AD NUMBER

ADB276328

NEW LIMITATION CHANGE

TO

Approved for public release, distribution unlimited

FROM

Distribution authorized to U.S. Gov't. agencies only; Specific Authority; Nov 2001. Other requests shall be referred to AFRL/SNRT, Rome, NY 13441-4514.

AUTHORITY

AFRL/IFOIP ltr, 15 Jun 2004

AFRL-SN-RS-TR-2001-244
Final Technical Report
November 2001

TWO-DIMENSIONAL PROCESSING FOR RADAR SYSTEMS

Jaime R. Roman, Dennis W. Davis, and Qingwen Zhang

AIR FORCE RESEARCH LABORATORY
SENSORS DIRECTORATE
ROME RESEARCH SITE
ROME, NEW YORK

AFRL-SN-RS-TR-2001-244 has been reviewed and is approved for publication.

APPROVED: James H. Michels

JAMES H. MICHELS
Project Engineer

FOR THE DIRECTOR: Robert G. Polce

ROBERT G. POLCE, Chief
Rome Operations Office
Sensors Directorate

DESTRUCTION NOTICE - For classified documents, follow the procedures in DOD 5200.22M. Industrial Security Manual or DOD 5200.1-R, Information Security Program Regulation. For unclassified limited documents, destroy by any method that will prevent disclosure of contents or reconstruction of the document.

If your address has changed or if you wish to be removed from the Air Force Research Laboratory Rome Research Site mailing list, or if the addressee is no longer employed by your organization, please notify Air Force Research Laboratory/SNRT, 26 Electronic Pky, Rome, NY 13441-4514. 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

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	NOVEMBER 2001	Final Apr 97 - Apr 99

4. TITLE AND SUBTITLE
TWO-DIMENSIONAL PROCESSING FOR RADAR SYSTEMS

5. FUNDING NUMBERS
C - F30602-97-C-0085
PE - 65502F
PR - 3005
TA - 72
WU - 16

6. AUTHOR(S)
Jaime R. Roman, Dennis W. Davis, and Qingwen Zhang

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

8. PERFORMING ORGANIZATION REPORT NUMBER
SSC-TR-01-01

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory/SNRT
26 Electronic Pky
Rome New York 13441-4514

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFRL-SN-RS-TR-2001-244

11. SUPPLEMENTARY NOTES
Air Force Research Laboratory Project Engineer: James H. Michels/SNRT/(315) 330-4432

12a. DISTRIBUTION AVAILABILITY STATEMENT
DISTRIBUTION AUTHORIZED TO U.S. GOVERNMENT AGENCIES ONLY; SPECIFIC AUTHORITY: DOD FAR SUPPLEMENT 252.227-7018 NOV 01. OTHER REQUESTS FOR THIS DOCUMENT SHALL BE REFERRED TO AFRL/SNRT, ROME, NY 13441-4514.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

REPORT DEVELOPED UNDER SBIR

This report presents algorithms and methodolgies for space-time adaptive processing (STAP) and detection in airborne surveillance phased array radar systems. Two approaches are considered for the know-signal case: weight vectors for multichannel one-dimensional (MC1D) data representations, and parametric models for single-channel (scalar) two-dimensional (SC2D) data representations. For the MC1D approach, a generic architecture that covers a wide variety of weight vector algorithms and associated detection rules is presented. Algorithms covered include the classical matched filter (MF) and three new methods which are low-dimensionality alternatives to the MF. The new methods can be configured in a variety of ways based on the selection of a handful of key parameters. The SC2D parametric approach is a novel extension of the parametric adaptive matched filter (PAMF) pioneered by Rangaswamy and Michels (1997). In this context the radar channel output data is represented as a scalar 2-D system, and parametric 2-D models are utilized to represent the channel output data under the target-absent hypothesis. Robust linear model identification algorithms are applied to estimate the 2-D model parameters. All algorithms introduced herein admit adaptive (data-based) formulations, and offer reduced computational requirements in relation to the classical MF. Future work includes performance and computational load assessments in relation to other methods.

14. SUBJECT TERMS	15. NUMBER OF PAGES
SBIR Report, Phased Array Radar, Two-Dimensional Models, Space-Time Adaptive Processing, Model-Based Processing, Multichannel Time Series, Adaptive Arrays	126
	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	UL

Standard Form 298 (Rev. 2-89) (EG)
Prescribed by ANSI Std. 239.18
Designed using Perform Pro, WHS/DIOR, Oct 94

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1.0 INTRODUCTION

This report is a summary of the work carried out by the Scientific Studies Corporation (SSC) team in Phase II of the "Two-Dimensional Processing for Radar Systems" Small Business Innovation Research (SBIR) program for the U. S. Air Force Research Laboratory, Sensors Directorate, Rome Research Site (AFRL/SNRT). As such, the report includes also the work carried out by the University of Central Florida (UCF) under contract to SSC. Dr. Qingwen Zhang, post-doctoral researcher at UCF, and Prof. W. B. Mikhael, Chairman of the Electrical and Computer Engineering Department at UCF, executed the subcontract to SSC, and their efforts are hereby acknowledged.

The work reported herein was carried out in the context of space-time adaptive processing (STAP) for airborne surveillance radar systems in support of an in-house research effort at AFRL/SNRT. However, this work has application in other areas, such as communication systems, active sonar array systems, optical sensor systems, non-destructive inspection (NDI) systems, geophysical array systems, mine detection systems, and medical technology.

This report covers two related, but distinct, technical thrusts of the program. The first thrust involves a generic STAP architecture that covers a wide variety of algorithms and their associated detection rules. In particular, this architecture covers the classical matched filter (MF), as well as three new algorithms for the calculation of STAP weights. These novel algorithms are low-dimensionality options to the MF, and constitute a major technical contribution of the program. Furthermore, these algorithms admit straightforward adaptive configurations for the unknown-covariance case, just as the adaptive matched filter (AMF) is the unknown-covariance

configuration of the MF. All three new algorithms can be configured in a variety of ways, based on the selection of key parameters, and the generic architecture admits all such variations. Thus, the generic architecture provides new insights into the structure of STAP algorithms and detection rules. This architecture constitutes another technical contribution of the program.

The second technical thrust involves the model-based multichannel detection formulation in the context of a two-dimensional (2-D) representation for space-time processes in general, and airborne surveillance phased array radar systems in particular. For the phased array STAP and detection problem, such a formulation requires a 2-D parametric model for the channel output process under the target-absent hypothesis. The identified 2-D parametric model is used to design a whitening filter for the interference process, as required for the parametric adaptive matched filter (PAMF) methodology pioneered by Rangaswamy and Michels (1997). Formulation of a 2-D PAMF methodology and its algorithmic implementation constitute a third technical contribution of the program. The 2-D PAMF offers major advantages over the 2-D innovations-based detection architecture (IBDA) methodology formulated and demonstrated in Phase I of this two-phase program in the context of 2-D STAP (Román and Davis, 1997). The 2-D IBDA introduced in Phase I is a 2-D formulation of the IBDA methodology pioneered by Metford and Haykin (1985) for the scalar one-dimensional (1-D) case, and extended by Michels (1991) to the multichannel 1-D case. Advantages of the 2-D PAMF over the 2-D IBDA include a simpler structure, since a whitening filter for the target process is unnecessary. Another advantage is the availability of multiple detection rules that can be utilized in the context of an extended 2-D PAMF. Each alternative detection rule offers distinct features, such as constant false alarm rate (CFAR) detection.

1.1 Airborne Surveillance Phased Array Radar Problem Statement

A state-of-the-art phased array radar surveillance platform in a modern scenario has multiple functions, including target detection, target tracking and track association, and possibly target identification. The focus herein is on moving target detection, which is a precursor to the other functions. A typical airborne surveillance radar scenario involves a linear array radar consisting of J equally-spaced, identical antenna elements (or identical beamformed subarrays) in a side-looking configuration on an airborne platform moving at a constant speed in level flight. The array is aligned with the aircraft's longitudinal axis, and the aircraft velocity makes a crab angle γ with the aircraft's longitudinal axis. The radar array is radiating a coherent pulse train of N-pulse duration, at a constant radiation frequency, and at a constant pulse repetition frequency (PRF). Each antenna element (or group of elements) is referred to as a channel, and the ith channel output (after pulse compression, demodulation, and sampling) corresponding to a single range resolution cell (gate) is a complex-valued discrete-time sequence denoted as $\{x_i(n) \mid n = 0, 1, \ldots, N-1\}$. The J scalar sequences are concatenated to form a vector sequence $\{\underline{x}(n) \mid n = 0, 1, \ldots, N-1\}$. Process $\{\underline{x}(n)\}$ is assumed to be stationary, ergodic, zero-mean, and Gaussian-distributed.

The received signal in an airborne surveillance phased array radar can be referred to as a space-time process, wherein the spatial connotation arises from the spatial diversity of the antenna array elements. In general, this received signal contains a moving target component, as well as receiver (broadband) noise, jammer noise (broadband interference), and ground clutter (narrowband interference) components. The system's objective is to detect the moving target given the noise- and interference-corrupted received signal. This problem admits a dual-hypothesis

formulation, with $\mathbf{H}_0$ representing the null hypothesis (target not present), and $\mathbf{H}_1$ representing the alternative hypothesis (target present). In a typical scenario, the moving target detection processor makes a detection decision each coherent processing interval (CPI) for each range gate, which requires processing a finite-duration sequence $\{\underline{x}(n) \mid n = 0, 1, \ldots, N\text{-}1\}$ for each range gate.

To date, most of the development and analysis of optimum joint-domain adaptive algorithms and sub-optimum block-covariance algorithms for STAP involves the 1-D multichannel (vector) representation of the radar space-time signal. Such past work encompasses target detection and interference rejection in airborne surveillance radar arrays, as represented in the work of Brennan and Reed (1973), Jaffer et al. (1991), and Ward (1994). The algorithms discussed by these authors are referred to often as the conventional (or classical) STAP algorithms. More recently, Michels (1991) and Román and Davis (1993a; 1993b) have adopted the 1-D vector representation using multichannel auto-regressive (AR) and state-space models, respectively, for joint-domain innovations-based detection.

1.2 Maximum Correlation And Other STAP And Detection Methods In The Context Of A Generic Detection Architecture

A summary of each novel STAP algorithm is provided next. In the first new STAP algorithm the principle of orthogonal projections (OP) is applied to generate a sequence of residuals. This OP algorithm is distinct from the least-squares predictive-transform (LSPT) algorithm of Guerci and Feria (1996) in three important aspects. First, the LSPT includes a transform step, whereas the OP does not. Second, the LSPT is formulated as the optimal solution to the improvement factor criterion, which leads to a detection rule involving the minimum eigenvalue of the residual (prediction innovation) covariance matrix. This minimum

eigenvalue criterion is associated naturally with the transform step. In contrast, the OP utilizes the MF detection rule, which involves inner products of the residuals. Third, the LSPT is a block processing method in the predictive step as well as in the transform step. In the predictive step of the LSPT (the step in common with the OP method), the full-dimension (spatial and temporal) array output vector is segmented into two sub-vectors, and one sub-vector is utilized to predict the other. The dimensions of the two sub-vectors add up to the full dimension of the array output vector (from the dimensionality-reduction standpoint, a wise choice is to select each of the two sub-vectors to be of dimension equal to one-half the full-dimension, as considered in Section 4.2 for one option of the OP algorithm). In contrast, in its most general form the OP is a hybrid processing method, with block-processing as well as sequential-processing characteristics. This is a consequence of the fact that the dimensions of each of the two sub-vectors can be selected such that their sum is less than the full-dimension. Via this mechanism the OP method introduced herein attains a higher level of dimensionality reduction than the LSPT.

One particular configuration of the OP algorithm leads to an implementation identical with the parametric adaptive matched filter (PAMF) using the Yule-Walker (YW) and Levinson-Durbin (LD) AR model identification algorithms (Section 4.1).

The second novel algorithm is based on the concept of canonical variables and canonical correlations. This algorithm is referred to as maximum correlation (MC) since it is based on maximization of the correlation between the "past" and the "future" of the array output random process. The MC method is an extension of earlier work by SSC in the specific context of sidelobe canceling for adaptive arrays (Román, 1998a). Specifically, Román (1998a) demonstrated that the conventional

sidelobe canceler problem can be formulated in the context of canonical correlations, and the solution attained is equivalent to the well-known minimum mean-square-error solution. It is possible to show that the MC approach is equivalent also to the reduced-dimension generalized sidelobe canceler (GSC) formulated by Goldstein and Reed (1997). Since the multistage Wiener filter (MWF) (Goldstein et al., 1998) is a sequence of orthogonal decompositions of the same form as the GSC, it is reasonable to expect that an equivalence exists between the MC method presented herein (which is a generalization of the MC method for the sidelobe canceler) and the MWF. Further investigation of this issue remains for future work.

The MC algorithm is a hybrid processing method, with block-processing and sequential-processing characteristics. The method also allows for configurations with significant reduction in dimensionality, in relation to the MF and the LSPT. This is due to the fact that the MC algorithm has two mechanisms for attaining reductions in dimensionality. In addition, the MC formulation results in simple expressions for entropy and mutual information of the array output process (Section 5.1). These concepts facilitate the development of probabilistic criteria for dimensionality reduction (Section 5.2).

The third new algorithm consists of the application of the orthogonal projection principle to the array output data after transformation onto the canonical variables basis, and therefore is referred to as orthogonal projection using canonical variables (OPCV). This algorithm combines the best features of the OP and MC methods, and is more powerful than either. In particular, the OPCV includes an optimal mechanism for dimensionality reduction of the data residual, whereas the OP lacks such a mechanism. Furthermore, the OPCV data residual is less than or equal to the MC data residual for all possible configurations. Beyond these

features, the OPCV provides unique insight into the structure of orthogonal projection in the context of STAP.

All three algorithms (OP; MC; OPCV) admit configurations that are fully optimal with respect to the associated criterion, but their most important practical aspect is to provide alternatives to the optimum MF in their reduced-dimensionality configurations.

The algorithmic discussion presented herein is from the analytical point of view, as opposed to the computational point of view. That is, a formulation is presented for each algorithm, without addressing the issue of numerical implementation. Also, all algorithmic formulations are presented for the known-covariance case in order to emphasize the theoretical aspects, as well as to simplify the notation. However, each algorithm can be implemented adaptively, and adaptive issues are considered briefly for each algorithm. An important issue in the context of adaptive systems is that dimensionality reduction leads to a reduction in secondary data requirements, and/or an increase in detection performance, in relation to the AMF.

1.3 STAP and Detection Via the Two-Dimensional Representation

As stated previously, prior work in model-based STAP has been focused on the 1-D vector representation of the channel output process. Alternatively, a space-time signal can be represented as a 2-D scalar process. The 2-D representation is essential for visualizing and understanding the spectral energy and correlation characteristics of the total signal, and of the ground clutter component in particular. This 2-D representation also offers algorithmic and modeling advantages. With respect to model-based detection for the airborne surveillance phased array radar STAP problem, 1-D models offer dynamic and static degrees of freedom in the temporal axis, but only static degrees of freedom in the

spatial axis. In contrast, 2-D models offer dynamic and static degrees of freedom in both axes (space and time). Thus, a model-based detection methodology using 2-D scalar models is inherently better-suited to the cases wherein channel-to-channel correlation exhibits a complicated structure. Such cases occur physically when the platform and scenario parameters lead to a non-integer-valued clutter ridge slope parameter, and/or to non-zero misalignment between the array longitudinal axis and the platform velocity vector, and/or to non-zero internal clutter motion. These conditions are common in airborne surveillance radar scenarios.

The 2-D, least-square, frequency domain (2D-LS-FD) technique formulated by Mikhael and Yu (1994) was adopted in Phase I as the baseline 2-D model identification method. This technique approximates a given 2-D complex-valued field with a 2-D rational function model of the auto-regressive moving-average (ARMA) class. Of course, the ARMA class includes the AR and the moving-average (MA) classes. In the 2D-LS-FD technique the ARMA coefficients are obtained as the solution to a set of linear equations. Excellent results have been obtained in the image noise canceling problem (Mikhael and Yu, 1994), which is related to clutter cancellation in radar space-time processing. The 2D-LS-FD technique was applied successfully in Phase I to clutter modeling. However, detection results obtained in Phase II using the 2D-LS-FD technique in the context of a 2-D PAMF were very poor in relation to those obtained with other methods, and were significantly below those obtained using the AMF implemented via sample matrix inversion. This poor performance is a result of the high noise present in the frequency-domain representation (magnitude as well as phase) of the channel output process, as required by the algorithm (in the 2D-LS-FD, model identification is carried out in the frequency domain). As a consequence, other 2-D model identification techniques were considered. UCF and SSC settled on

the 2-D AR least-squares (2DARLS) algorithm as the preferred technique to apply in the context of the 2-D PAMF. The 2DARLS is conceptually simple and provides robust model identification performance. For example, the closely-related 1-D multichannel AR LS algorithm has out-performed a suite of algorithms in recent studies (Román et al., 2000). In addition, the 2DARLS admits various numerically stable and efficient software implementations (Appendix A). The 2-D PAMF formulation and its architecture using the 2DARLS identification algorithm is another major contribution of this program.

1.4 Report Overview

The generic PAMF architecture is introduced first in Section 2.0, along with notation and definitions that are used in other sections. Then the MF, OP, MC, and OPCV STAP algorithms are summarized in Sections 3.0 through 6.0, respectively. Model-based multichannel detection in the context of the 2-D representation of the channel output vector process is discussed in Section 7.0, including the 2-D PAMF formulation and the 2DARLS identification algorithm. A summary and suggestions for further work are presented in Section 8.0. Appendix A summarizes the LS identification algorithm for multichannel AR processes, and presents computational options for its software implementation in the context of the PAMF for STAP in airborne surveillance phased array radar systems. These procedures and software options apply to the 2DARLS with straightforward modifications.

2.0 GENERIC BLOCK STAP AND DETECTION ARCHITECTURE

A standard airborne surveillance radar system scenario and side-looking linear array configuration with typical STAP parameters is assumed herein. As stated earlier, the number of channels is J, the number of pulses in a CPI is N, and $\{\underline{x}(n) \mid n = 0, 1, \ldots, N-1\}$ is the received vector sequence for the range bin to be processed to attain a detection decision. This data set is referred to as the primary data. Also of relevance is the secondary data set, which is a collection of channel output sequences for K range bins in the neighborhood of the primary data. In the context of secondary data sets, a "neighborhood" is defined in many ways, and any standard approach suffices for our purposes provided that the secondary data set is selected to be representative of the null hypothesis condition (target not present) and statistically-independent of the primary data. This secondary data set is denoted herein as $\{\underline{x}_k(n|\mathbf{H}_0) \mid n = 0, 1, \ldots, N-1; k = 1, 2, \ldots, K\}$. For notational simplicity, it is assumed herein that N is an even number. This is hardly a restriction in the practical sense since in most radar systems N is selected to be a power of two.

The generic STAP architecture proposed herein is presented in Figure 2-1, and the delay stack and advance stack blocks that appear in this figure are defined in Figures 2-2 and 2-3, respectively. The steering vector sequence $\{\underline{e}(n) \mid n = 0, 1, \ldots, N-1\}$ is formed by un-stacking the so-called JN-element steering vector $\underline{e}$ into N J-element vectors (this un-stacking is the inverse of an Nth order advance stack operator). That is,

$$\underline{e} = \begin{bmatrix} \underline{e}(0) \\ \underline{e}(1) \\ \vdots \\ \underline{e}(N-1) \end{bmatrix} \tag{2-1a}$$

$$\underline{e}(n) = \left(e^{j2\pi n f_d}\right) \begin{bmatrix} 1 \\ e^{j2\pi f_s} \\ \vdots \\ e^{j2\pi(J-1)f_s} \end{bmatrix} \qquad n = 0, 1, \ldots, N\text{-}1 \tag{2-1b}$$

where f_d and f_s denote normalized Doppler and spatial frequencies, respectively.

With respect to Figure 2-1, stacked (or block) vectors $\underline{x}_{\mathcal{P}:P}(n)$, $\underline{e}_{\mathcal{P}:P}(n)$, and $\underline{x}_{\mathcal{F}:Q}(n)$, and $\underline{e}_{\mathcal{F}:Q}(n)$ are defined as indicated in Figures 2-2 and 2-3, respectively. That is, $\underline{x}_{\mathcal{P}:P}(n)$ and $\underline{e}_{\mathcal{P}:P}(n)$ are JP-element vectors, and $\underline{x}_{\mathcal{F}:Q}(n)$ and $\underline{e}_{\mathcal{F}:Q}(n)$ are JQ-element vectors. Subscript $\mathcal{P}$ denotes that vector $\underline{x}_{\mathcal{P}:P}(n)$ represents the <u>past</u> of the process $\{\underline{x}(n)\}$, with respect to time instant n, and subscript $\mathcal{F}$ denotes that vector $\underline{x}_{\mathcal{F}:Q}(n)$ represents the <u>future</u> of the process $\{\underline{x}(n)\}$, with respect to time instant n. Data item $\underline{x}(n)$ can be included either in the past or in the future (as selected herein), without loss of generality. Of course, these considerations apply equally for the steering sequence, $\{\underline{e}(n)\}$. For simplicity, the discrete-time argument, n, is dropped from the past and future block vectors in instances where the intended meaning is clear.

Integers P and Q have specific meaning in each weight calculation algorithm, as discussed in the sections that follow. In order to simplify notation, and without loss of generality, it is assumed herein that

$$P \geq Q \tag{2-2}$$

Also, the following constraint must be satisfied by the integers P and Q:

$$P + Q \leq N \tag{2-3}$$

This constraint states that the number of blocks in the data block vectors is, at most, equal to the number of available data sequence elements. Equality is required, for example, in the LSPT algorithm (Guerci and Feria, 1996); however, in the OP and MC algorithms high levels of dimensionality reduction are attained with $P + Q << N$.

The data-based block vectors have block covariances that are defined in terms of the elements of the matrix auto-covariance sequence (ACS). Let $R_{xx}(m)$ denote the mth lag of the data ACS under the null hypothesis, defined as

$$R_{xx}(m) = E\left[\underline{x}(n|\mathbf{H}_0)\underline{x}^H(n-m|\mathbf{H}_0)\right] \tag{2-4}$$

Unless stated otherwise, hereinafter all covariance matrices are assumed to be defined for the null hypothesis condition, and all explicit notational indications of the null hypothesis are omitted. Then, the block covariance matrices are defined as

$$\mathcal{R}_{\mathcal{P}:P,P} = E\left[\underline{x}_{\mathcal{P}:P}\,\underline{x}^H_{\mathcal{P}:P}\right] = \begin{bmatrix} R_{xx}(0) & R_{xx}(1) & \cdots & R_{xx}(P-1) \\ R^H_{xx}(1) & R_{xx}(0) & \cdots & R_{xx}(P-2) \\ \vdots & \vdots & \ddots & \vdots \\ R^H_{xx}(P-1) & R^H_{xx}(P-2) & \cdots & R_{xx}(0) \end{bmatrix} \tag{2-5}$$

$$\mathcal{R}_{\mathcal{F}:Q,Q} = E\left[\underline{x}_{\mathcal{F}:Q}\,\underline{x}^H_{\mathcal{F}:Q}\right] = \begin{bmatrix} R_{xx}(0) & R^H_{xx}(1) & \cdots & R^H_{xx}(Q-1) \\ R_{xx}(1) & R_{xx}(0) & \cdots & R^H_{xx}(Q-2) \\ \vdots & \vdots & \ddots & \vdots \\ R_{xx}(Q-1) & R_{xx}(Q-2) & \cdots & R_{xx}(0) \end{bmatrix} \tag{2-6}$$

$$(2\text{-}7) \qquad \mathcal{R}_{\mathcal{F}:Q,\mathcal{P}:P} = E\left[\underline{x}_{\mathcal{F}:Q}\,\underline{x}^{H}_{\mathcal{P}:P}\right] = \begin{bmatrix} R_{xx}(1) & R_{xx}(2) & \cdots & R_{xx}(P) \\ R_{xx}(2) & R_{xx}(3) & \cdots & R_{xx}(P+1) \\ \vdots & \vdots & \ddots & \vdots \\ R_{xx}(Q) & R_{xx}(Q+1) & \cdots & R_{xx}(P+Q-1) \end{bmatrix}$$

In these definitions, the Zapf Chancery font size 14 is used for the data block covariances, and the subscripts denote whether the matrix is a past block covariance ($\mathcal{P}$), or a future block covariance ($\mathcal{F}$), or a future-to-past block cross-covariance ($\mathcal{F};\mathcal{P}$). The subscripts (P and Q) also denote the number of block rows (subscript preceding the comma) and the number of block columns (subscript following the comma). Block covariances $\mathcal{R}_{\mathcal{P}:P,P}$ and $\mathcal{R}_{\mathcal{F}:Q,Q}$ are square, Hermitian, as well as block Hermitian and block Toeplitz (a block Toeplitz matrix is a matrix in which the (i,j)th block element is a function of i-j). Block covariance $\mathcal{R}_{\mathcal{F}:Q,\mathcal{P}:P}$ is rectangular (unless $P = Q$) as well as block Hankel (a block Hankel matrix is a matrix in which the (i,j)th block element is a function of i+j). $\mathcal{R}_{\mathcal{F}:Q,\mathcal{P}:P}$ is also block symmetric when $P = Q$.

In the most general version, weight matrices V and W in Figure 2-1 are dimensioned JQxL and JPxL, respectively, and both the data residual vector $\underline{\varepsilon}(n)$ and the steering residual vector $\underline{u}(n)$ are L-element column vectors, with $L \leq JQ$. Weight matrices V and W are generated by each of the STAP algorithms considered herein as the solution to distinct optimality criteria. From Figure 2-1, the data and steering residual vector sequences are given as

$$(2\text{-}8) \qquad \underline{\varepsilon}(n) = W^{H}\underline{x}_{\mathcal{F}:Q}(n) - V^{H}\underline{x}_{\mathcal{P}:P}(n) = \underline{\alpha}(n) - \underline{\beta}(n) \qquad n = P, P+1, \ldots, N-Q$$

$$(2\text{-}9) \qquad \underline{u}(n) = W^{H}\underline{e}_{\mathcal{F}:Q}(n) - V^{H}\underline{e}_{\mathcal{P}:P}(n) \qquad n = P, P+1, \ldots, N-Q$$

respectively. The first available residual is at time instant P because $P \geq Q$, and the last available residual is at time instant N-Q because the future block vector, $\underline{x}_{\mathcal{F}:Q}(n)$, has Q block elements. Vector sequences $\{\underline{\alpha}(n)\}$ and $\{\underline{\beta}(n)\}$ are intermediate variables that have meaning only for some algorithms.

Weight matrix C in Figure 2-1 is dimensioned LxL, and it is required in some algorithms in order to normalize and de-correlate the residual vector $\underline{\varepsilon}(n)$ so that its covariance matrix is equal to an identity matrix. Then, the normalized (unit variance) and uncorrelated (both spatially and temporally) data residual vector sequence and the fully-processed steering residual vector sequence are given as

(2-10) $$\underline{v}(n) = C^H \underline{\varepsilon}(n) \qquad n = P, P+1, \ldots, N-Q$$

(2-11) $$\underline{s}(n) = C^H \underline{u}(n) \qquad n = P, P+1, \ldots, N-Q$$

respectively. Weight matrix C is determined such that the following condition is satisfied:

(2-12) $$C^H R_{\varepsilon\varepsilon}(0)\, C = C^H \Omega\, C = I_L$$

where $R_{\varepsilon\varepsilon}(0)$ and Ω denote the covariance matrix of the residual $\underline{\varepsilon}(n)$,

(2-13a) $$\Omega = R_{\varepsilon\varepsilon}(0) = E[\,\underline{\varepsilon}(n)\,\underline{\varepsilon}^H(n)\,]$$

(2-13b) $$\Omega = W^H \mathcal{R}_{\mathcal{F}:Q,Q} W - V^H \mathcal{R}_{\mathcal{P}:P,\mathcal{F}:Q} W - W^H \mathcal{R}_{\mathcal{F}:Q,\mathcal{P}:P} V + V^H \mathcal{R}_{\mathcal{P}:P,P} V$$

The closed form expression in Relation (2-13b) is very useful.

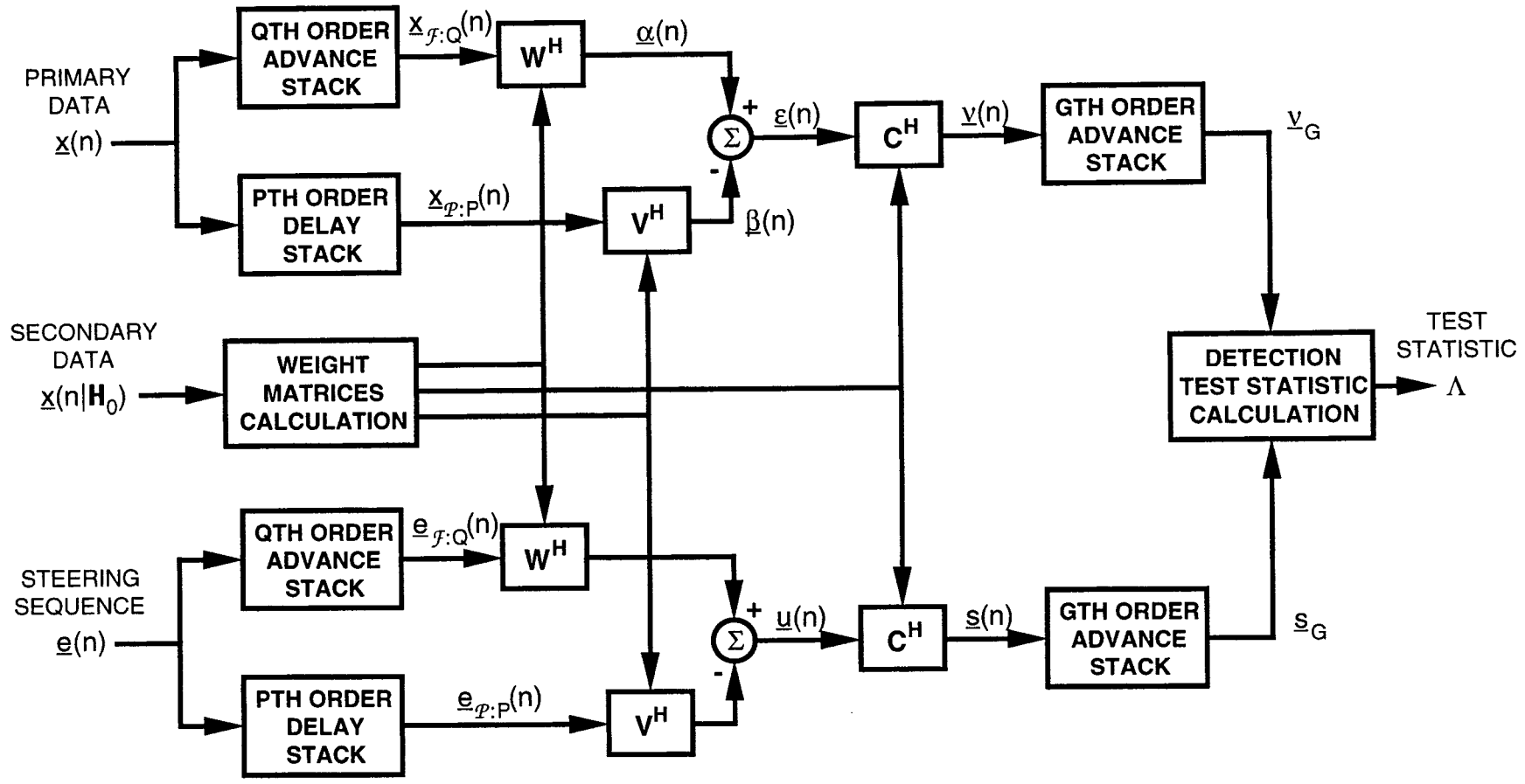

Figure 2-1. Generic STAP and detection architecture.

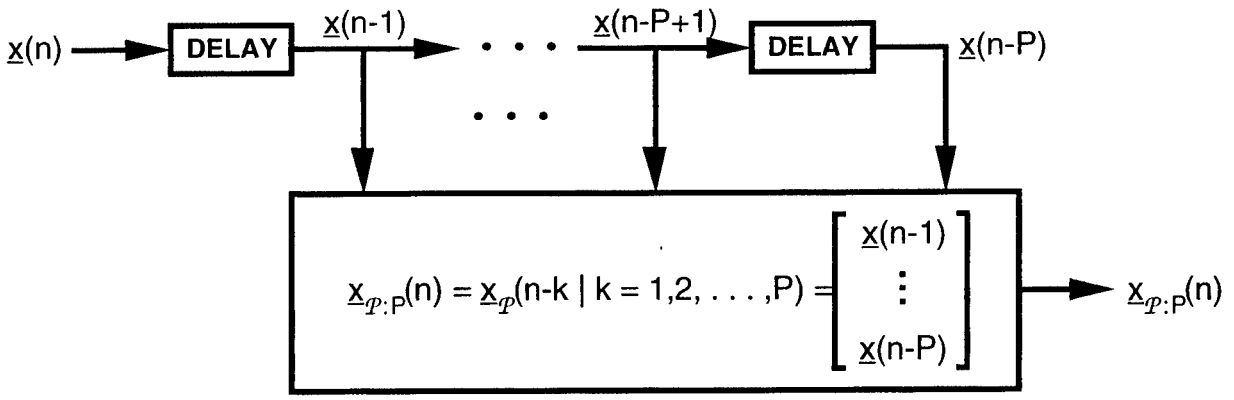

Figure 2-2. Pth order delay stack definition.

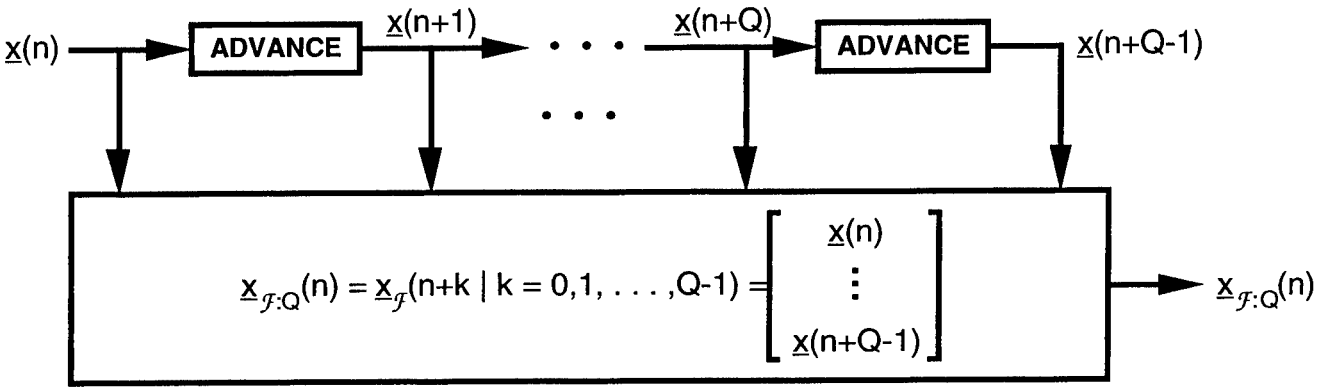

Figure 2-3. Qth order advance stack definition.

A weight matrix $\mathbf{C}$ that satisfies Relation (2-12) can be determined from the factors in a decomposition of the covariance Ω. Notice that Relation (2-12) implies Ω is non-singular. This is a reasonable requirement because a singular Ω is indicative of an ill-posed problem. Now consider any factorization of Ω into the product of a matrix $\mathbf{B}$ and its Hermitian transpose. That is,

$$\Omega = \mathbf{B}\mathbf{B}^H \tag{2-14}$$

Then, it is straightforward to show that Relation (2-12) is satisfied with a weight matrix $\mathbf{C}$ determined as

$$\mathbf{C}^H = \mathbf{B}^{-1} \tag{2-15}$$

The singular value decomposition (SVD), the LDL decomposition, the Cholesky factorization, or any other such factorization can be used to generate the matrix factor $\mathbf{B}$. Each matrix factor type has unique analytical and numerical properties.

Each of the two sequences $\{\underline{v}(n)\}$ and $\{\underline{s}(n)\}$ is stacked in block vector form according to the convention in Figure 2-3 to generate GL-element block column vectors $\underline{v}_G$ and $\underline{s}_G$, respectively. Integer G denotes the number of elements in the residual vector sequence; that is,

$$G = N - P - Q + 1 \tag{2-16}$$

where the nominal N-element input data sequence is assumed.

Parameters P and Q are key to the definition of each algorithm, and can be viewed as analogous to model order in parametric models (Michels et al., 1998; Román et al., 1997, 1998). That is, the specific values selected for these parameters establish the particular structure of each algorithm

configuration, and thus impact performance directly. In the OP and MC STAP algorithms considered herein, parameter L (which denotes the number of elements in the residual vectors) also plays a similar role, whereas for the MF algorithm it is specified as JQ. For the OP and MC algorithms, the selection of parameter L directly impacts detection performance as well as computational requirements (via reductions in dimensionality).

The detection test statistic block in Figure 2-1 is defined (for the Gaussian-distributed data case considered herein) as

$$\Lambda = \frac{\left|\underline{s}_G^H \underline{v}_G\right|^2}{\underline{s}_G^H \underline{s}_G} = \frac{\left|\sum_{n=P}^{N-Q} \underline{s}^H(n)\,\underline{v}(n)\right|^2}{\sum_{n=P}^{N-Q} \underline{s}^H(n)\,\underline{s}(n)} \tag{2-17}$$

This test statistic is of the same form as the one proposed by Rangaswamy and Michels (1997) for the PAMF. However, the PAMF test statistic differs in two important aspects: first, the PAMF residuals are J-element vectors, and second, the PAMF residual sequence is of duration $N - P_{MI}$ (where P_{MI} is a function of the model identification algorithm used in conjunction with the PAMF). Test statistic Λ is compared with a threshold $\mathcal{T}_D$ in order to select between the null and alternative hypotheses. The threshold is calculated such that a pre-determined false alarm rate is met, following the Neyman-Pearson (NP) detection criterion (the NP criterion is to maximize the probability of detection at a fixed probability of false alarm).

2.1 Constant False Alarm Rate (CFAR) Issues

A pre-determined threshold can be used for a detector with the constant false alarm rate (CFAR) property, wherein detection

performance is independent of the covariance matrix of the total disturbance (clutter, jamming, and receiver noise). In theory, only a few select detectors have the CFAR property. One such detector is the MF, with a detection rule similar in form to the middle expression in Equation (2-17). This similarity suggests that some of the detectors discussed herein that use Equation (2-17) may have the CFAR property, or exhibit CFAR-like performance over a range of conditions. This point of view is supported by the argument developed next.

Consider the term $\left|\underline{s}_G^H \underline{v}_G\right|^2$ (the numerator of Equation (2-17)) under the null hypothesis, and assume that the data residual vector sequence $\{\underline{\varepsilon}(n)\}$ is temporally uncorrelated. Then, given Equations (2-10) through (2-16), the expected value of the numerator of the test statistic Λ is

(2-18a) $$E\left[\left|\underline{s}_G^H \underline{v}_G\right|^2\right] = E\left[\underline{s}_G^H \underline{v}_G \underline{v}_G^H \underline{s}_G\right] = \underline{s}_G^H E\left[\underline{v}_G \underline{v}_G^H\right] \underline{s}_G$$

(2-18b) $$E\left[\left|\underline{s}_G^H \underline{v}_G\right|^2\right] = \underline{s}_G^H \begin{bmatrix} \hat{C}^H \Omega \hat{C} & [0] & \cdots & [0] \\ [0] & \hat{C}^H \Omega \hat{C} & \cdots & [0] \\ \vdots & \vdots & \ddots & \vdots \\ [0] & [0] & \cdots & \hat{C}^H \Omega \hat{C} \end{bmatrix} \underline{s}_G$$

where $\hat{C}$ denotes an appropriate estimate of the weight matrix C. The non-diagonal block elements in the right-hand-side of Equation (2-18b) are zero-valued LxL matrices because the data residual sequence $\{\underline{\varepsilon}(n)\}$ is assumed to be uncorrelated in time. If $\hat{C}$ is a sufficiently-accurate estimate of C, then Equation (2-12) implies that each block diagonal element of the matrix on the right-hand-side of Equation (2-18b) approximates an LxL identity matrix. That is,

$$\hat{I}_L = \hat{C}^H \Omega \hat{C} \tag{2-19}$$

where $\hat{I}_L$ denotes an estimate of the LxL identity matrix. Then the block covariance of the data residual block vector $\underline{v}_G$ is an estimate of the GLxGL identity matrix, and the expectation of the numerator of the detection test statistic becomes

$$E\left[\left|\underline{s}_G^H \underline{v}_G\right|^2\right] = \underline{s}_G^H \hat{I}_{GL} \underline{s}_G \approx \underline{s}_G^H \underline{s}_G \tag{2-20}$$

where $\hat{I}_{GL}$ is an estimate of the GLxGL identity matrix. The rightmost term in Equation (2-20) is the same as the denominator of the detection test statistic, Λ. Thus, the expected value of the detection test statistic is unity, irrespective of the structure of the covariance matrix of the total disturbance. Such is the requisite condition for CFAR performance.

In adaptive implementations of the STAP methods (wherein all covariance matrices are substituted by their estimates), the key assumptions in the above-outlined argument are that: a) the data residual vector sequence $\{\underline{\varepsilon}(n)\}$ is uncorrelated, and b) $\hat{C}$ is a sufficiently-accurate estimate of C. In addition, CFAR requires that the temporal whitening of the residual and the estimate of weight matrix C be attained with estimator structures that are independent of the disturbance covariance matrix. The OP and MC STAP algorithms discussed herein have the potential for generating an uncorrelated data residual vector as well as an accurate estimate of the weight matrix (see Remark 2.2.4 in Section 2.2). However, the requirement of estimator independence from the disturbance covariance matrix is difficult to meet. In addition, the theoretical assessment of that issue is difficult to attain also. Nevertheless, the OP and MC STAP methods and associated

detector rules have the potential for CFAR-like performance at least over a range of disturbance conditions.

2.2 Remarks

The remarks below address various relevant issues, including the preferred approach for generating the data residual covariance matrix, and variations of the generic STAP architecture.

Remark 2.2.1. A variation to the structure defined in Equations (2-8) and (2-9) consists of having the input data sequence of duration $N+P+Q$, wherein P additional sequence elements can be viewed as being "negative time" sequence elements, and Q sequence elements are added as "positive time" sequence elements. Specifically, the data sequence is of the form $\{\underline{x}(n) \mid -P, \ldots, -1, 0, 1, \ldots, N-1, N, N+1, \ldots N+Q-1\}$, and similarly for the steering sequence. In such a structure the residual sequences are both of duration N, with initial time 0. The key distinction in this variation is that it utilizes more input data. That difference involves practical implications that must be taken into account, specially in comparative performance evaluations. The case where the input data sequence is of duration N is the baseline in this work.

Remark 2.2.2. The architecture in Figure 2-1 can be generalized further by allowing the stacking of the primary data into two column vectors of any possible combination of number of elements, rather than as discrete multiples of J. For example, $\underline{x}_{\mathcal{F}:Q}$ can be selected to be a scalar, irrespective of the value of J, and likewise, $\underline{x}_{\mathcal{P}:P}$ can be selected to be a column vector such that its number of elements is in the range $\{1, 2, \ldots, J(N-1)\}$. This generalization manifests itself in some of the algorithms as allowing the number of elements in the residual vectors to be selected among the allowable range of discrete values, rather than as a multiple of J. Such generalization may offer advantages in

some applications, but the inherent structure of the space-time data appears to favor the partition modulo J. Nevertheless, this option has attractive features and should be investigated further.

The selection of $\underline{x}_{\mathcal{J}:Q}$ as a scalar, and of $\underline{x}_{\mathcal{P}:P}$ as a J(N-1)-element column vector is a generalization of the reduced-dimension GSC of Goldstein and Reed (1997). One distinction is that the GSC structure includes a blocking matrix, whereas the architecture in Figure 2-1 does not.

Remark 2.2.3. Another generalization to the architecture presented herein is to utilize, in an *ad hoc* manner, other expressions for the detection rule. Two such candidates are the adaptive coherence (AC) proposed independently by Scharf and McWhorter (1996) and by Conte et al. (1995; 1996), and the generalized likelihood ratio test (GLRT) proposed by Kelly (1986).

Remark 2.2.4. In practical situations wherein the true covariance Ω is unknown, an estimate must be utilized to generate weight matrix C. The OP and MC STAP algorithms include formulas to estimate Ω, and such an estimate is referred to herein as the model residual covariance. Alternatively, a maximum likelihood (ML) estimate can be generated as an ensemble average over the residual sequences of the secondary data set. The ML estimate is referred to herein as the sample residual covariance. Simulation-based analyses reported elsewhere (Román, 1998b) for the PAMF in the context of airborne surveillance phased array radar systems, indicate that using the sample residual covariance (instead of the model residual covariance) to design the weight matrix C leads to reduced threshold variability and higher probability of detection. It is reasonable to expect that such performance extends to the OP and MC methods; however, appropriate simulation-based analyses should be carried out.

3.0 MATCHED FILTER (MF)

The MF detection algorithm, proposed independently by Cai and Wang (1990), Chen and Reed (1991), and Robey et al. (1992), is obtained via maximization of the generalized likelihood assuming the total disturbance covariance is known. Its adaptive variation, the AMF, is obtained by using the ML estimate in place of the unknown covariance. Both methods, MF and AMF, have the CFAR property for Gaussian-distributed disturbance.

The MF algorithm fits into the architecture in Figure 2-1 as follows. For the MF, the key integers are Q and P, which assume the following values:

(3-1) $Q = N$

(3-2) $P = 0$

In this algorithm integer L is determined directly by J and Q as

(3-3) $L = JQ = JN$

Also, given P and Q, it follows from Equations (2-16) and (3-1)-(3-2), and from Figure 2-1, that

(3-4) $G = 1$

(3-5) $V = [0]$

Weight matrix W is dimensioned JNxJN, and is obtained as

(3-6) $W = \mathcal{R}_{\mathcal{F}:N,N}^{-1/2}$

where $\mathcal{R}_{\mathcal{F}:N,N}^{-1/2}$ denotes the matrix inverse square-root of the block covariance of the total disturbance. For a positive-definite covariance, the matrix square-root is unique, and can be generated via the SVD. However, most practical implementations would avoid calculation of the matrix inverse square-root.

Consider now the data and steering sequence residuals. For this algorithm each of these two is a single JN-element vector (not a sequence); specifically,

(3-7) $$\underline{\varepsilon} = \mathcal{R}_{\mathcal{F}:N,N}^{-1/2}\,\underline{x}_N$$

(3-8) $$\underline{u} = \mathcal{R}_{\mathcal{F}:N,N}^{-1/2}\,\underline{e}_N$$

And the data residual covariance under the null hypothesis is given as

(3-9) $$\Omega = I_{JN}$$

since W is the whitening filter for the JN-element data vector $\underline{x}_N$. It follows from Equation (2-12) that weight matrix C is also the JNxJN identity,

(3-10) $$C = I_{JN}$$

Finally, the test statistic for this algorithm attains the form

(3-11) $$\Lambda = \frac{\left|\underline{u}^H \underline{\varepsilon}\right|^2}{\underline{u}^H \underline{u}} = \frac{\left|\underline{e}_N^H \mathcal{R}_{\mathcal{F}:N,N}^{-1}\,\underline{x}_N\right|^2}{\underline{e}_N^H \mathcal{R}_{\mathcal{F}:N,N}^{-1}\,\underline{e}_N}$$

From Equations (3-7) through (3-11), it is clear that the "matched filter" nomenclature is due to the inner product in the numerator

of Equation (3-11). The denominator is a normalization factor which gives the CFAR property for Gaussian-distributed disturbance.

In the AMF, block covariance $\mathcal{R}_{\mathcal{F}:N,N}$ is replaced by its ML estimate, denoted as $\hat{\mathcal{R}}_{\mathcal{F}:N,N}$. Block covariance $\hat{\mathcal{R}}_{\mathcal{F}:N,N}$ is referred to as the sample covariance matrix, and in most typical cases it is generated as an ensemble average over the secondary data (recall that the secondary data is assumed to be representative of the null hypothesis). The data residual covariance under the null hypothesis is then of the form

$$\Omega_{AMF} = \hat{\mathcal{R}}_{\mathcal{F}:N,N}^{-1/2} \mathcal{R}_{\mathcal{F}:N,N} \hat{\mathcal{R}}_{\mathcal{F}:N,N}^{-1/2} \tag{3-10}$$

Notice that Ω_{AMF} approaches the identity matrix as the sample covariance matrix approaches the true covariance.

4.0 ORTHOGONAL PROJECTION (OP)

In this algorithm the whitening of the data residual sequence is sought by determining the orthogonal projection of the space spanned by $\underline{x}_{\mathcal{F}:Q}$, denoted as $\mathcal{X}^+$, onto the space spanned by $\underline{x}_{\mathcal{P}:P}$, denoted as $\mathcal{X}^-$. Since the data is zero-mean and Gaussian-distributed, an orthogonal projection in these vector spaces is equivalent to conditional expectation. Let $\hat{\underline{x}}_{\mathcal{F}:Q}(n|\mathcal{X}^-)$ denote the expectation of $\underline{x}_{\mathcal{F}:Q}(n)$ conditioned on $\underline{x}_{\mathcal{P}:P}(n)$. For appropriately-selected integers P and Q, this conditional expectation is of the form

$$\hat{\underline{x}}_{\mathcal{F}:Q}(n|\mathcal{X}^-) = E\left[\underline{x}_{\mathcal{F}:Q}\,\underline{x}_{\mathcal{P}:P}^H\right]\left(E\left[\underline{x}_{\mathcal{P}:P}\,\underline{x}_{\mathcal{P}:P}^H\right]\right)^{-1}\underline{x}_{\mathcal{P}:P}(n) \tag{4-1a}$$

$$\hat{\underline{x}}_{\mathcal{F}:Q}(n|\mathcal{X}^-) = \mathcal{R}_{\mathcal{F}:Q,\mathcal{P}:P}\,\mathcal{R}_{\mathcal{P}:P,P}^{-1}\,\underline{x}_{\mathcal{P}:P}(n) \tag{4-1b}$$

where Definitions (2-5) and (2-7) have been invoked. Given the orthogonal projection (4-1), an uncorrelated residual sequence can be generated based on the architecture in Figure 2-1.

For this algorithm the key parameters are P and Q also. These parameters can assume a range of integer values, and each pair of values results in an algorithm with unique performance characteristics. Thus, it is more appropriate to view this algorithm as a class, rather than a single algorithm. Class members are defined by the specific integer values adopted for the key parameters in the formulation, and those values must satisfy the following conditions:

$$1 \le Q \le P \le N-1 \tag{4-2}$$

$$P + Q \le N \tag{4-3}$$

In this algorithm parameter L is determined directly by J and Q as

(4-4) $$L = JQ$$

Since integer G satisfies Equation (2-16) for any valid P and Q pair, taking into consideration Equations (4-2) and (4-3) together with Equation (2-16) leads to the following permissible range of values for G (as a function of P and Q):

(4-5) $$1 \le G \le N - 1$$

Matrix W is the JQxJQ identity,

(4-6) $$W = I_{JQ}$$

And the JPxJQ matrix V and the JQxJQ matrix C are given as

(4-7) $$V^H = \mathcal{R}_{\mathcal{F}:Q,\mathcal{P}:P}\, \mathcal{R}^{-1}_{\mathcal{P}:P,P}$$

(4-8) $$C^H = B^{-1}$$

respectively, where matrix B is obtained via the factorization of the data residual covariance matrix, Ω, as in Equation (2-14). Expression (4-7) for weight matrix V follows from Equation (4-1) and Figure 2-1.

For this algorithm the JQ-element data and steering residuals are

(4-9) $$\underline{\varepsilon}(n) = \underline{x}_{\mathcal{F}:Q}(n) - \underline{\hat{x}}_{\mathcal{F}:Q}(n|\mathcal{X}^-) = \underline{x}_{\mathcal{F}:Q}(n) - \mathcal{R}_{\mathcal{F}:Q,\mathcal{P}:P}\, \mathcal{R}^{-1}_{\mathcal{P}:P,P}\, \underline{x}_{\mathcal{P}:P}(n) \quad n = P, \ldots, N\text{-}Q$$

(4-10) $$\underline{u}(n) = \underline{e}_{\mathcal{F}:Q}(n) - \mathcal{R}_{\mathcal{F}:Q,\mathcal{P}:P}\, \mathcal{R}^{-1}_{\mathcal{P}:P,P}\, \underline{e}_{\mathcal{P}:P}(n) \quad n = P, \ldots, N\text{-}Q$$

respectively. Also, the normalized and uncorrelated (temporally as well as spatially) JQ-element data and steering residuals are

(4-11) $$\underline{v}(n) = C^H \underline{\varepsilon}(n) \qquad n = P, P+1, \ldots, N-Q$$

(4-12) $$\underline{s}(n) = C^H \underline{u}(n) \qquad n = P, P+1, \ldots, N-Q$$

From Equations (2-13b), (4-6), and (4-7), it follows that

(4-13) $$\Omega = \mathcal{R}_{\mathcal{F}:Q,Q} - \mathcal{R}_{\mathcal{F}:Q,\mathcal{P}:P}\, \mathcal{R}^{-1}_{\mathcal{P}:P,P}\, \mathcal{R}^{H}_{\mathcal{F}:Q,\mathcal{P}:P} = \mathcal{R}_{\mathcal{F}:Q,Q} - V^H \mathcal{R}^{H}_{\mathcal{F}:Q,\mathcal{P}:P}$$

Matrix Ω generated according to Equation (4-13) is the model residual covariance referred to in Remark 2.2.4.

The detection test statistic for this algorithm attains the form given in Equation (2-17), wherein the block form of the fully-processed data and steering vector sequences, $\{\underline{v}(n)\}$ and $\{\underline{s}(n)\}$, respectively, is used also.

In its most general form (when $Q \leq P << N$) the OP method is a hybrid processing method, since it has both block and sequential processing aspects. That is, the matrix weight V is generated and applied to the array output in a block mode, and this is repeated several times to generate a sequence of residual vectors.

One advantage of this configuration in relation to the MF is that the dimensionality of the required inverse operation is reduced from JN for the MF to JP for the OP. This presents a major reduction in the computational burden, although several matrix operations are required besides the JPxJP matrix inverse. Such operations include matrix-matrix products and matrix-vector products involving JQxJP and JPxJP matrices and JQ-element

vectors, one matrix addition of JPxJP matrices, and two additions of JQ-element vectors.

In the context of the radar problem considered herein, an adaptive OP (AOP) algorithm is defined by using ensemble averages (ML estimates) over the secondary data in place of unavailable block covariance matrices. Analogously, the ML estimate of the residual covariance matrix, Ω, is generated as an ensemble average over the residual sequences of the secondary data set. The ML estimate is preferred over the model estimate, Equation (4-13), since it yields better detection performance results in other contexts (Román, 1998b).

In the adaptive context, dimensionality reduction induces a reduction in the number of elements required for the secondary data set. This is of relevance in practical cases where the scenario places constraints on the selection of the secondary data, which is often the case for typical N and J values. In addition, due to such dimensionality reduction the AOP can exhibit better detection performance than the AMF using the same secondary data, or equivalent performance using a smaller set of secondary data. This is due to the fact that the accuracy of the ML estimate of a covariance matrix increases as the number of secondary data used to generate the estimate increases.

As stated earlier, each specific choice of key integers leads to a distinct algorithm. Two such cases of interest are presented in Sections 4.1 and 4.2 under known-covariance conditions (for simplicity). The selected cases represent practical boundaries for algorithm configuration options: fully-sequential and fully-block.

4.1 Parametric Adaptive Matched Filter With Yule-Walker (PAMF-YW) And Levinson-Durbin (PAMF-LD) Algorithms

The PAMF for Gaussian-distributed processes has been formulated by Michels et al. (1998) using a sequential whitening filter based on AR models, and using the Strand-Nuttall (SN) model identification algorithm to estimate the AR parameters. This detection approach is referred to as the PAMF-SN. A related detection approach is to utilize the Yule-Walker (YW) algorithm to estimate the AR parameters required in an AR-based PAMF. Such a detection approach is referred to herein as the PAMF-YW. It is demonstrated herein that the PAMF-YW results from a particular selection of the key parameters P and Q in the OP approach. In addition, it is demonstrated that the multichannel Levinson-Durbin (LD) algorithm for solving the augmented YW equations leads to the PAMF-LD detection approach. Thus, both the PAMF-LD and the PAMF-YW are implemented with the detection architecture in Figure 2-1.

Equations (4-2) and (4-3) present the allowable range of values for the P and Q parameters in the OP algorithm. For the present case, select P and Q as

(4-14) $Q = 1$

(4-15) $1 \leq P \leq N-1$

As shown below, in the present context P is the order of the AR system selected to represent the input data. Its specific value is selected based on performance considerations. Given these values for P and Q, it follows from Equations (4-4)-(4-6) that

(4-16) $L = J$

(4-17) $G = N - P$

(4-18) $$W = I_J$$

Weight matrices V and C and covariance matrix Ω are defined next.

The equivalence between the OP method and the PAMF-YW rests on an alternative interpretation of the role of weight matrix V. Weight matrix V is JPxJ in this case, and is obtained using Equation (4-7) for $Q = 1$; namely,

(4-19) $$V^H = \mathcal{R}_{\mathcal{F}:1,\mathcal{P}:P}\, \mathcal{R}^{-1}_{\mathcal{P}:P,P}$$

An algebraically-equivalent expression is obtained by applying the Hermitian transpose to both sides, and then pre-multiplying both sides by $\mathcal{R}_{\mathcal{P}:P,P}$:

(4-20) $$\mathcal{R}_{\mathcal{P}:P,P}\, V = \mathcal{R}^H_{\mathcal{F}:1,\mathcal{P}:P}$$

Now let the unknown matrix V be a partitioned matrix of the form

(4-21) $$V = -\begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_P \end{bmatrix}$$

where each matrix in the set $\{A_k \mid k = 1, 2, \ldots, P\}$ is JxJ. Given this interpretation of weight matrix V, it is straightforward to recognize that Equation (4-20) is the well-known Yule-Walker equation of time series analysis. It is easy to recognize also that the matrices $\{A_k\}$ are the matrix parameters of a multichannel AR system of the form

$$\underline{x}(n) = -\sum_{k=1}^{P} A_k^H \underline{x}(n-k) + \underline{w}(n) \qquad n = 0, 1, \ldots, N\text{-}1 \tag{4-22}$$

where $\{\underline{w}(n)\}$ denotes the temporally-uncorrelated input (or driving) sequence with covariance matrix Ω. System (4-22) is referred to as an AR system of order P, and is denoted as AR(P).

Using Equations (4-9), (4-10), and (4-21), the J-element data and steering residual sequences are determined as

$$\underline{\varepsilon}(n) = \underline{x}(n) - \hat{\underline{x}}(n|n-1) = \underline{x}(n) - V^H \underline{x}_{\mathcal{P}:P}(n) \qquad n = P, P+1, \ldots, N\text{-}1 \tag{4-23a}$$

$$\underline{\varepsilon}(n) = \underline{x}(n) + \sum_{k=1}^{P} A_k^H \underline{x}(n-k) = \sum_{k=0}^{P} A_k^H \underline{x}(n-k) \qquad n = P, P+1, \ldots, N\text{-}1 \tag{4-23b}$$

$$\underline{u}(n) = \underline{e}(n) - V^H \underline{e}_{\mathcal{P}:P}(n) \qquad n = P, P+1, \ldots, N\text{-}1 \tag{4-24a}$$

$$\underline{u}(n) = \underline{e}(n) + \sum_{k=1}^{P} A_k^H \underline{e}(n-k) = \sum_{k=0}^{P} A_k^H \underline{e}(n-k) \qquad n = P, P+1, \ldots, N\text{-}1 \tag{4-24b}$$

respectively. Notice in Equation (4-23a) that the more common and simpler notation $\hat{\underline{x}}(n|n-1)$ has replaced $\hat{\underline{x}}_{\mathcal{F}:Q}(n|\mathcal{X}^-)$. Equations (4-23) and (4-24) are in the form of a multichannel moving average (MA) system, with inputs $\{\underline{x}(n)\}$ and $\{\underline{e}(n)\}$, respectively, and outputs $\{\underline{\varepsilon}(n)\}$ and $\{\underline{u}(n)\}$, respectively. An equivalent interpretation is that these expressions constitute the whitening filter associated with the AR model in Equation (4-22). This is a statement of the fact that an MA system and an AR system are system inverses of each other. Next, from Equations (2-6), (2-7), (4-13), (4-19), and (4-21), the data residual covariance Ω is determined as

$$(4\text{-}25) \qquad \Omega = \mathcal{R}_{\mathcal{F}:1,1} - V^H \mathcal{R}^H_{\mathcal{F}:1,\mathcal{P}:P} = R_{xx}(0) + \sum_{k=1}^{P} A_k^H R_{xx}(-k)$$

Finally, the uncorrelated and normalized J-element data and steering residual sequences are obtained as in Equations (4-11) and (4-12), since the JxJ weight matrix C is obtained as in Equation (4-8), with B a factorization of Ω as in Equation (2-14), and Ω calculated as in Equation (4-25).

Concatenation of Equations (4-20) and (4-25) leads to the augmented Yule-Walker equation. Specifically,

$$(4\text{-}26) \qquad \left[\begin{array}{c|c} \mathcal{R}_{\mathcal{F}:1,1} & \mathcal{R}_{\mathcal{F}:1,\mathcal{P}:P} \\ \hline \mathcal{R}^H_{\mathcal{F}:1,\mathcal{P}:P} & \mathcal{R}_{\mathcal{P}:P,P} \end{array}\right] \left[\begin{array}{c} I_J \\ \hline -V \end{array}\right] = \left[\begin{array}{c} \Omega \\ \hline O_{JP,J} \end{array}\right]$$

Equation (4-26) can be solved using the multichannel Levinson-Durbin (LD) algorithm. Such an approach applied to the detection context considered herein thus leads to the PAMF-LD detection algorithm.

In the context of this sub-section, the OP method is fully-sequential, and the application of the weight matrix V in Figure 2-1 is viewed as a recursive filter, instead of a block filter.

4.2 Matched Maximum-Dimension Orthogonal Projection

Consider the constraints for P and Q in Equations (4-2) and (4-3), and select for this case

$$(4\text{-}27) \qquad Q = P = \frac{N}{2}$$

These values result in fully-block configuration, which is the opposite of the fully-sequential configuration of Section 4.1 above. Then, from Equations (2-16) and (4-4)-(4-6),

$$L = \frac{JN}{2} \tag{4-28}$$

$$G = 1 \tag{4-29}$$

$$W = I_{JN/2} \tag{4-30}$$

These parameter values indicate that the data and steering residuals sequences consist of only one L-element vector.

From Equations (4-7), (4-8), and (4-13) with $P = N/2$, the JPxJP weight matrix V, the JPxJP weight matrix C, and the JPxJP covariance matrix Ω are obtained as follows,

$$V^H = \mathcal{R}_{\mathcal{F}:P,\mathcal{P}:P}\, \mathcal{R}^{-1}_{\mathcal{P}:P,P} \tag{4-31}$$

$$\Omega = \mathcal{R}_{\mathcal{F}:P,P} - \mathcal{R}_{\mathcal{F}:P,\mathcal{P}:P}\, \mathcal{R}^{-1}_{\mathcal{P}:P,P}\, \mathcal{R}^{H}_{\mathcal{F}:P,\mathcal{P}:P} = \mathcal{R}_{\mathcal{F}:P,P} - V^H \mathcal{R}^{H}_{\mathcal{F}:P,\mathcal{P}:P} \tag{4-32}$$

$$C^H = B^{-1} \tag{4-33}$$

$$BB^H = \Omega \tag{4-34}$$

An augmented set of equations can be defined by concatenating Equations (4-31) and (4-32), as in Section 4.1. But, in contrast with the case handled in Section 4.1, the case considered in this section does not appear to have alternative interpretations.

The OP algorithm configuration in this sub-section is a block processing method. That is, weight matrix V is applied once per

CPI per range cell (the residual sequence has one vector element). In this fully-block configuration the past and future vectors have the same dimensionality, so the conditional expectation in Equation (4-1) requires the inverse and product of square matrices of the same dimensions, as indicated in Equation (4-31). The residual covariance matrix is generated also involving only operations on square matrices of the same dimensions, as indicated in Equation (4-32). This results in the minimum number of computations for a block processing method (Guerci and Feria, 1996).

5.0 MAXIMUM CORRELATION (MC)

The MC algorithm for adaptive arrays has been proposed by Román (1998a) in the context of adaptive sidelobe canceling using complex-valued data, and is based on the concept of canonical variables and canonical correlations formulated by Hotelling (1936) for real-valued random variables. That methodology is extended herein to STAP applications in general, and enhanced to include a detection test. One of the contributions in (Román, 1998a) is the non-trivial extension of the canonical variables concept to handle the case of complex-valued data.

Hotelling's canonical correlations formulation is as follows. Given two sets of scalar random variables (or equivalently, two random vectors) find a basis for each set such that in the transformed basis the first variable in each set is maximally correlated with the first variable in the other set, and uncorrelated with all the other variables in the two sets. And also that the second variable in each transformed set is maximally correlated with the second variable in the other transformed set, and uncorrelated with all the other variables in the two transformed sets. And so on until all the variables in the two sets have been considered. In addition, each variable in the two transformed sets is normalized to unit variance. The variables in the transformed basis are the canonical variables, and the normalized and optimized correlations are the canonical correlations. The maximization required repeatedly in the procedure is straightforward for real-valued random variables, but the problem formulation requires a modification for complex-valued random variables because the maximum of a complex-valued variable is non-unique (Román, 1998a).

In the STAP application context as formulated herein, block vectors $\underline{\mathbf{X}}_{\mathcal{P}:\mathrm{P}}$ and $\underline{\mathbf{X}}_{\mathcal{F}:\mathrm{Q}}$ are the vectors to be transformed into

canonical variables, and weight matrices W and V in Figure 2-1 are the transformation matrices. The foundation for the MC algorithm is that the residual sequence $\{\underline{\varepsilon}(n)\}$ is uncorrelated in time because it is generated as the difference between maximally-correlated random variables.

As before, integers P and Q define the number of sub-vectors in the stack vectors $\underline{x}_{\mathcal{P}:P}(n)$ and $\underline{x}_{\mathcal{F}:Q}(n)$, respectively; that is, $\underline{x}_{\mathcal{P}:P}(n)$ is a JP-element vector, and $\underline{x}_{\mathcal{F}:Q}(n)$ is a JQ-element vector. Parameters P and Q can be selected to be within a wide range of values, but are considered herein to satisfy the following conditions (recall Assumption (2-2) and Constraint (2-3)):

(5-1) $\quad 1 \le Q \le P \le N-1$

(5-2) $\quad P + Q \le N$

Recall that Constraint (5-2) is due to the fact that the number of data points is fixed. Selection of P and Q is based on the typical trade-off between computational and performance issues (see Section 5.2 for additional comments). For example, small-valued P and Q allow for dramatic dimensionality reduction in relation to the MF, but can entail a noticeable loss in detection performance in relation to the full-dimension case ($P + Q = N$).

The MC algorithm includes another variable parameter, the dimensionality of the data and steering residuals, whose selection allows for additional dimensionality reduction beyond that provided by appropriate selection of P and Q. In the generic architecture of Figure 2-1, weight matrices V and W are dimensioned JPxL and JQxL, respectively, and the data and steering residuals, $\underline{\varepsilon}(n)$ and $\underline{u}(n)$, respectively, are L-element column vectors, where L is selected such that

(5-3) $1 \le L \le JQ$

Selection of L is based on dimensionality and performance considerations. Criteria for the selection of L are based on information theory concepts and utilize intermediate results from the algorithm (see Section 5.2).

In summary, there are two mechanisms for dimensionality reduction in the MC algorithm: first, selection of P and Q such that $Q \le P < N$ (and $P << N$ can provide satisfactory performance); and second, selection of L such that $L < JQ$.

Formulation of the method and establishment of the general solution is accomplished best by assuming full dimensionality for the residual sequences; that is, with $L = JQ$. Then, the two JQ-element vector sequences $\{\underline{\alpha}(n) \mid n = P, P+1, \ldots, N-Q\}$ and $\{\underline{\beta}(n) \mid n = P, P+1, \ldots, N-Q\}$ in Figure 2-1 are defined in terms of two intermediate sequences, $\{\underline{\upsilon}(n) \mid n = P, P+1, \ldots, N-Q\}$ and $\{\underline{\mu}(n) \mid n = P, P+1, \ldots, N-Q\}$, as follows:

(5-4a)
$$\underline{\upsilon}(n) = \begin{bmatrix} \upsilon_1(n) \\ \upsilon_2(n) \\ \vdots \\ \upsilon_{JQ}(n) \end{bmatrix} = \begin{bmatrix} \underline{w}_1^H \\ \underline{w}_2^H \\ \vdots \\ \underline{w}_{JQ}^H \end{bmatrix} \begin{bmatrix} \underline{x}(n) \\ \underline{x}(n+1) \\ \vdots \\ \underline{x}(n+Q-1) \end{bmatrix} \qquad n = P, P+1, \ldots, N-Q$$

(5-4b)
$$\underline{\upsilon}(n) = W^H \underline{x}_{\mathcal{F}:Q}(n) \qquad n = P, P+1, \ldots, N-Q$$

(5-5)
$$\underline{\alpha}(n) = \underline{\upsilon}(n) = W^H \underline{x}_{\mathcal{F}:Q}(n) \qquad n = P, P+1, \ldots, N-Q$$

(5-6a) $$\underline{\mu}(n) = \begin{bmatrix} \mu_1(n) \\ \vdots \\ \mu_{JQ}(n) \\ \hline \mu_{JQ+1}(n) \\ \vdots \\ \mu_{JP}(n) \end{bmatrix} = \begin{bmatrix} \underline{v}_1^H \\ \vdots \\ \underline{v}_{JQ}^H \\ \hline \underline{h}_1^H \\ \vdots \\ \underline{h}_{J(P-Q)}^H \end{bmatrix} \begin{bmatrix} \underline{x}(n-1) \\ \underline{x}(n-2) \\ \vdots \\ \underline{x}(n-P) \end{bmatrix} \qquad n = P, P+1, \ldots, N-Q$$

(5-6b) $$\underline{\mu}(n) = \begin{bmatrix} \underline{\mu}_V(n) \\ \hline \underline{\mu}_H(n) \end{bmatrix} = \begin{bmatrix} V^H \\ \hline H^H \end{bmatrix} \underline{x}_{\mathcal{P}:P}(n) \qquad n = P, P+1, \ldots, N-Q$$

(5-7) $$\underline{\beta}(n) = \underline{\mu}_V(n) = V^H \underline{x}_{\mathcal{P}:P}(n) \qquad n = P, P+1, \ldots, N-Q$$

where each $\underline{w}_i(n)$, for $i = 1, 2, \ldots, JQ$, is a JQ-element column vector, each $\underline{v}_i(n)$, for $i = 1, 2, \ldots, JQ$, is a JP-element column vector, and each $\underline{h}_i(n)$, for $i = 1, 2, \ldots, J(P-Q)$, is a JP-element column vector. In the context of this general case with full-dimensionality ($Q < P < N$ and $L = JQ$), W is JQxJQ, V is JPxJQ, and H is JPxJ(P-Q); also, all three matrices have full rank. The sequence $\{\underline{\mu}_H(n) \mid n = P, P+1, \ldots, N-Q\}$ of J(P-Q)-element vectors lacks a role in the STAP architecture because it is selected by the MC algorithm such that it spans a subspace orthogonal to $\underline{\upsilon}(n)$.

The residual vector $\underline{\varepsilon}(n)$ is a JQ-element column vector, and the data residual vector sequence is

(5-8a) $$\underline{\varepsilon}(n) = \underline{\alpha}(n) - \underline{\beta}(n) = \underline{\upsilon}(n) - \underline{\mu}_V(n) \qquad n = P, P+1, \ldots, N-Q$$

(5-8b) $$\underline{\varepsilon}(n) = W^H \underline{x}_{\mathcal{F}:Q}(n) - V^H \underline{x}_{\mathcal{P}:P}(n) \qquad n = P, P+1, \ldots, N-Q$$

As before, G is used to denote the number of elements in the residual vector sequence. Since P and Q are specified only to lie within a range, then

(5-9) $$G = N - P - Q + 1$$

Definitions analogous to those in Equations (5-4) through (5-8) can be introduced for the steering sequence path in the architecture of Figure (2-1). However, the MC algorithm weight matrices V and W are designed based on probabilistic concepts, and applied analogously in the steering sequence path of Figure 2-1. Also, notice that in the MC approach both weight matrices V and W play a role. This is in contrast with the MF, wherein $V = [0]$, and with the OP, wherein $W = I_{JQ}$.

Equations (5-8) suggests that the data residual sequence is uncorrelated in time if the weight matrices V and W (as well as matrix H) are selected such that: a) the JQ elements of $\underline{\mu}_V(n)$ are maximally correlated with the corresponding elements of $\underline{\upsilon}(n)$ and uncorrelated with the others, b) the J(P-Q) elements of $\underline{\mu}_H(n)$ are uncorrelated with the elements of $\underline{\upsilon}(n)$, c) the elements of $\underline{\upsilon}(n)$ are pair-wise uncorrelated, and d) the elements of $\underline{\mu}(n)$ are pair-wise uncorrelated. Such is the objective of the MC approach, as presented in detail next.

Maximum Correlation Formulation (L = JQ Configuration). Generate the weight matrices V and W, and matrix H, such that at each instant of time n, the conditions stated next are true for the sequences $\{\underline{\upsilon}(n)\}$ and $\{\underline{\mu}(n)\}$. First, select $\underline{w}_1$ and $\underline{v}_1$ such that $\upsilon_1(n)$ is maximally correlated with $\mu_1(n)$, and both $\upsilon_1(n)$ and $\mu_1(n)$ have unit variance. Second, select $\underline{w}_2$ and $\underline{v}_2$ such that $\upsilon_2(n)$ is maximally correlated with $\mu_2(n)$, and $\upsilon_2(n)$ and $\mu_2(n)$ are both uncorrelated with $\upsilon_1(n)$ and with $\mu_1(n)$, and both $\upsilon_2(n)$ and $\mu_2(n)$ have unit variance. The ith condition (for $2 \le i \le JQ$) is to select $\underline{w}_i$ and $\underline{v}_i$ such that $\upsilon_i(n)$ is

maximally correlated with $\mu_i(n)$, and $\upsilon_i(n)$ and $\mu_i(n)$ are both uncorrelated with $\upsilon_1(n)$, $\upsilon_2(n)$, . . . , $\upsilon_{i-1}(n)$, and with $\mu_1(n)$, $\mu_2(n)$, . . . , $\mu_{i-1}(n)$, and both $\upsilon_i(n)$ and $\mu_i(n)$ have unit variance. The (JQ+1)th condition is to select $\underline{h}_1$ such that $\mu_{JQ+1}(n)$ has unit variance and is uncorrelated with $\upsilon_1(n)$, $\upsilon_2(n)$, . . . , $\upsilon_{JQ}(n)$, and with $\mu_1(n)$, $\mu_2(n)$, . . . , $\mu_{JQ}(n)$. Lastly, the JPth condition is to select $\underline{h}_{J(P-Q)}$ such that $\mu_{JP}(n)$ has unit variance and is uncorrelated with $\upsilon_1(n)$, $\upsilon_2(n)$, . . . , $\upsilon_{JQ}(n)$, and with $\mu_1(n)$, $\mu_2(n)$, . . . , $\mu_{JP-1}(n)$. Carrying out the required optimization at each step (maximizing the complex-valued correlation coefficient as in [Román, 1998a]) leads to the following compact set of matrix equations,

$$E\left[\underline{\upsilon}(n)\underline{\upsilon}^H(n)\right] = W^H E\left[\underline{x}_{\mathcal{F}:Q}(n)\underline{x}^H_{\mathcal{F}:Q}(n)\right]W = W^H \mathcal{R}_{\mathcal{F}:Q,Q}\, W = I_{JQ} \tag{5-10}$$

$$E\left[\underline{\mu}(n)\underline{\mu}^H(n)\right] = \begin{bmatrix} V^H \\ H^H \end{bmatrix} E\left[\underline{x}_{\mathcal{P}:P}(n)\underline{x}^H_{\mathcal{P}:P}(n)\right]\begin{bmatrix} V & H \end{bmatrix} = \begin{bmatrix} V^H \\ H^H \end{bmatrix} \mathcal{R}_{\mathcal{P}:P,P}\begin{bmatrix} V & H \end{bmatrix} = I_{JP} \tag{5-11}$$

$$E\left[\underline{\upsilon}(n)\underline{\mu}^H(n)\right] = W^H E\left[\underline{x}_{\mathcal{F}:Q}(n)\underline{x}^H_{\mathcal{P}:P}(n)\right]\begin{bmatrix} V & H \end{bmatrix} = W^H \mathcal{R}_{\mathcal{F}:Q,\mathcal{P}:P}\begin{bmatrix} V & H \end{bmatrix} \tag{5-12a}$$

$$E\left[\underline{\upsilon}(n)\underline{\mu}^H(n)\right] = R_{\upsilon\mu}(0) = R_{\upsilon\mu} = \begin{bmatrix} \rho_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \rho_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \rho_{JQ} & 0 & \cdots & 0 \end{bmatrix} \tag{5-12b}$$

$$E\left[\underline{\upsilon}(n)\underline{\mu}^H(n)\right] = W^H \mathcal{R}_{\mathcal{F}:Q,\mathcal{P}:P}\left[V \,\vdots\, H\right] = \left[\begin{array}{cccc:ccc} \rho_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \rho_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \rho_{JQ} & 0 & \cdots & 0 \end{array}\right] \tag{5-13}$$

$$1 \geq \rho_1 \geq \rho_2 \geq \ldots \geq \rho_{JQ} \geq 0 \tag{5-14}$$

where ρ_i is the correlation coefficient between $\upsilon_i(n)$ and $\mu_i(n)$, and Equation (5-13) follows trivially from Equations (5-12). The unit variance conditions (Equations (5-10) and (5-11)) are introduced without loss of generality because the correlation coefficient between two random variables is invariant to a scale factor applied to either one of the two variables or to both of the variables.

Variables $\{\upsilon_i(n) \mid i = 1, 2, \ldots, JQ\}$ and $\{\mu_i(n) \mid i = 1, 2, \ldots, JP\}$ constitute the complete set of canonical variables, and the correlation coefficients $\{\rho_i \mid i = 1, 2, \ldots, JQ\}$ are the canonical correlations for the random vectors $\underline{x}_{\mathcal{P}:P}$ and $\underline{x}_{\mathcal{F}:Q}$ (recall that $Q \leq P$ is assumed).

In summary, Equations (5-10) through (5-14) constitute the relations that must be solved for the unknown canonical correlations and weight matrices.

Maximum Correlation Solution ($L = JQ$ Configuration). Weight matrices V and W, matrix H, and the canonical correlations $\{\rho_i\}$ that satisfy Equations (5-10) through (5-14) are obtained as follows. Given the array output data block covariances defined in Equations (2-5)-(2-7), form the coherence matrix $R_{\theta\delta}$ as

$$(5\text{-}15) \qquad R_{\theta\delta} = E\left[\underline{\theta}(n)\underline{\delta}^H(n)\right] = \mathcal{R}_{\mathcal{F}:Q,Q}^{-1/2}\, \mathcal{R}_{\mathcal{F}:Q,\mathcal{P}:P}\, \mathcal{R}_{\mathcal{P}:P,P}^{-1/2}$$

where $\underline{\theta}(n) = \mathcal{R}_{\mathcal{F}:Q,Q}^{-1/2}\, \underline{x}_{\mathcal{F}:Q}(n)$ and $\underline{\delta}(n) = \mathcal{R}_{\mathcal{P}:P,P}^{-1/2}\, \underline{x}_{\mathcal{P}:P}(n)$ are dummy variables. As inferred in Equation (5-15), $R_{\theta\delta}$ is the JQxJP cross-covariance between the two white random vectors, $\underline{\theta}(n)$ and $\underline{\delta}(n)$ (both have identity covariance matrix). Now apply the SVD to the coherence matrix $R_{\theta\delta}$ to obtain a factorization of the form

$$(5\text{-}16) \qquad R_{\theta\delta} = T_1 D T_2^H$$

where T_1 is a JQxJQ unitary matrix, T_2 is a JPxJP unitary matrix, and D is a real-valued JQxJP matrix with zero-valued elements everywhere except along the main diagonal (recall that $P \geq Q$), and the elements along the main diagonal are arranged in order of descending magnitude. The real-valued elements along the main diagonal are the JQ canonical correlations between the past and future vectors, arranged in descending order. That is,

$$D = \left[\begin{array}{cccc:ccc} \rho_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \rho_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \rho_{JQ} & 0 & \cdots & 0 \end{array}\right] = \left[\begin{array}{c:c} D_1 & D_2 \end{array}\right] \tag{5-17}$$

$$1 \geq \rho_1 \geq \rho_2 \geq \ldots \geq \rho_{JQ} \geq 0 \tag{5-18}$$

with D_1 a JQxJQ diagonal real-valued matrix, and D_2 a JQxJ(P-Q) null matrix. In the cases where $P = Q$, matrix D is square and diagonal. Also, it follows from Equations (5-12b) and (5-17) that $R_{\upsilon\mu} = D$.

Matrices T_1 and T_2 play a role that is obscured in the above formulation. These matrices transform the intermediate, dummy variables $\underline{\theta}(n)$ and $\underline{\delta}(n)$ into the canonical variables $\underline{\upsilon}(n)$ and $\underline{\mu}(n)$, respectively. Specifically, $\underline{\upsilon}(n) = T_1^H \underline{\theta}(n)$ and $\underline{\mu}(n) = T_2^H \underline{\delta}(n)$.

Based on the above development, weight matrices W and V and matrix H are obtained as

$$W = \mathcal{R}_{\mathcal{F}:Q,Q}^{-1/2} T_1 \tag{5-19}$$

$$V = \mathcal{R}_{\mathcal{P}:P,P}^{-1/2} T_{2V} \tag{5-20}$$

(5-21) $$H = \mathcal{R}_{\mathcal{P}:P,P}^{-1/2} T_{2H}$$

where T_{2V} and T_{2H} are JPxJQ and JPxJ(P-Q) partitions, respectively, of the unitary matrix T_2; namely,

(5-22) $$T_2 = \left[\underline{t}_{2,1} \quad \underline{t}_{2,2} \quad \cdots \quad \underline{t}_{2,JQ} \,\vdots\, \underline{t}_{2,JQ+1} \quad \cdots \quad \underline{t}_{2,JP}\right] = \left[T_{2V} \,\vdots\, T_{2H}\right]$$

The covariance matrix of the data residual is generated using Equation (2-13) and Equations (5-10) through (5-18).

For the MC method the residual covariance is a simple function of the non-zero canonical correlations; specifically,

(5-23) $$\Omega = 2I_{JQ} - 2D_1 = 2(I_{JQ} - D_1) = 2\begin{bmatrix} 1-\rho_1 & 0 & \cdots & 0 \\ 0 & 1-\rho_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1-\rho_{JQ} \end{bmatrix}$$

where D_1 is a JQxJQ diagonal sub-matrix of D defined in Equation (5-17). The diagonal form of Ω simplifies the final block filtering step in the detection architecture because weight matrix C is then diagonal also (Equations (2-14) and (2-15)). That is,

(5-24) $$C^H = B^{-1} = \Omega^{-1/2}$$

where $\Omega^{-1/2}$ is the inverse square-root matrix of the data residual covariance matrix. The normalized and uncorrelated (temporally and spatially) JQ-element data and steering residuals for the MC algorithm are

(5-25) $$\underline{v}(n) = C^H \underline{\varepsilon}(n) \qquad n = P, P+1, \ldots, N-Q$$

(5-26) $\underline{s}(n) = C^H \underline{u}(n)$ $n = P, P+1, \ldots, N-Q$

The form of Ω provides insight into the issue of dimensionality reduction, and suggests a criterion for dimensionality reduction (Section 5.2).

The detection test statistic for this algorithm attains the form given in Equation (2-17), wherein the block form of the fully-processed data and steering vector sequences, $\{\underline{v}(n)\}$ and $\{\underline{s}(n)\}$, respectively, is used also.

Maximum Correlation Solution ($L < JQ$ Configuration). Consider now the condition where the data residual vector dimensionality is $L < JQ$. This condition could be forced by the need to reduce the computational requirements, or may be driven by the recognition, utilizing some dimensionality-reducing measure, that the effect of neglecting the canonical variables beyond $L+1$ is negligible (see Section 5.2). Under such conditions only a subset of the canonical variables is utilized, even though the complete set of canonical variables and canonical correlations is generated by the algorithm. Specifically, for $L < JQ$, only the variables $\{\upsilon_i(n) \mid i = 1, 2, \ldots, L\}$ and $\{\mu_i(n) \mid i = 1, 2, \ldots, L\}$ and the correlation coefficients $\{\rho_i \mid i = 1, 2, \ldots, L\}$ are retained. This implies the formulas for the calculation of the weight matrices must be modified.

For the case $L < JQ$ the L-element vector sequences $\{\underline{\alpha}(n) \mid n = P, P+1, \ldots, N-Q\}$ and $\{\underline{\beta}(n) \mid n = P, P+1, \ldots, N-Q\}$ in Figure 2-1 are defined as

(5-27) $$\underline{\alpha}(n) = \begin{bmatrix} \upsilon_1(n) \\ \vdots \\ \upsilon_L(n) \end{bmatrix} = \begin{bmatrix} \underline{w}_1^H \\ \vdots \\ \underline{w}_L^H \end{bmatrix} \underline{x}_{\mathcal{F}:Q}(n) = W^H \underline{x}_{\mathcal{F}:Q}(n) \qquad n = P, P+1, \ldots, N-Q$$

$$\underline{\beta}(n) = \begin{bmatrix} \mu_1(n) \\ \vdots \\ \mu_L(n) \end{bmatrix} = \begin{bmatrix} \underline{v}_1^H \\ \vdots \\ \underline{v}_L^H \end{bmatrix} \underline{x}_{\mathcal{P}:P}(n) = V^H \underline{x}_{\mathcal{P}:P}(n) \qquad n = P, P+1, \ldots, N-Q \tag{5-28}$$

Also, it is convenient to define a partition in each one of the two matrices T_1 and T_2 of the form

$$T_1 = \left[\underline{t}_{1,1} \;\; \underline{t}_{1,2} \;\; \cdots \;\; \underline{t}_{1,L} \;\vdots\; \underline{t}_{1,L+1} \;\; \cdots \;\; \underline{t}_{1,JQ} \right] = \left[T_{1A} \;\vdots\; T_{1B} \right] \tag{5-29}$$

$$T_2 = \left[\underline{t}_{2,1} \;\; \underline{t}_{2,2} \;\; \cdots \;\; \underline{t}_{2,L} \;\vdots\; \underline{t}_{2,L+1} \;\; \cdots \;\; \underline{t}_{2,JQ} \;\; \cdots \;\; \underline{t}_{2,JP} \right] = \left[T_{2A} \;\vdots\; T_{2B} \right] \tag{5-30}$$

where T_{1A} and T_{1B} are JQxL and JQxJ(Q-L) partitions, respectively, of the unitary matrix T_1; also, T_{2A} and T_{2B} are JPxL and JPxJ(P-L) partitions, respectively, of the unitary matrix T_2. The partitions in Equation (5-30) are distinct from those in Equation (5-22). However, notice that $T_{2A} = T_{2V}$ and $T_{2B} = T_{2H}$ when $L = JQ$. Given these definitions, weight matrices W and V are obtained as

$$W = \mathcal{R}_{\mathcal{F}:Q,Q}^{-1/2} T_{1A} \tag{5-31}$$

$$V = \mathcal{R}_{\mathcal{P}:P,P}^{-1/2} T_{2A} \tag{5-32}$$

As before, the covariance matrix of the data residual is generated using Equation (2-13) and Equations (5-10) through (5-18), which results in,

$$\Omega = 2I_L - 2D_{11} = 2(I_L - D_{11}) = 2 \begin{bmatrix} 1-\rho_1 & 0 & \cdots & 0 \\ 0 & 1-\rho_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1-\rho_L \end{bmatrix} \tag{5-33}$$

where D_{11} is an LxL diagonal sub-matrix of D_1; namely,

(5-34a) $$D_1 = \begin{bmatrix} D_{11} & [0] \\ [0] & D_{12} \end{bmatrix}$$

(5-34b) $$D_{11} = \text{diag}\left[\{\rho_1, \rho_2, \ldots, \rho_L\}\right]$$

(5-34c) $$D_{12} = \text{diag}\left[\{\rho_{L+1}, \rho_{L+2}, \ldots, \rho_{JQ}\}\right]$$

Weight matrix C is calculated using Equation (5-24), with Ω as in Equation (5-33). The detection test statistic is generated also via Equation (2-17). This completes the formulas for $L < JQ$.

The calculations in the MC algorithm involve matrices of dimensions JQ and JP, with $Q \leq P$. In contrast, the MF algorithm requires the inverse of a JNxJN Hermitian matrix. This implies a dramatic reduction in the computational requirements for typical values of J and N, specially if $P << N$. In addition, the calculations in the final block filtering step (weight matrix C) are reduced further when $L < JQ$ is feasible (Section 5.2).

In the radar array context considered herein, an adaptive MC (AMC) algorithm is configured by using ensemble averages (ML estimates) over the secondary data in place of unavailable block covariance matrices. Analogously, the ML estimate of the residual covariance matrix, Ω, is generated as an ensemble average over the residual sequences of the secondary data set. Such ML estimate is preferred over the appropriate model estimate, either Equation (5-23) or Equation (5-33), since it yields better detection performance results in other contexts (Román, 1998b).

In the adaptive context, dimensionality reduction induces a reduction in the number of elements required for the secondary

data set, as for the OP algorithm. This is of relevance in practical cases where the scenario places constraints on the selection of the secondary data, which is often the case for typical N and J values. In addition, due to such dimensionality reduction the AMC can exhibit better detection performance than the AMF using the same secondary data, or equivalent performance using a smaller set of secondary data, as for the OP algorithm.

In its most general form (when $Q \leq P << N$) the MC method is a hybrid processing method, since it has both block and sequential processing aspects. Specifically, the matrix weights, W and V, are generated and applied to the array output in a block mode, and this is repeated several times to generate a sequence of residual vectors. A parametric, fully-sequential implementation of the MC method is obtained by generating a multichannel (multi-input, multi-output) state variable model (SVM) using one of the two sets of canonical variables as the state variables. The SVM matrix parameters are generated via operations on the past and future auto- and cross-covariances. Such an approach has been formulated and tested extensively by SSC for the case $Q = P$, and is referred to as the PAMF using canonical correlations (CC), or PAMF-CC (Román and Davis, 1993b; Román et al., 1997, 1998). The PAMF admits several variations wherein other linear model types and/or model identification algorithms are utilized, as evidenced by the developments in Section 4.1.

As with the OP algorithm, the MC algorithm admits a wide variety of configurations based on the selection of the parameters P and Q. One such configuration is the PAMF-CC mentioned above. Another is the matched, maximum-dimension, fully-block algorithm obtained when $Q = P = N/2$. At the other end of the spectrum is the sidelobe canceler, wherein $JQ = 1$ and $JP = N-1$ (Román, 1998a). Further analysis of these variations remains for future work.

5.1 Information Theory And Canonical Correlations

Information theory plays a key role in the interpretation of canonical variables, and in the formulation of a dimensionality-reduction criterion (implemented via determination of parameter L). The basic information theory concepts referred to herein are discussed in detail in various texts, including Beckmann (1967) and Thomas (1969).

The amount of information about a vector random process provided by a realization of the process is referred to as self-information. Self-information of a realization of a random vector $\underline{x}$, denoted herein by $\mathfrak{I}(\underline{x})$, is defined as

$$\mathfrak{I}(\underline{x}) = -\log_b[p(\underline{x})] = \log_b\left[\frac{1}{p(\underline{x})}\right] \tag{5-35}$$

where $p(\underline{x})$ denotes the probability density function (PDF) of $\underline{x}$, and the logarithm base b is arbitrary. Notice that $\mathfrak{I}(\underline{x})$ is a random variable. The average self-information of a vector random process is the entropy of the process. It follows that the entropy of a random vector $\underline{x}$, denoted herein by $\mathcal{H}(\underline{x})$, is defined as

$$\mathcal{H}(\underline{x}) = \int_{-\infty}^{\infty}\cdots\int \mathfrak{I}(\underline{x})\, p(\underline{x})\, d\underline{x} = -\int_{-\infty}^{\infty}\cdots\int \log_b[p(\underline{x})]\, p(\underline{x})\, d\underline{x} \tag{5-36}$$

The entropy of a random vector $\underline{x}$ given a realization of another random vector $\underline{z}$ is referred to as the conditional entropy. Conditional entropy of $\underline{x}$ conditioned on $\underline{z}$, denoted herein by $\mathcal{H}(\underline{x}|\underline{z})$, is defined as

$$\mathcal{H}(\underline{x}|\underline{z}) = -\int_{-\infty}^{\infty}\cdots\int \log_b[p(\underline{x}|\underline{z})]\, p(\underline{x},\underline{z})\, d\underline{x}\, d\underline{z} \tag{5-37}$$

where $p(\underline{x},\underline{z})$ and $p(\underline{x}|\underline{z})$ denote the joint and conditional PDFs, respectively, of $\underline{x}$ and $\underline{z}$. Notice that $\mathcal{H}(\underline{x}|\underline{z})$ is a random variable also. The joint entropy of a random vector $\underline{x}$ and a random vector $\underline{z}$, denoted herein by $\mathcal{H}(\underline{x},\underline{z})$, is defined as

$$\mathcal{H}(\underline{x},\underline{z}) = -\int_{-\infty}^{\infty}\cdots\int \log_b[p(\underline{x},\underline{z})]\, p(\underline{x},\underline{z})\, d\underline{x}\, d\underline{z} \tag{5-38}$$

Joint entropy and conditional entropy satisfy the following key relations,

$$\mathcal{H}(\underline{x},\underline{z}) = \mathcal{H}(\underline{x}) + \mathcal{H}(\underline{z}|\underline{x}) = \mathcal{H}(\underline{z}) + \mathcal{H}(\underline{x}|\underline{z}) \tag{5-39}$$

The last concept of relevance to the objectives herein is the average mutual information between two random vectors, $\underline{x}$ and $\underline{z}$. Most often this concept is referred to simply as mutual information, and is a measure of the information in $\underline{x}$ about $\underline{z}$, or equivalently, the information in $\underline{z}$ about $\underline{x}$. Mutual information between the random vectors $\underline{x}$ and $\underline{z}$, denoted herein by $I(\underline{x} \leftrightarrow \underline{z})$, is defined as

$$I(\underline{x} \leftrightarrow \underline{z}) = \mathcal{H}(\underline{x}) - \mathcal{H}(\underline{x}|\underline{z}) = \mathcal{H}(\underline{z}) - \mathcal{H}(\underline{z}|\underline{x}) \tag{5-40}$$

Using Equation (5-39) it is straightforward to show that

$$I(\underline{x} \leftrightarrow \underline{z}) = \mathcal{H}(\underline{x}) + \mathcal{H}(\underline{z}) - \mathcal{H}(\underline{x},\underline{z}) \tag{5-41}$$

Notice that mutual information is an entropy-type quantity.

In the context of interest herein, the above-defined measures are determined for the past and future processes, $\underline{x}_{\mathcal{P}:P}$ and $\underline{x}_{\mathcal{F}:Q}$, respectively, individually and jointly. Recall that the array

output process $\{\underline{x}(n)\}$ has mean zero and is Gaussian-distributed. The complex Gaussian PDF for a zero-mean, M-element random vector $\underline{x}$ is of the form

$$p(\underline{x}) = \pi^{-M} |\mathcal{R}|^{-M} \exp\left[-\underline{x}^H \mathcal{R}^{-1} \underline{x}\right] \tag{5-42}$$

where $\mathcal{R}$ is the MxM covariance matrix of $\underline{x}$,

$$\mathcal{R} = E\left[\underline{x}\,\underline{x}^H\right] \tag{5-43}$$

and $|\mathcal{R}|$ denotes the determinant of $\mathcal{R}$. This PDF expression is required to represent each of three cases. First, the PDF of $\underline{x}_{\mathcal{P}:P}$, with $M = JP$ and covariance matrix $\mathcal{R}_{\mathcal{P}:P,P}$; second, the PDF of $\underline{x}_{\mathcal{F}:Q}$, with $M = JQ$ and covariance matrix $\mathcal{R}_{\mathcal{F}:Q,Q}$; and third, the joint PDF of $\underline{x}_{\mathcal{P}:P}$ and $\underline{x}_{\mathcal{F}:Q}$, with $M = J(P+Q)$ and covariance matrix

$$E\left[\begin{bmatrix} \underline{x}_{\mathcal{P}:P}(n) \\ \underline{x}_{\mathcal{F}:Q}(n) \end{bmatrix} \begin{bmatrix} \underline{x}^H_{\mathcal{P}:P}(n) & \underline{x}^H_{\mathcal{F}:Q}(n) \end{bmatrix}\right] = \begin{bmatrix} \mathcal{R}_{\mathcal{P}:P,P} & \mathcal{R}_{\mathcal{P}:P,\mathcal{F}:Q} \\ \mathcal{R}_{\mathcal{F}:Q,\mathcal{P}:P} & \mathcal{R}_{\mathcal{F}:Q,Q} \end{bmatrix} \tag{5-44}$$

Then, adopting the natural (base e) logarithm option, the entropy measures for the three processes are obtained as

$$\mathcal{H}(\underline{x}_{\mathcal{P}:P}) = JP + JP \ln[\pi] + \ln\left[\left|\mathcal{R}_{\mathcal{P}:P,P}\right|\right] \tag{5-45}$$

$$\mathcal{H}(\underline{x}_{\mathcal{F}:Q}) = JQ + JQ \ln[\pi] + \ln\left[\left|\mathcal{R}_{\mathcal{F}:Q,Q}\right|\right] \tag{5-46}$$

$$\begin{aligned} \mathcal{H}(\underline{x}_{\mathcal{P}:P}, \underline{x}_{\mathcal{F}:Q}) = {} & JP + JQ + JP \ln[\pi] + JQ \ln[\pi] + \ln\left[\left|\mathcal{R}_{\mathcal{P}:P,P}\right|\right] \\ & + \ln\left[\left|\mathcal{R}_{\mathcal{F}:Q,Q}\right|\right] + \sum_{i=1}^{JQ} \ln\left[1-\rho_i^2\right] \end{aligned} \tag{5-47}$$

Given these entropy measures, the mutual information between the past and future of the array output process $\{\underline{x}(n)\}$ is determined as (via Equation (5-41))

$$I(\underline{x}_{\mathcal{P}:P} \leftrightarrow \underline{x}_{\mathcal{F}:Q}) = -\sum_{i=1}^{JQ} \ln\left[1-\rho_i^2\right] \tag{5-48}$$

Mutual information is a measure of the correlation between two processes (recall that the $\{\rho_i\}$ are correlation coefficients). If the past and future processes are uncorrelated, $\rho_i = 0$ for $i = 1, 2, \ldots, JQ$, and the mutual information is zero. If the past and future processes are fully-correlated, $\rho_i = 1$ for $i = 1, 2, \ldots, JQ$, and the mutual information is infinite. Equation (5-48) combines all the information about the correlation of the two processes in a single, scalar measure. This compact, powerful, and simple relation is fundamental to dimensionality reduction.

For convenience and notational simplicity, let η represent the mutual information measure, $I(\underline{x}_{\mathcal{P}:P} \leftrightarrow \underline{x}_{\mathcal{F}:Q})$, and also define a set of variables $\{\kappa_i \mid i = 1, 2, \ldots, JQ\}$ as

$$\kappa_i = -\ln\left[1-\rho_i^2\right] \qquad i = 1, 2, \ldots, JQ \tag{5-49}$$

The $\{\kappa_i\}$ are referred to herein as the <u>information coefficients</u> for the process $\{\underline{x}(n)\}$. As demonstrated in the next section, these parameters provide a means for dimensionality reduction in the context of the MC algorithm. Similarly, they provide a means for model order reduction in the parametric, sequential implementation of the MC algorithm (Román et al., 1998).

5.2 Dimensionality Reduction Criteria

The MC algorithm reduces the dimensionality of the problem in relation to the MF approach in two ways. First, the MC algorithm operates on matrices with JQxJQ, or JQxJP, or JPxJP elements, whereas the MF algorithm requires the inverse of a JNxJN matrix. In most cases this reduction is significant because P and Q are selected such that $Q \leq P << N$. Second, further reduction is achieved by retaining only the L most significant canonical variables, with $L < JQ$.

These two dimensionality-reducing mechanisms are linked in a subtle way. Specifically, if P and Q are large-valued, then it is likely that more canonical correlations are negligible (L is small) than when P and Q are small-valued. Viewed in a different way, if many canonical correlations are negligible for the P and Q values selected, then it is likely that P and Q are too large, and a smaller-valued pair can be adopted with negligible loss of information.

Dimensionality reduction via the data residual vector dimensionality parameter, L, is the focus of this section. Such reduction is accomplished with the aid of criteria based on the canonical correlations. Criteria of this type have been discussed by Román and Davis (1993a; 1993b) in the context of model order selection for a state variable parametric model for the ground clutter in an airborne surveillance phased array radar system. Two of the criteria discussed by Román and Davis (1993a; 1993b) are of relevance herein. In addition, two criteria based on the residual covariance matrix are proposed.

Canonical Correlations Criterion. This criterion consists of examining all the canonical correlations to determine the integer i_o for which $\{\rho_i \mid i = 1, 2, \ldots, i_o\}$ are non-negligible, and $\{\rho_i \mid i = i_o+1, i_o+2, \ldots$

, JQ} are negligible (the so-called "knee-in-the-curve"). Then, L is selected equal to i_o. If the canonical correlations exhibit a continuous variation (fail to exhibit "knee-in-the-curve" behavior), then the value of i_o is selected such that a pre-set limit to the sum of the canonical correlations is met. Variations to this approach can be defined wherein the behavior of a function of the canonical correlations (square; information coefficients; etc.) is considered instead. Such is the context of the other two criteria herein.

Residual Covariance Matrix Criteria. Two distinct criteria can be defined based on the residual covariance matrix, Ω. One criterion is based on the trace of Ω, whereas the other is based on its determinant.

Consider the residual covariance matrix for the full-dimensionality configuration ($L = JQ$), and let ζ denote the trace of Ω normalized by the scalar term $2JQ$; that is,

$$\zeta = \frac{1}{2JQ} tr[\Omega] = 1 - \frac{1}{JQ} \sum_{i=1}^{JQ} \rho_i \qquad (5\text{-}50)$$

where $tr[\bullet]$ denotes the trace operator. The first criterion function is then defined as the normalized total residual variance (as measured by the trace operator) for the first L-element subset of the canonical variables; namely,

$$\zeta(L) = \frac{1}{\zeta}\left(1 - \frac{1}{L}\sum_{i=1}^{L}\rho_i\right) = \frac{1 - \frac{1}{L}\sum_{i=1}^{L}\rho_i}{1 - \frac{1}{JQ}\sum_{i=1}^{JQ}\rho_i} \qquad 1 \le L \le JQ \qquad (5\text{-}51)$$

Notice that $\zeta(1) \leq \ldots \leq \zeta(JQ\text{-}1) < \zeta(JQ) = 1$. Further, if the canonical correlations are strictly-monotonic and non-zero, $1 \geq \rho_1 > \rho_2 > \ldots > \rho_{JQ} > 0$, then $\zeta(L)$ is a monotonically-increasing function (notice that $\rho_1 = 1$ is allowed for strict monotonicity).

For the second criterion, consider again the residual covariance matrix for the full-dimensionality configuration ($L = JQ$), and let ξ represent the determinant of Ω normalized by the scalar term 2^{JQ}; that is,

$$\xi = \frac{1}{2^{JQ}} \det[\Omega] = \prod_{i=1}^{JQ} (1-\rho_i) \tag{5-52}$$

where $\det[\bullet]$ denotes the determinant operator. The second criterion function is then defined as this quantity normalized by the partial residual variance (as measured by the determinant operator) obtained with a configuration consisting of the first L-element subset of the canonical variables; namely,

$$\xi(L) = \frac{\xi}{\prod_{i=1}^{L} (1-\rho_i)} = \frac{\prod_{i=1}^{JQ} (1-\rho_i)}{\prod_{i=1}^{L} (1-\rho_i)} = \prod_{i=L+1}^{JQ} (1-\rho_i) \qquad 1 \leq L \leq JQ \tag{5-53}$$

Notice that $\xi(1) \leq \ldots \leq \xi(JQ\text{-}1) < \xi(JQ) = 1$ if all $\rho_i < 1$. Function $\xi(L)$ is a monotonically-increasing function if the canonical correlations are strictly-monotonic, non-unity, and non-zero, $1 > \rho_1 > \rho_2 > \ldots > \rho_{JQ} > 0$. This criterion is inadequate when at least one canonical correlation is unity. Such cases, however, represent pathological conditions.

For both criteria, the value of the criterion function at argument L is indicative of the portion of the total residual

variance (the variance for the full-dimensional configuration, $L = JQ$) that is retained with the first L-element subset of the canonical variables. Each criterion thus provides a measure of the degree of information retained in a configuration of dimensionality L for the residual vector. The larger the value of L, the higher the degree of information retained in the configuration. Thus, for each criterion, specification of a percentage threshold allows selection of the parameter L as the index associated with the total residual variance (as measured by either the trace or the determinant) that exceeds the pre-specified threshold.

Mutual Information Criterion. Based on the development and definitions of Section 5.1, the mutual information between the past and future of the process $\{\underline{x}(n)\}$ is the sum of the JQ information coefficients (recall Equation (5-49)),

$$\eta = \sum_{i=1}^{JQ} \kappa_i \tag{5-54}$$

A related quantity, the normalized mutual information for the first L-element subset of the canonical variables, is obtained as

$$\eta(L) = \frac{1}{\eta} \sum_{i=1}^{L} \kappa_i = \frac{\sum_{i=1}^{L} \kappa_i}{\sum_{i=1}^{JQ} \kappa_i} \tag{5-55}$$

The value of this function represents the fraction of the mutual information in the past about the future that is retained with the first L-element subset of the canonical variables. Thus, specification of a percentage threshold for the mutual information allows selection of the parameter L as the index associated with

the normalized mutual information that exceeds the pre-specified threshold.

Román et al. (1997) applied the normalized mutual information criterion to state space model identification using simulated and measured airborne phased array radar data. The effects of variations in scenario parameters such as crab angle, platform velocity, and clutter-to-noise ratio (CNR) were considered for the simulated data. The results presented therein indicate that highly-representative state variable models can be identified using low-dimensionality covariance matrices.

Remarks. In summary, another important feature of the MC algorithm is the fact that a simple and powerful criterion is used to reduce dimensionality or assess information losses. That is, either determine the number of canonical variables to keep for achieving a pre-specified level of a probabilistic measure (total variance; mutual information), or assess the loss of correlation information suffered by keeping only L canonical variables.

The impact (if any) on detection performance of the AMC due to large reductions in dimensionality remains to be assessed. Preliminary work carried out for the related PAMF-CC suggests that considerable reductions in dimensionality are possible while surpassing the performance of the AMF (Román et al., 1998).

6.0 ORTHOGONAL PROJECTION WITH CANONICAL VARIABLES (OPCV)

Application of the orthogonal projection principle to the past and future canonical variables results in an important extension to the MC formulation. This enhanced formulation is based on determination of the orthogonal projection of the space spanned by the "future" canonical variables onto the space spanned by the "past" canonical variables. Such projection is an optimal prediction of the future given the past, and the resulting prediction error vector constitutes the residual vector in the generic STAP architecture. The orthogonal projection with canonical variables (OPCV) algorithm is an application of orthogonal projection in the context of a structure wherein the array output data (past and future vectors) are represented in the canonical variables basis. As such, the OPCV for the full-dimensional residual case ($L = JQ$) is equivalent analytically to the OP. However, an important distinction between the OPCV and the OP is that the OPCV includes an optimal mechanism for dimensionality reduction (selection of $L < JQ$), which the OP lacks.

Full-Dimension Residual Configuration ($L = JQ$). As indicated in Section 4.0, an orthogonal projection in the vector spaces considered herein is equivalent to conditional expectation. Also, recall from Section 5.0 that $\underline{\upsilon}(n)$ and $\underline{\mu}(n)$ denote, respectively, the JQ-element future and JP-element past canonical variables for the array output process. Now let $\mathcal{M}$ denote the space spanned by $\underline{\mu}(n)$ (notice that $\mathcal{M} = X^-$), and let $\underline{\hat{\upsilon}}(n|\mathcal{M})$ denote the expectation of the future canonical variables conditioned on the past canonical variables. Then, for appropriately-selected integers P and Q,

$$\underline{\hat{\upsilon}}(n|\mathcal{M}) = E\left[\underline{\upsilon}(n)\underline{\mu}^H(n)\right]\left(E\left[\underline{\mu}(n)\underline{\mu}^H(n)\right]\right)^{-1}\underline{\mu}(n) \tag{6-1a}$$

$$(6\text{-}1b) \qquad \underline{\hat{\upsilon}}(n|\mathcal{M}) = R_{\upsilon\mu}\,\underline{\mu}(n) = \left[\, D_1 \;\vdots\; D_2 \,\right]\underline{\mu}(n) = \left[\, D_1 \;\vdots\; [0] \,\right]\left[\begin{array}{c} \underline{\mu}_V(n) \\ \hline \underline{\mu}_H(n) \end{array}\right]$$

$$(6\text{-}1c) \qquad \underline{\hat{\upsilon}}(n|\mathcal{M}) = D_1\,\underline{\mu}_V(n)$$

where $\underline{\mu}_V(n)$ and $\underline{\mu}_H(n)$ are JQ-element and J(P-Q)-element vectors, respectively, as defined in Equations (5-6), and where Equations (5-11), (5-12), and (5-17) have been invoked. Consider now the OPCV algorithm in the context of the generic STAP architecture (Figure 2-1). In this context, the JQ-element OPCV variables $\underline{\alpha}(n)$ and $\underline{\beta}(n)$ are defined in terms of the canonical variables as

$$(6\text{-}2) \qquad \underline{\alpha}(n) = \underline{\upsilon}(n)$$

$$(6\text{-}3) \qquad \underline{\beta}(n) = \underline{\hat{\upsilon}}(n|\mathcal{M}) = D_1\,\underline{\mu}_V(n)$$

respectively. The OPCV JQ-element data and steering residuals are generated as

$$(6\text{-}4a) \qquad \underline{\varepsilon}(n) = \underline{\alpha}(n) - \underline{\beta}(n) = \underline{\upsilon}(n) - \underline{\hat{\upsilon}}(n|\mathcal{M}) \qquad n = P, \ldots, N\text{-}Q$$

$$(6\text{-}4b) \qquad \underline{\varepsilon}(n) = \underline{\upsilon}(n) - D_1\,\underline{\mu}_V(n) = W_{MC}^H\,\underline{x}_{\mathcal{F}:Q}(n) - D_1\,V_{MC}^H\,\underline{x}_{\mathcal{P}:P}(n) \qquad n = P, \ldots, N\text{-}Q$$

$$(6\text{-}5) \qquad \underline{u}(n) = W_{MC}^H\,\underline{e}_{\mathcal{F}:Q}(n) - D_1\,V_{MC}^H\,\underline{e}_{\mathcal{P}:P}(n) \qquad n = P, \ldots, N\text{-}Q$$

respectively, and where W_{MC} and V_{MC} denote the weight matrices for the MC algorithm (Equations (5-19) and (5-20), respectively). From these expressions it follows that the OPCV algorithm weight matrices W and V are defined as

$$(6\text{-}6) \qquad W = W_{MC} = \mathcal{R}_{\mathcal{F}:Q,Q}^{-1/2}\,T_1$$

(6-7) $$V = V_{MC} D_1 = \mathcal{R}_{\mathcal{T}:P,P}^{-1/2} T_{2V} D_1$$

respectively. Next, the OPCV residual covariance matrix is obtained from Equations (2-13b), (6-6), and (6-7), as

(6-8) $$\Omega = I_{JQ} - D_1^2 = \begin{bmatrix} 1-\rho_1^2 & 0 & \cdots & 0 \\ 0 & 1-\rho_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1-\rho_{JQ}^2 \end{bmatrix}$$

Covariance matrix Ω generated according to Equation (6-8) is the model residual covariance referred to in Remark 2.2.4.

The normalized and uncorrelated (temporally and spatially) JQ-element data and steering residuals for the OPCV algorithm are

(6-9) $$\underline{v}(n) = C^H \underline{\varepsilon}(n) \qquad n = P, P+1, \ldots, N-Q$$

(6-10) $$\underline{s}(n) = C^H \underline{u}(n) \qquad n = P, P+1, \ldots, N-Q$$

The JQxJQ weight matrix C is a diagonal matrix specified as

(6-11) $$C^H = B^{-1} = \Omega^{-1/2}$$

where $\Omega^{-1/2}$ is the inverse square-root matrix of the data residual diagonal covariance matrix. The detection test statistic is generated using Equation (2-17). This completes the formulas for the $L = JQ$ configuration.

Reduced-Dimension Residual Configuration ($L < JQ$). Cases wherein the data residual vector dimensionality is $L < JQ$ arise due to processing and/or performance requirements, as stated in Section

5.0. In the OPCV algorithm formulation for such cases an orthogonal projection is desired from a subset of the future canonical variables onto the past canonical variables. Specifically, for $L < JQ$, only the variables $\{\upsilon_i(n) \mid i = 1, 2, \ldots, L\}$ and $\{\mu_i(n) \mid i = 1, 2, \ldots, L\}$, and the correlation coefficients $\{\rho_i \mid i = 1, 2, \ldots, L\}$ are retained. Thus, the formulas for the calculation of the weight matrices must be modified. Determination of the value of L is considered in Section 6.2.

For the case $L < JQ$ it is convenient to define an $LxJQ$ selector matrix S_L as

$$S_L = \left[I_L \quad [0]_{L,JQ-L} \right] \tag{6-12}$$

Then, the sub-vector of $\underline{\upsilon}(n)$ formed from the first L future canonical variables is denoted as

$$\underline{\upsilon}_L(n) = S_L \underline{\upsilon}(n) = \begin{bmatrix} \upsilon_1(n) \\ \vdots \\ \upsilon_L(n) \end{bmatrix} \qquad 1 \le L < JQ \tag{6-13}$$

Now let $\hat{\underline{\upsilon}}_L(n|\mathcal{M})$ denote the expectation of the first L future canonical variables conditioned on the past canonical variables. Then, for appropriately-selected integers P and Q,

$$\hat{\underline{\upsilon}}_L(n|\mathcal{M}) = E\left[\underline{\upsilon}_L(n)\underline{\mu}^H(n)\right]\left(E\left[\underline{\mu}(n)\underline{\mu}^H(n)\right]\right)^{-1}\underline{\mu}(n) = S_L R_{\upsilon\mu}\,\underline{\mu}(n) \tag{6-14a}$$

$$\hat{\underline{\upsilon}}_L(n|\mathcal{M}) = \left[I_L \quad [0]_{L,JQ-L} \right] \begin{bmatrix} D_{11} & [0]_{L,JQ-L} & [0]_{L,J(P-Q)} \\ [0]_{JQ-L,L} & D_{12} & [0]_{JQ-L,J(P-Q)} \end{bmatrix} \underline{\mu}(n) \tag{6-14b}$$

$$\hat{\underline{\upsilon}}_L(n|\mathcal{M}) = D_{11}\,\underline{\mu}_L(n) \tag{6-14c}$$

where Equations (5-11), (5-12), (5-17), (5-34), and (6-12) have been invoked, and where $\underline{\mu}_L(n)$ is a vector of the first L past canonical correlations,

$$\underline{\mu}_L(n) = \begin{bmatrix} \mu_1(n) \\ \vdots \\ \mu_L(n) \end{bmatrix} \qquad 1 \le L < JQ \tag{6-15}$$

Given these relations, the L-element OPCV variables $\underline{\alpha}(n)$ and $\underline{\beta}(n)$ are defined as

$$\underline{\alpha}(n) = \underline{\upsilon}_L(n) \tag{6-16}$$

$$\underline{\beta}(n) = \underline{\hat{\upsilon}}_L(n|\mathcal{M}) = D_{11}\underline{\mu}_L(n) \tag{6-17}$$

respectively, and the OPCV L-element data and steering residuals are generated as

$$\underline{\varepsilon}(n) = \underline{\alpha}(n) - \underline{\beta}(n) = \underline{\upsilon}_L(n) - \underline{\hat{\upsilon}}_L(n|\mathcal{M}) \qquad n = P, \ldots, N\text{-}Q \tag{6-18a}$$

$$\underline{\varepsilon}(n) = \underline{\upsilon}_L(n) - D_{11}\underline{\mu}_L(n) = W_{MCL}^H \underline{x}_{\mathcal{F}:Q}(n) - D_{11} V_{MCL}^H \underline{x}_{\mathcal{P}:P}(n) \qquad n = P, \ldots, N\text{-}Q \tag{6-18b}$$

$$\underline{u}(n) = W_{MCL}^H \underline{e}_{\mathcal{F}:Q}(n) - D_{11} V_{MCL}^H \underline{e}_{\mathcal{P}:P}(n) \qquad n = P, \ldots, N\text{-}Q \tag{6-19}$$

respectively. In these expressions, W_{MCL} and V_{MCL} denote the JQxL and JPxL, respectively, weight matrices for the reduced-dimensionality MC algorithm (Equations (5-31) and (5-32), respectively). It follows that the reduced-dimensionality OPCV algorithm weight matrices W and V are defined as

$$W = W_{MCL} = \mathcal{R}_{\mathcal{F}:Q,Q}^{-1/2} T_{1A} \tag{6-20}$$

(6-21) $$V = V_{MCL} D_{11} = \mathcal{R}_{\mathcal{P}:P,P}^{-1/2} T_{2A} D_{11}$$

respectively. These weight matrices are substituted in Equation (2-13) to obtain the LxL OPCV residual covariance matrix as

(6-22) $$\Omega = I_L - D_{11}^2 = \begin{bmatrix} 1-\rho_1^2 & 0 & \cdots & 0 \\ 0 & 1-\rho_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1-\rho_L^2 \end{bmatrix}$$

Weight matrix C is calculated using Equation (6-11), with Ω as in Equation (6-22). The detection test statistic is generated also via Equation (2-17). This completes the formulas for $L < JQ$.

The OPCV algorithm admits an adaptive formulation also, which is referred to as the adaptive OPCV (AOPCV). In the AOPCV each unavailable block covariance matrix is replaced by its ML estimate, obtained as an ensemble average over the secondary data. As is the case for the other two algorithms, the ML estimate of the residual covariance matrix (obtained as an ensemble average over the residuals for the secondary data set) is preferred over the model residual covariance estimate (generated via either Equation (6-8) or Equation (6-22) using the estimated canonical correlations).

Since the OPCV algorithm is based on the canonical variables basis, the comments made for the MC algorithm in Section 5.0 regarding issues such as adaptability, dimensionality reduction, implementation aspects, computational issues, and algorithm structure issues are valid also for the OPCV. However, issues unique to the OPCV are discussed in the sections that follow.

6.1 Optimality Of The OPCV

Yohai and Garcia Ben (1980) have demonstrated that the canonical variables constitute the optimal solution to a reduced-dimensionality linear prediction problem. In the context of the notation and formulation adopted herein, the problem addressed in (Yohai and Garcia Ben, 1980) is stated as follows. Given the JQ-element future vector $\underline{x}_{\mathcal{F}:Q}(n)$ and the JP-element past vector $\underline{x}_{\mathcal{P}:P}(n)$, then, for $L < JQ$, determine a JPxL linear transformation T on the past process of the form

$$\underline{z}(n) = T^H \underline{x}_{\mathcal{P}:P}(n) \tag{6-23}$$

such that an orthogonal projection of the space spanned by $\underline{x}_{\mathcal{F}:Q}(n)$ onto the space spanned by $\underline{z}(n)$, denoted as $\mathcal{Z}$, minimizes the criterion

$$J(T) = \det\left[E\left[\left[\underline{x}_{\mathcal{F}:Q}(n) - \hat{\underline{x}}_{\mathcal{F}:Q}(n|\mathcal{Z})\right]\left[\underline{x}_{\mathcal{F}:Q}(n) - \hat{\underline{x}}_{\mathcal{F}:Q}(n|\mathcal{Z})\right]^H \right] \right] \tag{6-24}$$

where $\hat{\underline{x}}_{\mathcal{F}:Q}(n|\mathcal{Z})$ is the orthogonal projection,

$$\hat{\underline{x}}_{\mathcal{F}:Q}(n|\mathcal{Z}) = E\left[\underline{x}_{\mathcal{F}:Q}(n)\underline{z}^H(n)\right]\left(E\left[\underline{z}(n)\underline{z}^H(n)\right]\right)^{-1}\underline{z}(n) \tag{6-25}$$

Notice that $\underline{z}(n)$ as defined in Equation (6-23) is an L-element vector, which implies that $\mathcal{Z} \subset \mathcal{X}^-$. This completes the statement of the problem formulated by Yohai and Garcia Ben (1980).

The optimization problem stated above admits as one solution that matrix T be selected as the JPxL weight matrix V_{MCL}, the past canonical variable transformation for the case $L < JQ$ (see Equation (5-32)), and the transformed variables are the first L past

canonical variables. Specifically, one optimal solution, denoted by superscript "o", is

$$T^o = V_{MCL} = \mathcal{R}_{\mathcal{P}:P,P}^{-1/2} T_{2A} \tag{6-26}$$

$$\underline{z}^o(n) = \underline{\mu}_L(n) = V_{MCL}^H \underline{x}_{\mathcal{P}:P}(n) = T_{2A}^H \mathcal{R}_{\mathcal{P}:P,P}^{-1/2} \underline{x}_{\mathcal{P}:P}(n) \tag{6-27}$$

Other equivalent optimal solutions are generated by selecting the transformation matrix to be of the form $T = V_{MCL}F$, where F is a full-rank LxL matrix. More specifically, matrix F is any general full-rank matrix if $\rho_L > \rho_{L+1}$, and is a linear function of the singular vectors for repeated canonical correlations if $\rho_{q+1} = \cdots = \rho_L$ for some q such that $1 < q < L$. The minimum value of the criterion $J(T)$ (Equation 6-24)), denoted by $J^o(T)$, is

$$J^o(T) = \det\left[\mathcal{R}_{\mathcal{F}:Q,Q}\right] \prod_{i=1}^{L} \left(1-\rho_i^2\right) \tag{6-28}$$

This completes the solution presented by Yohai and Garcia Ben (1980), with minor modifications to fit the context herein.

In summary, two important points are established by the Yohai-Garcia Ben result. First, the past canonical variables constitute an optimal basis, with respect to Criterion (6-24), to represent an L-dimensional proper sub-space of $\mathcal{X}^-$. Second, the solution provided by the past canonical variables has the simplest form, since all other solutions involve an additional matrix factor, F.

The Yohai-Garcia Ben result is significant in the context of the reduced-dimensionality OPCV because it states that the selection of the variables $\underline{\alpha}(n)$ and $\underline{\beta}(n)$ as in Equations (6-16) and (6-17), respectively, is optimal according to Criterion (6-24).

Specifically, with $\underline{x}_{\mathcal{I}:Q}(n)$ expressed as $\underline{\upsilon}_L(n)$, and $\underline{x}_{\mathcal{P}:P}(n)$ expressed as $\underline{\mu}_L(n)$, the Yohai-Garcia Ben result states that $\underline{z}(n) = \underline{\mu}_L(n)$, and Criterion (6-24) becomes $J(T) = \det[\Omega]$ since $\Omega = E\left[\underline{\varepsilon}(n)\underline{\varepsilon}^H(n)\right]$ when the data residual $\underline{\varepsilon}(n)$ is generated via Equation (6-18). With Ω given as in Equation (6-22), the optimal criterion value is then

$$J^0(T) = \prod_{i=1}^{L}\left(1-\rho_i^2\right) \qquad \text{FOR OPCV ALGORITHM} \tag{6-29}$$

Equation (6-29) results from Equation (6-28) when $\underline{x}_{\mathcal{I}:Q}(n)$ is expressed as $\underline{\upsilon}_L(n)$ (that is, in canonical variables basis) because the covariance matrix of $\underline{\upsilon}_L(n)$ is I_L.

The optimal criterion value in Equation (6-29) can serve as a criterion for reducing the dimensionality of the data residual vector in the OPCV algorithm (selecting L), which is one of the two ways in which dimensionality reductions can be attained with the MC and OPCV algorithms (see Section 5.2). It turns out that one of the two residual covariance matrix criteria types proposed in Section 5.2 leads to an identical expression. This issue is explored next, in Section 6.2.

6.2 Dimensionality Reduction Criteria

Since the OPCV algorithm is based on the MC algorithm, the issues and criteria for dimensionality reduction discussed in Section 5.2 are applicable also to the OPCV. The exception involves the criteria based on the residual covariance matrix, Ω, because each algorithm has a different residual covariance matrix. Thus, only the residual covariance matrix criteria are discussed herein. As in Section 5.2, two distinct criteria are defined based on Ω: one based on its trace, the other based on its determinant. Both criteria have the same analytical form as their

MC counterparts. A key distinction, however, is that the MC criteria are functions of ρ_i, whereas the OPCV criteria are functions of ρ_i^2.

Residual Covariance Matrix Trace Criterion. Consider the residual covariance matrix for the full-dimensionality configuration ($L = JQ$), and let ζ denote the trace of Ω normalized by the scalar term JQ; that is,

$$\zeta = \frac{1}{JQ}\,\mathrm{tr}[\Omega] = 1 - \frac{1}{JQ}\sum_{i=1}^{JQ}\rho_i^2 \tag{6-30}$$

The first criterion function is defined as the normalized total residual variance (as measured by the trace operator) for the first L-element subset of the canonical variables; namely,

$$\zeta(L) = \frac{1}{\zeta}\left(1 - \frac{1}{L}\sum_{i=1}^{L}\rho_i^2\right) = \frac{1 - \frac{1}{L}\sum_{i=1}^{L}\rho_i^2}{1 - \frac{1}{JQ}\sum_{i=1}^{JQ}\rho_i^2} \qquad 1 \le L \le JQ \tag{6-31}$$

Notice that $\zeta(1) \le \ldots \le \zeta(JQ\text{-}1) < \zeta(JQ) = 1$. Further, if the canonical correlations are strictly-monotonic and non-zero, $1 \ge \rho_1 > \rho_2 > \ldots > \rho_{JQ} > 0$, then $\zeta(L)$ is a monotonically-increasing function (notice that $\rho_1 = 1$ is allowed for strict monotonicity).

Residual Covariance Matrix Determinant Criterion. As before, consider the residual covariance matrix for the full-dimensionality configuration ($L = JQ$), and let ξ represent the determinant of Ω,

$$\xi = \det[\Omega] = \prod_{i=1}^{JQ}\left(1-\rho_i^2\right) \tag{6-32}$$

The second criterion function is defined as this quantity normalized by the partial residual variance (as measured by the determinant operator) obtained with a configuration consisting of the first L-element subset of the canonical variables; namely,

$$\xi(L) = \frac{\xi}{\prod_{i=1}^{L}\left(1-\rho_i^2\right)} = \frac{\prod_{i=1}^{JQ}\left(1-\rho_i^2\right)}{\prod_{i=1}^{L}\left(1-\rho_i^2\right)} = \prod_{i=L+1}^{JQ}\left(1-\rho_i^2\right) \qquad 1 \le L \le JQ \tag{6-33}$$

Notice that $\xi(1) \le \ldots \le \xi(JQ-1) < \xi(JQ) = 1$ if all $\rho_i < 1$. Function $\xi(L)$ is a monotonically-increasing function if the canonical correlations are strictly-monotonic, non-unity, and non-zero, $1 > \rho_1 > \rho_2 > \ldots > \rho_{JQ} > 0$. As for its MC counterpart, this criterion is inadequate when at least one canonical correlation is unity; however, the occurrence of canonical correlations of exactly unity value is an unlikely event due to uncorrelated noise and computational errors. Notice that the un-normalized criterion ξ of Equation (6-32) with JQ replaced by L is identical to the minimum value of the Yohai-Garcia Ben criterion for the case when the array output data is in canonical variables basis, Equation (6-29).

Remarks. For both criteria, the value of the criterion function at argument L is indicative of the portion of the total residual variance (the variance for the full-dimensional configuration, L = JQ) that is retained with the first L-element subset of the canonical variables. Each criterion thus provides a measure of the degree of information retained in a configuration of dimensionality L for the residual vector. The larger the value of

L, the higher the degree of information retained in the configuration. Thus, for each criterion, specification of a percentage threshold allows selection of the parameter L as the index associated with the total residual variance (as measured by either the trace or the determinant) that exceeds the pre-specified threshold.

6.3 Residual Covariance Matrix Comparison: MC And OPCV

The OPCV residual is less than or equal to the MC residual for all configurations ($L \leq JQ$). This is expected since the orthogonal projection is the best linear predictor of a JQ-element vector $\underline{x}_{\mathcal{F}:Q}(n)$ based on a JP-element vector $\underline{x}_{\mathcal{P}:P}(n)$ for a variety of criteria, including minimization of each of the following two functions of the prediction error, $\underline{\varepsilon}(n)$,

$$(6\text{-}34a)\quad J_2(T) = E\left[\|\underline{\varepsilon}(n)\|_2^2\right] = E\left[\left\|\underline{x}_{\mathcal{F}:Q}(n) - \hat{\underline{x}}_{\mathcal{F}:Q}(n|\mathcal{X}^-)\right\|_2^2\right] = E\left[\left\|\underline{x}_{\mathcal{F}:Q}(n) - T^H\underline{x}_{\mathcal{P}:P}(n)\right\|_2^2\right]$$

$$(6\text{-}34b)\quad J_2(T) = \mathrm{tr}\left[E\left[\underline{\varepsilon}(n)\underline{\varepsilon}^H(n)\right]\right] = \mathrm{tr}\left[E\left[\left[\underline{x}_{\mathcal{F}:Q}(n) - T^H\underline{x}_{\mathcal{P}:P}(n)\right]\left[\underline{x}_{\mathcal{F}:Q}(n) - T^H\underline{x}_{\mathcal{P}:P}(n)\right]^H\right]\right]$$

$$(6\text{-}35a)\quad J_D(T) = \det\left[E\left[\underline{\varepsilon}(n)\underline{\varepsilon}^H(n)\right]\right]$$

$$(6\text{-}35b)\quad J_D(T) = \det\left[E\left[\left[\underline{x}_{\mathcal{F}:Q}(n) - \hat{\underline{x}}_{\mathcal{F}:Q}(n|\mathcal{X}^-)\right]\left[\underline{x}_{\mathcal{F}:Q}(n) - \hat{\underline{x}}_{\mathcal{F}:Q}(n|\mathcal{X}^-)\right]^H\right]\right]$$

$$(6\text{-}35c)\quad J_D(T) = \det\left[E\left[\left[\underline{x}_{\mathcal{F}:Q}(n) - T^H\underline{x}_{\mathcal{P}:P}(n)\right]\left[\underline{x}_{\mathcal{F}:Q}(n) - T^H\underline{x}_{\mathcal{P}:P}(n)\right]^H\right]\right]$$

where T denotes a JPxJQ linear transformation, and $\|\bullet\|_2$ denotes the vector Euclidean norm (or 2-norm). The orthogonal projection solution is

(6-36) $$T^H = E\left[\underline{x}_{\mathcal{T}:Q}(n)\underline{x}^H_{\mathcal{P}:P}(n)\right]\left(E\left[\underline{x}_{\mathcal{P}:P}(n)\underline{x}^H_{\mathcal{P}:P}(n)\right]\right)^{-1}$$

Minimality of the OPCV algorithm in relation to the MC algorithm is demonstrated for each of these two criteria.

Consider first the Euclidean norm (covariance matrix trace) criterion, $J_2(T)$, for the general case $L \le JQ$. For this criterion it is necessary to demonstrate that $tr[\Omega_{MC}] \ge tr[\Omega_{OPCV}]$. This inequality is true if the following condition is satisfied,

(6-37) $$2L - 2\sum_{i=1}^{L}\rho_i \ge L - \sum_{i=1}^{L}\rho_i^2 \qquad 1 \le L \le JQ$$

Straightforward manipulation of Inequality (6-37) leads to the expression

(6-38) $$\sum_{i=1}^{L}\left(\rho_i^2 - 2\rho_i + 1\right) \ge 0 \qquad 1 \le L \le JQ$$

Inequality (6-38) is satisfied if $\rho_i^2 - 2\rho_i + 1 \ge 0$ is satisfied for each value of $i = 1, 2, \ldots, L$. The left-hand-side of this expression is a quadratic polynomial with double root at $\rho_i = 1$, with value 1 at $\rho_i = 0$, and positive-valued over the domain of definition for the canonical correlations: $0 \le \rho_i \le 1$ for $i = 1, 2, \ldots, L$. Thus, condition $tr[\Omega_{MC}] \ge tr[\Omega_{OPCV}]$ is indeed true.

Consider next the covariance matrix determinant criterion, $J_D(T)$, for the general case $L \le JQ$. For this criterion it is necessary to demonstrate that $det[\Omega_{MC}] \ge det[\Omega_{OPCV}]$. This inequality is true if the following condition is satisfied,

$$\prod_{i=1}^{L} 2(1-\rho_i) \geq \prod_{i=1}^{L} \left(1-\rho_i^2\right) \qquad 1 \leq L \leq JQ \tag{6-39}$$

Straightforward manipulation of Inequality (6-39) leads to this equivalent expression

$$\prod_{i=1}^{L} \frac{\rho_i + 1}{2} \leq 1 \qquad 1 \leq L \leq JQ \tag{6-40}$$

Inequality (6-40) is satisfied if $\rho_i \leq 1$ is satisfied for each value of $i = 1, 2, \ldots, L$, which is satisfied indeed over the domain of definition for the canonical correlations. Thus, condition $\det[\Omega_{MC}] \geq \det[\Omega_{OPCV}]$ is true.

In summary, the OPCV algorithm residual is less than or equal to the MC algorithm residual for all configurations ($L \leq JQ$), and according to both optimization criteria.

An important point to notice is that the MC algorithm could out-perform the OPCV algorithm due to software implementation and computational accuracy issues, even though it is less optimal in theory. For example, if the numerical stability of the estimates of the canonical correlations is poor, then the OPCV weight matrix V computed via Equation (6-7) is less accurate than the MC weight matrix V computed via Equation (5-20). This issue should be pursued further via simulation-based analyses.

6.4 Past-Only OPCV Configuration

The Yohai-Garcia Ben result discussed in Section 6.1 suggests a variation of the OPCV algorithm in which only the past block vector is transformed onto the canonical variables basis. In such a configuration, referred to herein as the Past-Only OPCV

configuration, only the past canonical transformation matrix, V_{MC}, is generated and applied to the past block vector, $\underline{x}_{\mathcal{P}:P}(n)$. Thus, the weight matrices W and V in Figure 2-1 for the Past-Only OPCV configuration of reduced dimensionality for the residual vector ($L < JQ$) are

$$W = I_{JQ} \tag{6-41}$$

$$V = V_{MCL} V_{MCL}^H \mathcal{R}_{\mathcal{F}:Q,\mathcal{P}:P}^H = \mathcal{R}_{\mathcal{P}:P,P}^{-1/2} T_{2A} T_{2A}^H \mathcal{R}_{\mathcal{P}:P,P}^{-1/2} \mathcal{R}_{\mathcal{F}:Q,\mathcal{P}:P}^H \tag{6-42}$$

respectively. In these expressions, the JPxL weight matrix V_{MCL} is as determined via Equation (5-32). Comparing these weight matrices with the corresponding ones for the OPCV configuration (Equations (6-20) and (6-21)), notice that weight matrix V has a simpler form in the OPCV configuration, whereas the reverse is true for weight matrix W.

Consider now the covariance matrix of the data residual. Using Equation (2-13), Ω is obtained as

$$\Omega = \mathcal{R}_{\mathcal{F}:Q,Q} - \mathcal{R}_{\mathcal{F}:Q,\mathcal{P}:P} \mathcal{R}_{\mathcal{P}:P,P}^{-1/2} T_{2A} T_{2A}^H \mathcal{R}_{\mathcal{P}:P,P}^{-1/2} \mathcal{R}_{\mathcal{F}:Q,\mathcal{P}:P}^H \tag{6-43}$$

Straightforward algebraic manipulations result in the following equivalent expression for Ω,

$$\Omega = \mathcal{R}_{\mathcal{F}:Q,Q}^{1/2} T_1 \begin{bmatrix} I_L - D_{11}^2 & [0] \\ [0] & I_{JQ-L} \end{bmatrix} T_1^H \mathcal{R}_{\mathcal{F}:Q,Q}^{1/2} \tag{6-44}$$

Equation (6-44) leads to a simple form for the determinant of Ω; namely,

$$\det[\Omega] = \det\left[\mathcal{R}_{\mathcal{F}:Q,Q}\right] \prod_{i=1}^{L}\left(1-\rho_i^2\right) \qquad (6\text{-}45)$$

This expression is identical to the optimal value for the criterion of the Yohai-Garcia Ben result discussed in Section 6.1, which is expected since the criterion expression (Equation (6-24)) becomes $J(T) = \det[\Omega]$ for the Past-Only OPCV configuration.

Weight matrix C and the detection statistic are generated via formulas analogous to those stated in Section 6.0 for the reduced-dimension residual configuration.

The Past-Only OPCV may have advantages over the OPCV in terms of numerical accuracy and/or computational load because only one block vector is transformed. Generation of the model residual covariance for the Past-Only OPCV requires more computations than for the OPCV; however, generation of the residual covariance is an irrelevant issue in the context of the generic architecture because the approach preferred for either algorithm is to use the sample residual covariance (which involves the same procedure for either algorithm). Further comparisons between OPCV and the Past-Only OPCV remains as a task for future work.

6.5 OP And OPCV ($L = JQ$ Configuration) Equivalence

Consider the OPCV algorithm for the full-dimension residual configuration; that is, with $L = JQ$. For this case the OPCV is equivalent to the OP with the array output data represented in the canonical variables basis, as shown next.

In the context of the generic STAP architecture in Figure 2-1, assume the array output data is represented in the canonical variables basis. That is, the JQ-element block vector $\underline{x}_{\mathcal{F}:Q}(n)$ is

replaced by the JQ-element vector $\underline{\upsilon}(n)$, and the JP-element block vector $\underline{x}_{\mathcal{P}:P}(n)$ is replaced by the JP-element vector $\underline{\mu}(n)$. Under these conditions, the OP weight matrix W_{OP} is the JQxJQ identity, as before (Equation (4-6)),

$$W_{OP} = I_{JQ} \tag{6-46}$$

and the JPxJQ OP weight matrix V_{OP} is (via Equation (4-7)),

$$V_{OP}^{H} = R_{\upsilon\mu} = [D_1 \,\vdots\, [0]] \tag{6-47}$$

where D_1 is as defined previously. The JQxJQ OP data residual covariance matrix is obtained as (via Equation (4-13)),

$$\Omega_{OP} = I_{JQ} - D_1^2 = \begin{bmatrix} 1-\rho_1^2 & 0 & \cdots & 0 \\ 0 & 1-\rho_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1-\rho_{JQ}^2 \end{bmatrix} \tag{6-48}$$

which is identical to the OPCV data residual covariance matrix, Equation (6-8). Thus, the OP with the array output data represented in the canonical variables basis is equivalent to the OPCV for the full-dimension residual configuration, $L = JQ$.

The advantages of the OPCV over the OP include that the OPCV structure allows for dimensionality reduction in the residual vector, and that one approach to dimensionality reduction with the OPCV is based on the optimal solution to a well-posed minimization problem, as shown in Sections 6.1 and 6.2. In addition, the OPCV provides unique insight into the structure of orthogonal projections in the context of STAP applications.

7.0 TWO-DIMENSIONAL ARLS AND PAMF DETECTION FOR STAP

The auto-regressive (AR) subclass of 2-D, linear, shift-invariant parametric time series models was selected to represent the channel output under the null hypothesis, with its attendant increase in modeling degrees-of-freedom (independent dynamic and static modeling capability along each axis) over the 1-D vector representation. The complete time series model class (MA, AR, and ARMA models) was considered, but the AR subclass and the 2DARLS model identification algorithm in particular provided major computational, software implementation, and modeling performance advantages. Since each member of the AR model subclass is invertible, the 2-D AR least-squares (2DARLS) algorithm inherently generates the whitening filter used in the 2-D PAMF detection architecture presented herein.

7.1 Two-Dimensional ARLS Model Identification

The channel output vector sequence $\{\underline{x}(n)\}$ can be viewed as a scalar 2-D sequence (Dudgeon and Mersereau, 1984). Let channel J be the temporal and spatial reference for the array, and define the following association,

$$x_{J-k}(n) = x(n,k) \qquad 0 \le n \le N-1;\ 0 \le k \le J-1 \tag{7-1}$$

The process $\{x(n,k)\}$ has a scalar 2-D power spectrum denoted as $\{S_{xx}(f_d,f_s)\}$, and a scalar 2-D auto-covariance sequence (ACS) denoted as $\{r_{xx}(m,\ell)\}$, where (m,ℓ) is the lag pair corresponding to the normalized frequency pair (f_d,f_s).

The association defined by Equation (7-1) was adopted for Phase I and its usage is continued in Phase II because it is consistent with the relation between multichannel 1-D and scalar 2-D systems demonstrated by Therrien (1981). Alternatives to

Equation (7-1) can be defined, such as $\{x_{k+1}(n) = x(n,k) \mid n = 0, 1, \ldots, N\text{-}1; k = 0, 1, \ldots, J\text{-}1\}$. This specific alternative definition corresponds to the MATLAB software default array definition convention.

Model-based detection as considered herein is predicated on the representation of the channel output process $\{x(n,k)\}$ as the output of a 2-D time series model driven by white noise. Furthermore, the time series model output is corrupted by additive white noise. To be precise, such a representation is approximate in the case of radar data. Nevertheless, in practice time series models have been shown to provide a good fit to radar data. Thus, the 2-D process $\{x(n,k)\}$ is assumed to be represented as

$$x(n,k) = y(n,k) + w(n,k) \tag{7-2}$$

where $y(n,k)$ is the output of a linear, shift-invariant, 2-D AR time series model, and $w(n,k)$ is a 2-D, zero-mean, Gaussian-distributed, white noise process. Processes $\{y(n,k)\}$ and $\{w(n,k)\}$ are independent.

Consider now the 2-D AR representation for the process $\{y(n,k)\}$. A 2-D $AR(P_d,P_s)$ process $\{y(n,k)\}$ is defined as

$$y(n,k) = - \sum_{\substack{i_d=1 \\ (i_d,i_s)\neq(0,0)}}^{P_d} \sum_{i_s=1}^{P_s} a^*(i_d,i_s)\, y(n-i_d,k-i_s) + u(n,k) \tag{7-3}$$

where the input $\{u(n,k)\}$ is a 2-D, zero-mean, Gaussian-distributed, white noise process with variance σ_u^2, $\{a(i_d,i_s) \mid i_d = 0, 1, \ldots, P_d; i_s = 0, 1, \ldots, P_s; (i_d,i_s) \neq (0,0)\}$ are complex-valued, constant coefficients referred to as the AR parameters, and P_d, P_s are the model order parameters along the time (Doppler) and space dimensions, respectively. The transfer function associated with model (7-3) is obtained as the

2-D Z-transform of Equation (7-3). Specifically, the transfer function is a 2-D, complex-valued, rational function of the form

$$T(z_d,z_s)=\frac{b_o}{A(z_d,z_s)}=\frac{b_o}{\sum_{p_d=0}^{P_d}\sum_{p_s=0}^{P_s}a^*(p_d,p_s)z_d^{-p_d}z_s^{-p_s}} \tag{7-4}$$

$$a(0,0)=1 \tag{7-5}$$

with complex-valued variables z_d and z_s, where $A(z_d,z_s)$ is the AR scalar 2-D polynomial, and b_o is a real-valued scalar coefficient. As indicated in Equation (7-5), the leading coefficient of $A(z_d,z_s)$ is unity. This follows from Equation (7-3), and is the 2-D version of a monic polynomial in 1-D. The 2-D frequency response of model (7-3), denoted herein as $T(f_d,f_s)$, is obtained by restricting the complex variables z_d and z_s to the unit surface in the complex 2-D plane,

$$z_d=e^{j2\pi f_d} \tag{7-6}$$

$$z_s=e^{j2\pi f_s} \tag{7-7}$$

with f_d and f_s the normalized temporal (Doppler) and spatial frequencies, respectively.

In the context of 2-D PAMF processing for surveillance phased array radar systems, the total interference (jamming and ground clutter) is represented by a process $\{y(n,k)\}$ of the form (7-3), and the receiver noise is represented by a white noise process $\{w(n,k)\}$ as defined above. Thus, it is required to identify model order and coefficient parameters, (P_d,P_s) and $\{a(i_d,i_s)\}$, respectively, given the secondary data. Having identified the model parameters which represent the channel output process, the associated 2-D whitening (inverse) filter is available directly. Specifically, given the

channel output 2-D sequence, $\{x(n,k)\}$, the whitening filter residual sequence, $\{\varepsilon(n,k)\}$, is obtained as

$$\varepsilon(n,k) = \sum_{i_d=0}^{P_d} \sum_{i_s=0}^{P_s} a^*(i_d,i_s)\, x(n-i_d,k-i_s) \tag{7-8}$$

with $a(0,0) = 1$. Notice that the whitening filter is a 2-D MA system, with $\{a(i_d,i_s)\}$ as the MA coefficients (an MA system is the system inverse of an AR system, and vice-versa).

A 2-D model of the form in Equation (7-3) is referred to as a first-quadrant system since only values in the first quadrant of the output 2-D plane (except for initial conditions) are used to generate the system output. Model (7-3) is causal, in loose analogy with the 1-D case, since only past outputs and present inputs are used to generate the present output. Similarly, its system inverse in Equation (7-8) is causal also. In the phased array radar space-time problem causality along the time axis is an inherent feature of the channel output. In contrast, the issue of causality along the spatial axis appears ambiguous because all channels generate an output at each temporal sampling instant. The ambiguity is removed by considering the phase reference point to be at array element $k = 0$. Model (7-3) is also recursively computable (Dudgeon and Mersereau, 1984), which simplifies hardware implementation. Other region of support options, such as the non-symmetric half plane (NSHP), are of interest and should be considered in future efforts.

Model (7-3) is causal and causally-invertible. Thus, if the noise $\{w(n,k)\}$ in Equation (7-2) is zero and model (7-3) represents the channel output exactly, then $\{x(n,k)\} = \{y(n,k)\}$ (see Equation (7-2)) and the residual $\{\varepsilon(n,k)\}$ of $\{x(n,k)\}$ is a true innovations sequence. For any other conditions, the filter residual $\{\varepsilon(n,k)\}$ approximates an innovations.

Two-dimensional time series models have several features distinct from their 1-D counterparts. The three distinctions most relevant to the problem considered herein are summarized next. The first key distinction is that most 2-D polynomials are not factorable. Thus, a polynomial $A(z_d, z_s)$ identified by an algorithm is likely to be unfactorizable. This complicates key issues such as stability determination. In 2-D systems, poles and zeros can occur as functions rather than as an isolated point (Dudgeon and Mersereau, 1984), and a 2-D system is stable if the loci of the pole is inside the unit circle. It is important to note that all cases modeled in Phase II using the 2DARLS algorithm resulted in a stable 2-D model, which is also true for the 1-D multichannel cases considered. The second major distinction is that 2-D models offer more dynamic as well as static modeling degrees-of-freedom. Recall that a 1-D vector system has isolated (single-point) poles and zeros only for the temporal dimension (a 1-D multichannel AR system does have system zeros), whereas the 2-D AR system in Equation (7-3) has poles for both time and space, and those poles are generalized into loci rather than isolated points (as stated above). The space dimension in a 1-D vector AR system is non-dynamic (has no poles) and with fixed value J (in the context considered herein), whereas in a 2-D scalar AR system the space dimension is dynamic and with model order P_s that can be different from J. When $P_s < J$, the dimensionality of the LS equations to be solved is smaller for 2-D scalar AR systems than for 1-D vector AR systems; however, the increased dimensionality of the LS equations in a 1-D multichannel AR system (when $P_s < J$) does not result in an increase in dynamic degrees-of-freedom (number of poles). The third distinction of relevance is that causality and region-of-support issues offer various alternatives for 2-D systems, in contrast with a single option for 1-D systems. Region-of-support options must be considered keeping in mind that causality is to be preserved if the innovations property is desired.

Estimation of the parameters for a 2-D scalar time series models has been addressed by several authors (see [Marple, 1987], [Dudgeon and Mersereau, 1984], [Therrien, 1986] and the references therein). For AR models most algorithms require generation of the ACS of the process, and the algorithms are extensions of the 1-D case. In contrast with the 1-D case, for 2-D systems the maximum entropy method (MEM) is not equivalent to the AR method. Further, the MEM parameters are obtained via optimization procedures, with their attendant convergence and other such difficulties. Most 2-D model identification algorithms are formulated in a stochastic framework and require the ACS (in practice, only an estimate of the ACS is available). One exception is the 2DARLS method selected in Phase II and summarized next. The 2DARLS is based on minimizing the mean-squared error in fitting a model to data. As such, it can be viewed as a statistical approach. Both approaches (stochastic/probabilistic and statistical) lead to identical results in the case of Gaussian-distributed data. Thus, the statistical framework is more general because it can be applied in cases where the data satisfies a distribution other than Gaussian.

In the presentation below the 2DARLS identification algorithm is formulated for the ideal case wherein the noise process $w(n,k)$ is zero. In other words, Equation (7-2) is assumed to reduce to $x(n,k) = y(n,k)$, so the formulation can be defined in terms of the channel output process, $x(n,k)$. However, Equation (7-2) is a better representation of conditions in practical scenarios. Fortunately, the 2DARLS algorithm is robust to the presence of additive white noise.

Consider the problem of fitting an $AR(P_d,P_s)$ model to the zero-mean, stationary (at least in the wide sense), 2-D finite-duration sequence $\{x(n,k) \mid n = 0, 1, \ldots, N\text{-}1; k = 0, 1, \ldots, J\text{-}1\}$. That is, it is desired to obtain a set of constant, complex-valued, scalar

coefficients $\{a(i_d,i_s) \mid i_d = 0, 1, \ldots, P_d;\ i_s = 0, 1, \ldots, P_s;\ (i_d,i_s) \neq (0,0)\}$, and a scalar variance σ_u^2 such that the random process $\{x(n,k)\}$ (of which the given sequence is a particular finite-duration realization) satisfies the following relation,

$$x(n,k) = -\sum_{\substack{i_d=1 \\ (i_d,i_s)\neq(0,0)}}^{P_d} \sum_{i_s=1}^{P_s} a^*(i_d,i_s)\, x(n-i_d,k-i_s) + u(n,k) \tag{7-9}$$

where $\{u(n,k)\}$ is the zero-mean, complex-valued, input white sequence with variance $\sigma_u^2 = E[u(n,k)\, u^*(n,k)]$. In the 2DARLS algorithm the unknown parameters, σ_u^2 and $\{a(i_d,i_s) \mid i_d = 0, 1, \ldots, P_d;\ i_s = 0, 1, \ldots, P_s;\ (i_d,i_s) \neq (0,0)\}$, are estimated via minimization of the square of the error in fitting model (7-9) for a fixed model order pair (P_d,P_s) to a given set of secondary data.

The 2DARLS method formulation is as follows. Consider first the case where the secondary data consists of only one element; that is, $K = 1$. Let $\varepsilon(n,k)$ denote the forward linear prediction error variable of the $AR(P_d,P_s)$ model,

$$\varepsilon(n,k) = x(n,k) + \sum_{\substack{i_d=1 \\ (i_d,i_s)\neq(0,0)}}^{P_d} \sum_{i_s=1}^{P_s} a^*(i_d,i_s)\, x(n-i_d,k-i_s) \qquad P_d \leq n \leq N\text{-}1;\ P_s \leq k \leq J\text{-}1 \tag{7-10}$$

The function to be minimized is the following real-valued scalar function of the forward linear prediction error variable,

$$J(\underline{a}) = \sum_{n=P_d}^{N-1} \sum_{k=P_s}^{J-1} \varepsilon^*(n,k)\varepsilon(n,k) = \sum_{n=P_d}^{N-1} \sum_{k=P_s}^{J-1} \varepsilon(n,k)\varepsilon^*(n,k) \tag{7-11}$$

where $\underline{a}$ is the following $((P_d+1)(P_s+1)-1)$-element column vector,

(7-12) $$\underline{a} = [a(P_d,P_s)\ a(P_d,P_s\text{-}1)\ \ldots\ a(P_d,0)\ \ldots\ a(0,P_s)\ a(0,P_s\text{-}1)\ \ldots\ a(0,1)]^T$$

In standard optimization theory, function $\mathbf{J}(\underline{a})$ is referred to as the Jacobian. The forward linear prediction squared error, denoted as Σ_f, is defined as total squared error in the linear fit, and for the special case of scalar sequences it is obtained as

(7-13) $$\Sigma_f = \sum_{n=P_d}^{N-1} \sum_{k=P_s}^{J-1} \varepsilon^*(n,k)\varepsilon(n,k) = \sum_{n=P_d}^{N-1} \sum_{k=P_s}^{J-1} \varepsilon(n,k)\varepsilon^*(n,k)$$

It follows from Equations (7-11) and (7-13) that $\Sigma_f = \mathbf{J}(\underline{a})$. If the $AR(P_d,P_s)$ model is an appropriate fit to the data, the prediction squared error is an estimate of the variance of the input noise. Equivalently, the prediction squared error is also an estimate of the variance of the prediction error; specifically,

(7-14) $$\hat{\sigma}_\varepsilon^2 = \hat{\sigma}_u^2 = \frac{1}{(N-P_d)(J-P_s)}\Sigma_f$$

In order to solve the minimization problem postulated above, it is convenient to define the following $((P_d+1)(P_s+1))$-element column vector,

(7-15) $$\underline{x}(n\text{-}P_d{:}n,k\text{-}P_s{:}k) = [x(n\text{-}P_d,k\text{-}P_s)\ \ldots\ x(n\text{-}P_d,k)\ \ldots\ x(n,k\text{-}P_s)\ \ldots\ x(n,k)]^T$$

Given these Definitions (7-12) and (7-15), the forward linear prediction error in Equation (7-10) can be expressed as,

(7-16) $$\varepsilon(n,k) = \left[\underline{a}^H \quad 1\right] \underline{x}(n{:}n-P_d,k{:}k-P_s)$$

and the function to be minimized, $\mathbf{J}(\underline{a})$, is expressed as

$$(7\text{-}17)\qquad \mathbf{J}(\underline{a}) = \begin{bmatrix} \underline{a}^H & 1 \end{bmatrix} \left[\sum_{n=P_d}^{N-1} \sum_{k=P_s}^{J-1} \underline{x}(n{:}n-P_d, k{:}k-P_s)\, \underline{x}^H(n{:}n-P_d, k{:}k-P_s) \right] \begin{bmatrix} \underline{a} \\ 1 \end{bmatrix}$$

Now let a $((P_d+1)(P_s+1))x((P_d+1)(P_s+1))$ matrix $\mathbf{R}$ represent the double summation of outer-products (dyads) in Equation (7-17); that is,

$$(7\text{-}18)\qquad \mathbf{R} = \sum_{n=P_d}^{N-1} \sum_{k=P_s}^{J-1} \underline{x}(n{:}n-P_d, k{:}k-P_s)\, \underline{x}^H(n{:}n-P_d, k{:}k-P_s)$$

Notice that matrix $\mathbf{R}$ is Hermitian. Now further define the following partitions in $\mathbf{R}$,

$$(7\text{-}19)\qquad \mathbf{R} = \begin{bmatrix} R_{11} & \underline{r}_{12} \\ \underline{r}_{12}^H & r_{22} \end{bmatrix}$$

where R_{11} is $((P_d+1)(P_s+1)-1)x((P_d+1)(P_s+1)-1)$, $\underline{r}_{12}$ is $((P_d+1)(P_s+1)-1)x1$, and r_{22} is a scalar. Now the function to be minimized is represented simply as

$$(7\text{-}20)\qquad \mathbf{J}(\underline{a}) = \underline{a}^H R_{11} \underline{a} + \underline{r}_{12}^H \underline{a} + \underline{a}^H \underline{r}_{12} + r_{22}$$

which is a real-valued scalar function of the complex-valued coefficient vector, $\underline{a}$.

Standard optimization theory applied to function $\mathbf{J}(\underline{a})$ in Equation (7-20) leads to an equation of the form

$$(7\text{-}21)\qquad R_{11} \underline{a} = -\underline{r}_{12}$$

The set of linear equations (7-21) can be solved for $\underline{a}$ using any one of various techniques (including Cholesky factorization since matrix R_{11} is Hermitian). Such a solution is of the form

(7-22) $$\underline{a} = R_{11}^{-1}\underline{r}_{12}$$

Once the optimum coefficients are available, the forward linear prediction error can be obtained using Equations (7-13) and (7-20); specifically,

(7-23) $$\Sigma_f = \mathbf{J}^o(\underline{a}) = \underline{r}_{12}^H\underline{a} + r_{22} = r_{22} - \underline{r}_{12}^H R_{11}^{-1}\underline{r}_{12}$$

where $\mathbf{J}^o(\underline{a})$ denotes the optimum value of the Jacobian. Thus, the standard approach to the ARLS method is to implement Equations (7-22) and (7-23).

A preferred alternative (from accuracy and computational considerations) to the approach outlined above is presented in Appendix A for the 1-D multichannel ARLS. With straightforward modifications, the approach in Appendix A applies also to the 2DARLS, and is summarized next. The first step is to combine Equations (7-21) and (7-23) into a single matrix-vector equation,

(7-24) $$\mathbf{R}\begin{bmatrix}\underline{a}\\ 1\end{bmatrix} = \begin{bmatrix}R_{11} & \underline{r}_{12}\\ \underline{r}_{12}^H & r_{22}\end{bmatrix}\begin{bmatrix}\underline{a}\\ 1\end{bmatrix} = \begin{bmatrix}\underline{0}\\ \Sigma_f\end{bmatrix}$$

From the structure of matrix $\mathbf{R}$ as presented in Equation (7-18), it follows that matrix $\mathbf{R}$ admits a factorization of the form

(7-25) $$\mathbf{R} = \mathbf{X}^H\mathbf{X}$$

where $\mathbf{X}$ is the data matrix of the normal equations in least-squares linear prediction. Appendix A states the conditions that must be satisfied for matrix $\mathbf{X}$ to have full rank. Assuming those conditions are satisfied, then matrix $\mathbf{X}$ admits a QR decomposition of the form

(7-26) $$\mathbf{X} = \mathbf{QU}$$

where $\mathbf{Q}$ is a unitary matrix, and $\mathbf{U}$ is a complex-valued, full-rank, upper-triangular matrix. It follows from Equations (7-25) and (7-26) that

(7-27) $$\mathbf{R} = \mathbf{X}^H\mathbf{X} = \mathbf{U}^H\mathbf{U}$$

Matrix $\mathbf{U}$ admits a partitioning analogous to that in Equation (7-19); that is,

(7-28) $$\mathbf{U} = \begin{bmatrix} \mathbf{U}_{11} & \underline{u}_{12} \\ [0] & u_{22} \\ [0] & [0] \end{bmatrix}$$

where $\mathbf{U}_{11}$ is a full-rank upper-triangular matrix, u_{22} is a non-zero scalar, and $\underline{u}_{12}$ is a column vector with no particular properties. It follows from Equations (7-24), (7-27), and (7-28) that

(7-29) $$\begin{bmatrix} \mathbf{U}_{11}^H\mathbf{U}_{11} & \mathbf{U}_{11}^H\underline{u}_{12} \\ \underline{u}_{12}^H\mathbf{U}_{11} & \underline{u}_{12}^H\underline{u}_{12} + u_{22}^* u_{22} \end{bmatrix} \begin{bmatrix} \underline{a} \\ 1 \end{bmatrix} = \begin{bmatrix} \underline{0} \\ \Sigma_f \end{bmatrix}$$

Equation (7-29) separates into the following two equations,

(7-30) $$\mathbf{U}_{11}^H\mathbf{U}_{11}\underline{a} = -\mathbf{U}_{11}^H\underline{u}_{12}$$

(7-31) $$\Sigma_f = \underline{u}_{12}^H\mathbf{U}_{11}\underline{a} + \underline{u}_{12}^H\underline{u}_{12} + u_{22}^* u_{22}$$

Elimination of $\mathbf{U}_{11}^H$ from both sides of Equation (7-30) leads to

(7-32) $$\mathbf{U}_{11}\underline{a} = -\underline{u}_{12}$$

and then $\underline{a}$ can be obtained as

(7-33) $$\underline{a} = -U_{11}^{-1}\underline{u}_{12}$$

Substitution of Equation (7-33) into Equation (7-31) results in

(7-34) $$\Sigma_f = u_{22}^{*}\, u_{22}$$

Equations (7-33) and (7-34) constitute the desired solution. A preferred alternative to the inverse calculation in Equation (7-33) is to solve Equation (7-32) using back-substitution since matrix U_{11} is upper-triangular. Back-substitution is both accurate and efficient. Thus, the major computational load involved in solving for $\underline{a}$ and Σ_f is the calculation of the QR decomposition.

For the case where the secondary data set has more than one element, $K > 1$, the four processing options identified in Appendix A apply directly. However, analyses carried out to date indicate that the block approaches generate better parameter estimates.

7.2 Two-Dimensional PAMF Detection for STAP Applications

The 2-D PAMF is an extension of the 1-D multichannel PAMF. In fact, the structure is analogous to that described by Rangaswamy and Michels (1997) and by Román et al. (1998), as suggested by inspection of the block diagram in Figure 7-1. However, the entries of the blocks in Figure 7-1 are different from those corresponding to the PAMF. Most significantly, in the 2-D PAMF the parameters are estimated using the 2DARLS algorithm, and the whitening filter is a 2-D scalar filter which carries out spatial and temporal whitening jointly. In the PAMF the whitening is carried out in two steps: first temporal and then spatial.

With respect to Figure 7-1, the unit-variance 2-D sequence $\{v(n,k)\}$ is obtained from the whitened sequence $\{\varepsilon(n,k)\}$ as

$$v(n,k) = \frac{1}{\hat{\sigma}_\varepsilon}\varepsilon(n,k) \qquad 0 \le n \le N\text{-}P_d\text{-}1;\ 0 \le k \le J\text{-}P_s\text{-}1 \tag{7-35}$$

and the notation $v(n,k) \leftrightarrow \underline{v}$ indicates conversion from 2-D sequence into column vector, following any convention for the mapping of the sequence elements into a vector. Finally, the test statistic is calculated as

$$\Lambda = \frac{\left|\underline{s}^H \underline{v}\right|^2}{\underline{s}^H \underline{s}} \tag{7-36}$$

Notice that this expression is analogous to the vector form of Equation (2-17), but the vector variables have very distinct meaning.

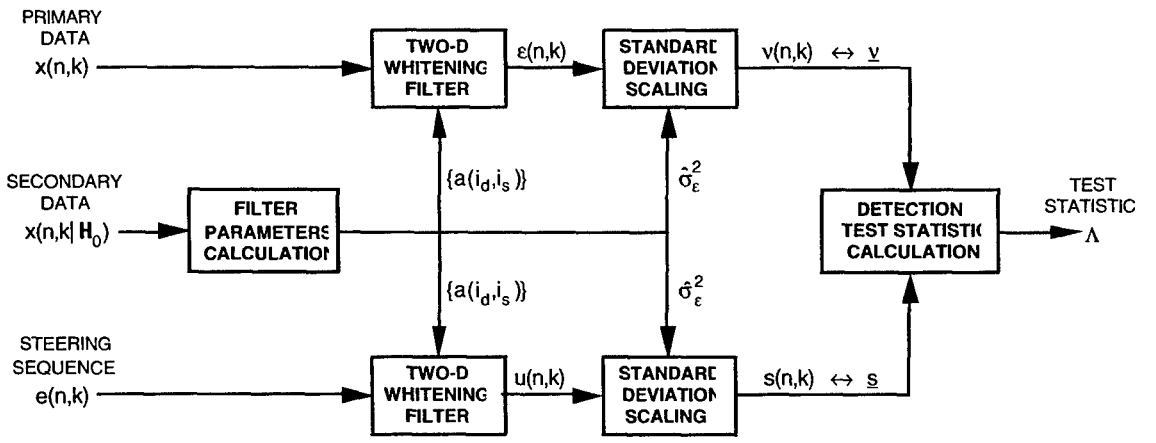

Figure 7-1. Two-dimensional PAMF STAP and detection architecture.

8.0 SUMMARY AND FUTURE WORK

The generic STAP and detection architecture introduced herein covers the MF as well as three new STAP algorithms, orthogonal projection (OP), maximum correlation (MC), and OP using canonical variables (OPCV). These algorithms have potential for significant dimensionality reduction, while being based on the optimization of probabilistic criteria.

Future work includes further development of the theoretical relationships between the STAP methods herein and others, such as the multistage Wiener filter. Another aspect of future work is the software-based analysis of detection performance of the three methods for a variety of configurations under known-covariance as well as unknown-covariance conditions. For each STAP algorithm several distinct test statistics can be defined, and each test statistic exhibits different performance properties, including CFAR. The application of alternative test statistics should be investigated. In the context of the adaptive (unknown-covariance) formulation of the algorithms, the utilization of structured covariance estimates instead of sample covariance estimates should be investigated also.

The 2-D PAMF constitutes a significant extension to the PAMF that may yield computational and or performance advantages in relation to the PAMF-LS (which is the leading candidate among the alternative PAMF implementations). Computational advantages are possible specially for cases wherein the number of channels is large because the PAMF-LS channel requires identification of P JxJ matrices, whereas the 2DARLS requires identification of P_d P_sxP_s matrices with $P_d = P$ and $P_s \leq J$. Performance advantages may arise due to the enhanced modeling capability of the 2DARLS along the spatial axis.

Future work for the 2-D PAMF includes software-based analysis of detection performance, specially in relation to other methods, including the 1-D multichannel PAMF-LS. Such analyses should include also the application of alternative test statistics. Recent work by SSC personnel in detection of anti-tank mines using ground penetrating radar demonstrates the potential for wide applicability of the 2-D PAMF and other related methods to images.

APPENDIX A. AUTO-REGRESSIVE LEAST-SQUARES MODEL IDENTIFICATION

This appendix presents the least-squares (LS) identification algorithm for multichannel auto-regressive (AR) processes. Also presented are the options adopted for its software implementation in the context of the parametric adaptive matched filter (PAMF) for space-time adaptive processing (STAP) in airborne surveillance phased array radar systems. In addition, simulation analysis results are presented for the algorithmic options discussed herein. The presentation herein serves as documentation for the SSC-generated MATLAB function *arls*. The algorithmic discussions are brief since details are available in texts such the one by Marple (1987). However, software implementation aspects unavailable elsewhere are discussed herein.

A.1 Multichannel Least-Squares Formulation

Consider the problem of fitting an AR model of order P to a zero-mean, stationary (at least in the wide sense), finite-duration sequence $\{\underline{x}(n) \mid n = 0, 1, \ldots, N\text{-}1\}$, with $\underline{x}(n) \in \mathbb{C}^J$. That is, it is desired to obtain a set of constant matrix coefficients $\{A(k) \mid k = 1, 2, \ldots, P\}$, with $A(k) \in \mathbb{C}^{JxJ}$, and a covariance matrix $R_{uu} \in \mathbb{C}^{JxJ}$ such that the random process $\{\underline{x}(n)\}$ (of which the given sequence is a particular finite-duration realization) satisfies the following equation,

$$\underline{x}(n) = -\sum_{k=1}^{P} A^H(k)\,\underline{x}(n-k) + \underline{u}(n) \qquad \text{(A-1)}$$

where $\{\underline{u}(n)\}$ is the zero-mean input white sequence, with $\underline{u}(n) \in \mathbb{C}^J$ and $R_{uu} = E[\underline{u}(n)\,\underline{u}^H(n)]$. If the process $\{\underline{x}(n)\}$ satisfies Equation (A-1), then it is said to be an AR process of order P, and is denoted as $AR(P)$.

The unknown parameters, R_{uu} and $\{A(k)\}$, are estimated via minimization of the error squared. That is, let $\underline{\varepsilon}(n)$ denote the forward linear prediction error of the AR(P) model,

$$\underline{\varepsilon}(n) = \underline{x}(n) + \sum_{k=1}^{P} A^H(k)\,\underline{x}(n-k) \qquad n = P, P+1, \ldots, N-1 \tag{A-2}$$

Then, the function (Jacobian) to be minimized is of the form

$$\mathbf{J} = \mathrm{tr}\left[\Sigma_f\right] = \sum_{n=P}^{N-1} \underline{\varepsilon}^H(n)\,\underline{\varepsilon}(n) \tag{A-3}$$

where the forward linear prediction error matrix, Σ_f, is defined as

$$\Sigma_f = \sum_{n=P}^{N-1} \underline{\varepsilon}(n)\,\underline{\varepsilon}^H(n) \tag{A-4}$$

If the AR(P) system is an appropriate fit to the data, the prediction error matrix is an estimate of the covariance matrix of the input noise; specifically,

$$\hat{R}_{uu} = \frac{1}{N-P}\,\Sigma_f \tag{A-5}$$

Marple (1987) shows that minimization of the function **J** leads to the following block matrix linear equation,

$$\left[\begin{array}{ccc:c} R_x(0,0) & \cdots & R_x(0,P-1) & R_x(0,P) \\ \vdots & \ddots & \vdots & \vdots \\ R_x(P-1,0) & \cdots & R_x(P-1,P-1) & R_x(P-1,P) \\ \hdashline R_x(P,0) & \cdots & R_x(P,P-1) & R_x(P,P) \end{array}\right] \left[\begin{array}{c} A(P) \\ \vdots \\ A(1) \\ \hdashline I_J \end{array}\right] = \left[\begin{array}{c} [0] \\ \vdots \\ [0] \\ \hdashline \Sigma_f \end{array}\right] \tag{A-6}$$

where $I_J \in \Re^{JxJ}$ is an identity matrix, and $R_x(i,j) \in \mathbb{C}^{JxJ}$ is of the form

$$R_x(i,j) = \sum_{n=0}^{N-1-P} \underline{x}(n+i)\underline{x}^H(n+j) \qquad (i,j) = (0,0), \ldots, (P,P) \tag{A-7}$$

The purpose of the partitions introduced explicitly in Equation (A-6) is made clear in Section A.2. In compact notation, Equation (A-6) can be expressed as

$$\mathbf{R}\begin{bmatrix}\mathbf{A}\\ I_J\end{bmatrix} = \begin{bmatrix}R_{11} & R_{12}\\ R_{21} & R_{22}\end{bmatrix}\begin{bmatrix}\mathbf{A}\\ I_J\end{bmatrix} = \begin{bmatrix}[\mathbf{0}]\\ \Sigma_f\end{bmatrix} \tag{A-8}$$

with $\mathbf{R} \in \mathbb{C}^{J(P+1)xJ(P+1)}$, $R_{11} \in \mathbb{C}^{JPxJP}$, $R_{12} \in \mathbb{C}^{JPxJ}$, $R_{21} \in \mathbb{C}^{JxJP}$, $R_{22} \in \mathbb{C}^{JxJ}$, and

$$\mathbf{A} = \begin{bmatrix} A(P) \\ A(P-1) \\ \vdots \\ A(1) \end{bmatrix} \tag{A-9}$$

Equations (A-6) and (A-8) are of the same structure as the corresponding multichannel Yule-Walker normal equation, except that matrix $\mathbf{R}$ is Hermitian only (whereas the corresponding matrix in the Yule-Walker formulation is both Hermitian and block Toeplitz). The diagonal sub-matrices, R_{11} and R_{22}, are Hermitian also. Furthermore, for appropriate values of N and P, matrices $\mathbf{R}$, R_{11}, and R_{22} are positive definite.

The formulation presented above is referred to as the covariance (or non-windowed) method of least-squares linear prediction. This terminology arises from its early usage in speech processing applications, and is misleading in the context of modern stochastic algorithms. Specifically, $\mathbf{R}$ is neither a

structured covariance matrix, nor a sample (maximum likelihood estimate) covariance matrix (except for the case where $P = N-1$, as noted in Section A.3). It is important to note that the ordering of the AR coefficients in block matrix $\mathbf{A}$ and the partitioning and definitions in Equations (A-6)-(A-9) differ from the usual conventions (Marple, 1987). This variation is adopted herein based only on considerations of the software implementation of the method, and there is strict equivalence between the results obtained using either notation.

The structure of matrix $\mathbf{R}$ as presented in Equations (A-6) and (A-7) allows a factorization of the form

$$\mathbf{R} = \mathbf{X}^H\mathbf{X} \tag{A-10}$$

where $\mathbf{X} \in \mathbb{C}^{(N-P)\times J(P+1)}$ is the data matrix of the normal equations in the covariance method of least-squares linear prediction, and is defined as

$$\mathbf{X} = \begin{bmatrix} \underline{x}^H(0) & \underline{x}^H(1) & \cdots & \underline{x}^H(P) \\ \underline{x}^H(1) & \underline{x}^H(2) & \cdots & \underline{x}^H(P+1) \\ \vdots & \vdots & \ddots & \vdots \\ \underline{x}^H(P) & \underline{x}^H(P+1) & \cdots & \underline{x}^H(2P) \\ \vdots & \vdots & \vdots & \vdots \\ \underline{x}^H(N-1-P) & \underline{x}^H(N-P) & \cdots & \underline{x}^H(N-1) \end{bmatrix} \tag{A-11}$$

The form in Equation (A-11) corresponds to the case where $N >> P$.

A.2 Normal Equations Solutions

The normal equations (A-6) (or equivalently, (A-8)) can be solved using a variety of techniques, including the singular value decomposition (SVD). However, two specific techniques have features that are important in the context of interest herein.

One technique operates directly on the data matrix $\mathbf{X}$, whereas the other exploits the Hermitian property of matrix $\mathbf{R}$. Each technique generates one set of AR(P) coefficients given one finite-duration realization of the array output process. Both techniques require matrix $\mathbf{R}$ to be positive definite in order to generate a complete solution (both $\mathbf{A}$ and Σ_f). Conditions are established which insure that $\mathbf{R}$ is positive definite. In addition, the positive semi-definite case is considered also.

A.2.1 DATA-BASED SOLUTION

Consider first the case where the number of rows in the data matrix $\mathbf{X}$ is equal to or larger than the number of columns; that is, when the following condition is satisfied:

(A-12) $$N - P \geq J(P + 1) = JP + J$$

Condition (A-12) insures that $\text{rank}[\mathbf{X}] = J(P + 1)$ with probability one due to the randomness of the data in the context of airborne surveillance radar applications. With Condition (A-12) satisfied, the QR decomposition of $\mathbf{X}$ is of the form

(A-13) $$\mathbf{X} = \mathbf{QU}$$

where $\mathbf{Q} \in \mathbb{C}^{(N-P)x(N-P)}$ is a unitary matrix, and $\mathbf{U} \in \mathbb{C}^{(N-P)xJ(P+1)}$ is an upper-triangular matrix with $\text{rank}[\mathbf{U}] = J(P + 1)$. It follows from Equations (A-10) and (A-13) that

(A-14) $$\mathbf{R} = \mathbf{X}^H\mathbf{X} = \mathbf{U}^H\mathbf{U}$$

Matrix $\mathbf{U}$ admits a partitioning analogous to that in Equation (A-8); that is,

$$\text{(A-15)} \qquad \mathbf{U} = \begin{bmatrix} \mathbf{U}_{11} & \mathbf{U}_{12} \\ [0] & \mathbf{U}_{22} \\ [0] & [0] \end{bmatrix}$$

where $\mathbf{U}_{11} \in \mathbb{C}^{JPxJP}$ is upper-triangular with $\text{rank}[\mathbf{U}_{11}] = JP$, $\mathbf{U}_{22} \in \mathbb{C}^{JxJ}$ is upper-triangular with $\text{rank}[\mathbf{U}_{22}] = J$, and $\mathbf{U}_{12} \in \mathbb{C}^{JPxJ}$ with no particular properties. It follows from Equations (A-8), (A-14), and (A-15) that

$$\text{(A -16)} \qquad \begin{bmatrix} \mathbf{U}_{11}^H \mathbf{U}_{11} & \mathbf{U}_{11}^H \mathbf{U}_{12} \\ \mathbf{U}_{12}^H \mathbf{U}_{11} & \mathbf{U}_{12}^H \mathbf{U}_{12} + \mathbf{U}_{22}^H \mathbf{U}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{A} \\ \mathbf{I}_J \end{bmatrix} = \begin{bmatrix} [\mathbf{0}] \\ \Sigma_f \end{bmatrix}$$

Equation (A-16) separates into the following two equations,

$$\text{(A-17)} \qquad \mathbf{U}_{11}^H \mathbf{U}_{11} \mathbf{A} = -\mathbf{U}_{11}^H \mathbf{U}_{12}$$

$$\text{(A-18)} \qquad \Sigma_f = \mathbf{U}_{12}^H \mathbf{U}_{11} \mathbf{A} + \mathbf{U}_{12}^H \mathbf{U}_{12} + \mathbf{U}_{22}^H \mathbf{U}_{22}$$

Pre-multiplication of both sides of Equation (A-17) by $\mathbf{U}_{11}^{-H}$ leads to

$$\text{(A-19)} \qquad \mathbf{U}_{11} \mathbf{A} = -\mathbf{U}_{12}$$

and then $\mathbf{A}$ can be obtained as

$$\text{(A-20)} \qquad \mathbf{A} = -\mathbf{U}_{11}^{-1} \mathbf{U}_{12}$$

Substitution of Equation (A-20) into Equation (A-18) then results in

$$\text{(A-21)} \qquad \Sigma_f = \mathbf{U}_{22}^H \mathbf{U}_{22}$$

Equations (A-20) and (A-21) constitute the desired solution.

Equation (A-20) expresses the solution for $\mathbf{A}$ in direct form, which is convenient for obtaining the solution for Σ_f. However, an alternative approach to a matrix inverse calculation followed by a matrix-matrix multiplication is to solve Equation (A-19) via back-substitution because U_{11} is upper-triangular. Since U_{11} is positive definite, this alternative is the most efficient computationally (Golub and Van Loan, 1989). The increase in computational efficiency over the direct inverse approach is proportional to the number of columns of U_{12}.

As stated above, Condition (A-12) insures that the data matrix $\mathbf{X}$ has rank equal to its smaller dimension (column rank), and thus U_{11} and U_{22} are both positive definite. This condition leads to an upper bound for model order (P), given N and J. The upper bound condition is expressed as

$$P \le P_{max} = \text{fix}\left[\frac{N-J}{J+1}\right] \qquad \text{BOUND FOR BOTH } \mathbf{A} \text{ AND } \Sigma_f \tag{A-22}$$

where the operator fix[•] truncates a real-valued number. This bound is very tight, and implies that $N \ge 2J+1$ so that $P_{max} \ge 1$.

A more relaxed bound is possible if only the AR coefficients are desired, or if the values of the parameters J, N, and P result in a data matrix $\mathbf{X}$ with fewer rows than columns (such is indeed the case when $N \approx J$, even if P is small). In such case the condition that must be satisfied is

$$JP + J > N - P \ge JP \tag{A-23}$$

where the second inequality follows from $N - P \ge \text{rank}[R_{11}] = JP$. Condition (A-23) insures that $\text{rank}[\mathbf{X}] = N - P \ge JP$ with probability one

due to the randomness of the data. Assuming Condition (A-23) is satisfied, the form of matrix $\mathbf{U}$ in the QR decomposition of $\mathbf{X}$ is

$$\text{(A-24)} \qquad \mathbf{U} = \begin{bmatrix} \mathbf{U}_{11} & \mathbf{U}_{12} \\ [0] & \mathbf{U}_{22} \end{bmatrix}$$

where $\mathbf{U}_{11} \in \mathbb{C}^{JP \times JP}$ is upper-triangular with $\text{rank}[\mathbf{U}_{11}] = JP$, and $\mathbf{U}_{12} \in \mathbb{C}^{JP \times J}$ with no particular properties, as before. However, now $\mathbf{U}_{22} \in \mathbb{C}^{(N-P-JP) \times J}$ with $\text{rank}[\mathbf{U}_{22}] = N - P - JP < J$. The structure of $\mathbf{U}_{22}$ is still upper-triangular, but the number of rows is less than the number of columns. In fact, when Condition (A-23) is met with equality on the right (that is, $N - P = JP$), matrix $\mathbf{U}_{22}$ is non-existent. Nevertheless, the AR coefficients are obtained by solving Equation (A-19), as before.

The second inequality ($N - P \geq JP$) in Condition (A-23) leads to another upper bound for model order, given N and J. In this case, the upper bound condition is expressed as

$$\text{(A-25)} \qquad P \leq P_{max} = \text{fix}\left[\frac{N}{J+1}\right] \qquad \text{BOUND FOR } \mathbf{A} \text{ ONLY}$$

This bound is less tight than Bound (A-22), and implies that $N \geq J + 1$ so that $P_{max} \geq 1$.

A.2.2 COVARIANCE-BASED SOLUTION

As in Section A.2.1, consider first the case where Condition (A-12) is satisfied. Since $\mathbf{X}$ has full rank in this case, it follows that $\mathbf{R}$ is positive definite ($\text{rank}[\mathbf{R}] = JP + J$) with probability one due to the randomness of the data (using the SVD it is easy to show that $\text{rank}[\mathbf{R}] = \text{rank}[\mathbf{X}]$). Furthermore, $\mathbf{R}_{11}$ and $\mathbf{R}_{22}$ are Hermitian and positive definite also ($\text{rank}[\mathbf{R}_{11}] = JP$; $\text{rank}[\mathbf{R}_{22}] = J$).

Based on the partitions defined in Equations (A-6) and (A-8), the normal equations separate into two equations,

(A-26) $$\mathbf{R}_{11}\mathbf{A} = -\mathbf{R}_{12}$$

(A-27) $$\Sigma_f = \mathbf{R}_{22} + \mathbf{R}_{21}\mathbf{A}$$

As for the data-based solution, Equation (A-26) is utilized to solve for $\mathbf{A}$, and then Σ_f is obtained via Equation (A-27). Equation (A-26) can be solved via a matrix inverse calculation followed by a matrix-matrix multiplication. However, since $\mathbf{R}_{11}$ is Hermitian and positive definite, a more efficient approach (based on the Cholesky decomposition) is possible.

The Cholesky decomposition of a Hermitian, positive definite matrix $\mathbf{R}_{11}$ is of the form

(A-28) $$\mathbf{R}_{11} = \mathbf{C}^H\mathbf{C}$$

where $\mathbf{C} \in \mathbb{C}^{JP \times JP}$ is upper-triangular with real- and positive-valued elements along the main diagonal, and with $\text{rank}[\mathbf{C}] = JP$. It follows from Equations (A-26) and (A-28) that

(A-29) $$\mathbf{C}^H\mathbf{C}\mathbf{A} = -\mathbf{R}_{12}$$

Given the triangular structure of $\mathbf{C}$, matrix $\mathbf{A}$ is solved for using back-substitution twice. Equations (A-27)-(A-29) constitute the desired solution.

The model order upper bound in Equation (A-22) is valid also for the covariance approach since rank Condition (A-12) applies. Similarly, the bound in Equation (A-25) is valid also for the covariance approach when Condition (A-23) holds and only the AR coefficients are required ($\text{rank}[\mathbf{R}] = \text{rank}[\mathbf{X}]$ holds for all cases). When

Condition (A-23) is satisfied as $JP + J > N - P > JP$ (that is, without equality on the right), Equation (A-27) generates a non-zero value for Σ_f, but such solution is incorrect because it is generated with a rank deficiency in the covariance matrix. Furthermore, when Condition (A-23) is met with equality on the right (that is, $N - P = JP$), the matrix Σ_f generated via Equation (A-27) is the numerical null matrix because the rank deficiency is maximum (within the limits of Condition (A-23)). If Condition (A-23) is violated from the right (that is, if $N - P < JP$), then the rank defiency is sufficiently large to introduce error in the calculation of $\mathbf{A}$ also.

A.3 Surveillance Scenario Software Implementation Options

As in Section 1.1, consider an airborne side-looking configuration of a J-element linear phased array radar for moving target detection. In this context, the data to be processed is the return from N pulses as collected by the J array elements for each one of K_T range bins. This $JxNxK_T$ set of data is the so-called data cube. A detection decision is desired for each range bin in the data cube. Further, the decision approach is adaptive in order to account for possible lack of statistical stationarity in the data cube. The favored approach is sequential in nature; that is, each range bin is tested in a pre-determined order. At each decision instant, the range bin to be tested is referred to as the primary data. Also, a subset of $K < K_T$ range bins in a neighborhood near the primary data is utilized as training data for the processing algorithm and the detection rule. This data set is referred to as the secondary data. The range bins in the secondary data set are assumed to provide target-free, independent realizations of the array output process.

The PAMF processing and detection rule utilizes the secondary data to generate a whitening filter for the primary data, and then

applies a detection rule based on the matched filter concept. One possible configuration utilizes the least-squares (LS) algorithm to identify a set of AR(P) parameters that fit the secondary data, and then utilizes the parameters as a moving-average filter to process the primary data. The filter output constitutes a residual sequence to which the detection rule is applied. Such configuration is referred to as the PAMF-LS.

One significant aspect of the surveillance radar context is the availability of more than one realization of the data to be utilized for the model identification step of the PAMF-LS. Within such context, the two solution approaches in Section A.2 lead to three distinct software implementation options of interest. These options are summarized next, and a descriptive name is provided for each. The given names link each algorithm with its software implementation in a MATLAB-based function generated by SSC and named *arls*.

1. COEFFQR: QR decomposition of data matrices and model parameter averaging.
   a) Form K data matrices, one for each secondary data range bin.
   b) Generate K sets of model parameters (AR coefficients and prediction error matrices) via the data matrix QR decomposition approach.
   c) Average the K sets of model parameters to obtain the desired solution.

2. COEFFCHOLESKY: Cholesky factorization of covariance matrices and model parameter averaging.
   a) Generate K covariance matrices, one for each secondary data range bin.
   b) Generate K sets of model parameters via the covariance matrix Cholesky factorization approach.

c) Average the K sets of model parameters to obtain the desired solution.

3. DATABLKQR: QR decomposition of block data matrix.
   a) Form a block data matrix with block rows consisting of the K data matrices (one data matrix for each secondary data range bin).
   b) Generate the desired model parameter solution via the QR decomposition approach applied to the block data matrix.

4. COVARCHOLESKY: Covariance averaging and Cholesky factorization of averaged covariance.
   a) Generate K covariance matrices, one for each secondary data range bin.
   b) Average the K covariance matrices to obtain a single covariance matrix.
   c) Generate the desired model parameter solution via the covariance matrix Cholesky factorization approach.

Options 1 and 2 utilize the LS solution approaches in a single-realization context, so the algorithmic approaches are applied as discussed in Section A.2.1. The rank conditions and model order upper bounds in Section A.2.1 apply also. Options 3 and 4, however, lead to a different rank condition and model order upper bound.

Consider first Option 4. For this option, the covariance matrix **R** for Equation (A-8) is generated as the average of K single-range-bin covariance matrices; namely,

(A-30) $$\mathbf{R} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{R}_k = \frac{1}{K}\sum_{k=1}^{K}\mathbf{X}_k^H\mathbf{X}_k$$

where $\mathbf{X}_k$ denotes the data matrix for the kth range bin, and $\mathbf{R}_k$ is the covariance for the kth range bin. Covariance matrix $\mathbf{R}$ as defined in (A-30) is a valid covariance matrix for the normal equations (A-8) because a set of normal equations is valid for each of the individual single-range-bin covariance matrices. As stated previously, the covariance-based solution to the normal equations (involving both $\mathbf{A}$ and Σ_f) exists and is unique when $\mathbf{R}$ has full rank, $\text{rank}[\mathbf{R}] = JP + J$. The conditions under which $\mathbf{R}$ attains full rank are established next.

Under Option 4, the conditions under which $\mathbf{R}$ attains full rank involve the system parameters J, N, and P, as before, but K now plays a key role. Consider first the cases where $N - P \geq JP + J$. For these cases each $\mathbf{X}_k$ has full column rank, $\text{rank}[\mathbf{X}_k] = JP + J$, and consequently, each $\mathbf{R}_k$ has full rank, $\text{rank}[\mathbf{R}_k] = JP + J$ (see Section A.2.2). Matrix $\mathbf{R}$ has full rank also since it is an average of K independent, full-rank realizations of the covariance matrix.

Consider now the cases where $N - P < JP + J$. For these cases each $\mathbf{X}_k$ has full row rank, $\text{rank}[\mathbf{X}_k] = N - P$, and consequently, each $\mathbf{R}_k$ is rank-deficient, $\text{rank}[\mathbf{R}_k] = N - P$ (see Section A.2.2). However, the rank of $\mathbf{R}$ increases by $N - P$ for every $\mathbf{R}_k$ added to the sum, up to the maximum possible rank value of $JP + J$. In other words, $\text{rank}[\mathbf{R}] = \min[(N - P)K, (P + 1)J]$. Thus, the condition for $\mathbf{R}$ to have full rank under Option 4 is

(A-31) $$(N - P)K \geq (P + 1)J$$

An associated model order upper bound, as in Equations (A-22) and (A-25), is meaningless for this option. In fact, Condition (A-31)

allows the AR model order to be as large as $P = P_{max} = N - 1$, provided K is correspondingly large ($K \geq JN$ if $P = N - 1$). This set of values for the system parameters is unlikely to be adopted in the context of the PAMF-LS because experience shows that low model orders suffice to provide excellent detection performance in most cases. However, it is appropriate to note that for $P = N - 1$ and $K \geq JN$, matrix $\mathbf{R}$ generated as in Equation (A-30) is the so-called sample covariance matrix, and is the maximum likelihood estimate of the covariance matrix of vector $\underline{x} = \left[\underline{x}^H(0) \;\; \underline{x}^H(1) \;\; \cdots \;\; \underline{x}^H(N-1)\right]^H$.

Consider now Option 3, the QR decomposition of the block data matrix. Define the block data matrix $\mathbf{X} \in \mathbb{C}^{K(N-P)xJ(P+1)}$ as

$$\text{(A-32)} \qquad \mathbf{X} = \frac{1}{\sqrt{K}} \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_K \end{bmatrix}$$

It is simple to verify that this block data matrix is related to the averaged covariance matrix in Equation (A-30) as

$$\text{(A-33)} \qquad \mathbf{R} = \mathbf{X}^H\mathbf{X} = \frac{1}{K}\sum_{k=1}^{K} \mathbf{X}_k^H\mathbf{X}_k = \frac{1}{K}\sum_{k=1}^{K} \mathbf{R}_k$$

The leftmost equality is of the same form as Equation (A-10), so the approach defined in Section A.2.1 can be applied directly. Specifically, the QR decomposition of the block data matrix $\mathbf{X}$ is generated as in Equation (A-13), and the unitary matrix in the decomposition is eliminated from the formulation as in Equation (A-14). The solution for the model parameters is obtained using only the upper-triangular matrix factor in the decomposition, as in Equations (A-20) and (A-21). The condition for $\mathbf{X}$ to have full rank is identical to the condition for $\mathbf{R}$ to have full rank;

namely, Condition (A-31). This follows from the leftmost equality in Equation (A-33). The various comments regarding the lack of AR model order bound for Option 4 apply without modification to Option 3 also.

From a structural standpoint, Options 3 and 4 can be used over the widest set of conditions, and constitute the only alternatives in cases where $J \approx N$ (in such cases Condition (A-12) fails). The relaxation offered by Condition (A-23) for Options 1 and 2 is important in the application context considered herein because the model estimate of the residual covariance matrix is neglected in the preferred configuration of the PAMF-LS (the maximum likelihood estimate or the time-averaged estimate are used instead). In terms of accuracy, the four options generate practically identical results when Condition (A-12) is satisfied strongly; namely, $N - P >> JP + J$. Results from all four options are very similar also when Condition (A-12) is satisfied with equality; namely, $N - P = JP + J$. From a computational viewpoint, there appears to be a set of system parameter values (J, N, P, K) wherein each algorithmic option executes more efficiently than the other three. A detailed operational count for each option is required in order to address this issue appropriately.

The model order upper bound for the LS algorithm presented in [Román et al., 2000] is based on Condition (A-31), and is correct from an algebraic point of view. However, the solution approach adopted in the paper is the block method of Option 3 (DATABLKQR), and for both block methods it is most appropriate to view the issue of model order as a condition to be satisfied rather than as a bound, for the reasons stated in the paragraph following Condition (A-31), and summarized next. Namely, when $K > 1$ and a block method is used, even a small number of secondary data sets allows a solution for typical scenario parameters (J and N) and practical values of model order.

A.4 Simulation Analysis Results

Simulation-based analyses to compare the LS algorithmic approaches have been carried out for a variety of system parameter values and scenario conditions, and a subset of those results is presented herein. Of specific interest are results obtained for the cases considered in [Román et al., 2000], and the inclusion of ground clutter temporal correlation model type as another parameter variation. Specifically, two model types for clutter temporal correlation are considered: exponential-shaped, and Gaussian-shaped.

The results in [Román et al., 2000] were generated using the algorithmic approach of Option 3. In most analyses carried out by SSC, Options 1 and 2 generate very similar results, and Options 3 and 4 also generate similar results. Thus, results are presented only for Options 1 and 4. Furthermore, the cases considered represent only the $J << N$ Analysis, and only one value of output signal-to-interference-plus-noise ratio (SINR). The $J \approx N$ Analysis is considered in [Román et al., 2000].

The conditions and parameters for the $J << N$ Analysis are: $J = 4$ channels, $N = 32$ pulses, $\sigma_w = 1$ normalized receiver noise standard deviation, $CNR = 40$ dB clutter-to-noise ratio, and $f_{ts} = 0.0$ and $f_{td} = 0.3336$ target normalized spatial and Doppler frequencies, respectively. Three sets of scenario conditions and two values of secondary data size are considered. Case 1 is for $\gamma = 0$ deg crab angle and no jamming. Case 2 is for $\gamma = 20$ deg and no jamming. Case 3 is for $\gamma = 0$ deg and two point-source barrage jammers: one jammer is at $f_{js} = -0.35$ jammer normalized spatial frequency with JNR = 45 dB jammer-to-noise ratio, and the other jammer is at $f_{js} = 0.2$ with JNR = 50 dB. For each case two values of secondary data size are considered: $K = 2JN = 256$ (the Brennan rule-of thumb value for 3

dB performance), and $K = 2J = 8$. For this Monte Carlo (MC) analysis, $P_{FA} \approx 0.01$ probability of false alarm, using $N_{MC} = 50$ repetitions of $N_{PFA} = 2{,}000$ independent data realizations each. Also, $P = 3$ model order is used, as in [Román et al., 2000].

Tables A-1 and A-2 present probability of detection (P_D) results for the three simulation cases at SINR = 9 dB and the exponential-shaped model for ground clutter temporal correlation. The sample standard deviation of the P_D estimates, denoted as $SD[P_D]$, is presented also in both tables. In addition, both tables include detection results for the optimal matched filter (MF) and the constant false alarm rate (CFAR) adaptive matched filter (AMF). The MF results are analytical, whereas the CFAR AMF results are simulation-based. Table A-1 results are for $K = 2JN = 256$, and compare with the respective single points in the SINR vs. P_D curves of Figures 4, 6, and 8 in [Román et al., 2000]. Table A-2 results are for $K = 2J = 8$, and compare with the respective single points in the SINR vs. P_D curves of Figures 5, 7, and 9 in [Román et al., 2000].

Examination of the results in Tables A-1 and A-2 indicates that both PAMF-LS implementations out-perform the CFAR AMF, in terms of detection as well as variability (smaller standard deviation), for both values of K (the CFAR AMF values in both tables are for $K = 2JN = 256$). Notice also that both PAMF-LS implementations perform almost identically for the large K value, and the performance is close to that of the optimal MF. However, for the small K value the PAMF-LS using covariance averaging (Option 4) out-performs, also in terms of detection and variability, the PAMF-LS using coefficient averaging (Option 1). Simulation-based results obtained by SSC for other analyses and conditions reflect the same trend.

CASE		MF	CFAR AMF		PAMF-LS COEFF AVG (P = 3)		PAMF-LS COVAR AVG (P = 3)	
CRAB ANGLE (deg)	NO. OF JAMMERS	P_D	P_D	SD[P_D]	P_D	SD[P_D]	P_D	SD[P_D]
0	0	0.8635	0.4571	0.0397	0.8190	0.0118	0.8242	0.0104
20	0	0.8635	0.4610	0.0382	0.8206	0.0125	0.8219	0.0116
0	2	0.8635	0.4565	0.0375	0.8229	0.0125	0.8280	0.0118

Table A-1. Detection performance of Options 1 and 4 of the PAMF-LS for the three simulation cases with SINR = 9 dB, K = 2JN = 256, and the exponential-shaped model for clutter temporal correlation.

CASE		MF	CFAR AMF (K = 2JN)		PAMF-LS COEFF AVG (P = 3)		PAMF-LS COVAR AVG (P = 3)	
CRAB ANGLE (deg)	NO. OF JAMMERS	P_D	P_D	SD[P_D]	P_D	SD[P_D]	P_D	SD[P_D]
0	0	0.8635	0.4571	0.0397	0.7542	0.0270	0.7835	0.0690
20	0	0.8635	0.4610	0.0382	0.7573	0.0893	0.7971	0.0575
0	2	0.8635	0.4565	0.0375	0.7609	0.0812	0.7822	0.0661

Table A-2. Detection performance of Options 1 and 4 of the PAMF-LS for the three simulation cases with SINR = 9 dB, K = 2J = 8, and the exponential-shaped model for clutter temporal correlation.

Results of a second set of simulation runs with identical parameters and scenario conditions, except for the ground clutter temporal correlation model, are presented in Tables A-3 and A-4. For these tables the Gaussian-shaped model was adopted. Table A-3 results are for K = 2JN = 256, and Table A-4 results are for K = 2J = 8. Thus, Tables A-3 and A-4 relate to Tables A-1 and A-2, respectively.

Examination of the results in Tables A-3 and A-4 by themselves leads to identical conclusions as obtained for Tables A-1 and A-2, with the distinction that the performance of the PAMF-LS is closer to that of the MF. A comparison of Tables A-3 and A-4 with Tables A-1 and A-2, respectively, indicates that both PAMF-LS implementations perform better for the Gaussian-shaped clutter temporal correlation model for both values of K. This is in contrast to results observed in previous analyses involving other PAMF implementations. Specifically, both the PAMF in conjunction with the Strand-Nuttal AR model identification algorithm [Nuttall, 1976; Strand, 1977] and the PAMF in conjunction with the canonical correlations state variable model identification algorithm exhibit degraded performance for the Gaussian-shaped clutter temporal correlation model. Other simulation-based analyses carried out by SSC further support these observations and conclusions.

CASE		MF	CFAR AMF		PAMF-LS COEFF AVG (P = 3)		PAMF-LS COVAR AVG (P = 3)	
CRAB ANGLE (deg)	NO. OF JAMMERS	P_D	P_D	SD[P_D]	P_D	SD[P_D]	P_D	SD[P_D]
0	0	0.8635	0.4654	0.0507	0.8360	0.0113	0.8383	0.0103
20	0	0.8635	0.4622	0.0495	0.7925	0.0173	0.8007	0.0111
0	2	0.8635	0.4643	0.0498	0.8269	0.0129	0.8290	0.0114

Table A-3. Detection performance of Options 1 and 4 of the PAMF-LS for the three simulation cases with SINR = 9 dB, K = 2JN = 256, and the Gaussian-shaped model for clutter temporal correlation.

CASE		MF	CFAR AMF (K = 2JN)		PAMF-LS COEFF AVG (P = 3)		PAMF-LS COVAR AVG (P = 3)	
CRAB ANGLE (deg)	NO. OF JAMMERS	P_D	P_D	$SD[P_D]$	P_D	$SD[P_D]$	P_D	$SD[P_D]$
0	0	0.8635	0.4654	0.0507	0.7829	0.0670	0.8054	0.0517
20	0	0.8635	0.4622	0.0495	0.7417	0.0937	0.7932	0.0699
0	2	0.8635	0.4643	0.0498	0.7798	0.0724	0.8035	0.0626

Table A-4. Detection performance of Options 1 and 4 of the PAMF-LS for the three simulation cases with $SINR = 9\,dB$, $K = 2J = 8$, and the Gaussian-shaped model for clutter temporal correlation.

In closing, it is important to mention that the AR model order value utilized herein is the same value utilized in [Román et al., 2000], in order to allow direct comparison between both sets of results. Thus, the model order has been fixed without the benefit of optimization for the covariance-averaging algorithmic option and the Gaussian-shaped model for the ground clutter temporal correlation. Early results from an on-going simulation-based analysis indicate that $P = 2$ is possibly a better choice. In particular, for the $\gamma = 20\,deg$ case in Table A-3, $P_D = 0.8236$ and $SD[P_D] = 0.0101$ using $P = 2$, and for the $\gamma = 20\,deg$ case in Table A-4, the detection and variability results are very similar using $P = 2$. Based on statistical considerations, the lowest-order model that gives acceptable performance should be adopted.

REFERENCES

P. Beckmann

(1967) Probability in Communication Engineering, Harcourt, Brace & World, Inc., New York, NY.

L. E. Brennan and I. S. Reed

(1973) "Theory of adaptive radar," IEEE Transactions on Aerospace and Electronic Systems, Vol. AES-9, No. 2 (March), pp. 237-252.

L. Cai and H. Wang

(1990) "On adaptive filtering with the CFAR feature and its performance sensitivity to non-Gaussian interference," Procs. of the 24th Annual Conference on Information Sciences Computer Siences, pp. 558-563.

W.-S. Chen and I. S. Reed

(1991) "A new CFAR detection test for radar," Digital Signal Processing, Vol. 1, No. 1 (January), pp. 198-214.

E. Conte, M. Lops, and G. Ricci

(1995) "Asymptotically optimum radar detection in compound-Gaussian clutter," IEEE Transactions on Aerospace and Electronic Systems, Vol. 31, No. 2 (April), pp. 617-625.

(1996) "Adaptive matched filter detection in spherically invariant noise," IEEE Signal Processing Letters, Vol. 3, No. 8 (August), pp. 248-250.

D. E. Dudgeon and R. M. Mersereau

(1984) Multidimensional Digital Signal Processing, Prentice-Hall, Englewood Cliffs, NJ.

J. S. Goldstein and I. S. Reed
(1997) "Theory of partially adaptive radar," IEEE Transactions on Aerospace and Electronic Systems, Vol. 33, No. 4 (October), pp. 1309-1325.

J. S. Goldstein, I. S. Reed, and L. L. Sharf
(1998) "A multistage representation of the Wiener filter based on orthogonal projections," IEEE Transactions on Information Theory, Vol. 44, No. 7 (November), pp. 2943-2959.

G. H. Golub and C. F. Van Loan
(1989) Matrix Computations, Second Edition, The Johns Hopkins University Press, Baltimore, MD.

J. R. Guerci and E. H. Feria
(1996) "Application of a least-squares predictive-transform modeling methodology to space-time adaptive array processing," IEEE Transactions on Signal Processing, Vol. 44, No. 7 (July), pp. 1825-1833.

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

A. G. Jaffer, M. H. Baker, W. P. Ballance, and R. J. Staub
(1991) Adaptive Space-Time Processing Techniques for Airborne Radar, RL Technical Report No. RL-TR-91-162, Rome Laboratory, Rome, NY.

E. J. Kelly
(1986) "An adaptive detection algorithm," IEEE Transactions on Aerospace and Electronic Systems, Vol. AES-22, No. 1 (March), pp. 115-127.

S. L. Marple, Jr.
(1987) Digital Spectral Analysis With Applications, Prentice-Hall, Inc., Englewood Cliffs, NJ.

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, Rome Laboratory Technical Report No. RL-TR-91-269, Rome Laboratory, Rome, NY.

J. H. Michels, T. Tsao, B. Himed, and M. Rangaswamy
(1998) "Space-time adaptive processing (STAP) in airborne radar applications," Procs. of the IASTED International Conf. on Signal Processing and Communications, Canary Islands, Spain, Feb. 11-14, pp. 445-450.

W. B. Mikhael and H. Yu
(1994) "A linear approach for two-dimensional, frequency domain least square signal and system modeling," IEEE Transactions on Circuits and Systems - II: Analog and Digital Signal Processing, Vol. 41, No. 12 (December), pp. 786-795.

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

M. Rangaswamy and J. H. Michels
(1997) "A parametric multichannel detection algorithm for correlated non-Gaussian random processes," Procs. of the IEEE National Radar Conf., Syracuse, NY, May 13-15, pp. 349-354.

F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg
(1992) "A CFAR adaptive matched filter detector," IEEE Transactions on Aerospace and Electronic Systems, Vol. AES-28, No. 1 (January), pp. 208-216.

J. R. Román
(1998a) Adaptive Sidelobe Canceling Using Complex-Valued Canonical Variables, AFRL Technical Report No. AFRL-SN-RS-TR-1998-46, Air Force Research Laboratory, Sensors Directorate, Rome Research Site, Rome, NY (March).
(1998b) CFAR for the PAMF Detector, SSC Technical Communication No. SSC-TC-98-01 for Air Force Research Laboratory Contract No. F30602-96-C-0085, Scientific Studies Corporation, Palm Beach Gardens, FL (March).

J. R. Román and D. W. Davis
(1993a) Multichannel System Identification and Detection Using Output Data Techniques, Rome Laboratory Technical Report No. RL-TR-93-141, Rome Laboratory, Rome, NY.
(1993b) State-Space Models for Multichannel Detection, Rome Laboratory Technical Report No. RL-TR-93-146, Rome Laboratory, Rome, NY.
(1997) Two-Dimensional Processing for Radar Systems, Rome Laboratory Technical Report No. RL-TR-97-127, Rome Laboratory, Rome, NY.

J. R. Román, D. W. Davis, and J. H. Michels

(1997) "Multichannel parametric models for airborne phased array clutter," Procs. of the IEEE National Radar Conf., Syracuse, NY, May 13-15, pp. 72-77.

(1998) "Parametric-based space-time adaptive processing and detection in airborne surveillance radar systems," Procs. of the IASTED International Conference on Signal and Image Processing (SIP '98), Las Vegas, NV, October 28-31, pp. 290-296.

J. R. Román, M. Rangaswamy, D. W. Davis, Q. Z. Zhang, B. Himed, and J. H. Michels

(2000) "Parametric adaptive matched filter for airborne radar applications," IEEE Transactions on Aerospace and Electronic Systems, Vol. 36, No. 2 (April), pp. 677-692.

L. L. Scharf and L. T. McWhorter

(1996) "Adaptive matched subspace detectors and adaptive coherence," Procs. of the 30th Asilomar Conf. on Signals, Systems, and Computers, Pacific Grove, CA, Nov.

O. N. Strand

(1977) "Multichannel complex maximum entropy (auto-regressive) spectral analysis," IEEE Transactions on Automatic Control, Vol. AC-22, No. 4, pp. 634-640.

C. W. Therrien

(1981) "Relations Between 2-D and Multichannel Linear Prediction," IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-29, No. 3 (June), pp. 454-456.

(1986) The Analysis of Multichannel Two-Dimensional Random Signals, Naval Postgraduate School Technical Report No. NPS62-87-002, Naval Postgraduate School, Monterey, CA.

J. B. Thomas

(1969) An Introduction to Statistical Communication Theory, J. Wiley & Sons, Inc., New York, NY.

J. Ward

(1994) Space-Time Adaptive Processing for Airborne Radar, Technical Report No. TR-1015 (December), contract no. F19628-95-C-0002, Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA.

V. J. Yohai and M. S. Garcia Ben

(1980) "Canonical variables as optimal predictors," The Annals of Statistics, Vol. 8, No. 4 (July), pp. 865-869.

*U.S. GOVERNMENT PRINTING OFFICE: 2002/710-038/10200

DEPARTMENT OF THE AIR FORCE
AIR FORCE RESEARCH LABORATORY (AFMC)

15 Jun 04

MEMORANDUM FOR DTIC-OCQ
ATTN: Larry Downing
Ft. Belvoir, VA 22060-6218

FROM: AFRL/IFOIP

SUBJECT: Distribution Statement Change

1. The following documents (previously limited by SBIR data rights) have been reviewed and have been approved for Public Release; Distribution Unlimited:

ADB226867, "Multichannel System Identification and Detection Using Output Data Techniques", RL-TR-97-5, Vol 1.

ADB176689, "Multichannel System Identification and Detection Using Output Data Techniques", RL-TR-93-141.

ADB198116, "Multichannel Detection Using Higher Order Statistics", RL-TR-95-11.

ADB232680, "Two-Dimensional Processing for Radar Systems", RL-TR-97-127.

ADB276328, "Two-Dimensional Processing for Radar Systems", AFRL-SN-RS-TR-2001-244.

2. Please contact the undersigned should you have any questions regarding this memorandum. Thank you very much for your time and attention to this matter.

KELLY CAGWIN
STINFO Officer
Information Directorate
315-330-7094/DSN 587-7094