degree of loss of optimality needs to be ascertained. Such a determination is a topic for future research. A related issue involves filter initialization transient effects. Since the steady-state filter gain is used, it may be necessary to neglect the first N_i filter outputs for each data batch. Determination of the value N_i can be carried out via analysis and simulation, and is also a topic for future research.

Anderson and Moore (1979) show that the filter estimation error for an innovations model is zero at all times. That is,

$$\hat{\underline{\alpha}}(n+1|n) = \underline{\alpha}(n+1) \tag{4-2}$$

Correspondingly, the filter estimation error correlation matrix is zero also. This can be inferred from the parallelism between the innovations model (2-29) and the filter representation (4-1). Thus, knowledge of the filter implies knowledge of the innovations model, and viceversa.

5.0 LIKELIHOOD RATIO DETECTION

A detection methodology for complex-valued multichannel Gaussian processes has been developed by Michels (1991) in the context of innovations-based detection. This approach has been generalized recently to include a class of non-Gaussian processes known as spherically-invariant random processes (SIRPs) and using linear estimators (Rangaswamy, Weiner, and Michels, 1993). Michels' methodology can be applied directly to the innovations sequence generated by the approach formulated herein. For brevity, only the likelihood ratio equation is presented here.

As discussed in Section 4.0, a Kalman filter (one-step predictor) is determined for each of the two hypotheses based on processing the multichannel data. The model order for the alternative hypothesis ($\mathbf{H}_1$) filter is chosen to be larger than the model order for the null hypothesis ($\mathbf{H}_0$) filter. For each hypothesis filter, denote the innovations sequence, Equation (4-1b), as

$$\underline{\varepsilon}(n|\mathbf{H}_i) = \underline{x}(n) - \underline{\hat{x}}(n|n-1;\mathbf{H}_i) = \underline{x}(n) - \mathbf{H}^H\underline{\hat{\alpha}}(n|n-1;\mathbf{H}_i) \qquad i = 0, 1 \tag{5-1}$$

The steady-state correlation matrix of the innovations is denoted as $\Omega(\mathbf{H}_i)$, and is defined in Equation (3-60).

Let $\Theta(\mathbf{H}_0,\mathbf{H}_1)$ denote the multichannel likelihood ratio as defined by Michels (1991) for the Gaussian signal case. Then, the log-likelihood ratio (LLR) can be expressed as follows,

$$\ln\left[\Theta(\mathbf{H}_0,\mathbf{H}_1)\right] = \sum_{n=n_o}^{N_T}\left[\ln\left[\frac{|\Omega(\mathbf{H}_0)|}{|\Omega(\mathbf{H}_1)|}\right] + \underline{\varepsilon}^H(n|\mathbf{H}_0)\,\Omega^{-1}(\mathbf{H}_0)\,\underline{\varepsilon}(n|\mathbf{H}_0) - \underline{\varepsilon}^H(n|\mathbf{H}_1)\,\Omega^{-1}(\mathbf{H}_1)\,\underline{\varepsilon}(n|\mathbf{H}_1)\right] \tag{5-2}$$

The LLR is compared to a threshold, $\mathcal{T}$, which is calculated adaptively to maintain a constant false alarm rate (CFAR),

$$\ln[\Theta(\mathbf{H}_0,\mathbf{H}_1)] = \begin{cases} \geq \mathcal{T} & \text{select } \mathbf{H}_1 \\ < \mathcal{T} & \text{select } \mathbf{H}_0 \end{cases} \tag{5-3}$$

A candidate CFAR approach with demonstrated good performance calculates the median of a set of the LLR values from a number of adjacent range cells (at the same azimuth) on both sides of the cell in question, and scales the calculated median value by a pre-determined constant to provide the desired false alarm rate (Metford and Haykin, 1985).

The LLR expression has to be modified if optimal time-varying filters are used instead of the steady-state filters. In such cases the modification is straightforward, and involves replacing the steady-state correlation matrices of the two innovations by their time-varying values.

Alternative expressions for the log-likelihood ratio can be generated based on factorization of the innovations correlation matrix and spatial whitening of the innovations process. This includes Cholesky factorization, LDU decomposition, and SVD. The first two techniques have been described by Michels (1991), and lead to simplified LLR expressions. The SVD technique is derived here.

Consider the steady-state innovations correlation matrices for each of the two hypotheses and carry out an SVD on each correlation matrix. This results in the following decompositions:

$$\Omega(\mathbf{H}_i) = \mathbf{V}_i \Sigma_i \mathbf{V}_i^H \qquad i = 0, 1 \tag{5-4}$$

where matrix V_i is a JxJ unitary matrix, and Σ_i is a diagonal matrix with real-valued, positive elements arranged along the diagonal in decreasing order of magnitude (it is assumed herein that the correlation matrix of the innovations sequence has full rank). That is,

$$\Sigma_i = \begin{bmatrix} \sigma_{i1}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{i2}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{iJ}^2 \end{bmatrix} \qquad i = 0, 1 \tag{5-5a}$$

$$\sigma_{i1}^2 \geq \sigma_{i2}^2 \geq \ldots \geq \sigma_{iJ}^2 > 0 \qquad i = 0, 1 \tag{5-5b}$$

Since matrix V_i is unitary, the determinant and inverse functions of $\Omega(\mathbf{H}_i)$ are obtained easily as

$$\Omega^{-1}(\mathbf{H}_i) = V_i \Sigma_i^{-1} V_i^H \qquad i = 0, 1 \tag{5-6}$$

$$|\Omega(\mathbf{H}_i)| = \prod_{k=1}^{J} \sigma_{ik}^2 \qquad i = 0, 1 \tag{5-7}$$

Now make a linear transformation on the innovations sequence using the unitary matrix V_i, to obtain

$$\underline{v}(n|\mathbf{H}_i) = V_i^H \underline{\varepsilon}(n|\mathbf{H}_i) \qquad i = 0, 1 \tag{5-8}$$

The transformed innovations sequence, $\{\underline{v}(n|\mathbf{H}_i)\}$, is uncorrelated spatially and temporally (recall that $\{\underline{\varepsilon}(n|\mathbf{H}_i)\}$ is uncorrelated temporally), with correlation matrix Σ_i. Transformation of a J-dimensional vector by a unitary matrix rotates the vector in the J-dimensional space, but does not alter its magnitude. Thus, the

spatial whitening transformation does not alter the variance of the elements of the innovations vector.

Substituting Equations (5-4) through (5-8) into Equation (5-2) results in the following LLR expression

$$\ln[\Theta(H_0,H_1)] = \sum_{n=n_o}^{N_T} \sum_{k=1}^{J} \left[\ln\left[\frac{\sigma_{0k}^2}{\sigma_{1k}^2}\right] + \frac{|v_k(n|H_0)|^2}{\sigma_{0k}^2} - \frac{|v_k(n|H_1)|^2}{\sigma_{1k}^2} \right] \tag{5-9}$$

where $v_k(n|H_i)$ denotes the kth element of $\underline{v}(n|H_i)$. This LLR is of the same form as the LLR derived by Michels (1991) for spatial whitening of the innovations using an LDU decomposition.

6.0 SOFTWARE SIMULATION

The identification and filtering algorithms described in the preceding sections have been programmed in FORTRAN 77 for Apple Macintosh processors. Support software for the validation and execution of the routines has been generated also. The support software includes signal generation routines, auxiliary routines for validation, and code for miscellaneous calculations. The identification algorithm makes use of the SVD. An SVD subroutine for complex-valued matrices was obtained from a version of the LINPACK software package (Dongarra et al., 1979). Separate code was written and exercised to validate the LINPACK routines before incorporation into the main algorithm code. The signal generation code uses a Gaussian random number generator obtained from the text by Press et al. (1989). Sample realizations generated by this code were tested for whiteness and gaussianity.

6.1 Software Validation

Code validation was carried out in two steps. First, all subroutines and select segments of code were validated individually. Second, the complete package was validated using examples generated for that purpose. The examples consisted of system models with a simple structure so that the computer output could be predicted. Both real-valued and complex-valued examples were generated. One particular example used is the second-order system defined by the following matrix parameters (for a system model of the form (2-2)):

$$F = \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix}$$

$$H^H = G = D^H = Q = C = I_2$$

This model was used to generate a random vector sequence to validate various aspects of the software. For example, defining matrix F with $f_{11} = f_{12} = f_{22} = 0$ and $f_{21} = 1$ generates an output vector sequence that consists of white noise in each output channel, but the two channels are correlated from one instant to the next (the correlation is due to the coupling induced by the non-zero (2,1) element of F). The output of the identification program should indicate a first-order model, with the first diagonal element of matrix $R_{\beta\mu}$ approximately equal to 0.7071, and low values for the remaining diagonal elements. This was the result obtained. Complex-valued test cases using this sample model were generated by letting F be a diagonal matrix with the desired complex-valued poles along the diagonal.

During validation and testing it was observed that system poles along the real axis are more difficult to estimate than poles with an imaginary component. This is common to most identification algorithms. It was observed also that poles close to the unit axis (in the complex z plane) are estimated more accurately than poles close to the origin. This is due to the fact that the closer that a pole is to the origin, the faster the decay of its response to an excitation.

6.2 Analyses and Simulation Results

The software has been exercised also with cases generated using multichannel AR models provided by the program monitor at RL, Dr. James H. Michels. These cases consist of signal only, clutter only, signal + noise, clutter + noise, and signal + clutter + noise. In all cases the signal, clutter, and noise processes are statistically independent of each other.

Signal AR Model

The signal model is a complex-valued, two-input, two-output AR model of order 2 with the following matrix parameters,

$$\underline{y}_s(n) = - A_s^H(1)\underline{y}_s(n-1) - A_s^H(2)\underline{y}_s(n-2) + \underline{\mu}_s(n)$$

$$A_s^H(1) = \begin{bmatrix} 1.6290 - j\,1.4241\times10^{-7} & 1.3733\times10^{-5} + j\,3.8202\times10^{-13} \\ 1.3733\times10^{-5} + j\,3.8202\times10^{-13} & 1.6290 - j\,1.4241\times10^{-7} \end{bmatrix}$$

$$A_s^H(2) = \begin{bmatrix} 0.80996 - j\,1.4162\times10^{-7} & 1.5259\times10^{-5} - j\,9.0949\times10^{-13} \\ 1.5259\times10^{-5} - j\,9.0949\times10^{-13} & 0.80996 - j\,1.4162\times10^{-7} \end{bmatrix}$$

The input to the signal AR recursion, $\{\underline{\mu}_s(n)\}$, is a zero-mean, unit variance white noise sequence with a spatial correlation structure defined as

$$Q_s = \begin{bmatrix} 0.13038 & 0.12907 \\ 0.12907 & 0.13038 \end{bmatrix}$$

This two-input, two-output AR model corresponds to a fourth-order state space model in an innovations representation (as described in Appendix A), with poles at the following locations in the complex z-plane:

True Signal Model Poles: $-0.81451 \pm j\ 0.38282$
$-0.81449 \pm j\ 0.38281$

This AR system was defined by Michels to have a very high channel-to-channel correlation (~0.99), which indicates that a lower-order model could represent the signal information. Specifically, a second-order state space model can represent the signal information well. Notice that the pole locations are almost repeated roots, which indicates that the two channels are almost

repeated roots, which indicates that the two channels are almost identical. Therefore, given a high-level of channel-to-channel correlation, a reduced-order model should perform adequately.

The AR process $\{\underline{y}_s(n)\}$ is corrupted by a zero-mean, unit-variance white noise sequence $\{\underline{w}(n)\}$ to give the noise-corrupted channel output sequence as

$$\underline{x}_s(n) = \underline{y}_s(n) + \underline{w}(n)$$

For this noise model and the signal model given above, the signal-to-noise ratio (SNR) is approximately 3 dB.

Consider the problem of representing the AR signal in additive white noise with a state space model (see Appendix A). The channel output noise, $\{\underline{w}(n)\}$, alters the parameters of the state space model designed for the AR signal $\{\underline{y}_s(n)\}$ only, but $\{\underline{x}_s(n)\}$ can be represented as the output of a state space model. That is, $\{\underline{y}_s(n)\}$ is represented as the output of an innovations model, but the model for $\{\underline{x}_s(n)\}$, which includes the additive noise $\{\underline{w}(n)\}$, is not an innovations model (there is an innovations model for $\{\underline{x}_s(n)\}$, but it is different from the innovations model for $\{\underline{y}_s(n)\}$). This is a manifestation of the well-known fact that an AR process corrupted by additive output white noise is no longer an AR process. In contrast, the state space model remains a valid representation of the signal even after the addition of a new noise source. The AR model class is a subset of the state space model class; thus, the state space model class can be expected to provide a better fit than the AR model class for a wide range of systems and applications where independent measurement noise is present. Additionally, state space identification algorithms can be expected to deliver comparable performance results using a lower

equivalent model order than algorithms based on time series models.

Clutter AR Model

The clutter model is a complex-valued, two-input, two-output AR model of order 2 with the following matrix parameters,

$$\underline{y}_c(n) = - A_c^H(1)\underline{y}_c(n-1) - A_c^H(2)\underline{y}_c(n-2) + \underline{u}_c(n)$$

$$A_c^H(1) = \begin{bmatrix} -1.0430 & 0.0 \\ 0.0 & -1.0430 \end{bmatrix}$$

$$A_c^H(2) = \begin{bmatrix} 0.4900 & 0.0 \\ 0.0 & 0.4900 \end{bmatrix}$$

The input to the clutter AR recursion, $\{\underline{u}_c(n)\}$, is a zero-mean, unit variance white noise sequence with a spatial correlation structure defined as

$$Q_c = \begin{bmatrix} 1.5502 & 0.0 \\ 0.0 & 1.5502 \end{bmatrix}$$

This two-input, two-output AR model corresponds to a fourth-order state space model in an innovations representation (see Appendix A), with poles at the following locations in the complex z-plane:

True Clutter Model Poles: $0.5215 \pm j\ 0.4669$
$0.5215 \pm j\ 0.4669$

The clutter AR coefficient values, the noise covariance values, and the diagonal structure of this AR system indicate that the two channels are uncorrelated. Thus, a fourth-order state space model can represent the clutter information well. Notice that the pole

locations are repeated roots, which indicates that the two channels are identical.

The clutter AR process $\{\underline{y}_c(n)\}$ is corrupted by the zero-mean, unit-variance white noise sequence $\{\underline{w}(n)\}$ to give the noise-corrupted channel output sequence as

$$\underline{x}_c(n) = \underline{y}_c(n) + \underline{w}(n)$$

For this noise model and clutter model the clutter-to-noise ratio (CNR) is approximately 6 dB.

Selected Simulation Results

The identification and filtering software was exercised with the signal plus noise sequence, $\{\underline{x}_s(n)\}$. Calculated values of the various model order criteria indicate that a second-order state space model is a good approximation to this system, as expected. Plots for two different criteria are presented in Figures 6-1 and 6-2 (all plots herein are for single-realization cases). Specifically, Figure 6-1 shows the canonical correlations, and Figure 6-2 shows the log parameters of Equation (3-76). In both figures the abscissa represents model order. Notice that the log parameters provide an easier determination of model order than the canonical correlations. This has been observed to be the case in most examples considered thus far. The same assessment is true also for the other two criteria that are related to these two criteria (normalized running sum of canonical correlations and normalized mutual information, respectively). The plot of the squared canonical correlations criterion is very similar to the plot of the log parameters, Figure 6-2. This also has been observed in most examples considered thus far.

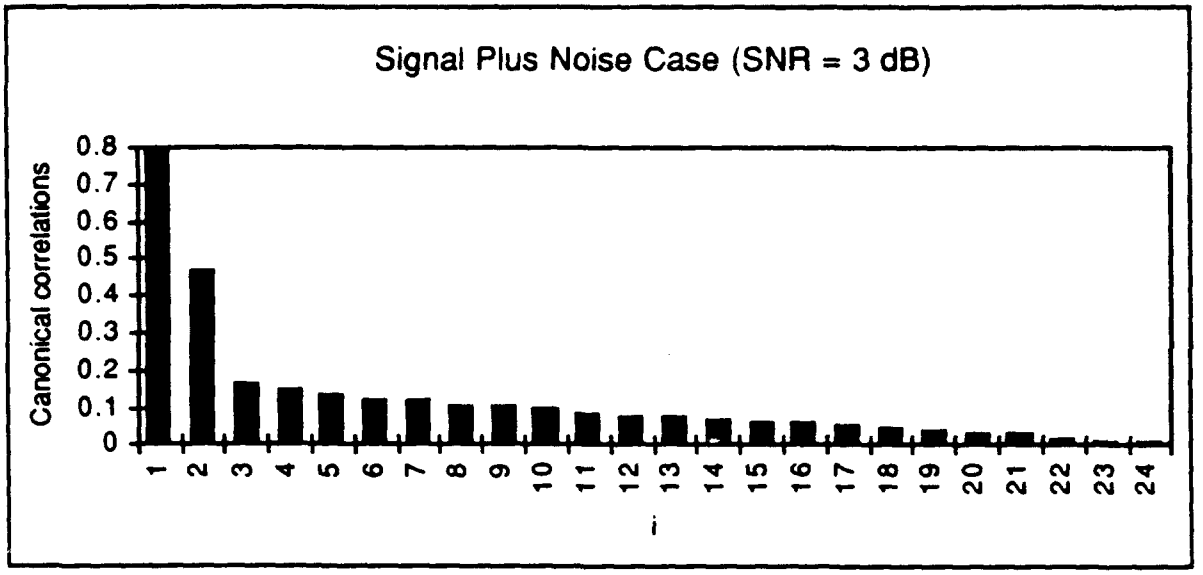


Figure 6-1. Canonical correlations for the signal plus noise case (SNR = 3 dB conditions).

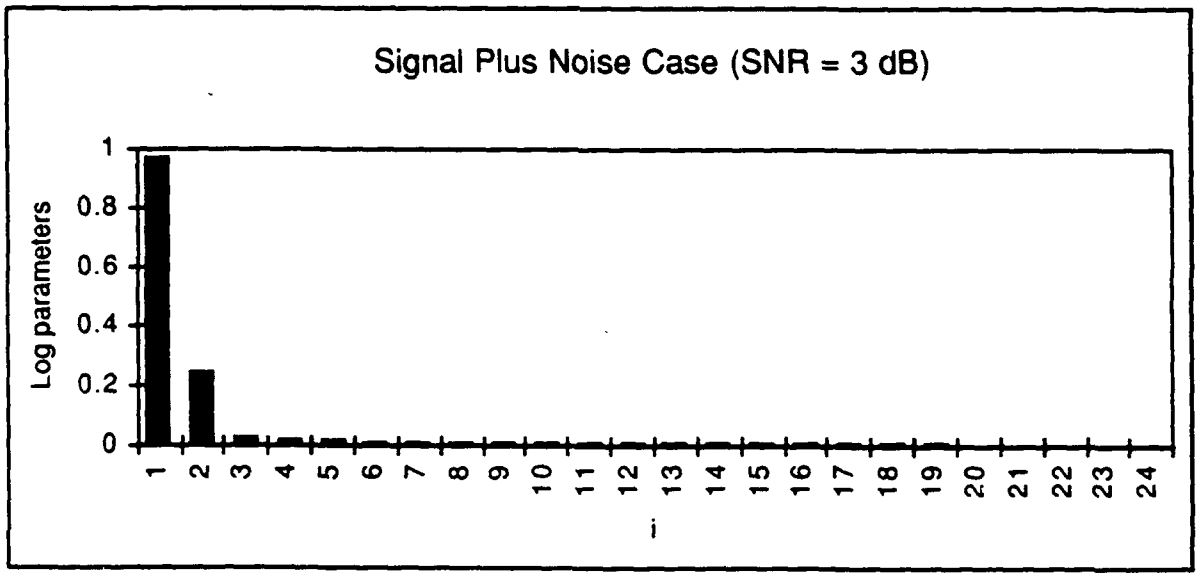


Figure 6-2. Log parameters criterion for the signal plus noise case (SNR = 3 dB).

Based on the above discussions, model order 2 was selected for the analyses and simulations involving the AR signal in white noise. Figure 6-3 is a plot of the real and imaginary parts of the first element of a single realization of the innovations vector process, $\{\varepsilon_1(n)\}$, generated using a filter of order 2. The filter parameters were identified using a total of 25 output correlation matrix lags, including the zero-lag correlation matrix. This corresponds to $L = 12$ in the block Hankel matrix. The output correlation matrix lags were estimated using a single realization of the output process with a duration of $N_T = 2,500$ output sequence vectors. Only the first 500 points are shown in Figure 6-3 (representing one-fifth of the available results), but these points are representative of the total innovations process. The innovations sequence appears to be unbiased, with a calculated sample mean of

$$\bar{\varepsilon}(n) = \begin{bmatrix} 0.0327 - j\,0.0081 \\ 0.0063 + j\,0.0157 \end{bmatrix}$$

Notice also the high degree of "whiteness" exhibited by the innovations. The second element of the innovations vector sequence, $\{\varepsilon_2(n)\}$, behaves similarly.

The zero-lag innovations correlation matrix identified by the software using Equation (3-61) is

$$\Omega = \begin{bmatrix} 1.5328 & 0.5089 + j\,0.0050 \\ 0.5089 - j\,0.0050 & 1.5326 \end{bmatrix}$$

and agrees very well with the sample correlation values. Several simulation runs were made using multiple sample realizations of the same length and filter order two. In all cases the results indicate clearly a white innovations process.

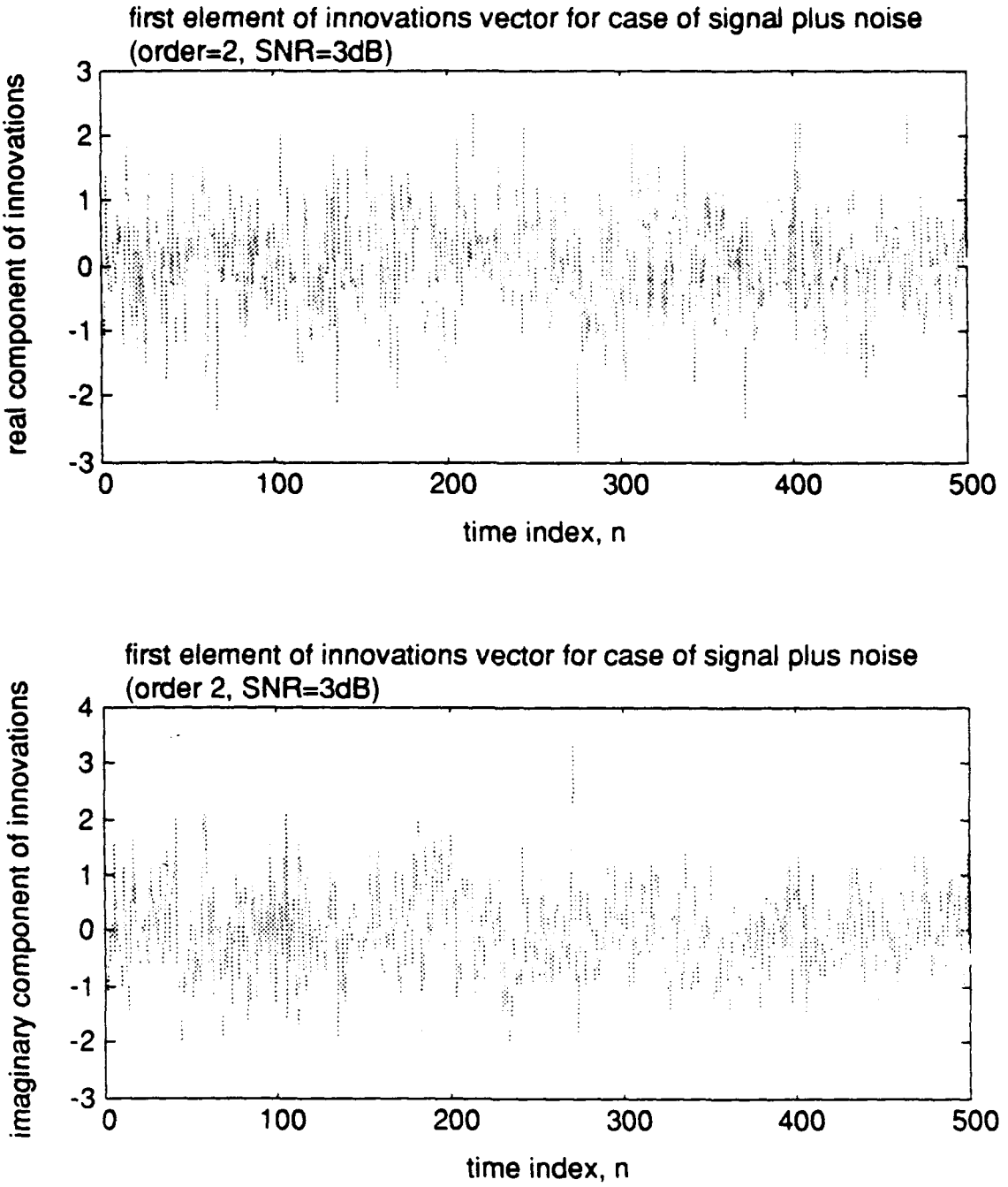


Figure 6-3. Real and imaginary parts of the first element of innovations sequence vector for the case of signal plus noise (SNR = 3 dB conditions).

Identification algorithm performance can be assessed by examining the roots of the identified innovations model system matrix, $\mathbf{F}$. The scatter plots in Figure 6-4, which correspond to results obtained for ten distinct realizations, illustrate the parameter identification capability of the algorithm. These scatter plots show the ten identified root pairs, all in close proximity to the true roots (recall that the true roots are located at $-0.8145 \pm j\ 0.3828$). All the identified roots are at a distance less than 1.5% of the true values, and most are much closer than that.

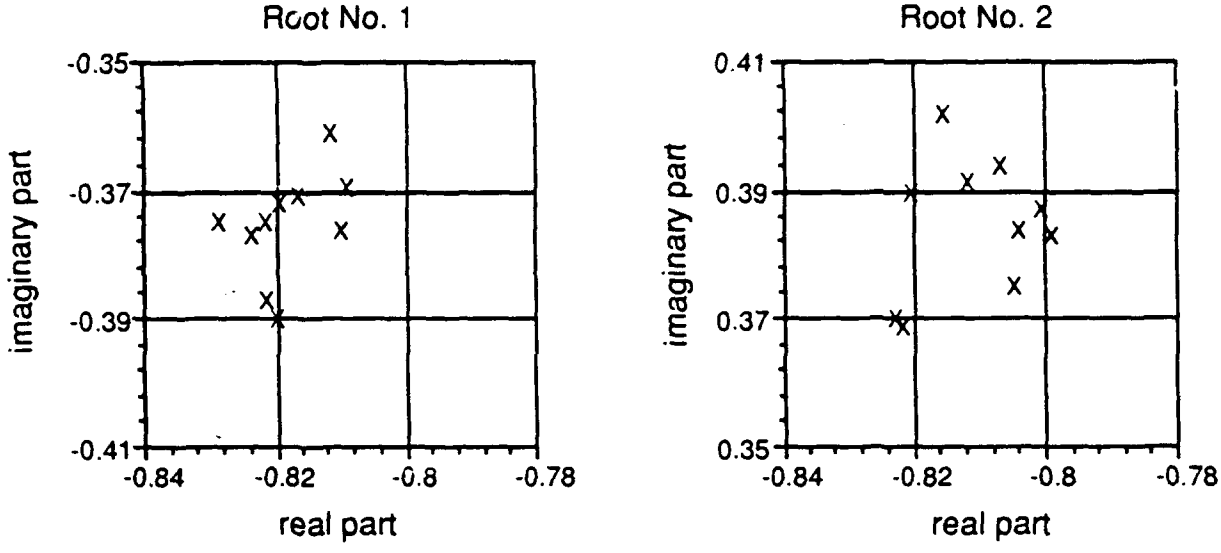


Figure 6-4. Scatter plot of real and imaginary parts of identified model poles for ten distinct realizations of signal plus noise (SNR = 3 dB conditions).

The software was used also to model and analyze the clutter plus noise sequence, $\{\underline{x}_c(n)\}$. For this case at a CNR of 20 dB, the concensus of the model order criteria indicate a fourth-order state-space model, as expected. A plot of the normalized running sum of the canonical correlations (parameter NRS_i) is presented in Figure 6-5, and a plot of the normalized mutual information is presented in Figure 6-6 (both of these figures present single-

realization cases). Notice in Figure 6-6 that there is a significant increase in mutual information as the model order is increased up to fourth-order, but for fifth-order and beyond the increase in mutual information is small compared to the prior increases. Such is not the case with the NRS_i criterion, as evident in Figure 6-5. The plot for the normalized running sum of the canonical correlations squared (parameter $NRSS_i$) is very similar to the plot of the normalized mutual information (Figure 6-6). As in the case of signal plus noise, criteria which involve the canonical correlations in a linear manner are not as useful as criteria based on nonlinear functions of the canonical correlations. These results together with the knowledge of the lack of channel correlation indicate that a fourth-order model is a good approximation to this system.

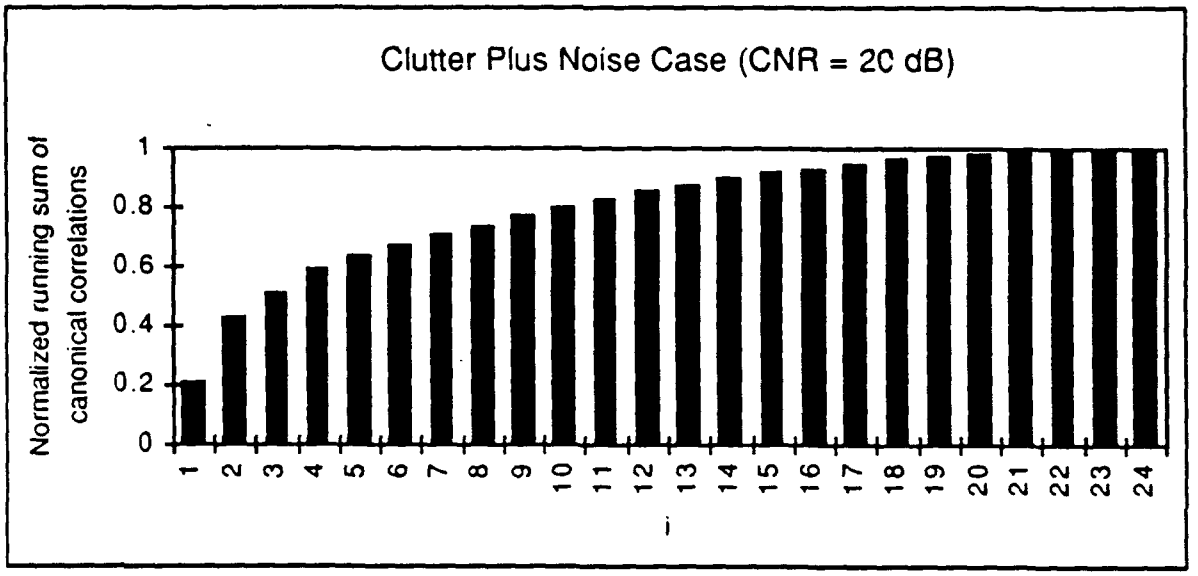


Figure 6-5. Normalized running sum of canonical correlations criterion for the clutter plus noise case (CNR = 20 dB).

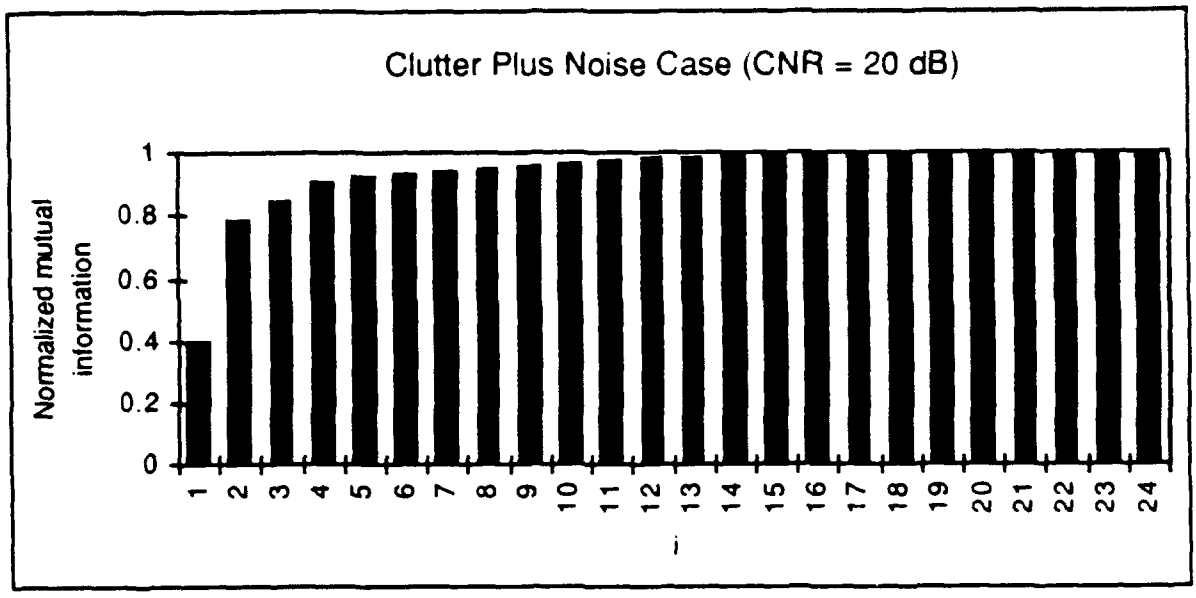


Figure 6-6. Normalized mutual information criterion for the clutter plus noise case (CNR = 20 dB).

Based on the above discussion, model order four was selected for the state space representation of the clutter AR process in additive white noise. Plots of the real and imaginary parts of the first element of the innovations vector process, $\{\varepsilon_1(n)\}$, are presented in Figure 6.7. These results were generated using a fourth-order filter and 6 dB CNR conditions. The other simulation parameters are the same as in the signal plus noise case. Specifically, a total of 25 output correlation matrix lags, including the zero-lag correlation matrix, were used to identify the filter parameters. This corresponds to $L = 12$ in the block Hankel matrix. Also, the output correlation matrix lags were estimated using a single realization of the output process with a duration of $N_T = 2,500$ output sequence vectors. Only the first 500 points are shown in Figure 6-7 (representing one-fifth of the available innovations sequence in this run), but these points are representative of the total innovations sequence. Both components (real and imaginary) of the sequence $\{\varepsilon_1(n)\}$ are unbiased, as indicated by the sample mean,

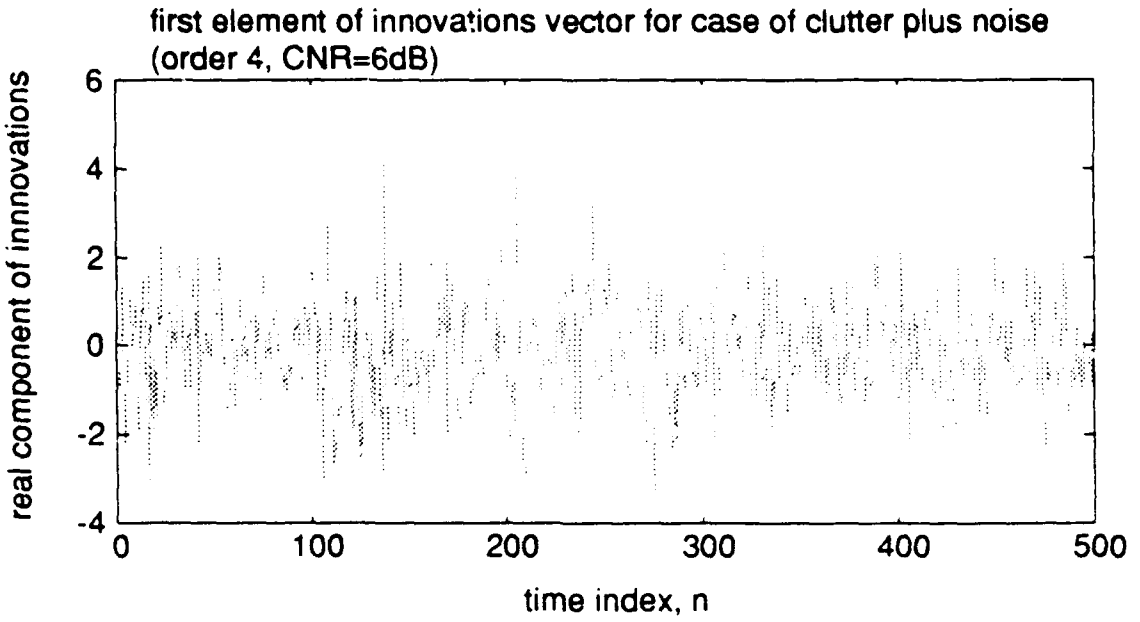
$$\bar{\varepsilon}(n) = \begin{bmatrix} 0.0586 + j\,0.0321 \\ -0.1025 - j\,0.0271 \end{bmatrix}$$

An estimate of the real and imaginary parts of the sample auto-correlation function of $\{\varepsilon_1(n)\}$ of Figure 6-7 is given in Figure 6-8. The behaviour of the real part is representative of a white innovations sequence: an impulse at lag $n = 0$, and approximately equal to zero everywhere else. The imaginary part exhibits low-amplitude oscillations about zero, also as expected of a white innovations. Several distinct output sequence realizations were generated and processed using the same parameters, and the performance was similar in all cases. The zero-lag innovations correlation matrix estimated using Equation (3-61) is

$$\Omega = \begin{bmatrix} 3.2694 & -0.0888 - j\,0.0073 \\ -0.0888 + j\,0.0073 & 3.1369 \end{bmatrix}$$

Element (1,1) of Ω agrees within less than 1% with the sample correlation value of 3.290 + j 0.0 indicated in Figure 6-8. The behaviour of $\{\varepsilon_2(n)\}$ is similar.

Figure 6-9 presents scatter plots of the roots of the fourth-order system characteristic equation for ten realizations. The roots are clustered about the values of the true repeated roots, $0.5215 \pm j\ 0.4669$, which are close to the center of the plots shown. The largest root estimation error is approximately 12.7%. This error is larger than the worst error in the signal plus noise case, and is due to the greater difficulty in estimating faster modes.



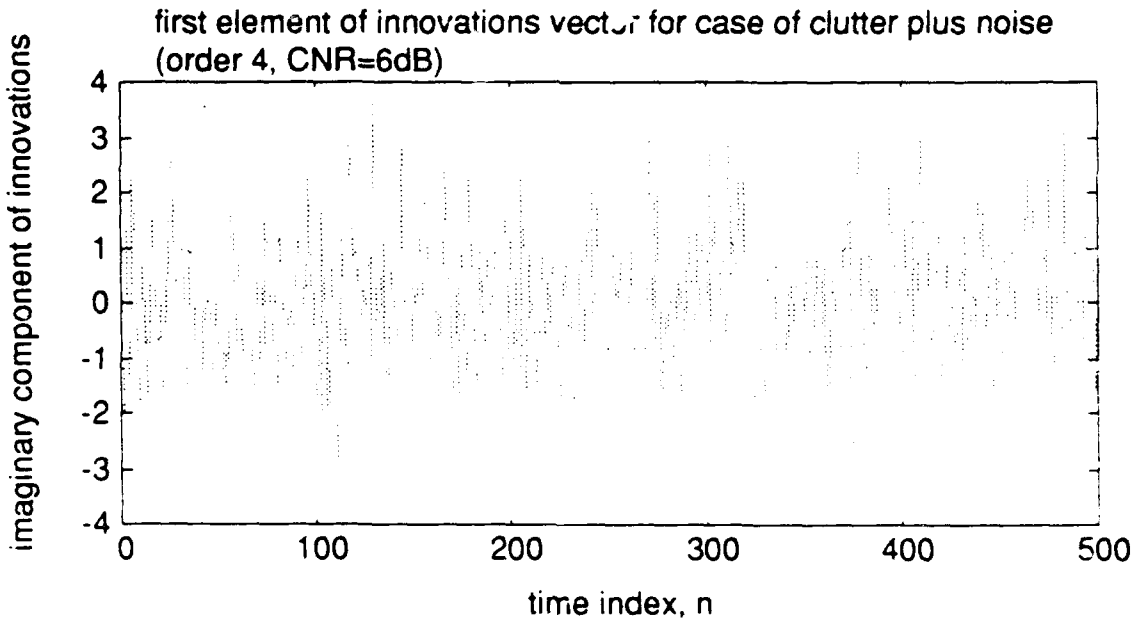
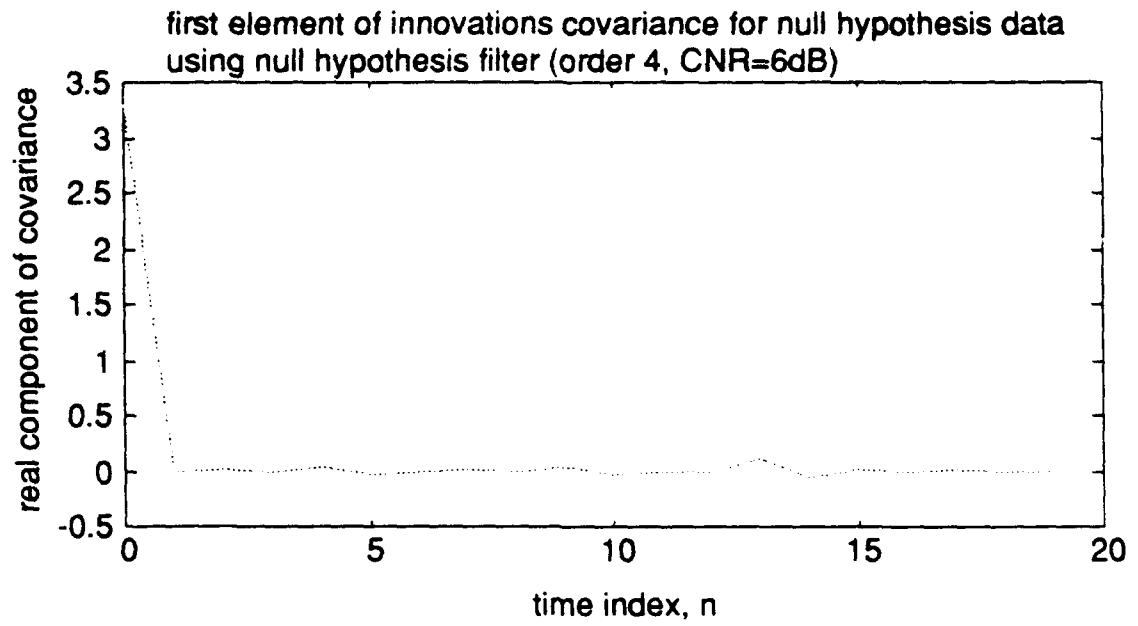


Figure 6-7. Real and imaginary parts of the first element of the innovations sequence vector for the case of clutter plus noise (CNR = 6 dB conditions).



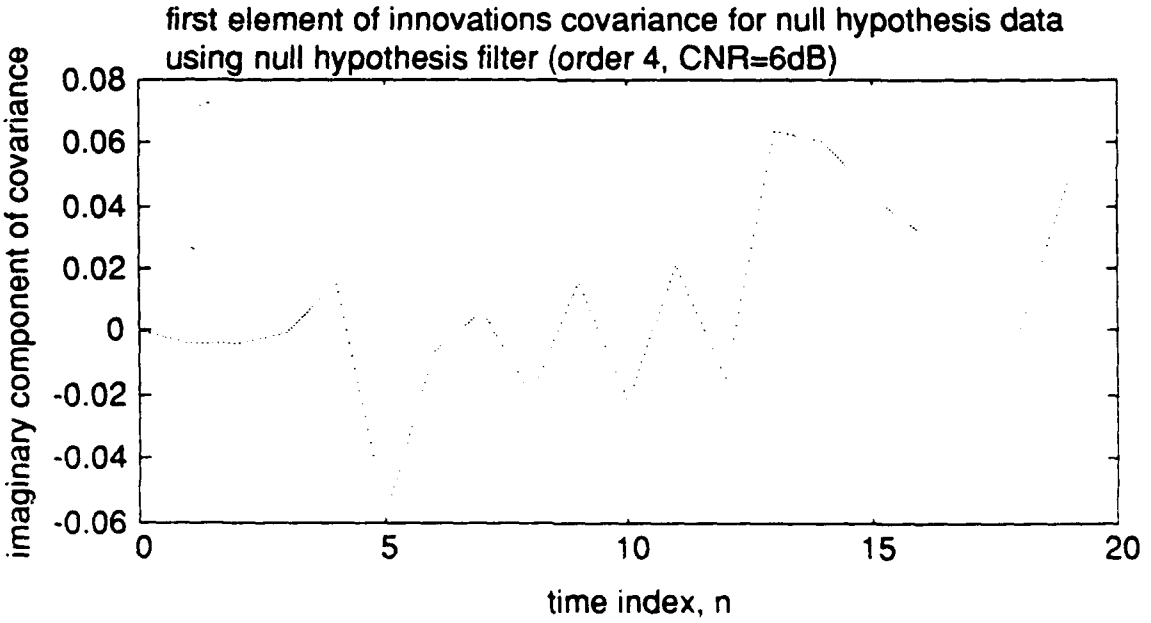


Figure 6-8. Real and imaginary parts of the auto-correlation function of the first element of the innovations sequence vector for the case of clutter plus noise (CNR = 6 dB).

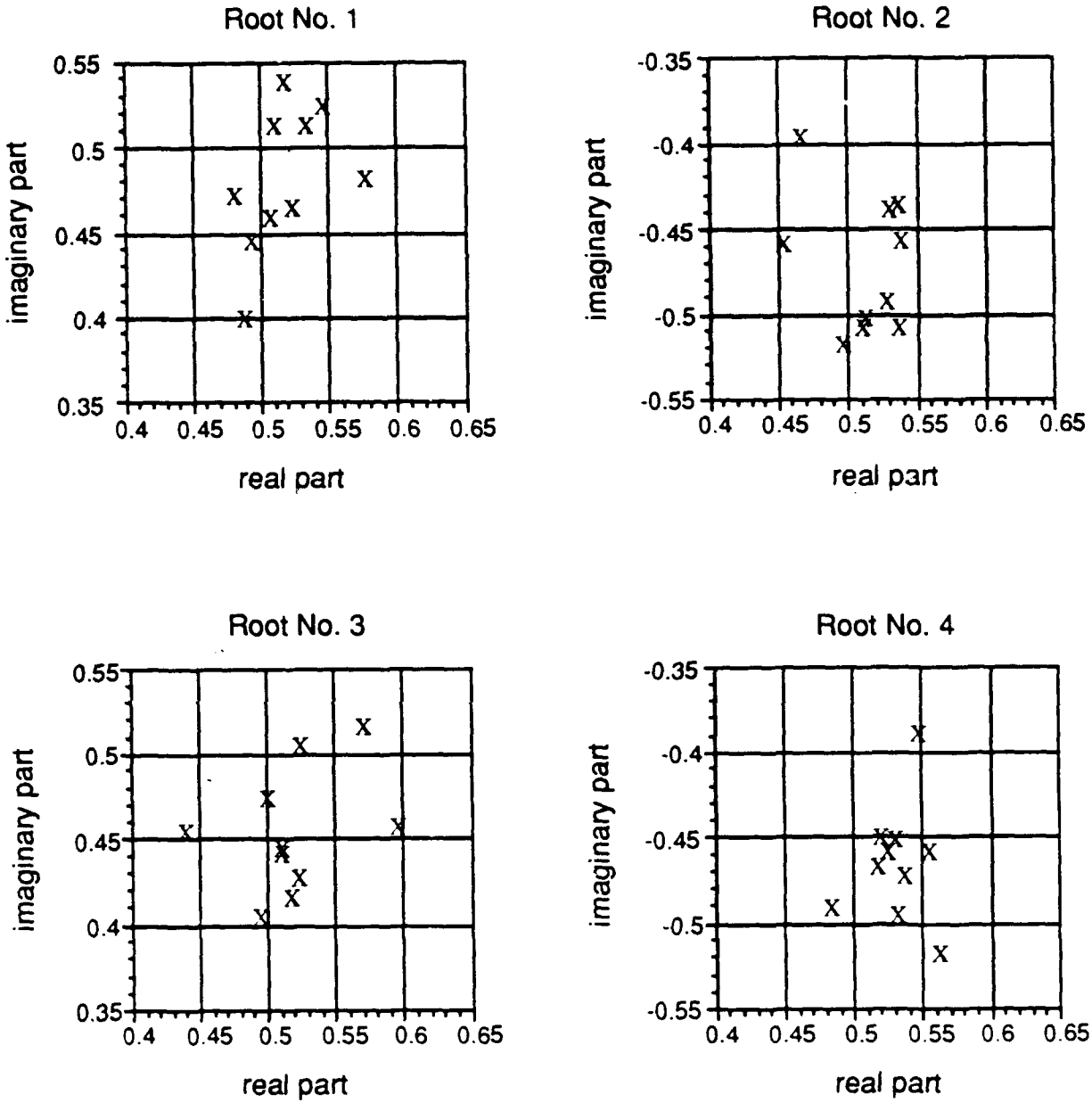


Figure 6-9. Scatter plot of real and imaginary parts of identified model poles for ten distinct realizations of clutter plus noise.

Similar biased behaviour has been observed in other estimation results (Michels, 1992b), as well as in detection performance results (Michels, 1992a) obtained using time series (AR) models. For the state-space approach pursued herein, unbiased estimates with reduced variance can be obtained by averaging several individual estimates and/or by increasing the duration of the output process realization.

Various simulations were carried out to obtain a first-order assessment of the discrimination capability of the innovations-based methodology using the canonical correlations algorithm. One set of simulations involved designing a Kalman filter for each hypothesis, processing data corresponding to each of the two hypotheses using both filters, and analyzing the resulting four filter output sequences (two filters, and each filter processes data sets corresponding to each of the two hypotheses). These results are presented next. As before, all plots correspond to single-realization cases.

Consider first the case of processing data from each of the two hypotheses using a null hypothesis filter, corresponding to clutter + noise only. For this case the filter order is four, as mentioned earlier in the clutter plus noise model discussion. Results are presented herein for two sets of conditions: (a) SNR = 3 dB and CNR = 6 dB; and (b) SNR = 3 dB and CNR = 20 dB. For each set of conditions the procedure described next was followed.

- A realization of the clutter + noise process of duration $N_T = 2,500$ was generated and 25 output correlation matrix lags were estimated. These correlation lags were processed to design a fourth-order Kalman filter. The resulting filter is the filter for the null hypothesis (signal not present).

- The null hypothesis filter was applied to a clutter + noise process sequence of duration N_T = 2,500, and the sample correlation matrix sequence of the filter output sequence was calculated. The real and imaginary parts of the (1,1) element of the resulting sample correlation matrix sequence are plotted in Figure 6-8 for CNR = 6 dB conditions, and Figure 6-10 for CNR = 20 dB conditions. Both sets of figures are representative of the auto-correlation of a white innovations sequence, as expected (both sets of figures show low-level energy content at the higher lags).

- The null hypothesis filter was applied to a combined signal + clutter + noise process sequence (alternative hypothesis case) of duration N_T = 2,500, and the sample correlation matrix sequence of the filter output sequence was calculated. In this case, however, the sequence is not a true innovations sequence because the filter is not optimal for this process. The real and imaginary parts of the (1,1) element of the resulting sample correlation matrix sequence are plotted in Figure 6-11 for CNR = 6 dB conditions, and Figure 6-12 for CNR = 20 dB conditions. Both of these figures show a marked deviation from the expected auto-correlation for a white innovations sequence.

In the discussions and results presented above the (2,2) element of the sample correlation matrix is not referred to. This is due to the fact that its behaviour is very similar to the behaviour of the (1,1) element.

In continuation of the first-order assessment of the discrimination capability of the canonical correlations approach, consider now the case of processing data from each of the two hypotheses using an alternative hypothesis filter, corresponding to the combined process of signal + clutter + noise. Since the signal and clutter are uncorrelated in this set of examples, a sixth-order state space model is required for the combined process. As before, results are presented for two sets of conditions: (a) SNR = 3 dB and CNR = 6 dB; and (b) SNR = 3 dB and CNR = 20 dB. For each set of conditions the procedure described next was followed (all plots are for single-realization cases).

- A realization of the combined signal + clutter + noise vector process of duration $N_T = 2,500$ was generated, and 25 lags of the output correlation matrix sequence were estimated. These lags were processed to design a sixth-order Kalman filter. The resulting filter is the filter for the alternative hypothesis (signal present).

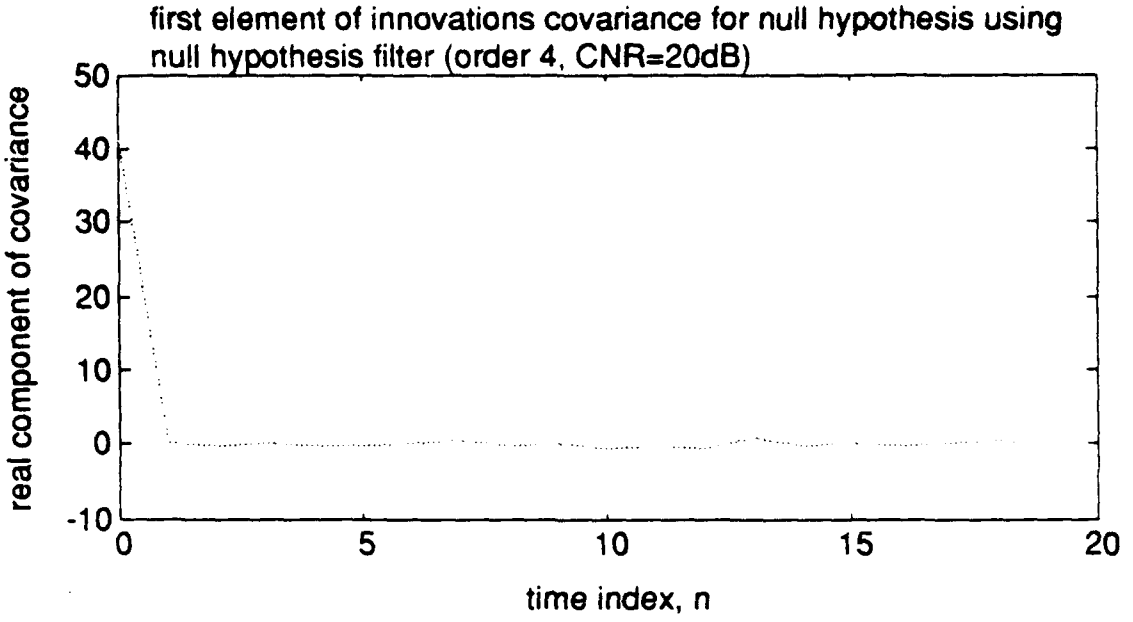
- The alternative hypothesis filter was applied to a combined process sequence of duration $N_T = 2,500$, and the sample correlation matrix sequence of the filter output sequence was calculated. The real part of the (1,1) element of the resulting sample correlation matrix sequence is plotted in Figure 6-13 for CNR = 6 dB conditions, and Figure 6-14 for CNR = 20 dB conditions. Both figures present correlation sequences which correspond to white innovations sequences, as expected (both figures show low-level energy content at the higher lags).

- The alternative hypothesis filter was applied to a clutter + noise process sequence (null hypothesis case)

of duration N_T = 2,500, and the sample correlation matrix sequence of the filter output was calculated. In this case, however, the sequence is not a true innovations sequence because the filter is not optimal for this process. The real part of the (1,1) element of the resulting sample correlation matrix sequence is plotted in Figure 6-15 for CNR = 6 dB conditions, and Figure 6-16 for CNR = 20 dB conditions. The correlation sequence in each of the figures corresponds to a colored process, and not to a white innovations sequence. Such is the expected result.

Figures 6-13 through 6-16 do not include the imaginary part of the sample correlation sequence because it is similar to the imaginary part of the sample correlation sequence presented in the preceding figures. Also, in all cases the behaviour of the (2,2) element is very similar to that of the (1,1) element of the sample correlation matrix sequence, as before.

These results indicate that the innovations-based detection methodology using the canonical correlations identification algorithm can discriminate between data corresponding to each of the two hypotheses. That is, a filter designed for the alternative hypothesis (signal + clutter + noise) generates a true innovations sequence given a signal + clutter + noise channel process, and generates a colored output given a clutter + noise channel process. Analogously, a filter designed for the null hypothesis (clutter + noise) generates a true innovations sequence given a clutter + noise channel process, and generates a colored output given a signal + clutter + noise channel process.



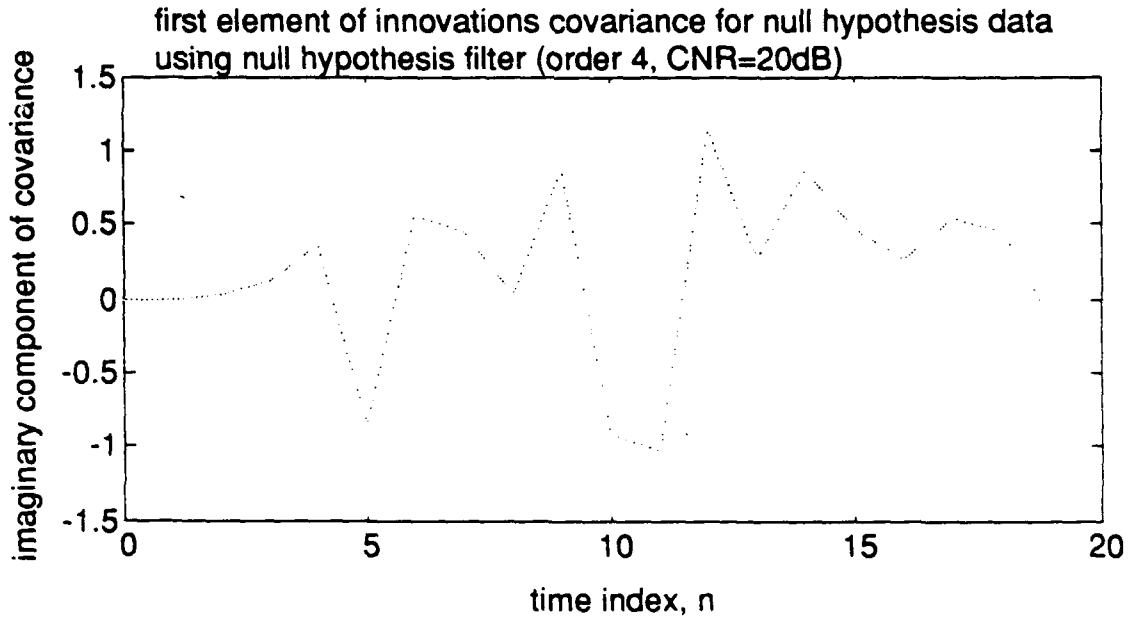
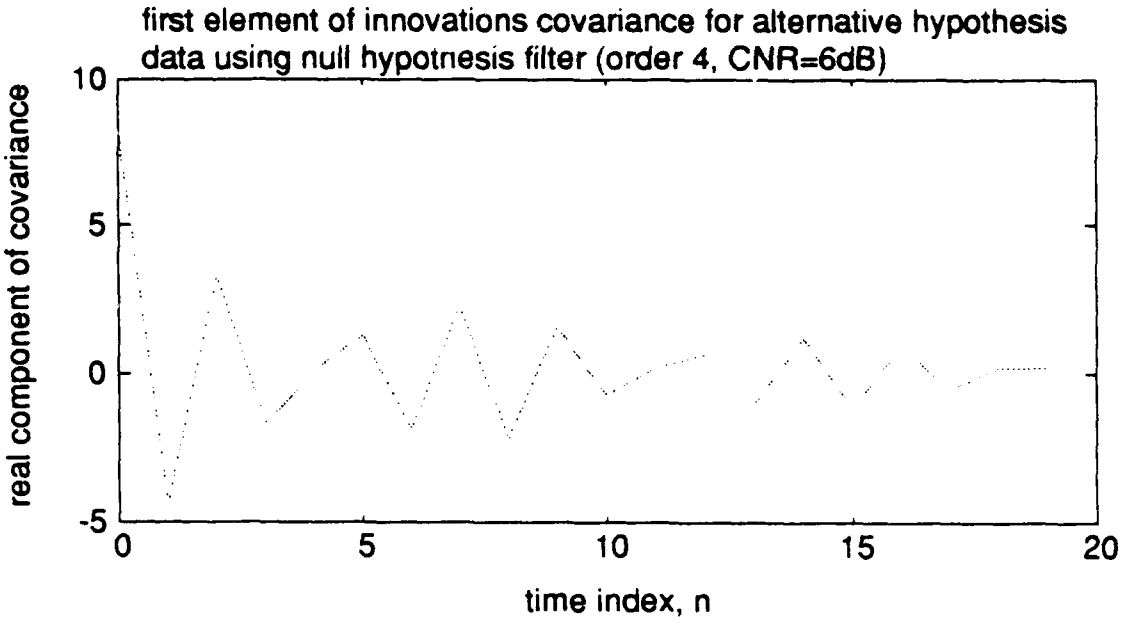


Figure 6-10. Real and imaginary parts of the auto-correlation function of the (1,1) element of the innovations sequence vector for the case of null hypothesis data using the null hypothesis filter (CNR = 20 dB conditions).



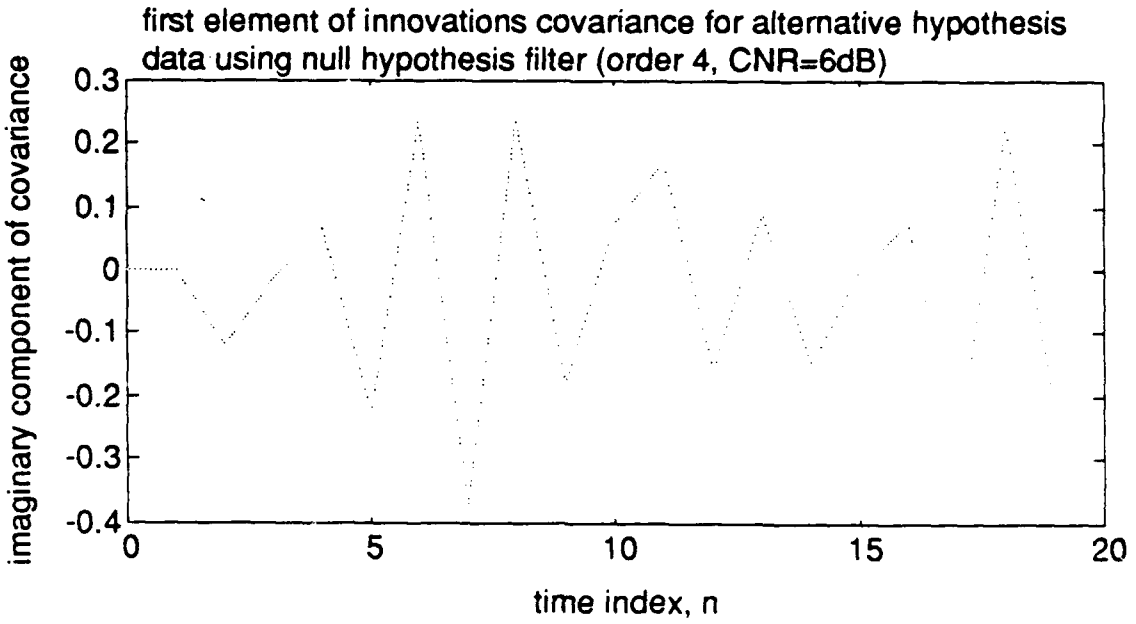
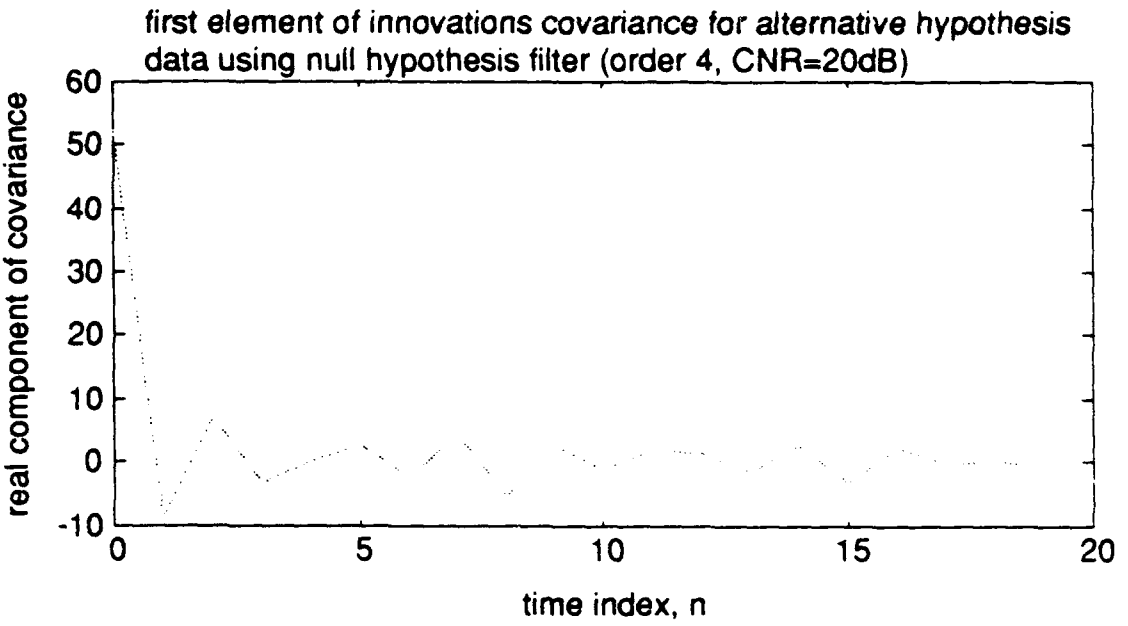


Figure 6-11. Real and imaginary parts of the auto-correlation function of the (1,1) element of the filter output vector for the case of alternative hypothesis data using the null hypothesis filter (CNR = 6 dB conditions).



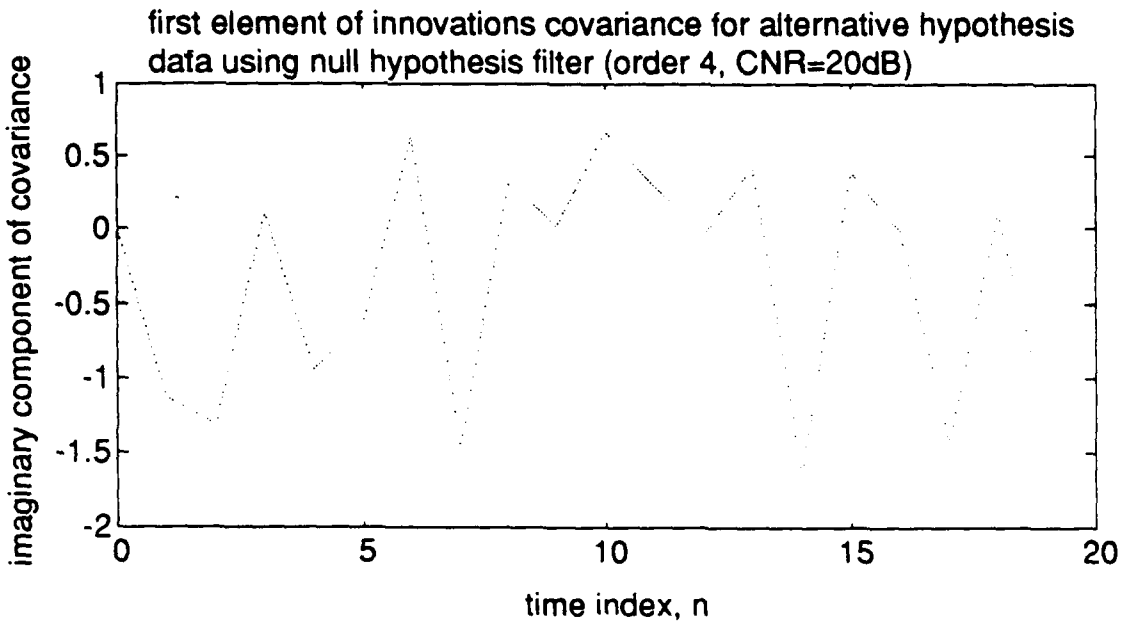


Figure 6-12. Real and imaginary parts of the auto-correlation function of the (1,1) element of the filter output vector for the case of alternative hypothesis data using the null hypothesis filter (CNR = 20 dB conditions).

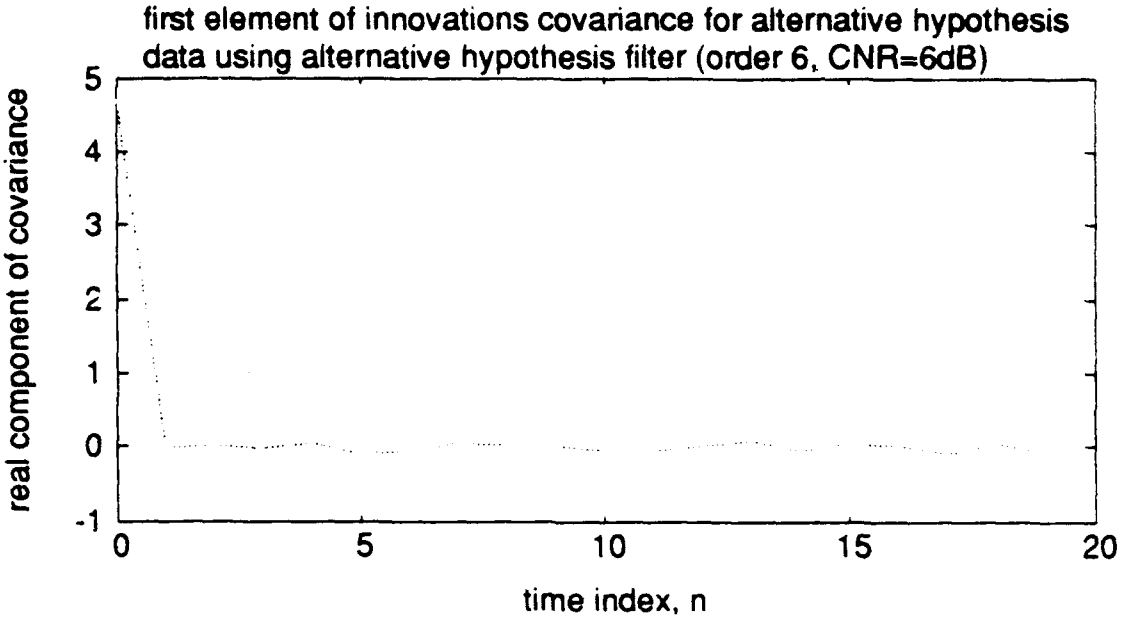


Figure 6-13. Real part of the correlation function of the (1,1) element of the innovations sequence vector for alternative hypothesis data using alternative hypothesis filter (CNR = 6 dB).

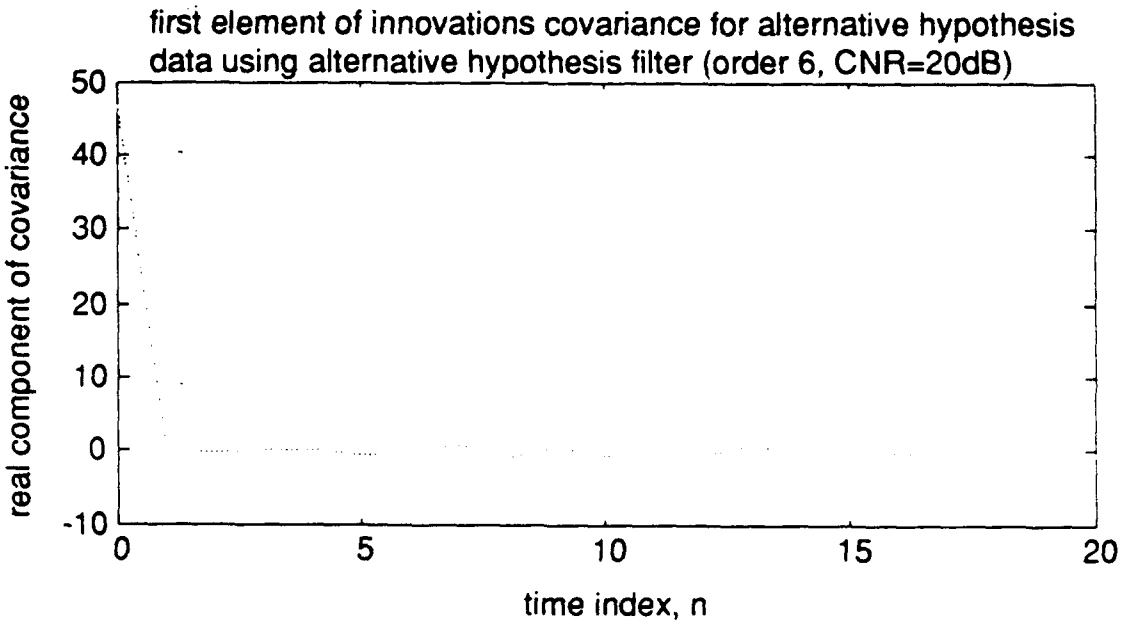


Figure 6-14. Real part of the correlation function of the (1,1) element of the innovations sequence vector for alternative hypothesis data using alternative hypothesis filter (CNR = 20 dB).

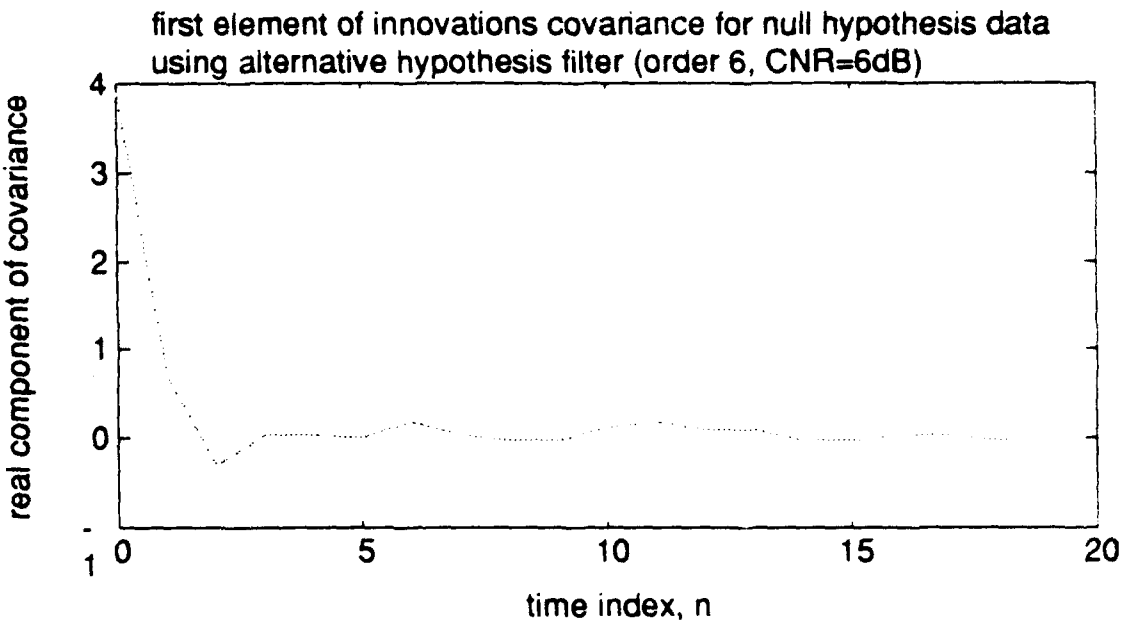


Figure 6-15. Real part of the correlation function of the (1,1) element of the filter output vector for null hypothesis data using the alternative hypothesis filter (CNR = 6 dB conditions).

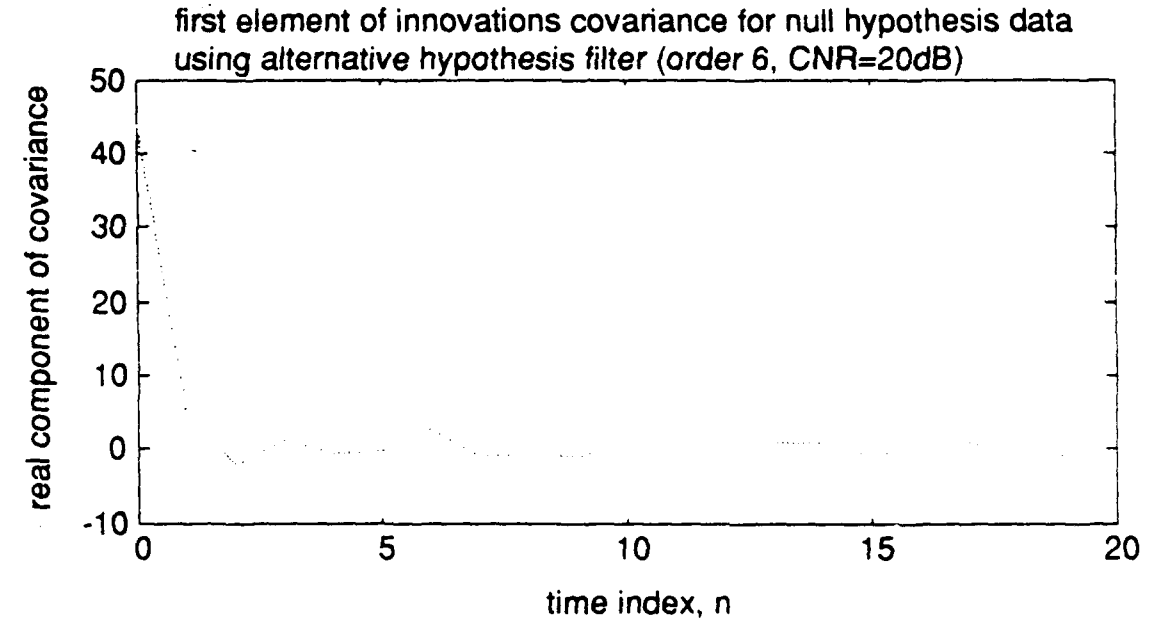


Figure 6-16. Real part of the correlation function of the (1,1) element of the filter output vector for null hypothesis data using the alternative hypothesis filter (CNR = 20 dB conditions).

7.0 CONCLUSIONS AND RECOMMENDATIONS

The work carried out in this program emphasized the development and analysis of a state space methodology and algorithm for the model-based multichannel detection problem in the context of radar system applications. Application of state space techniques for multichannel detection in radar systems is one novel aspect of the work reported here. The state space model class is richer than the time series model class that is used often in radar system applications. And, as demonstrated in this work, the state space model class can be used to represent effectively multichannel radar signals.

Another novel aspect of the work is the utilization in the detection methodology of the canonical correlations algorithm developed by Desai et al. (1985), which in turn is based on the work of Akaike (1974; 1975). This algorithm was adopted in the program for the multichannel radar output modeling and parameter identification functions. In the process, the algorithm was extended to the case of complex-valued radar system data, and an alternative derivation of the algorithm was developed which is based on the SVD technique. The SVD is a robust and stable numerical technique. Thus, the algorithm offers numerical and performance advantages over other techniques.

A computer simulation was developed to validate the algorithm and methodology, and to serve as a testbed for evaluation of the algorithm in radar system applications. The simulation can be exercised with internally-generated sample multichannel output data, or with externally-provided data. Extensive tests were carried out to validate the code.

Simulation-based analyses carried out to date demonstrate the feasibility of the SSC state space approach for multichannel identification and detection in radar system applications. The algorithm has demonstrated the capability to discriminate between signal plus clutter plus noise and clutter plus noise in an innovations-based detection algorithm formulation for the multichannel detection problem. Several cases have been analyzed at various SNR and CNR levels, and in all cases simulated thus far discrimination has been demonstrated.

In the process of completing the work reported here several areas have been identified for further research and development in future programs. These areas are summarized below.

Processor Requirements Definition

Determination of the true potential of the SSC approach for radar system applications requires the establishment of a detailed set of requirements for various radar problems such as space/time processing in a radar array system and the fusion of data from multiple distinct radar systems.

Additional Analyses and Detailed Algorithm Formulation

The analyses listed below are required to generate a detailed algorithm definition for the requirements, and to provide a precise assessment of the SSC approach in the context of radar system applications requirements.

- The innovations model matrix parameters $\mathbf{F}$, Γ, and $\mathbf{H}$ can be estimated using different equations. These alternative approaches need to be evaluated and traded with respect to computational efficiency and accuracy.

- Model order selection criteria for on-line and off-line decisions need to be evaluated and traded further. This includes the ones discussed in Section 3.3.

- The steady-state Kalman filter was used in this program to generate the innovations sequence. Alternatively, the time-varying Kalman filter can be used. The loss in performance, if any, incurred by using the steady-state approximation needs to be evaluated. A related issue is the duration of the transient effect in the case of the steady-state filter.

- Key implementation parameters for radar system applications need to be established. This includes the minimum required channel output sequence duration, and the block dimension of the block Hankel matrix.

- Identification and detection performance should be compared with that of other methods. This includes methods based on time series models.

Once these technical issues are addressed, a detailed architecture design can be defined.

Real-Time Processor Architecture Design

A real-time implementation architecture for the algorithm should be developed, and a candidate hardware implementation identified. Specifically, the following issues should be addressed.

- Generation of an architecture design that best meets the features of the detailed algorithm design and the established processor requirements. The result may be an architecture with features different from those in existing processors, and which is likely to consist of various fundamental architectures (systolic; vector; parallel arrays; etc.).

- Analysis of state-of-the-art processors to determine which contemporary and next-generation VLSI components best match the optimized architecture design and the requirements.

In addressing these issues the emphasis should be on the most computation-intensive tasks of the algorithm.

APPENDIX A. STATE SPACE REPRESENTATION OF TIME SERIES MODELS

Consider a discrete-time, time-invariant, complex-valued, zero-mean, random process $\{\underline{x}(n)\}$ defined as the output of the following state space system model

(A-1a) $$\underline{y}(n+1) = F\underline{y}(n) + G\underline{u}(n)$$

(A-1b) $$\underline{x}(n) = H^H\underline{y}(n) + D^H\underline{w}(n)$$

Vector recursive processes such as moving-average (MA), auto-regressive (AR), and auto-regressive moving-average (ARMA) processes can be modeled with state variable models (SVMs) of the form (A-1). The discussion herein is limited to the particular case where the matrix coefficients of the recursion are square matrices, and the number of output coefficients is equal to the number of input coefficients. The generation of a minimal-order SVM for a vector recursive process involves the properties of polynomial matrix pairs and canonical forms for multiple input, multiple output SVMs.

In contrast, minimal-order SVMs for scalar recursive processes (MA, AR, ARMA) can be generated in a straightforward manner given the recursion coefficients. The SVM generic form appropriate for modeling scalar recursive processes is

(A-2a) $$\underline{y}(n+1) = F\underline{y}(n) + \underline{g}u(n)$$

(A-2b) $$x(n) = \underline{h}^H\underline{y}(n) + d^*w(n)$$

This SVM is a single-input, single-output system.

A.1 Scalar MA Process Model

A scalar MA process of order M is defined as

$$x(n) = \sum_{k=0}^{M} b_k^* u(n-k)$$

$$x(n) = b_0^* u(n) + b_1^* u(n-1) + b_2^* u(n-2) + \ldots + b_M^* u(n-M)$$

where $\{u(n)\}$ is a zero-mean white noise sequence. This recursion can be modeled with a state-space system of the form (A-2) with input sequence $\{u(n)\}$, and state vector with elements that are determined by the input sequence,

$$\underline{y}(n) = \begin{bmatrix} y_1(n) \\ \vdots \\ y_{M-1}(n) \\ y_M(n) \end{bmatrix} = \begin{bmatrix} u(n-1) \\ \vdots \\ u(n-M+1) \\ u(n-M) \end{bmatrix} \qquad \forall\, n$$

The output noise sequence is also equal to the input noise sequence,

$$w(n) = u(n) \qquad \forall\, n$$

which means that the input and output noise sequences in the state space model are completely correlated. Model parameters (F, $\underline{g}$, $\underline{h}$, d) are defined as

$$F = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} = \begin{bmatrix} \underline{0}_M^T \\ I_{M-1} \;\; \underline{0}_{M-1} \end{bmatrix}$$

$$\underline{g} = \underline{i}_1 = \begin{bmatrix} 1 \\ \underline{0}_{M-1} \end{bmatrix}$$

$$\underline{h}^H = \underline{b}^H$$

$$d^* = b_0^*$$

with $\underline{b}^H$ denoting a $1 \times M$ vector defined by M of the MA recursion coefficients,

$$\underline{b}^H = \begin{bmatrix} b_1^* & b_2^* & \dots & b_M^* \end{bmatrix}$$

The special form of matrix F is one of the possible four variations of the so-called companion matrix form. Also, the system parameters, the quadruple $(F, \underline{g}, \underline{h}, d)$, is a variation of the so-called controllable canonical form. These forms have the minimal number of non-zero elements (whereby the name "canonical") of all possible SVMs that model the scalar MA process.

Note that the definition of the state vector $\underline{y}(n)$ in terms of the sequence $\{u(n)\}$ inherently defines the initial condition vector, $\underline{y}(0)$. Once the initial condition vector is defined, the state propagation, Equation (A-2a), provides for continued generation of the output process.

Verification of the above-defined model proceeds as follows. The form of matrix F provides for continued "scrolling" of the input noise sequence as elements of $\underline{y}(n)$, for all n. Validation of the model follows from (A-2b) and the definition of $\underline{h}$, $w(n)$, and $\underline{y}(n)$. That is,

$$x(n) = \underline{h}^H \underline{y}(n) + d^* w(n) = \underline{b}^H \underline{y}(n) + b_0^* u(n)$$

Expanding the term $\underline{b}^H \underline{y}(n)$, and substitution of the definition of $\underline{y}(n)$ in terms of the sequence $\{u(n)\}$ results in

$$x(n) = b_0^* u(n) + b_1^* u(n-1) + b_2^* u(n-2) + \ldots + b_M^* u(n-M)$$

which is the MA process definition. Model validation can be carried out also using the transfer function concept, as summarized next.

Consider first the derivation of the transfer function from the MA process definition. Since the MA process is a discrete-time process, the appropriate tool for the determination of the transfer function is the z-transform. Application of the z-transform to the definition of the MA model results in the expression

$$X(z) = \sum_{k=0}^{M} b_k^* z^{-k} U(z)$$

where z denotes the transform variable, and $X(z)$ and $U(z)$ are the z-transforms of the sequences $\{x(n)\}$ and $\{u(n)\}$, respectively. The transfer function for this linear system is then defined as

$$T(z) = \frac{X(z)}{U(z)} = \sum_{k=0}^{M} b_k^* z^{-k}$$

This corresponds to the transfer function of an all-zero system, as is well known.

The transfer function for a single-input, single-output state variable model (A-2) is of the form

$$T(z) = \underline{h}^H[zI - F]^{-1}\underline{g} + d^*$$

The particular characteristics of matrix F and vector $\underline{g}$ lead to a very simple expression for the product $[zI - F]^{-1}\underline{g}$; namely,

$$[zI - F]^{-1}\underline{g} = \frac{1}{\gamma(z)}\,\underline{\theta}(z)$$

where $\gamma(z)$ is the system characteristic polynomial (the determinant of matrix $[zI - F]$),

$$\gamma(z) = z^M$$

and $\underline{\theta}(z)$ is vector with elements of the form $\theta_i(z) = z^{i-1}$; that is,

$$\underline{\theta}^T(z) = \begin{bmatrix} z^{M-1} & \cdots & z^2 & z & 1 \end{bmatrix}$$

Substitution of these expressions and of $\underline{h}^H$ and d^* in the equation for the transfer function leads to the following result

$$T(z) = \frac{\underline{h}^H\underline{\theta}(z) + d^*\gamma(z)}{\gamma(z)} = \frac{\underline{b}^H\underline{\theta}(z) + b_0^* z^M}{z^M} = z^{-M}\left[b_0^* z^M + \underline{b}^H\underline{\theta}(z)\right]$$

$$T(z) = b_0^* + b_1^* z^{-1} + b_2^* z^{-2} + \ldots + b_M^* z^{-M} = \sum_{k=0}^{M} b_k^* z^{-k}$$

This result is identical to the transfer function expression derived from the definition of the MA process.

A.2 Scalar AR Process Model

A scalar AR process of order M is defined as

$$x(n) = -\sum_{k=1}^{M} a_k^* x(n-k) + u(n)$$

$$x(n) = -a_1^* x(n-1) - a_2^* x(n-2) - \ldots - a_M^* x(n-M) + u(n)$$

where $\{u(n)\}$ is a zero-mean white noise sequence. This recursion can be modeled with a state-space system of the form (A-2) with input sequence $\{u(n)\}$, and state vector with elements that are determined by the output sequence,

$$\underline{y}(n) = \begin{bmatrix} y_1(n) \\ \vdots \\ y_{M-1}(n) \\ y_M(n) \end{bmatrix} = \begin{bmatrix} x(n-1) \\ \vdots \\ x(n-M+1) \\ x(n-M) \end{bmatrix} \qquad \forall\, n$$

The output noise sequence is equal to the input noise sequence,

$$w(n) = u(n) \qquad \forall\, n$$

This implies complete correlation between the input and output noise sequences in the SVM (as in the case of the MA model). Model parameters (F, $\underline{g}$, $\underline{h}$, d) are defined as

$$F = \begin{bmatrix} -a_1^* & -a_2^* & \cdots & \cdots & \cdots & -a_{M-1}^* & -a_M^* \\ 1 & 0 & \cdots & \cdots & \cdots & 0 & 0 \\ 0 & 1 & \ddots & & & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & & \vdots & \vdots \\ \vdots & \vdots & & \ddots & \ddots & 0 & 0 \\ 0 & 0 & \cdots & \cdots & 1 & 0 & 0 \\ 0 & 0 & \cdots & \cdots & 0 & 1 & 0 \end{bmatrix}$$

$$F = \begin{bmatrix} -\underline{a}^H \\ I_{M-1} \quad \underline{0}_{M-1} \end{bmatrix}$$

$$\underline{g} = \underline{i}_1 = \begin{bmatrix} 1 \\ \underline{0}_{M-1} \end{bmatrix}$$

$$\underline{h}^H = -\underline{a}^H$$

$$d^* = 1$$

with $\underline{a}^H$ denoting a vector with elements equal to the AR recursion coefficients,

$$\underline{a}^H = \begin{bmatrix} a_1^* & a_2^* & \cdots & a_M^* \end{bmatrix}$$

The system parameters quadruple (F, $\underline{g}$, $\underline{h}$, d), is in controllable canonical form, as in the MA model case.

Note that the definition of the state vector $\underline{y}(n)$ in terms of the sequence $\{x(n)\}$ inherently defines the initial condition vector, $\underline{y}(0)$. Once the initial condition vector is defined, the state propagation, Equation (A-2a), provides for continued generation of the output process.

Verification of the above-defined model proceeds as follows. From (A-2a) and the definition of F, $\underline{y}(n)$, $\underline{y}(n+1)$, and $\underline{g}$, it follows that

$$y_M(n+1) = - a_1^* y_1(n) - a_2^* y_2(n) - \ldots - a_M^* y_M(n) + u(n)$$

$$y_M(n+1) = - \underline{a}^H \underline{y}(n) + u(n)$$

Also, it follows from (A-2b) and the definition of $\underline{h}$, $w(n)$, and $\underline{y}(n)$ that

$$x(n) = \underline{h}^H \underline{y}(n) + w(n) = - \underline{a}^H \underline{y}(n) + u(n)$$

which indicates that $x(n) = y_M(n+1)$. Then, expanding the term $- \underline{a}^H \underline{y}(n)$ and substitution of the definition of $\underline{y}(n)$ in terms of the sequence $\{x(n)\}$ results in

$$x(n) = - a_1^* x(n-1) - a_2^* x(n-2) - \ldots - a_M^* x(n-M) + u(n)$$

which is the AR process definition.

The transfer function approach can be used also to validate this SVM for scalar AR processes. Application of the z-transform to the definition of the AR model results in the expression

$$\sum_{k=0}^{M} a_k^* z^{-k} X(z) = U(z)$$

where $a_0 = 1$ is introduced for notational simplicity, and $X(z)$ and $U(z)$ are the z-transforms of the sequences $\{x(n)\}$ and $\{u(n)\}$, respectively. The transfer function for this linear system is then defined as

$$T(z) = \frac{X(z)}{U(z)} = \frac{1}{\sum_{k=0}^{M} a_k^* z^{-k}}$$

This corresponds to the transfer function of an all-pole system, as is well known.

Consider now the transfer function for the state variable model (A-2). In the present AR process case, the system characteristic polynomial is

$$\gamma(z) = z^M + a_1^* z^{M-1} + \ldots + a_{M-1}^* z + a_M^*$$

and the particular characteristics of matrix F and vector $\underline{g}$ lead the same simple expression for the product $[zI - F]^{-1}\underline{g}$; namely,

$$[zI - F]^{-1}\underline{g} = \frac{1}{\gamma(z)}\,\underline{\theta}(z)$$

where $\underline{\theta}(z)$ is as defined previously. Notice that the characteristic polynomial can be expressed as

$$\gamma(z) = z^M + \underline{a}^H \underline{\theta}(z)$$

Substitution of these expressions and of $\underline{h}^H$ and d^* in the equation for the transfer function leads to the following result

$$T(z) = \frac{\underline{h}^H \underline{\theta}(z) + d^* \gamma(z)}{\gamma(z)} = \frac{-\underline{a}^H \underline{\theta}(z) + \gamma(z)}{\gamma(z)} = \frac{z^M}{\gamma(z)}$$

$$T(z) = \frac{1}{z^{-M}\gamma(z)} = \frac{1}{\sum_{k=0}^{M} a_k^* z^{-k}}$$

This is identical to the transfer function expression derived from the definition of the AR process.

A.3 Scalar ARMA Process Model

A scalar ARMA process of order M is defined as

$$x(n) = -\sum_{k=1}^{M} a_k^* x(n-k) + \sum_{k=0}^{M} b_k^* u(n-k)$$

$$x(n) = -a_1^* x(n-1) - \ldots - a_{M-1}^* x(n-M+1) - a_M^* x(n-M) + b_0^* u(n) + b_1^* u(n-1) +$$

$$+ b_2^* u(n-2) + \ldots + b_M^* u(n-M)$$

where $\{u(n)\}$ is a zero-mean white noise sequence. This recursion can be modeled with a state-space system of the form (2) with input sequence $\{u(n)\}$, and output noise sequence equal to the input sequence,

$$w(n) = u(n) \qquad \forall\, n$$

This implies complete correlation between the input and output noise sequences in the SVM (as in the case of the MA and the AR models). Model parameters (F, $\underline{g}$, $\underline{h}$, d) are defined as

$$F = \begin{bmatrix} -a_1^* & -a_2^* & \cdots & \cdots & \cdots & -a_{M-1}^* & -a_M^* \\ 1 & 0 & \cdots & \cdots & \cdots & 0 & 0 \\ 0 & 1 & \ddots & & & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & & \vdots & \vdots \\ \vdots & \vdots & & \ddots & \ddots & 0 & 0 \\ 0 & 0 & \cdots & \cdots & 1 & 0 & 0 \\ 0 & 0 & \cdots & \cdots & 0 & 1 & 0 \end{bmatrix}$$

$$F = \begin{bmatrix} -\underline{a}^H \\ I_{M-1} \;\; \underline{0}_{M-1} \end{bmatrix}$$

$$\underline{g} = \underline{i}_1 = \begin{bmatrix} 1 \\ \underline{0}_{M-1} \end{bmatrix}$$

$$\underline{h}^H = \underline{b}^H - b_0^* \underline{a}^H$$

$$d^* = b_0^*$$

Here, as in the AR case, vector $\underline{a}^H$ has elements equal to the AR recursion coefficients,

$$\underline{a}^H = \begin{bmatrix} a_1^* & a_2^* & \cdots & a_M^* \end{bmatrix}$$

and vector $\underline{b}^H$ has elements defined by M of the MA recursion coefficients,

$$\underline{b}^H = \begin{bmatrix} b_1^* & b_2^* & \cdots & b_M^* \end{bmatrix}$$

The system parameters quadruple $(F, \underline{g}, \underline{h}, d)$, is in controllable canonical form, as in the MA and AR model cases.

State vector initial conditions, $\underline{y}(0)$, for this case are related to the input and output sequences in a more complex manner, and have to be selected appropriately. Once the initial condition vector is defined, the state propagation, Equation (A-2a), provides for continued generation of the output process.

The simplest approach to validate this model is via the transfer function approach. Application of the z-transform to the definition of the ARMA model results in the expression

$$\sum_{k=0}^{M} a_k^* z^{-k} X(z) = \sum_{k=0}^{M} b_k^* z^{-k} U(z)$$

where, as before, $X(z)$ and $U(z)$ are the z-transforms of the sequences $\{x(n)\}$ and $\{u(n)\}$, respectively, and $a_0 = 1$ is introduced for notational simplicity. The transfer function for this linear system is then defined as

$$T(z) = \frac{X(z)}{U(z)} = \frac{\sum_{k=0}^{M} b_k^* z^{-k}}{\sum_{k=0}^{M} a_k^* z^{-k}} = \frac{\sum_{k=0}^{M} b_k^* z^{M-k}}{\sum_{k=0}^{M} a_k^* z^{M-k}}$$

where the two polynomial ratio expressions (corresponding to inverse powers of z or direct powers of z) are equivalent, as indicated. This is a transfer function with both poles and zeros, as expected for an ARMA process.

Consider now the transfer function for the state variable model (A-2). For an ARMA process the system characteristic polynomial is

$$\gamma(z) = z^M + a_1^* z^{M-1} + \ldots + a_{M-1}^* z + a_M^*$$

which is equal to that for an AR process SVM model. As in the other two cases,

$$[zI - F]^{-1}\underline{g} = \frac{1}{\gamma(z)}\,\underline{\theta}(z)$$

given the particular features of matrix F and vector $\underline{g}$ ($\underline{\theta}(z)$ is as defined previously). Notice also that, as in the AR process case, the characteristic polynomial can be expressed as

$$\gamma(z) = z^M + \underline{a}^H\underline{\theta}(z)$$

Substitution of these expressions and of $\underline{h}^H$ and d^* in the equation for the transfer function leads to the following result

$$T(z) = \frac{\underline{h}^H\underline{\theta}(z) + d^*\gamma(z)}{\gamma(z)} = \frac{\left(\underline{b}^H - b_0^*\underline{a}^H\right)\underline{\theta}(z) + b_0^*\gamma(z)}{\gamma(z)} = \frac{b_0^* z^M + \underline{b}^H\underline{\theta}(z)}{\gamma(z)}$$

It is easy to verify that this result is identical to the transfer function expression derived from the definition of the ARMA process. That is,

$$T(z) = \frac{b_0^* z^M + \underline{b}^H\underline{\theta}(z)}{\gamma(z)} = \frac{\sum_{k=0}^{M} b_k^* z^{M-k}}{\sum_{k=0}^{M} a_k^* z^{M-k}}$$

where $a_0 = 1$, as before.

A.4 Models for Vector Recursive Processes

Vector recursive processes of the MA, AR, and ARMA type can be represented with SVMs of the type given herein. For vector recursive processes the appropriate notation is:

MA $$\underline{x}(n) = B_0^H \underline{u}(n) + B_1^H \underline{u}(n-1) + B_2^H \underline{u}(n-2) + \ldots + B_M^H \underline{u}(n-M)$$

AR $$\underline{x}(n) = - A_1^H \underline{x}(n-1) - A_2^H \underline{x}(n-2) - \ldots - A_M^H \underline{x}(n-M) + \underline{u}(n)$$

ARMA $$\underline{x}(n) = - A_1^H \underline{x}(n-1) - \ldots - A_{M-1}^H \underline{x}(n-M+1) - A_M^H \underline{x}(n-M) + B_0^H \underline{u}(n) + B_1^H \underline{u}(n-1) + B_2^H \underline{u}(n-2) + \ldots + B_M^H \underline{u}(n-M)$$

where each of the coefficient matrices is dimensioned JxJ. Also analogous to the scalar case, the corresponding transfer function matrices can be defined using the z-transform; which leads to

$$T_{MA}(z) = B(z)$$

$$T_{AR}(z) = A^{-1}(z)$$

$$T_{ARMA}(z) = A^{-1}(z)\, B(z)$$

where A(z) and B(z) are the following matrix polynomials in z,

$$A(z) = \sum_{k=0}^{M} A_k^H z^{-k}$$

$$B(z) = \sum_{k=0}^{M} B_k^H z^{-k}$$

with A_0 the JxJ identity matrix. The matrix pair $\{A(z), B(z)\}$ (including the cases with either $A(z) = I$ or $B(z) = I$) corresponding to a linear discrete-time system is referred to as a matrix polynomial representation or a matrix fraction description (MFD) for the system.

Departing from the time-domain definition for the vector recursive processes, the SVM for each of the three processes is of the same form as the corresponding scalar case SVM, with the following changes: a JxJ coefficient matrix in place of the corresponding coefficient scalar, a JxJ identity matrix (I_J) in place of each unit scalar, and a JxJ null matrix (O_J) in place of each zero-valued scalar. Specifically, the SVM for the ARMA vector process is:

$$F = \begin{bmatrix} -A_1^H & -A_2^H & \cdots & \cdots & \cdots & -A_{M-1}^H & -A_M^H \\ I_J & O_J & \cdots & \cdots & \cdots & O_J & O_J \\ O_J & I_J & \ddots & & & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & & \vdots & \vdots \\ \vdots & \vdots & & \ddots & \ddots & O_J & O_J \\ O_J & O_J & \cdots & \cdots & I_J & O_J & O_J \\ O_J & O_J & \cdots & \cdots & O_J & I_J & O_J \end{bmatrix}$$

$$G = \begin{bmatrix} I_J \\ O_J \\ \vdots \\ O_J \end{bmatrix}$$

$$H^H = \begin{bmatrix} B_1^H - A_1^H B_0^H & B_2^H - A_2^H B_0^H & \cdots & B_M^H - A_M^H B_0^H \end{bmatrix}$$

$$D^H = B_0^H$$

The SVM for the other vector processes (MA; AR) is obtained by substituting the correct values for the vector process coefficients in the above system parameters (that is, $A_i = O_J$ for an MA process; and $B_0 = I_J$ and $B_i = O_J$, $i \geq 1$ for an AR process). In all cases, the transfer function matrix is obtained from the SVM representation as

$$T(z) = H^H[zI - F]^{-1}G + D^H$$

A transfer function calculated according to this relation is equivalent to the transfer function calculated from the appropriate polynomial matrices.

The order (dimension of the state vector) of the resulting SVM for each of the three vector processes is $N = MJ$, since for each process the system matrix F consists of M block rows and M block columns, where each block in each row and column is a JxJ matrix. SVM order is important for practical and computational considerations. An SVM representation is of minimal order if no other SVM representation of lower order leads to the same transfer function matrix. In terms of the system parameters (F, G, H, D), the order of the SVM representation is determined by the rank of the controllability matrix or the rank of the observability matrix, whichever is smaller. Given the form of the matrix pair

(F,G) for all three cases, it is easy to verify that the controllability matrix has full rank for all three cases. However, the observability matrix has a simple form only for the MA SVM. The special form of the observability matrix for the MA case considered herein (with B_M a square matrix) indicates by inspection that the rank of the observability matrix is equal to MJ if and only if matrix B_M has full rank. Such a simple result is not available for the AR and the ARMA SVMs. Determination of the conditions on the coefficients of the polynomial matrices $A(z)$ and $B(z)$ for AR and ARMA vector processes that lead to an SVM representation of minimal order is a difficult problem. This is due to the fact that both AR and the ARMA vector processes lead to a transfer function matrix with elements which are, in general, a ratio of polynomials in z.

Model order and related issues for matrix polynomial representations have been discussed by several researchers. The results summarized next are available in the text by Rosenbrock (1970). Consider the matrix polynomial representation of a system, and assume that the determinant of $A(z)$ is different from zero to eliminate pathological cases. For an AR vector process, the order of the system is given by the degree of the determinant of $A(z)$. Thus, the SVM representation presented herein for vector AR processes is of minimal order if the determinant of $A(z)$ (with $A_0 = I_J$) has degree equal to MJ.

Several definitions need to be introduced prior to stating the relevant results regarding minimal order for ARMA vector processes. A square polynomial matrix is said to be <u>regular</u> when the matrix coefficient of the highest power of z is non-singular. The determinant of a regular polynomial matrix has maximum possible degree. A square polynomial matrix is said to be <u>unimodular</u> if its determinant is a non-zero constant. Unimodular

polynomial matrices have an inverse which is also a polynomial matrix. As an example, the polynomial matrix

$$Q(z) = Q_0 + Q_1 z^{-1} = \begin{bmatrix} 1 + z^{-1} & 3 + z^{-1} \\ 2 + z^{-1} & 4 + z^{-1} \end{bmatrix}$$

is unimodular because the determinant of $Q(z)$ is equal to -2. Notice that the inverse of $Q(z)$ is also a polynomial matrix,

$$Q^{-1}(z) = \frac{1}{-2} \begin{bmatrix} 4 + z^{-1} & -(3 + z^{-1}) \\ -(2 + z^{-1}) & 1 + z^{-1} \end{bmatrix}$$

as expected. Notice also that $Q(z)$ is not a regular matrix since Q_1 is singular.

Two polynomial matrices $A(z)$ and $B(z)$ are said to have a <u>common left divisor</u> $S(z)$ if

$$A(z) = S(z)P_A(z)$$

$$B(z) = S(z)P_B(z)$$

where $S(z)$, $P_A(z)$, and $P_B(z)$ are polynomial matrices. Finally, if all the common (left) divisors of two polynomial matrices $A(z)$ and $B(z)$ are unimodular, then the two matrices are said to be <u>relatively (left) prime</u>. That is, if $A(z)$ and $B(z)$ are relatively (left) prime, then the determinant of the polynomial matrix $S(z)$ in the above factorizations is a constant. This implies that the degree of the determinant of $P_A(z)$ is equal to the degree of the determinant of $A(z)$, and the degree of the determinant of $P_B(z)$ is equal to the degree of the determinant of $B(z)$. Furthermore, the determinant of $P_A(z)$ has no polynomial factors in common with the determinant of $P_B(z)$. A matrix polynomial pair $(A(z), B(z))$ with $A(z)$

and $\mathbf{B(z)}$ relatively (left) prime is an irreducible matrix polynomial representation for the system.

The relevant results for ARMA vector processes can be stated now. As in the AR case, for an ARMA vector process the determinant of $\mathbf{A(z)}$ (with $\mathbf{A_0 = I_J}$) must have degree equal to $\mathbf{MJ}$ for the SVM representation presented herein to be of minimal order. However, two additional conditions must be satisfied. Namely, matrix $\mathbf{B_M}$ must have full rank, and the polynomial matrices $\mathbf{A(z)}$ and $\mathbf{B(z)}$ must be relatively (left) prime. Full rank for matrix $\mathbf{B_M}$ implies that $\mathbf{B(z)}$ is a regular polynomial matrix. If $\mathbf{A(z)}$ and $\mathbf{B(z)}$ are not relatively prime, then the order of the system is reduced by the degree of the determinant of the greatest common (left) divisor of $\mathbf{A(z)}$ and $\mathbf{B(z)}$. This is related to the so-called pole/zero cancelations.

APPENDIX B. DETERMINISTIC REALIZATION ALGORITHMS

Deterministic realization algorithms are of relevance in this work because they provide insight into similar algebraic issues associated with stochastic realization problems due to the similarities in the factorization of the deterministic and stochastic Hankel matrices. Also, deterministic realization algorithms can be applied to obtain the matrix triple (F, Γ, H) of the innovations representation. However, stochastic realization algorithms (such as the canonical correlations algorithm of Section 3.0) are preferred because the state correlation matrix, Π, is identified also.

Two specific realization algorithms are presented below: Ho's algorithm and an algorithm using the singular value decomposition (SVD). Both algorithms are based on algebraic and factorization properties of the deterministic Hankel matrix for a discrete-time, time-invariant, linear system, as summarized next.

B.1 Deterministic Hankel Matrix Properties

Consider a discrete-time, time-invariant, Nth-order system of the form (2-2) where the state, input, and output vectors are deterministic and $D^H = [0]$,

(B-1a) $$\underline{y}(n+1) = F\underline{y}(n) + G\underline{u}(n) \qquad n \geq n_o$$

(B-1b) $$\underline{x}(n) = H^H\underline{y}(n) \qquad n \geq n_o$$

(B-1c) $$n_o = 0$$

As in the rest of this report, the input and output vectors are J-dimensional (in the general case the dimension of the input vector

can be different from the dimension of the output vector). System (B-1) is assumed to be completely reachable and completely observable, and thus has minimal-order. Complete reachability and observability also imply that the NxJL controllability matrix

$$\mathcal{C}_L = \begin{bmatrix} G & FG & \cdots & F^{L-1}G \end{bmatrix} \qquad L \geq N \tag{B-2}$$

and the JLxN observability matrix

$$\mathcal{O}_L = \begin{bmatrix} H^H \\ H^H F \\ \vdots \\ H^H F^{L-1} \end{bmatrix} \qquad L \geq N \tag{B-3}$$

both have rank equal to the system order, N.

The JLxJL deterministic block Hankel matrix for system (B-1) consists of JxJ block elements, with impulse response matrices {A(n)} assigned as the JxJ block elements according to the rule

$$H_{L,L}(\text{block } i, \text{block } j) = A(i+j-1) \qquad i, j = 1, 2, \ldots, L \tag{B-4}$$

for $L \geq N$. In expanded form, matrix $H_{L,L}$ is

$$H_{L,L} = \begin{bmatrix} A(1) & A(2) & \cdots & A(L) \\ A(2) & A(3) & \cdots & A(L+1) \\ \vdots & \vdots & \ddots & \vdots \\ A(L) & A(L+1) & \cdots & A(2L-1) \end{bmatrix} \tag{B-5}$$

Block Hankel matrices are block symmetric, but not element-by-element symmetric, in general. The impulse response sequence {A(n)}

for system (B-1) is given in terms of the system parameter matrices as

$$A(n) = H^H F^{n-1} G \qquad n \geq 1 \tag{B-6}$$

Using the Cayley-Hamilton theorem it is easy to show that the impulse response sequence satisfies a set of recursion relations of the form (Kalman et al., 1969),

$$A(N+k) = -a_1^* A(N+k-1) - \ldots - a_{N-1}^* A(k+1) - a_N^* A(k) \qquad k \geq 1 \tag{B-7}$$

where $\{a_i^*\}$ are the coefficients of the Nth-order characteristic equation of matrix F. Some systems have a minimal polynomial of degree $r < N$. For those systems an rth-order set of recursion relations of the form (B-7) are valid. However, for those systems the set of recursion relations (B-7) based on the characteristic polynomial are valid also.

Inspection of Equations (B-2)-(B-6) indicates that matrix $H_{L,L}$ admits a factorization of the form

$$H_{L,L} = O_L C_L \tag{B-8}$$

Given this factorization and the fact that for $L \geq N$ matrices C_L and O_L both have full rank equal to N, it follows from Sylvester's inequality for the rank of the product of two rectangular matrices (Gantmacher, 1960) that the rank of matrix $H_{L,L}$ is equal to N. In fact, Equation (B-8) also implies that

$$\text{rank}(H_{N+k,N+k}) = N \qquad k \geq 1 \tag{B-9}$$

for a system (B-1) of order N. Equation (B-7) and the block Hankel structure (the sequential arrangement of the matrices $\{A(n)\}$ as block elements of $H_{L,L}$) imply that the block columns (rows) of $H_{L,L}$ also satisfy a recursion of the form (B-7).

A column-shifted deterministic block Hankel matrix, denoted as $\overleftarrow{H}_{L,L}$, is defined by deleting the first block column of $H_{L,L}$ and inserting a new block column in a manner such as to preserve the Hankel structure (the same result is obtained by deleting the first block row and adding a new block row); that is,

$$\text{(B-10)} \qquad \overleftarrow{H}_{L,L} = \begin{bmatrix} A(2) & A(3) & \cdots & A(L+1) \\ A(3) & A(4) & \cdots & A(L+2) \\ \vdots & \vdots & \ddots & \vdots \\ A(L+1) & A(L+2) & \cdots & A(2L) \end{bmatrix} = \mathcal{O}_L F \mathcal{C}_L$$

The significance of the shifted block Hankel matrix is the form of its factorization, as indicated in Equation (B-10). The factorizations in Equations (B-8) and (B-10) are the basis for the realization algorithms presented herein, as well as others.

B.2 Ho's Realization Algorithm

Consider the block Hankel matrix $H_{L,L}$ of Equation (B-5). Apply a sequence of elementary right and left matrix operations (transformations) to the Hankel matrix $H_{L,L}$ in order to drive it to diagonal form, with unity elements along the diagonal. That is,

$$\text{(B-11)} \qquad T_2 H_{L,L} T_1^H = \begin{bmatrix} I_N & O_{N,JL-N} \\ O_{JL-N,N} & O_{JL-N,JL-N} \end{bmatrix}$$

Here T_1 and T_2 are non-singular matrices which represent the product of all the column operations and row operations, respectively, required to transform $H_{L,L}$ into diagonal form (B-11). It is always possible to carry out such elementary transformations because matrix $H_{L,L}$ has rank N. It follows from Equations (B-8) and (B-11) that

$$T_2 H_{L,L} T_1^H = (T_2 O_L)(C_L T_1^H) = \begin{bmatrix} I_N \\ O_{JL-N,N} \end{bmatrix} \begin{bmatrix} I_N & O_{N,JL-N} \end{bmatrix} \tag{B-12}$$

The explicit factorization of the diagonal matrix in Equation (B-12) indicates that T_2 transforms the observability matrix into a matrix with unity elements along the main diagonal and zeros elsewhere. Likewise, T_1 transforms the controllability matrix into a matrix with unity elements along the main diagonal and zeros elsewhere.

Given the factorizations in Equation (B-12), matrix G is obtained as the NxJ upper-left-hand submatrix of $T_2 H_{L,L}$,

$$T_2 H_{L,L} = (T_2 O_L) C_L = \begin{bmatrix} I_N \\ O_{JL-N,N} \end{bmatrix} \begin{bmatrix} G & FG & \cdots & F^{L-1}G \end{bmatrix} \tag{B-13a}$$

$$T_2 H_{L,L} = \begin{bmatrix} G & FG & \cdots & F^{L-1}G \\ & & O_{JL-N,JL} & \end{bmatrix} \tag{B-13b}$$

Similarly, matrix H^H is obtained as the JxN upper-left-hand submatrix of $H_{L,L} T_1^H$,

$$(B\text{-}14a) \qquad H_{L,L}T_1^H = O_L\left(C_L T_1^H\right) = \begin{bmatrix} H^H \\ H^H F \\ \vdots \\ H^H F^{L-1} \end{bmatrix} \begin{bmatrix} I_N & O_{N,JL-N} \end{bmatrix}$$

$$(B\text{-}14b) \qquad H_{L,L}T_1^H = \begin{bmatrix} H^H & \\ H^H F & \\ \vdots & O_{JL,JL-N} \\ H^H F^{L-1} & \end{bmatrix}$$

Finally, it follows from Equations (B-10) and (B-12) that matrix F is obtained as the $N \times N$ upper-left-hand submatrix of $T_2 \vec{H}_{L,L} T_1^H$,

$$(B\text{-}15a) \qquad T_2 \vec{H}_{L,L} T_1^H = \left(T_2 O_L\right)(F)\left(C_L T_1^H\right) = \begin{bmatrix} I_N \\ O_{JL-N,N} \end{bmatrix} [F] \begin{bmatrix} I_N & O_{N,JL-N} \end{bmatrix}$$

$$(B\text{-}15b) \qquad T_2 \vec{H}_{L,L} T_1^H = \begin{bmatrix} F & O_{N,JL-N} \\ O_{JL-N,N} & O_{JL-N,JL-N} \end{bmatrix}$$

This completes Ho's algorithm for the determination of a matrix triple (F, G, H) which realizes an impulse response matrix sequence $\{A(n)\}$.

In the above discussion it is assumed implicitly that the given impulse response matrix sequence corresponds to an Nth-order system of the form (B-1), and that the sequence is available without distortions due to noise or other such effects. If either of these two conditions is not satisfied, then the diagonalized block Hankel matrix (Equation (B-11)) will have non-zero elements beyond the Nth diagonal location. In such cases, model order is estimated by determination of the diagonal location beyond which

the diagonal elements represent zero or represent the contribution of noise. This requires appropriate selection of the block dimension of the Hankel matrix (L must be sufficiently large). One mechanism for verifying the cut-off diagonal element is to examine the norm of the transformation matrices. But this implies a large computational load. Other alternative model order selection criteria have difficulties also. These difficulties arise because the Hankel matrix is transformed to an identity, and because no constraints are imposed on the norm of the columns of the transformation matrices.

B.3 SVD-Based Realization Algorithm

An alternative implementation of Ho's algorithm has been proposed by Zeiger and McEwen (1974). Instead of carrying out elementary row and column operations on the Hankel matrix (Equation (B-11)), Zeiger and McEwen (1974) propose a singular value decomposition of the Hankel matrix. This offers two important advantages: first, the SVD is numerically robust even for matrices of large dimensions; second, the SVD provides an inherent mechanism for the determination of the model order, or determination of the best model fit for a selected model order (herein best is intended in the sense of minimizing the Frobenius norm of the difference between the given Hankel matrix and the Hankel matrix that corresponds to the selected model order). Golub (1969) provides a good summary of the SVD, its properties, and its applications.

Consider the block Hankel matrix $H_{L,L}$ of Equation (B-5). The singular value decomposition (SVD) of $H_{L,L}$ is a factorization of the form

(B-16) $$H_{L,L} = T_B \Delta_{JL} T_A^H$$

where T_A and T_B are unitary matrices, and Δ_{JL} is a diagonal matrix with non-negative, real-valued diagonal elements arranged in order of descending magnitude. That is, matrix Δ_{JL} is of the form

$$\text{(B-17a)} \qquad \Delta_{JL} = \begin{bmatrix} \Delta_N & O_{N,JL-N} \\ O_{JL-N,N} & O_{JL-N,JL-N} \end{bmatrix}$$

$$\text{(B-17b)} \qquad \Delta_N = \begin{bmatrix} \delta_1 & 0 & \cdots & 0 & 0 \\ 0 & \delta_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \delta_{N-1} & 0 \\ 0 & 0 & \cdots & 0 & \delta_N \end{bmatrix}$$

$$\text{(B-17c)} \qquad \delta_1 \geq \delta_2 \geq \ldots \geq \delta_{N-1} \geq \delta_N > 0$$

The factorization in Equation (B-16) is a generalization of the concept of the eigenvector/eigenvalue decomposition of a matrix, with the property that it is applicable also to non-square matrices. This decomposition is unique (except possibly for sign changes to the columns of the unitary matrices T_A and T_B). The diagonal elements of Δ_{JL} are referred to as the singular values of $H_{L,L}$, and the rank of matrix $H_{L,L}$ is equal to the number of non-zero singular values. The columns of T_B are the left singular vectors, and the columns of T_A are the right singular vectors of $H_{L,L}$. For an Nth-order system in noise-free conditions, there are N non-zero singular values, and $JL - N$ zero-valued singular values.

In the case where the matrix to be decomposed is Hermitian (not just block Hermitian), the SVD is an eigendecomposition. That is, for a Hermitian matrix, $T_A = T_B = T$, the columns of T

(singular vectors) are the eigenvectors, and the singular values are the eigenvalues.

Another important property of the SVD is that it provides a means for determining a JLxJL matrix M of specified rank k (with $k < N$) which best approximates the Hankel matrix in the sense of minimizing the Frobenius norm of the difference between the Hankel matrix and the desired matrix M. The desired optimal matrix approximation is of the form (B-16), with the modification that only the first k diagonal elements of submatrix Δ_N (Equation (B-17) are retained, and the diagonal elements beyond the kth one are set to zero.

Equation (B-16) can be factorized further by taking the matrix square root of Δ_{JL} to obtain (since Δ_{JL} is diagonal, its matrix square root is trivial)

$$H_{L,L} = T_B \Delta_{JL} T_A^H = \left(T_B \Delta_{JL}^{1/2}\right)\left(\Delta_{JL}^{1/2} T_A^H\right) = O_L C_L \tag{B-18}$$

Given the explicit factorization in Equation (B-18), it follows from Equation (B-2) that matrix G is given by the NxJ upper-left-hand submatrix of $\Delta_{JL}^{1/2} T_A^H$; similarly, from Equations (B-3) and (B-18), it follows that matrix H^H is given by the JxN upper-left-hand submatrix of $T_B \Delta_{JL}^{1/2}$.

Matrix F is obtained using Equations (B-10) and (B-16)-(B-18). Specifically,

$$F = \begin{bmatrix} \Delta_N^{-1/2} & O_{N,JL-N} \end{bmatrix} T_B^H \tilde{H}_{L,L} T_A \begin{bmatrix} \Delta_N^{-1/2} \\ O_{JL-N,N} \end{bmatrix} \tag{B-19}$$

This completes the SVD-based algorithm for the determination of a matrix triple (F, G, H) which realizes an impulse response matrix sequence {A(n)}.

As before, the above development assumed that the available impulse response matrix sequence corresponds to an Nth-order system of the form (B-1), and that the sequence is available without distortions due to noise or other such effects. If either of these two conditions is not satisfied, then there will be more than N non-zero singular values for the block Hankel matrix (Equations (B-16) and (B-17)). In such cases, model order is estimated by determination of the diagonal location beyond which the singular values represent zero or represent the contribution of noise. This requires appropriate selection of the block dimension of the Hankel matrix (L must be sufficiently large).

In the SVD-based algorithm, the columns of matrices T_A and T_B have unity norm (such is not the case for Ho's algorithm). That is, the magnitude of each singular value is representative of the importance of the contribution (in the sense of the Frobenius norm) of the singular value to the numerical representation of the operator $H_{L,L}$. Thus, model order determination using the singular values has a firm numerical and algebraic foundation, and can be carried out once the SVD is computed.

APPENDIX C. COMPLEX-VALUED CANONICAL CORRELATIONS

Hotelling (1936) introduced the concept of canonical variables and canonical correlations to establish a canonical relationship between two sets of random variables (or between two random vectors). In linear algebra, the term "canonical" is used to denote the element of an equivalence class which is represented with the minimum number of non-zero independent parameters. For example, the Jordan form is a canonical form for the equivalence class of square matrices under a similarity transformation. As defined by Hotelling (1936), the canonical variables embody the essence of the correlation structure among the random variables of the two given sets.

The canonical variables formulation is presented herein for the special case where the dimension of the two random vectors (the number of variables in each set) is the same because that is the case in the context of the multichannel detection application. Extension to the general case where the two vectors have different dimensions is straightforward.

Consider two complex-valued, zero-mean, L-dimensional random vectors $\underline{z}$ and $\underline{v}$ with auto- and cross-correlation matrices defined as

(C-1) $$R_{zz} = E[\underline{z}\,\underline{z}^H]$$

(C-2) $$R_{vv} = E[\underline{v}\,\underline{v}^H]$$

(C-3) $$R_{zv} = E[\underline{z}\,\underline{v}^H]$$

(C-4) $$R_{vz} = E[\underline{v}\,\underline{z}^H]$$

The canonical variables for $\underline{z}$ and $\underline{v}$ are two complex-valued, zero-mean, L-dimensional random vectors

(C-5) $$\underline{\mu} = [\mu_1 \quad \mu_2 \ldots \mu_L]^T$$

(C-6) $$\underline{\beta} = [\beta_1 \quad \beta_2 \ldots \beta_L]^T$$

such that the following conditions are satisfied:

i) $\underline{\mu} = T_1\underline{z}$

ii) $\underline{\beta} = T_2\underline{v}$

iii) μ_1 and β_1 have unit variance and are maximally correlated, with correlation coefficient ρ_1.

iv) for $i \leq L$, μ_i and β_i have unit variance and are maximally correlated, with correlation coefficient ρ_i; furthermore, μ_i is uncorrelated with $\mu_{i-1}, \mu_{i-2}, \ldots, \mu_1$, and β_i is uncorrelated with $\beta_{i-1}, \beta_{i-2}, \ldots, \beta_1$.

v) $1 \geq \rho_1 \geq \rho_2 \geq \ldots \geq \rho_L \geq 0$

The two linear transformations T_1 and T_2 introduced in conditions (i) and (ii) are complex-valued, full-rank, LxL matrices. Condition (v) implies that the positive-valued correlation coefficients are selected (the sign of the rows of matrices T_1 and T_2 can be selected in all cases such that the correlation coefficient of two random variables μ_i and β_i is positive-valued). The correlation coefficients $\{\rho_i\}$ are the canonical correlations for $\underline{z}$ and $\underline{v}$. Since the canonical variables $\{\mu_i\}$ and $\{\beta_i\}$ are covariance-normalized, their correlation coefficients are less than or equal to unity, as indicated in condition (v).

Conditions (iii)-(v) can be expressed in compact form using Conditions (i) and (ii) and Equations (C-1)-(C-4),

$$E[\underline{\mu}\,\underline{\mu}^H] = I_J = T_1 R_{zz} T_1^H \tag{C-7}$$

$$E[\underline{\beta}\,\underline{\beta}^H] = I_J = T_2 R_{vv} T_2^H \tag{C-8}$$

$$E[\underline{\beta}\,\underline{\mu}^H] = R_{\beta\mu} = T_2 R_{vz} T_1^H \tag{C-9}$$

$$R_{\beta\mu} = \begin{bmatrix} \rho_1 & 0 & \cdots & 0 \\ 0 & \rho_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \rho_L \end{bmatrix} \tag{C-10a}$$

$$1 \geq \rho_1 \geq \rho_2 \geq \ldots \geq \rho_L \geq 0 \tag{C-10b}$$

Equations (C-7)-(C-10) constitute an analytic formulation of the canonical correlations problem. Golub (1969) has shown that the solution for this problem in the case of real-valued variables can be obtained using the singular value decomposition (SVD). The extension to the case of complex-valued variables is straightforward, as carried out herein.

The first step in the development is to determine the matrix square root of each of the correlation matrices R_{zz} and R_{vv}. A matrix square root for a correlation matrix can be calculated using any one of several methods, and all methods lead to equivalent results in the context of the problem at hand. Alternative methods include the Cholesky decomposition and the SVD. Of the alternative methods, the SVD method is preferred herein because of its robust numerical properties, and because the

inverse of the square root matrix is determined easily given the SVD of the matrix. Additionally, the SVD is used for another purpose in the realization algorithm. Thus, the matrix square roots of correlations matrices R_{zz} and R_{vv} are obtained using the SVD as

$$R_{zz} = U_z S_z U_z^H = \left(U_z S_z^{1/2} U_z^H\right)\left(U_z S_z^{1/2} U_z^H\right) = R_{zz}^{1/2} R_{zz}^{1/2} \tag{C-11}$$

$$R_{vv} = U_v S_v U_v^H = \left(U_v S_v^{1/2} U_v^H\right)\left(U_v S_v^{1/2} U_v^H\right) = R_{vv}^{1/2} R_{vv}^{1/2} \tag{C-12}$$

Now transform the random vectors $\underline{z}$ and $\underline{v}$ to define two correlation-normalized random vectors as

$$\underline{\theta} = R_{zz}^{-1/2} \underline{z} = U_z S_z^{-1/2} U_z^H \underline{z} \tag{C-13}$$

$$\underline{\gamma} = R_{vv}^{-1/2} \underline{v} = U_v S_v^{-1/2} U_v^H \underline{v} \tag{C-14}$$

Given these definitions, it is easy to show that

$$E[\underline{\theta}\,\underline{\theta}^H] = I_J \tag{C-15}$$

$$E[\underline{\gamma}\,\underline{\gamma}^H] = I_J \tag{C-16}$$

$$E[\underline{\gamma}\,\underline{\theta}^H] = R_{\gamma\theta} = R_{vv}^{-1/2} R_{vz} R_{zz}^{-1/2} \tag{C-17}$$

Equations (C-15)-(C-17) are similar to Equations (C-7)-(C-9), but it is incorrect to assume that Equations (C-13) and (C-14) define the desired canonical transformations, and that $\underline{\theta}$ and $\underline{\gamma}$ are the desired canonical variables. Variables $\underline{\theta}$ and $\underline{\gamma}$ are not the canonical variables because their correlation matrix, $R_{\gamma\theta}$, is not

in diagonal form. However, the variables $\underline{\theta}$ and $\underline{\gamma}$ constitute an important intermediate transformation.

Consider now the cross-correlation matrix $R_{\gamma\theta}$ of Equation (C-17), and carry out an SVD on it to obtain

$$R_{\gamma\theta} = U_R \Delta_R V_R^H \tag{C-18}$$

U_R and V_R are unitary LxL matrices, and Δ_R is an LxL diagonal matrix with non-negative elements along the diagonal. The diagonal elements of Δ_R are bounded by unity and zero, and are arranged in order of decreasing magnitude, with the largest at the (1,1) location:

$$\Delta_R = \begin{bmatrix} \delta_1 & 0 & \cdots & 0 & 0 \\ 0 & \delta_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \delta_{L-1} & 0 \\ 0 & 0 & \cdots & 0 & \delta_L \end{bmatrix} \tag{C-19a}$$

$$1 \geq \delta_1 \geq \delta_2 \geq \ldots \geq \delta_L \geq 0 \tag{C-19b}$$

Given the decomposition in Equations (C-18) and (C-19), and the given the relations in Equations (C-13)-(C-17), the desired canonical variables and canonical correlations are obtained as

$$\underline{\mu} = V_R^H \underline{\theta} = V_R^H R_{zz}^{-1/2} \underline{z} \tag{C-20}$$

$$\underline{\beta} = U_R^H \underline{\gamma} = U_R^H R_{vv}^{-1/2} \underline{v} \tag{C-21}$$

$$R_{\beta\mu} = \Delta_R \tag{C-22a}$$

(C-22b) $\rho_i = \delta_i \qquad i = 1, 2, \ldots, L$

Since matrices U_R and V_R are unitary, the norms of vectors $\underline{\mu}$ and $\underline{\beta}$ are equal to the norms of vectors $\underline{\theta}$ and $\underline{\gamma}$, respectively. From relations (C-20) and (C-21) the transformation matrices are determined in a straightforward manner to be

(C-23) $T_1 = V_R^H R_{zz}^{1/2}$

(C-24) $T_2 = U_R^H R_{vv}^{1/2}$

Direct substitution verifies that Equations (C-7)-(C-10) are satisfied by this choice of transformation matrices.

An important relation can be inferred from Equations (C-17) and (C-18),

(C-25) $R_{vz} = R_{vv}^{1/2} R_{\theta} R_{zz}^{1/2} = R_{vv}^{1/2} U_R \Delta_R V_R^H R_{zz}^{1/2}$

This relation is useful in the validation of Equation (C-9) using the transformation matrices in Equations (C-23) and (C-24). It is useful also in the system identification (stochastic realization) algorithm of Section 3.0.

REFERENCES

H. Akaike

(1974) "Stochastic Theory of Minimal Realization," IEEE Transactions on Automatic Control, Vol. AC-19, No. 6 (December), pp. 667-674.

(1975) "Markovian Representation of Stochastic Processes by Canonical Variables," SIAM Journal on Control, Vol. 13, No. 1 (January), pp. 162-173.

B. D. O. Anderson and J. B. Moore

(1979) Optimal Filtering, Prentice-Hall, Englewood Cliffs, NJ.

J. V. Candy

(1976) Realization of Invariant System Descriptions From Markov Sequences, Ph. D. Dissertation, Department of Electrical Engineering, University of Florida, Gainesville, FL.

U. B. Desai, D. Pal, and R. D. Kirkpatrick

(1985) "A realization approach to stochastic model reduction," International Journal on Control, Vol. 42, No. 4, pp. 821-838.

J. J. Dongarra, C. B. Moler, J. R. Bunch, and G. W. Stewart

(1979) LINPACK Users' Guide, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

P. L. Faurre

(1976) "Stochastic realization algorithms," in System Identification: Advances and Case Studies, R. K. Mehra and D. G. Laniotis (eds.), Academic Press, New York, NY.

F. R. Gantmacher
(1960) The Theory of Matrices, Vol. 1, Chelsea Publishing Co., New York, NY.

I. M. Gelfand and A. M. Yaglom
(1959) "Calculation of the amount of information about a random function contained in another such function," American Mathematical Society Translations (2), Vol. 12, pp. 199-246.

G. H. Golub
(1969) "Matrix Decompositions and Statistical Calculations," in Statistical Computation, R. C. Milton and J. A. Nelder (eds.), Academic Press, New York, NY.

H. Hotelling
(1936) "Relations between two sets of variables," Biometrika, Vol. 28, pp. 321-377.

R. E. Kalman, P. L. Falb, and M. A. Arbib
(1969) Topics in Mathematical System Theory, McGraw-Hill Book Co., New York, NY.

P. A. S. Metford and S. Haykin
(1985) "Experimental analysis of an innovations-based detection algorithm for surveillance radar," IEE Proceedings, Vol. 132, Pt. F, No. 1 (February), pp. 18-26.

J. H. Michels
(1991) Multichannel Detection Using the Discrete-Time Model-Based Innovations Approach, RL Technical Report No. RL-TR-91-269, Rome Laboratory, Griffiss AFB, NY.

(1992a) "Detection of Partially Correlated Signals in Clutter Using a Multichannel Model-Based Approach," presented at the 1992 National Telesystems Conference, May 19-20, The George Washington University, Ashburn, VA.

(1992b) Considerations of the Error Variances of Time-Averaged Estimators for Correlated Processes, RL Technical Report No. RL-TR-92-339, Rome Laboratory, Griffiss AFB, NY.

B. C. Moore

(1981) "Principal component analysis in linear systems: Controllability, observability, and model reduction," IEEE Transactions on Automatic Control, Vol. AC-26, No. 1 (February), pp. 17-31.

A. H. Nuttall

(1976) "Multivariate Linear Predictive Spectral Analysis Employing Weighted Forward and Backward Averaging: A Generalization of Burg's Algorithm," Naval Underwater Systems Center Tech. Report No. TR-5501, New London, CT.

W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling

(1989) Numerical Recipes - The Art of Scientific Computing (FORTRAN Version), Cambridge Univ. Press, New York, NY.

M. Rangaswamy, D. D. Weiner, and J. H. Michels

(1993) "Multichannel Detection for Correlated Non-Gaussian Random Processes Based on Innovations," presented at the SPIE International Symposium on Optical Engineering and Photonics in Aerospace and Remote Sensing (Conference 1955), April 12-16, Orlando, FL.

H. H. Rosenbrock
(1970) State-space and Multivariable Theory, J. Wiley & Sons, Inc., New York, NY.

O. N. Strand
(1977) "Multichannel Complex Maximum Entropy (Auto-Regressive) Spectral Analysis," IEEE Transactions on Automatic Control, Vol. AC-22, No. 4 (August), pp. 634-640.

R. A. Wiggins and E. A. Robinson
(1965) "Recursive Solution to the Multichannel Filtering Problem," Journal of Geophysical Research, Vol. 70, No.8 (April 15), pp. 435-441.

H. P. Zeiger and A. J. McEwen
(1974) "Approximate Linear Realizations of Given Dimension via Ho's Algorithm," IEEE Transactions on Automatic Control, Vol. AC-19, No. 2 (April), pg. 153.

MISSION

OF

ROME LABORATORY

Rome Laboratory plans and executes an interdisciplinary program in research, development, test, and technology transition in support of Air Force Command, Control, Communications and Intelligence (C^3I) activities for all Air Force platforms. It also executes selected acquisition programs in several areas of expertise. Technical and engineering support within areas of competence is provided to ESD Program Offices (POs) and other ESD elements to perform effective acquisition of C^3I systems. In addition, Rome Laboratory's technology supports other AFSC Product Divisions, the Air Force user community, and other DOD and non-DOD agencies. Rome Laboratory maintains technical competence and research programs in areas including, but not limited to, communications, command and control, battle management, intelligence information processing, computational sciences and software producibility, wide area surveillance/sensors, signal processing, solid state sciences, photonics, electromagnetic technology, superconductivity, and electronic reliability/maintainability and testability.