AFIT/DS/ENG/99-04

DUAL CHANNEL MATCHED FILTERING
AND
SPACE-TIME ADAPTIVE PROCESSING

DISSERTATION

Scott D. Berger
Major, USAF

AFIT/DS/ENG/99-04

Approved for public release; distribution unlimited

The views expressed in this dissertation are those of the author and do not reflect the official policy or position of the Department of Defense or the United States Government.

AFIT/DS/ENG/99-04

DUAL CHANNEL MATCHED FILTERING
AND
SPACE-TIME ADAPTIVE PROCESSING

DISSERTATION

Presented to the Faculty of the Graduate School of Engineering

of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

Scott D. Berger, B.S.E.E.,M.S.E.E.

Major, USAF

June, 1999

Approved for public release; distribution unlimited

AFIT/DS/ENG/99-04

DUAL CHANNEL MATCHED FILTERING

AND

SPACE-TIME ADAPTIVE PROCESSING

Scott D. Berger, B.S.E.E.,M.S.E.E.

Major, USAF

Approved:

Dr. Byron M. Welsh Chairman, Advisory Committee	Date
Dr. Mark E. Oxley Member, Advisory Committee	Date
Dr. Vittal P. Pyati Member, Advisory Committee	Date
Dr. Yung K. Yeo Dean's Representative	Date

Dr. Robert A. Calico, Jr.
Dean, Graduate School of Engineering

Acknowledgements

I would first like to thank my advisor Dr. Byron Welsh for his outstanding guidance, strong sense of commitment to me, continuous words of encouragement, and scholarly fellowship which made this endeavor enjoyable. I would also like to thank the other members of my committee, Dr. Vittal Pyati and Dr. Mark Oxley, for their time and assistance. I extend a special thanks to Mr. Dave Doak for his fine computer support, Dr. William Baker for the many lively mathematical discussions, and the other numerous professors who took an interest and the time to respond to my e-mails. Its nice to know that people are still willing to help a stranger.

I am indebted to my sponsors at the Air Force Office of Scientific Research, Dr. Arje Nachman and Dr. Jon Sjogren, and Capt. Todd Hale of the Air Force Research Laboratory for their monetary support and review of my dissertation.

On a personal level, I would like to thank my family, especially my wonderful wife, Patti, for her support, encouragement, and understanding throughout this three year endeavor. Finally, I would like to dedicate this dissertation to my father, Allen C. Berger, who through example, showed me the importance of hard work and dedication and who continually pushed me to pursue higher levels of education.

Scott D. Berger

Table of Contents

Page

Acknowledgements iii

List of Figures vii

List of Tables x

Abstract xi

I. Introduction 1-1

1.1 Airborne Radar Surveillance 1-1

1.2 Space-Time Adaptive Processing 1-4

1.3 Previous Research 1-10

1.3.1 Non-Data Adaptive Transformations 1-11

1.3.2 Data Adaptive Transformations 1-13

1.3.3 Transformation Summary 1-16

1.4 Block Diagonalizing the Correlation Matrix 1-18

1.4.1 Block STAP Concept 1-18

1.4.2 Research Objectives 1-20

1.5 Organization 1-21

II. Dual Channel Secondary Data Requirements 2-1

2.1 Introduction 2-1

2.2 Notation and Previous Results 2-2

2.3 Dual Channel Normalized SINR 2-4

2.4 Reduced Secondary Data Requirements 2-12

2.5 Practical Aspects 2-15

2.6 Simulation Example 2-18

2.7 Summary 2-21

Page

III. Block Diagonalizing STAP Correlation Matrices 3-1

3.1 Introduction 3-1

3.2 STAP Correlation Matrices 3-2

3.3 DKLT and Simultaneous Diagonalization 3-8

3.3.1 Uniqueness of the DKLT 3-9

3.3.2 Diagonalizing the Family $\mathcal{S}$ 3-11

3.4 Invariant Subspaces 3-14

3.5 Simultaneous Block Diagonalization of the Family $\mathcal{S}$ 3-19

3.5.1 Observations and Basic Concepts 3-20

3.5.2 Quasi-Diagonalization 3-24

3.5.3 General Case 3-28

3.6 Centrosymmetric Clutter 3-34

3.6.1 Infinite Number of Clutter Patches and Constant Power 3-36

3.6.2 Finite Number of Clutter Patches and Symmetric Power 3-43

3.6.3 Potential Uses 3-46

3.7 Summary 3-47

IV. Transformation Selection 4-1

4.1 Introduction 4-1

4.2 Identity Transformation 4-2

4.3 Approximate Block Diagonalization 4-7

4.4 Similar Weight Vectors 4-9

4.5 Block STAP SINR Efficiency 4-13

4.5.1 Efficiency Analysis 4-15

4.5.2 Criterion Discussion 4-20

4.6 Summary 4-29

Page

V. Reduced-Rank Direct Form Transformation Selection 5-1

5.1 Introduction 5-1

5.2 Cross-Spectral Metric Generalized Sidelobe Canceler 5-2

5.3 SINR Metric Direct Form Processor 5-5

5.4 Simulation Results 5-8

5.5 Limitations and Relevance to Block STAP 5-13

5.6 Summary 5-19

VI. Conclusions and Recommendations 6-1

6.1 Results and Contributions 6-1

6.2 Directions for Future Research 6-3

Appendix A. CSM of the Noise Subspace Eigenvectors A-1

Bibliography BIB-1

Vita VITA-1

List of Figures

Figure Page

1.1. Pictorial view of the interference environment seen by an airborne surveillance radar. 1-5

1.2. Block diagram of a reduced-dimension (rank) direct form processor. 1-11

1.3. Block diagram of a reduced-dimension (rank) general sidelobe canceler. 1-17

1.4. Block diagram of the Block STAP processor. 1-20

2.1. The dual and single channel systems are equivalent in terms of SINR performance if the matrix $\mathbf{V} = [\mathbf{V}_1 \ \mathbf{V}_2]$ block diagonalizes the true correlation matrix $\mathbf{R}$. 2-2

2.2. Sample mean of ρ_{dual} (i.e., $\mathrm{E}\{\rho_{\text{dual}}\}$) based on 10,000 samples for each N and k overlaid with the approximate $\mathrm{E}\{\rho_{\text{dual}}\}$ computed using Eqn. (2.42) and $K = 2N$. The error bars are approximate 99.5% confidence intervals. 2-13

2.3. Sample Variance of ρ_{dual} (i.e., $\mathrm{V}\{\rho_{\text{dual}}\}$) based on 10,000 samples for each N and k overlaid with the approximate $\mathrm{V}\{\rho_{\text{dual}}\}$ computed using Eqn. (2.43) and $K = 2N$. 2-14

2.4. Cumulative probability distribution of the normalized SINR for the single channel and dual channel systems in a centrosymmetric interference plus noise environment. 2-20

3.1. Numerical evaluation of the standard identity of degree MN (first element of the first column only, see Eqn. (3.82)) as a function of β with the parameters listed in Table 3.1. 3-35

4.1. SINR performance of the identity matrix and DFT matrix configurations relative to the optimum processor. 4-9

4.2. SINR performance of the identity matrix and centrosymmetric matrix configurations relative to the optimum processor. 4-13

Figure Page

4.3. Plots showing the regions of the angle-Doppler plane where the SINR and $\|\mathbf{e}\|_2$ of the identity matrix configuration simultaneously exceed those of the centrosymmetric matrix configuration. 4-14

4.4. Eigenvalues of $\mathbf{G}$ sorted in ascending order for the identity, DFT, and centrosymmetric configurations, where the interference plus noise environment consisted of receiver noise, clutter, and three barrage noise jammers. 4-24

4.5. SINR performance of the identity matrix and centrosymmetric matrix configurations relative to the optimum processor, where the interference plus noise environment consisted of receiver noise, symmetric clutter, and anti-symmetric clutter with a gain of either 0 dB or 20 dB above the symmetric clutter patches. 4-26

4.6. Eigenvalues of $\mathbf{G}$ sorted in ascending order for the identity and centrosymmetric configurations, where the interference plus noise environment consisted of receiver noise, symmetric clutter, and anti-symmetric clutter with a gain of either 0 dB or 20 dB above the symmetric clutter patches. 4-27

5.1. Block diagram of a reduced-dimension (rank) generalized sidelobe canceler. 5-2

5.2. Block diagram of a reduced-dimension (rank) direct form processor. 5-5

5.3. SINR Performance of the SINR metric, Small, CSM GSC, and PC GSC methods as a function of the transformation rank for Case 1: SNR=15 dB, Jammers at normalized angles and JNRs of (0.25, 40 dB) and (±0.433, 30 dB), and CNR=55 dB. 5-14

5.4. SINR Performance of the SINR metric, Small, CSM GSC, and PC GSC methods as a function of the transformation rank for Case 4: SNR=15 dB, Jammers at normalized angles and JNRs of (0.25, 0 dB) and (±0.433, 0 dB), and CNR=20 dB. 5-14

5.5. SINR Performance of the SINR metric and CSM GSC methods as a function of the transformation rank for the high clutter and jamming environment (Cases 1-3). 5-15

Figure Page

5.6. SINR Performance of the SINR metric and CSM GSC methods as a function of the transformation rank for the low clutter and jamming environment (Cases 4-6). 5-15

5.7. Eigenvalues of $\mathbf{R}$ and SINR metric for each eigenvector of $\mathbf{R}$ for Case 1: SNR=15 dB, Jammers at normalized angles and JNRs of (0.25, 40 dB) and ($\pm$0.433, 30 dB), and CNR=55 dB. 5-16

5.8. Eigenvalues of $\mathbf{R}_b$ and CSM for each eigenvector of $\mathbf{R}_b$ for Case 1: SNR=15 dB, Jammers at normalized angles and JNRs of (0.25, 40 dB) and ($\pm$0.433, 30 dB), and CNR=55 dB. 5-16

5.9. Eigenvalues of $\mathbf{R}$ and SINR metric for each eigenvector of $\mathbf{R}$ for Case 4: SNR=15 dB, Jammers at normalized angles and JNRs of (0.25, 0 dB) and ($\pm$0.433, 0 dB), and CNR=20 dB. 5-17

5.10. Eigenvalues of $\mathbf{R}_b$ and CSM for each eigenvector of $\mathbf{R}_b$ for Case 4: SNR=15 dB, Jammers at normalized angles and JNRs of (0.25, 0 dB) and ($\pm$0.433, 0 dB), and CNR=20 dB. 5-17

List of Tables

Table Page

2.1. Predicted and simulated normalized SINR performance for the single channel and dual channel systems in a centrosymmetric interference plus noise environment. 2-20

3.1. Spatial frequencies used in the numerical evaluation of the standard identity of degree MN (See Eqn. (3.82)). 3-34

4.1. Simulation parameters. 4-9

4.2. Summary of simulation results for the identity, DFT, and centrosymmetric matrix configurations, where the interference plus noise environment consisted of receiver noise, clutter, and three barrage noise jammers. 4-23

4.3. Summary of simulation results for the identity and centrosymmetric matrix configurations, where the interference plus noise environment consisted of receiver noise, symmetric clutter, and anti-symmetric clutter with a gain of either 0 dB or 20 dB above the symmetric clutter patches. 4-28

5.1. Simulation parameters, where SNR, JNR, and CNR denote the signal-to-receiver noise, jammer-to-receiver noise, and total clutter-to-receiver noise ratios, respectively. 5-9

AFIT/DS/ENG/99-04

Abstract

We propose a dual channel matched filtering system that addresses two key challenges in the practical implementation of a single channel matched filtering system: secondary data support and computational cost. We derive an exact expression of the dual channel normalized signal-to-interference plus noise ratio (SINR) in terms of random variables with known distributions and approximate expressions of the mean and variance of the normalized SINR. Using these approximate expressions, we demonstrated that the dual channel system requires half the secondary data to achieve nearly the same SINR performance as an equivalent single channel system. With the dual channel system, two reduced dimension weight vectors are used in place of the larger single channel weight vector, offering the potential reduction in computational cost. The key to the dual channel system is the efficient block diagonalization of the interference plus noise correlation matrix with a fixed transformation. The dual channel system is a viable replacement for a single channel system in applications involving real, wide-sense stationary random processes. We investigate the application of this dual channel concept to the problem domain of space-time adaptive processing (STAP), referring to the system as Block STAP. We provide evidence that the family of STAP correlation matrices cannot be simultaneously block diagonalized with a fixed transformation and thus, the implementation of the Block STAP processor will be suboptimal. We propose a transformation selection criterion for minimizing the loss in SINR performance of a suboptimal Block STAP processor. Finally, we introduce the SINR metric and a new eigen-based, reduced-rank direct form STAP processor based on the SINR metric. The SINR metric is used to identify the eigenvectors of the correlation matrix that have the greatest impact on SINR performance of a direct form processor. If the rank reduction transformation is constructed from r eigenvectors of the correlation matrix, then the r eigenvectors with the largest SINR are the optimal set of eigenvectors in terms of minimizing the loss in SINR performance of an eigen-based, reduced-rank direct form processor.

DUAL CHANNEL MATCHED FILTERING
AND
SPACE-TIME ADAPTIVE PROCESSING

I. Introduction

1.1 Airborne Radar Surveillance

Effective airborne radar surveillance of the airspace above a battlefield is critical for both offensive and defensive operations. Without effective surveillance, friendly forces are blind to the number, position, and intention of enemy aircraft. Airborne radar surveillance provides an all weather detection (and location) capability against hostile aircraft with a greater radar horizon and less terrain masking effects than ground-based radar systems. Thus, an airborne radar system has the potential to detect hostile aircraft at longer ranges in comparison to a ground-based radar system. The earlier (long range) detection of hostile aircraft provides additional time to plan, coordinate, and allocate resources to effectively engage or avoid enemy forces. The detection of airborne targets represents a difficult problem, especially when the targets are slow, low-flying aircraft with small radar cross sections. The difficulty is further compounded when the enemy forces use electronic countermeasures (jamming) against the radar system. The source of the difficulty in detecting airborne targets is the presence of unwanted signals in the radar receiver.

The problem of detecting a target is essentially a power problem. In general, the power of the signal reflected by the target must exceed the power of the unwanted signals in the radar receiver for detection to occur. The unwanted signals can be divided into two categories: receiver noise and interference. Ever present and constant for a given radar system, receiver noise is generated by the radar components and ultimately, limits the detection performance of the radar system (i.e., the power of the receiver noise is the minimum power that target signal must exceed for detection). In contrast, interference signals vary in the number, type, and power level depending on the operating environment.

Interference signals are further divided into two categories: ground clutter and jamming. The signals reflected by the ground are referred to as ground clutter and enter the radar receiver through the mainlobe and sidelobes of the radar antenna. Mainlobe clutter has a relatively narrow bandwidth and is centered at a Doppler frequency determined by the velocity vector of the radar platform and pointing angle of the antenna. The mainlobe clutter also has a narrow angular spread as determined by the antenna beamwidth. The power level of mainlobe clutter is, in general, high because of the high gain of the antenna mainlobe and the large area of the ground illuminated by the mainlobe. Sidelobe clutter, in contrast, has a wide angular extent and hence, has a bandwidth that fills the entire Doppler processing bandwidth of the radar, since the antenna has sidelobes in all directions. The power level of sidelobe clutter is typically less than mainlobe clutter because of the reduced gain of the antenna sidelobes. However, the power level of the sidelobe clutter, in many cases, is relatively high in comparison to the target return power level. Jamming can also be divided into two categories: deception and noise. The goal of deception jamming to obscure the true target or overload the radar processor or radar operator with false targets. Deception jamming is typically employed to defeat the tracking capabilities of a radar system. A noise jammer transmits a signal that resembles the receiver noise with the goal of raising the receiver noise floor to prevent the detection of the target. Because noise jamming is typically employed against surveillance radars which are the focus of this research, we will not consider deception jamming. In general, noise jamming is a high power signal with a bandwidth that covers the bandwidth of the radar receiver and has a narrow angular extent. The objective of the radar engineer is to design the radar components and signal processing algorithms to eliminate the interference signals so that the target signal only competes with the receiver noise. The possibility of eliminating or minimizing the effects of the interference signals is severely limited in a conventional radar system, which we define as a radar system with a single element antenna and that uses Doppler processing.

The types of interference signals that must be eliminated will depend on the operating environment and the position and velocity of the target relative to the radar platform. Mainlobe clutter is a primary concern in look-down scenarios (i.e., the target's altitude

is less than the altitude of the radar platform). If the target has sufficient altitude, then the mainlobe clutter and target return will be well separated in time and range gating can be used to isolate the target return from the mainlobe clutter. At lower altitudes, the target return and mainlobe clutter will be nearly time coincident and hence, range gating cannot be used to isolate the target return from the mainlobe clutter (i.e., both signals fall within the same range gate). When the target return and mainlobe fall within the same range gate, Doppler processing can potentially be used to isolate the target return and mainlobe clutter. Because the ground and target are moving relative to the radar platform, the mainlobe clutter and target return signals will each experience a Doppler shift. A series of fixed filters, referred to as Doppler filters, are used to divide the Doppler processing bandwidth into several bands with a detection decision made in each band. When the difference between the relative velocity of the target and ground is large, the target return and mainlobe clutter will fall into different Doppler filters that are well separated, effectively isolating the two signals. If the target and ground have the same relative velocity, Doppler processing cannot isolate the two signals. Further, when the target and ground have nearly the same relative velocity, the mainlobe clutter can leak into the target Doppler filter through the filter's sidelobe and obscure the target. Although Doppler processing provides a means for eliminating or minimizing mainlobe clutter, it is not as effective against sidelobe clutter and noise jamming. The bandwidth of the sidelobe clutter and noise jamming is typically equal to or greater than the Doppler processing bandwidth and thus, Doppler filtering will not isolate the target return from the sidelobe clutter and noise jamming. A common approach used by radar engineers to counter the effects of sidelobe clutter and noise jamming is to trade mainlobe antenna gain for lower sidelobe levels which effectively reduces the power levels of the sidelobe clutter and noise jamming entering the radar receiver through the antenna sidelobes. However, reducing the mainlobe gain also reduces the target return power which could adversely effect detection. Additionally, maintaining the low sidelobe levels on an installed antenna is problematic. The presence of mainlobe clutter leakage and inability of Doppler processing to eliminate sidelobe clutter and noise jamming have a severe impact on the detection performance of a conventional radar system.

As noted earlier, the clutter and noise jamming signals also have an angular dependence which provides another domain for isolating the target signal from the interference signals. Shown in Fig. 1.1 is a pictorial view of the interference environment, highlighting the angular and Doppler dependencies of the interference signals. Recall that the Doppler shift of a particular patch of the ground is dependent on its relative velocity which is, in turn, dependent on the angle between the ground patch and the velocity vector of the radar platform. Thus, the clutter is concentrated on a line running across the angle-Doppler plane. The noise jamming is concentrated at a particular angle, but fills the entire Doppler processing bandwidth. The angular dependence of the interference signals cannot be exploited in a radar system that uses a single element antenna since the angle information is not available and the antenna pattern is fixed. In a radar system with a multi-element antenna, the angular dependence of the interference signals can be exploited since the angle information is encoded in the phase difference between the signals received in each antenna element. The advent of high-speed digital signal processors and multi-element (array) antennas has provided the radar engineer with the opportunity to develop new signal processing algorithms that more effectively eliminate the interference signals by exploiting both the Doppler and angular dependencies. These new algorithms are essentially a simultaneous combination of adaptive array (spatial) and adaptive Doppler (temporal) processing and are referred to as space-time adaptive processing.

1.2 Space-Time Adaptive Processing

Conceptually, space-time adaptive processing (STAP) is a two-dimensional (spatial and temporal) filtering operation. To eliminate the interference signals while enhancing the target signal, the filter has nulls in the direction of the interference signals and gain in the direction of the target signal. Under the assumption that the interference and noise are Gaussian signals, Brennan and Reed [6] have shown that maximizing the signal-to-interference plus noise ratio (SINR) is equivalent to maximizing the probability of detection. Further, Brooks and Reed [7] have shown that the maximum SINR filter, likelihood ratio processor, and minimum variance linear signal estimator (Wiener filter) only differ by a constant under the same Gaussian assumption. Thus, the objective is to

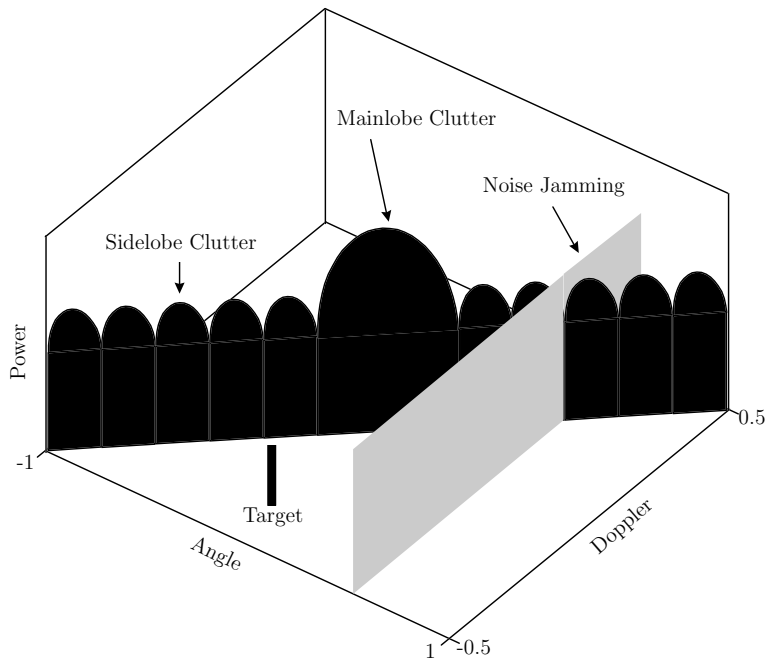


Figure 1.1 Pictorial view of the interference environment seen by an airborne surveillance radar.

construct a filter with nulls in the direction of the interference and gain in the direction of the target that maximizes the SINR. Because the radar system operates in a wide variety of interference environments, the filter must adapt to the interference environment to achieve the objective of maximizing the SINR. As discussed next, the maximum SINR (optimum) filter is constructed from the interference plus noise correlation matrix, which contains both the spatial and temporal information needed to cancel the interference, and a steering vector that defines the direction (target signal) of interest.

The airborne surveillance radar under consideration has a uniform linear array of N equally spaced antenna elements where the output from each antenna element is sampled after in-phase and quadrature down conversion to baseband. Since the primary function of the radar system is surveillance, the dwell time in any particular direction is finite which limits the number of samples available for each coherent processing interval (CPI). For a given range gate (range of interest) and CPI, the radar system processes M samples (pulses) from each of the antenna elements. Initially, the samples are stored in a $M \times N$

data matrix $\mathbf{X}$ which has the following form:

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,N} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ x_{M,1} & x_{M,2} & \cdots & x_{M,N} \end{bmatrix}, \tag{1.1}$$

where $x_{m,n}$ represents the complex sample from the mth pulse in the nth antenna element. The data matrix is then reorganized into a $MN \times 1$ data vector $\mathbf{x}$ such that the first N samples are the N antenna samples from the first pulse, the next N samples are from the second pulse, and so forth until the last N samples are from the Mth pulse. That is,

$$\mathbf{x}^T = \begin{bmatrix} x_{1,1} & \cdots & x_{1,N} & x_{2,1} & \cdots & x_{2,N} & \cdots & x_{M,1} & \cdots & x_{M,N} \end{bmatrix}, \tag{1.2}$$

where $\{\cdot\}^T$ denotes transpose. The data vector $\mathbf{x}$ represents the received signal which may or may not contain a target. When a target is present, the received signal is given by

$$\mathbf{x} = \alpha\mathbf{s} + \mathbf{n}, \tag{1.3}$$

where $\alpha\mathbf{s}$ is the target return signal, with the complex gain α defining the amplitude and initial phase of the target signal and the complex steering vector $\mathbf{s}$ defining the angle and Doppler directions to the target, and $\mathbf{n}$ is the interference plus noise signal vector. Obviously, in the absence of a target, the received signal is given by

$$\mathbf{x} = \mathbf{n}. \tag{1.4}$$

To detect the presence or absence of a target, the received signal is passed through a filter and then, the filter's output is compared to a given threshold value. The output of the filter defined by the weight vector $\mathbf{w}$ is given by

$$y = \mathbf{w}^H\mathbf{x}, \tag{1.5}$$

where $\{\cdot\}^H$ denotes complex conjugate transpose. If the magnitude of the filter output exceeds the threshold, then a target detection is declared. As noted earlier, the filter should maximize the SINR to maximize the probability of detection.

Assuming the interference and noise signals are zero-mean random processes, the SINR at the output of the filter with a target present is defined as [34]

$$\mathrm{SINR} = \frac{|\mathrm{E}\{y\}|^2}{\mathrm{V}\{y\}} = \frac{|\alpha|^2|\mathbf{w}^H\mathbf{s}|^2}{\mathbf{w}^H\mathbf{R}\mathbf{w}}, \tag{1.6}$$

where $\mathrm{E}\{\cdot\}$ and $\mathrm{V}\{\cdot\}$ denote the expected value and variance operations, respectively and $\mathbf{R}$ denotes the interference plus noise correlation matrix which is defined as

$$\mathbf{R} = \mathrm{E}\left\{\mathbf{n}\mathbf{n}^H\right\}. \tag{1.7}$$

The optimum filter (weight vector) in terms of maximizing the output SINR is given by [34]

$$\mathbf{w}_{opt} = \mathbf{R}^{-1}\mathbf{s}. \tag{1.8}$$

Note that the maximum SINR filter is also known as a matched filter. Substituting the optimum weight vector defined by Eqn. (1.8) into Eqn. (1.6) yields a maximum SINR of

$$\mathrm{SINR}_{\max} = |\alpha|^2\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s}. \tag{1.9}$$

Note that we will often incorporate the parameter α into the steering vector to simplify the discussion or presentation. The computation of the weight vector via Eqn. (1.8) is referred to as the direct matrix inversion (DMI) method. Two problems are immediately apparent with the DMI method. First, the computation of the optimum weight requires the inversion of the interference plus noise correlation matrix (or equivalently, finding the solution to the system of linear equations $\mathbf{R}\mathbf{w} = \mathbf{s}$) which is computationally expensive, especially when the dimension of $\mathbf{R}$ is large. If the radar system has N antenna elements and M samples per CPI, then the computation of the optimum weight vector requires on the order of $(MN)^3$ operations. Radar detection is a real-time application and the

computational cost of computing a new weight vector for each range gate and look-direction may become prohibitive as the product MN becomes large. Second, the DMI method requires knowledge of the true interference plus noise correlation matrix. Because the radar system operates in a variety of environments which are not known a priori, the true interference plus noise correlation matrix is not known a priori. To circumvent this problem, an estimate of the interference plus noise correlation matrix is often used in place of the true correlation matrix which introduces additional challenges.

When an estimate of the interference plus noise correlation matrix is used in place of the true correlation matrix, the weight vector is computed as

$$\hat{\mathbf{w}}_{opt} = \hat{\mathbf{R}}^{-1}\mathbf{s}, \tag{1.10}$$

where $\hat{\mathbf{R}}$ denotes the estimated correlation matrix. The computation of the weight vector via Eqn. (1.10) is referred to as the sample matrix inversion (SMI) method. Note that like the DMI method, the SMI method is also a computationally demanding task. Unlike the DMI method, the SMI method produces a random weight vector and hence, the output SINR is a random variable. Since output SINR from the SMI method is random, the SINR performance must be expressed in statistical terms. The average SINR performance of the SMI method is less than the DMI (optimal) SINR performance. The expected loss in SINR performance will depend on the quality of the estimated interference plus noise correlation matrix. Typically, the correlation matrix is estimated from a set of signal vectors, referred to as secondary data vectors, which only contain the interference plus noise signal. These secondary data vectors are obtained from range gates that surround the range gate of interest. To assess the performance of the SMI method and determine the secondary data requirements, Reed et. al. [34] defined two statistics: the SINR conditioned on $\hat{\mathbf{w}}_{opt}$ (or the conditioned SINR) and the normalized SINR. The SINR conditioned on $\hat{\mathbf{w}}_{opt}$ is defined as

$$\text{SINR}|\hat{\mathbf{w}}_{opt} = \frac{|\hat{\mathbf{w}}_{opt}^H\mathbf{s}|^2}{\text{E}\left\{|\hat{\mathbf{w}}_{opt}^H\mathbf{n}|^2\right\}}, \tag{1.11}$$

where the expectation is taken with respect to $\mathbf{n}$. The normalized SINR is defined as

$$\rho_{\text{smi}} = \frac{\text{SINR}|\hat{\mathbf{w}}_{opt}}{\text{SINR}_{\text{max}}} \tag{1.12}$$

which provides an indication of the SINR performance loss of the SMI method relative to the DMI method. When the maximum likelihood estimate of $\mathbf{R}$ is used and the secondary data vectors are independent and identically distributed (i.i.d.) observations of a zero-mean, MN-variate complex normal distribution with a correlation matrix of $\mathbf{R}$, Reed et. al [34] have shown that ρ_{smi} is a beta random variable with parameters $K + 2 - MN$ and $MN - 1$, where K denotes the number of secondary data vectors used to estimate the correlation matrix, N is the number of antenna elements, and M is the number of samples in a CPI. The expected value of ρ_{smi} represents the expected loss in SINR performance of the SMI method relative to the DMI method and is given as

$$\text{E}\left\{\rho_{\text{smi}}\right\} = \frac{K + 2 - MN}{K + 1}. \tag{1.13}$$

A common rule of thumb is that $2MN$ secondary data vectors are required to achieve acceptable performance, which assumes that an average SINR loss of 3 dB (i.e., $\rho_{\text{smi}} = 0.5$) is acceptable. This rule of thumb is derived by setting Eqn. (1.13) equal to 0.5 and solving for K which yields $K = 2MN - 3$. Also observe that if we fixed MN and let K approach infinity, the expected SINR loss is 0 dB (i.e., no loss in SINR performance). Boroson [5] further considered the issue of secondary data requirements, suggesting that the number of secondary data vectors should be increased to $3MN$ or $4MN$ to reduce the probability of having an SINR less than 0.5. Thus, in general, the number of secondary data vector required for acceptable SINR performance is proportional to the product MN. As the product MN becomes large, the requirement for MN secondary data vectors may become prohibitive. The MN secondary data vectors are drawn from MN different range gates. The requirement for a large number of range gates places additional demands on the radar system in terms of increased bandwidth and/or power which may not be available or possible [41]. Additionally, the secondary data should be homogeneous (similar) with the interference plus noise in the range gate of interest. It does not seem reasonable to expect

the interference plus noise from all MN range gates to resemble the interference plus noise in the range gate of interest, especially for large MN. Thus, adequate secondary data support is a key issue with the SMI method along with the high computational cost.

Summarizing, the DMI method produces the optimum filter (weight vector) for maximizing the SINR, but is computationally demanding and requires a priori knowledge of the true interference plus noise correlation matrix. By using an estimated interference plus noise correlation matrix, the SMI method exchanges the problem of knowing the true correlation matrix with the problem of adequate secondary data support. Like the DMI method, the SMI method is computationally demanding. Thus, two of the main research objectives of the STAP community are reducing the computational cost and the secondary data requirements.

1.3 Previous Research

For STAP, the computational cost and the secondary data requirements are proportional to the dimension of the weight vector (i.e., the product MN). Thus, a natural choice for reducing the computational cost and secondary data requirements is to reduce the dimension of the weight vector (filter). A $MN \times 1$ weight vector is said to have MN degrees of freedom (DOF) and in general, MN DOF are not needed to effectively suppress the interference. Reduced-dimension STAP methods attempt to exploit this fact while maintaining optimal or near optimal performance. The underlying concept of reduced-dimension STAP methods is the use of a transformation to reduce the dimension of the received signals. Shown in Fig. 1.2 is a block diagram for implementing reduced-dimension STAP. Suppose we want to reduce the dimension from MN to r, then $\mathbf{V}$ is a $MN \times r$ matrix, referred to as the reduction transformation, that is directly applied to the $MN \times 1$ data vector yielding a $r \times 1$ data vector. The $r \times 1$ weight vector $\mathbf{w}$ is then applied to the reduced-dimension data vector. The structure shown in Fig. 1.2 is referred to as the reduced-dimension direct form processor. Later, in Section 1.3.2, we discuss another implementation structure known as the generalized sidelobe canceler. The research literature available on reduced-dimension STAP is extensive, revealing a wide range of proposed methods. Regardless of the implementation structure, the various reduced-dimension STAP methods can roughly be divided

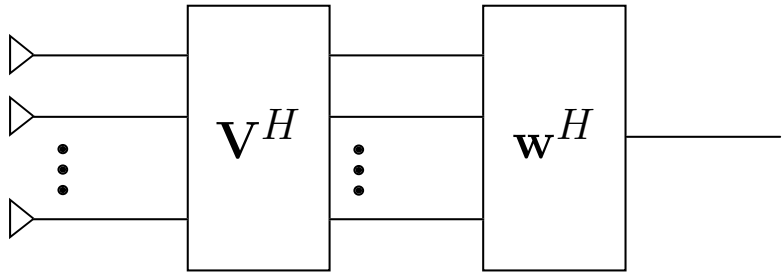


Figure 1.2 Block diagram of a reduced-dimension (rank) direct form processor.

into two categories based on the type of reduction transformation: non-data adaptive and data adaptive.

1.3.1 Non-Data Adaptive Transformations. With non-data adaptive transformation methods, a predetermined transformation is used to preprocess the data vector, yielding a reduced-dimension data vector. The adaptive (spatial and temporal) processing is then accomplished using the reduced-dimension data vector. One advantage of using a predetermined transformation is in the area of computational cost. Since the transformation is known, one can usually select or design a transformation that has an efficient implementation. However, using a predetermined transformation is also a disadvantage. The preprocessing done by the transformation essentially limits the number of DOF available in either the spatial or temporal domains. Because the radar system operates in a wide variety of interference environments, the number of DOF needed for a particular (spatial or temporal) domain are not known a priori. Thus, the potential exists that the predetermined transformation may reduce the available DOF below the level necessary to suppress the interference. Ward [41] provides an extensive review of non-data adaptive, reduced-dimension STAP methods based on the direct form processor. To highlight the concept of non-data adaptive transformation methods, we provide a quick review of two methods presented in Ward's report: element-space pre-Doppler STAP and beamspace pre-Doppler STAP.

In element-space pre-Doppler STAP, the CPI data is partitioned in the temporal (pulse) domain into Z overlapping sub-CPIs with K pulses in each sub-CPI. For the pth

sub-CPI, the reduction (preprocessing) transformation is given by

$$\mathbf{J}_p \otimes \mathbf{I}_N, \tag{1.14}$$

where

$$\mathbf{J}_p = \begin{bmatrix} \mathbf{0}_{p\times K} \\ \mathbf{I}_K \\ \mathbf{0}_{(M-K-p)\times K} \end{bmatrix} \tag{1.15}$$

is a $M \times K$ selection matrix and $\otimes$ denotes the Kronecker product. The notation $\mathbf{I}_m$ denotes a $m \times m$ identity matrix and $\mathbf{0}_{m\times n}$ denotes a $m \times n$ matrix of zeros. Each sub-CPI data vector consist of K pulse returns from all N antenna elements. A $KN \times 1$ weight vector is computed using the SMI method and applied to the $KN \times 1$ sub-CPI data vector. After the adaptive sub-CPI processing, the resulting Z outputs are processed in a conventional Doppler filter bank to extract the target Doppler. Note that the weight vector is computed using a $KN \times KN$ interference plus noise correlation matrix and a $KN \times 1$ steering vector, reducing the overall dimensionality from MN to KN. The computational cost is reduced roughly from $(MN)^3$ to $Z(KN)^3$ and the secondary data requirements from $2MN$ to $2KN$. However, because the CPI data vector is partitioned only in the temporal domain, element-space pre-Doppler STAP has a reduced number of temporal DOF and a full complement of spatial DOF. Thus, element-space pre-Doppler STAP can effectively handle noise jamming, but its capability to handle clutter is reduced.

In beamspace pre-Doppler STAP, the outputs for the antenna elements are combined to form K_s subapertures, effectively reducing the number of antenna elements from N to K_s and the dimension of the weight vector from MN to MK_s. When all the temporal samples are used, the reduction transformation is given by

$$\mathbf{I}_M \otimes \mathbf{G}, \tag{1.16}$$

where $\mathbf{G}$ is the $N \times K_s$(spatial) beamforming matrix. By replacing the identity matrix in Eqn. (1.16) with the matrix in Eqn. (1.15), one can combine element-space pre-Doppler

STAP with beamspace pre-Doppler STAP, leading to a greater reduction in the dimension of the weight vector and available DOF. A wide variety of options are available for selecting the beamforming matrix $\mathbf{G}$ including non-overlapping and overlapping subapertures. With some a priori knowledge of the interference (e.g., angle to each jammer), one can more intelligently select the beamforming matrix to minimize any loss in performance (See Ward's report for a complete discussion). Assuming the use of Eqn. (1.16), a $MK_s \times 1$ weight vector is computed using the SMI method and applied to the $MK_s \times 1$ data vector. The computational cost is reduced roughly from $(MN)^3$ to $(MK_s)^3$ and the secondary data requirements from $2MN$ to $2MK_s$. The effective reduction in the number of antenna elements reduces the spatial DOF available and thus, the capability of beamspace pre-Doppler STAP to handle noise jamming is less than element-space pre-Doppler STAP.

1.3.2 Data Adaptive Transformations. The problems with non-data adaptive transformation methods suggest that one could achieve better performance if the transformation is adapted to the interference environment. Data adaptive transformation methods decompose the interference plus noise correlation matrix to gain insight into the structure of the interference environment and then, use this insight to select the best transformation. Although the use of this insight will, in general, yield a better transformation than non-data adaptive methods, one must be concerned with the additional computational cost introduced by the decomposition. Further, since the transformation is not known a priori and is based on the interference environment, the possibility of efficiently implementing the transformation is low. Thus, one of the main concerns with data adaptive transformation methods is computational cost (i.e., ensuring that the computational cost are less than the full dimension SMI method). Additionally, one must also consider the secondary data support needed for a useful decomposition of the interference plus noise correlation matrix. As noted earlier, the number of DOF needed for a particular interference environment is, in general, less than the dimension of the weight vector. Through the decomposition process, data adaptive transformation methods attempt to determine the number of DOF needed and then, allocate the DOF to suppress the interference with the proper transformation.

We can gain insight into the required number and allocation of the DOF by examining the structure of the interference plus noise correlation matrix.

Eigenspace analysis plays a central role in decomposing the interference plus noise correlation matrix and understanding its structure. The interference plus noise correlation matrix is a positive definite, Hermitian matrix and thus, has a set of MN orthonormal eigenvectors and all the eigenvalues are positive and real [23:104]. The eigenvectors of the interference plus noise correlation matrix form an orthonormal basis of the MN-dimensional vector space over the field of complex numbers. Under the assumption that the interference and receiver noise are uncorrelated, the interference plus noise correlation matrix is simply the sum of the interference correlation matrix and the receiver noise correlation matrix. The receiver noise is typically modeled as spatially and temporally uncorrelated noise and hence, its correlation matrix is the identity matrix times a scalar equal to the variance (power) of the receiver noise. The interference correlation matrix is a positive semidefinite, Hermitian matrix and thus, has a set of MN orthonormal eigenvectors and all the eigenvalues are non-negative and real [23:104]. In fact, the eigenvectors of the interference correlation matrix are the eigenvectors of the interference plus noise correlation matrix, since the receiver noise correlation is the identity matrix times a scalar. Typically, the interference correlation matrix is a low rank matrix [41:83]. Let r denote the rank of the interference correlation matrix, then the interference correlation matrix has r non-zero (large) eigenvalues and $MN - r$ zero (small) eigenvalues. Based on the eigenvalues, the vector space is decomposed into two orthogonal subspaces. One subspace is referred to as the interference subspace and is spanned by the eigenvectors (principal components) associated with the largest eigenvalues (principal values). The other subspace is spanned by the remaining eigenvectors of the interference correlation matrix and is referred to as the noise subspace. The optimum weight vector lies in a subspace spanned by the steering vector (desired signal) and interference subspace eigenvectors [41:83-88]. If the span of the reduction transformation contains the signal plus interference subspace, then reduced-dimension processing provides the same performance as full dimension processing. Thus, the proper identification of the two subspaces plays a critical role in determining the reduction transformation with the rank of the interference subspace determining the

number of DOF needed and defining the space that the transformation must span. If the interference environment is complex and an estimate of the interference plus noise correlation matrix is used, then identifying the two subspaces becomes a difficult task.

Various eigenspace-based STAP methods have been proposed. In a direct application of the fact that the optimum weight vector lies in the signal plus interference subspace, Chang and Yeh [9] suggest selecting the principal components of the estimated signal plus interference plus noise correlation matrix with the number of principal components selected based on an estimate of the number of interference sources. Note by including the range gate of interest, the estimated correlation matrix is the sum of the signal, interference, and receiver noise correlation matrices and the previous discussion on the interference subspace is modified to include the signal (steering) vector. Thus, the vector space is decomposed into the direct sum of the signal plus interference subspace and the noise subspace. In contrast, the principal component inverse (PCI) method proposed by Kirsteins and Tufts [25], the orthogonal projection (OP) method proposed by Subbaram and Abend [37], and the two eigencancelers (minimum power and minimum norm) proposed by Haimovich [18] force the weight vector to lie in the noise subspace. Since the noise subspace is orthogonal to the interference subspace, the weight vector is orthogonal to the interference and thus, will cancel the interference. With the PCI, OP, and minimum norm eigencancelers methods, the projection into the noise subspace is constructed from the principal components (interference subspace). Note that although the above principal component methods do not explicitly use a reduction transformation, one could introduce an appropriate transformation to model these methods. For example, one could model the OP method with the reduced-dimension direct form processor shown in Fig. 1.2 by letting the columns of $\mathbf{V}$ equal the eigenvectors of the noise subspace. The above principal component methods can exhibit a sharp decrease in SINR performance if the number of principal components is underestimated since the selected principal components will not span the entire interference subspace [14].

Recently, Goldstein and Reed [11–14] have proposed a reduced-dimension (rank) generalized sidelobe canceler (GSC) which provides a graceful degradation in performance as the rank of the transformation is reduced below the rank of the interference subspace.

The basic structure of the reduced-dimension GSC is shown in Fig. 1.3. The upper branch forces the GSC to have a response in the spatial and Doppler directions defined by the steering vector $\mathbf{s}$. The lower branch provides an estimate of the noise in the upper branch and the final processing step of subtracting the lower branch from the upper branch reduces the output noise level. The $MN - 1 \times MN$ matrix $\mathbf{B}$, referred to as the blocking matrix, in the lower branch, annihilates the steering vector (i.e., $\mathbf{Bs} = \mathbf{0}$) and has full rank. The blocking matrix prevents the cancellation of signals received in the spatial and Doppler directions defined by the steering vector. The $MN - 1 \times r$ matrix $\mathbf{V}$ is the reduction transformation and the weight vector $\mathbf{w}_{gsc}$ is an unconstrained Wiener filter. Without the rank reduction transformation matrix $\mathbf{V}$, the weight vector for the Wiener filter is given as [14]

$$\mathbf{w}_{gsc} = \mathbf{R}_b^{-1}\mathbf{r}_{bd} \tag{1.17}$$

where $\mathbf{r}_{bd} = \mathbf{BRs}$ and $\mathbf{R}_b = \mathbf{BRB}^H$. One can show that the full dimension GSC and full dimension DMI direct form processor have the same SINR performance [14]. The unique aspect of the method proposed by Goldstein and Reed is the introduction of the output SINR as a cost function into the process of selecting the reduction transformation. With the Goldstein and Reed method, the columns of the reduction transformation are selected based on a cross-spectral metric (CSM) as opposed to selecting the eigenvectors associated with the principal values (largest eigenvalues). The CSM is used to identify the eigenvectors which have the greatest impact on the output SINR. The eigenvectors with the greatest impact become the columns of the reduction transformation. The concept of using the output SINR as a cost function for selecting the reduction transformation is not limited to the GSC. One can derive a metric similar to the CSM for the direct form processor to identify the best (i.e., minimize any loss in SINR performance) eigenvectors for constructing the reduction transformation [4] (See Chapter V).

1.3.3 Transformation Summary. To counter the high computational cost and secondary data requirements of full dimension STAP, which are primarily driven by the dimension of the weight vector, researchers have proposed a wide variety of reduced-

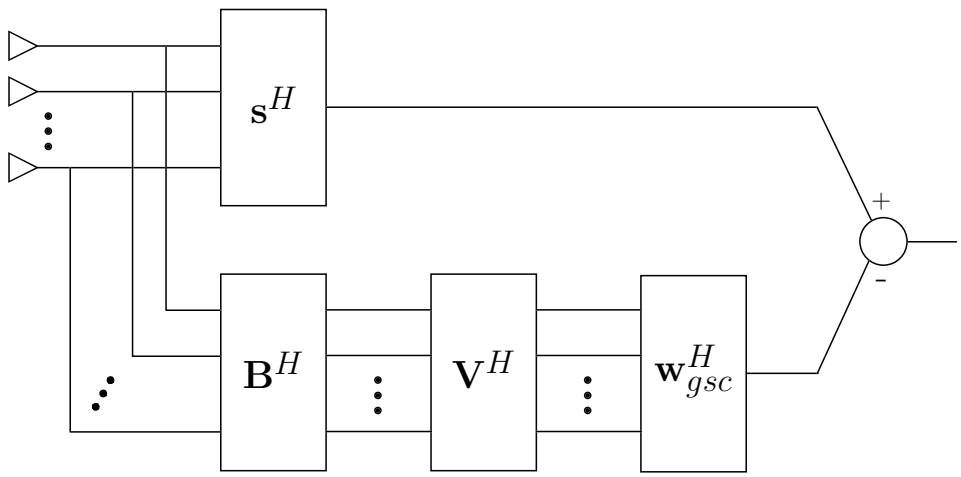


Figure 1.3 Block diagram of a reduced-dimension (rank) general sidelobe canceler.

dimension STAP methods. In general, these reduced-dimension STAP methods either explicitly or implicitly use a transformation to reduce the dimension of the data vector, leading to a reduction in the dimension of the interference plus noise correlation matrix and the weight vector. If the reduction transformation spans the signal plus interference subspace, the reduced-dimension STAP processor can achieve the same SINR performance as the full-dimension STAP processor. With non-data adaptive transformation methods, a predetermined transformation is used to preprocess the full dimension data vector, yielding a reduced dimension data vector. A non-data adaptive transformation provides the opportunity to design a transformation with an efficient implementation, but limits flexibility in terms of allocating the available DOF to counter the various interference environments confronted by the radar system. With data adaptive transformation methods, the transformation is adapted to the interference environment based on a decomposition of the interference plus noise correlation matrix. A data adaptive transformation offers the flexibility to allocate the available DOF, but does not provide the opportunity to design a transformation with an efficient implementation. The data adaptive transformation methods also incur the additional computational costs associated with the decomposition process.

1.4 Block Diagonalizing the Correlation Matrix

In this dissertation, we introduce and investigate a new STAP method based on the block diagonalization of the interference plus noise correlation matrix, that we refer to as Block STAP. Although presented from a radar perspective, the Block STAP concept can be applied to any application requiring a maximum SINR (matched) filter. For general applications, we refer to this proposed method as dual channel matched filtering. By block diagonalizing (two blocks on the diagonal) the interference plus noise correlation matrix, the optimum weight vector can be partitioned into two reduced-dimension weight vectors, where each is the solution to a system of linear equations of reduced dimension and can be computed independently of the other. The computation of the two reduced-dimension weight vectors is computationally less demanding than computing the full dimension weight vector directly. The implementation of the Block STAP processor also has a natural structure for parallel implementation, offering the potential for further computational savings. Each of the reduced-dimension weight vectors is computed with a reduced-dimension interference plus noise correlation matrix, leading to a reduction in secondary data requirements. Thus, the Block STAP method addresses two of the main research objectives, computational cost and secondary data requirements, of the STAP community while providing full dimension SINR performance. This research only considers the case where the block diagonalization of the interference plus noise correlation matrix yields two blocks on the diagonal. One could envision a more elaborate system with recursive block diagonalization yielding K blocks, allowing the optimum weight vector to be partitioned into K reduced-dimension weight vectors and for a greater reduction in the computational cost and secondary data requirements. Thus, one could view this research as the initial step in developing a divide and conquer STAP algorithm.

1.4.1 Block STAP Concept. The development of the Block STAP concept is relative straightforward. Suppose $\mathcal{S}$ is a family of correlation matrices that represent the interference plus noise environments of interest and that the unitary matrix $\mathbf{V}$ simultane-

ously block diagonalizes $\mathcal{S}$. That is,

$$\mathbf{V}^H\mathbf{R}\mathbf{V} = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_2 \end{bmatrix} \quad \text{for all } \mathbf{R} \in \mathcal{S}, \tag{1.18}$$

where $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are square matrices. Now, consider a direct form processor with a preprocessor defined by $\mathbf{V}$. Note that the introduction of $\mathbf{V}$ does not reduce the dimension of received signal vectors or change the output SINR. The optimum weight vector of this direct form process is given by [19]

$$\mathbf{w}_{opt} = \left(\mathbf{V}^H\mathbf{R}\mathbf{V}\right)^{-1}\mathbf{V}^H\mathbf{s} \quad \text{where } \mathbf{R} \in \mathcal{S}. \tag{1.19}$$

If we partition $\mathbf{V}$ such that $\mathbf{V} = [\mathbf{V}_1\ \mathbf{V}_2]$, then Eqn. (1.19) can be written as

$$\begin{aligned}
\mathbf{w}_{opt} &= \begin{bmatrix} \mathbf{V}_1^H\mathbf{R}\mathbf{V}_1 & \mathbf{V}_1^H\mathbf{R}\mathbf{V}_2 \\ \mathbf{V}_2^H\mathbf{R}\mathbf{V}_1 & \mathbf{V}_2^H\mathbf{R}\mathbf{V}_2 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{V}_1^H\mathbf{s} \\ \mathbf{V}_2^H\mathbf{s} \end{bmatrix} = \begin{bmatrix} \mathbf{V}_1^H\mathbf{R}\mathbf{V}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2^H\mathbf{R}\mathbf{V}_2 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{V}_1^H\mathbf{s} \\ \mathbf{V}_2^H\mathbf{s} \end{bmatrix}, \\
&= \begin{bmatrix} \left(\mathbf{V}_1^H\mathbf{R}\mathbf{V}_1\right)^{-1} & \mathbf{0} \\ \mathbf{0} & \left(\mathbf{V}_2^H\mathbf{R}\mathbf{V}_2\right)^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^H\mathbf{s} \\ \mathbf{V}_2^H\mathbf{s} \end{bmatrix} = \begin{bmatrix} \left(\mathbf{V}_1^H\mathbf{R}\mathbf{V}_1\right)^{-1}\mathbf{V}_1^H\mathbf{s} \\ \left(\mathbf{V}_2^H\mathbf{R}\mathbf{V}_2\right)^{-1}\mathbf{V}_2^H\mathbf{s} \end{bmatrix}, \\
&= \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \end{bmatrix},
\end{aligned} \tag{1.20}$$

where we have used the fact $\mathbf{V}$ block diagonalizes every member of $\mathcal{S}$. Thus, we can compute the optimum weight vector by computing the two reduced-dimension weight vectors $\mathbf{w}_1$ and $\mathbf{w}_2$. Finally, let $\mathbf{V}^H\mathbf{x}$ be transformed data vector from the range gate of interest, then the filter output is given as

$$\begin{aligned}
y &= \mathbf{w}_{opt}^H\mathbf{V}^H\mathbf{x} \\
&= \begin{bmatrix} \mathbf{w}_1^H & \mathbf{w}_2^H \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^H\mathbf{x} \\ \mathbf{V}_2^H\mathbf{x} \end{bmatrix} \\
&= \mathbf{w}_1^H\mathbf{V}_1^H\mathbf{x} + \mathbf{w}_2^H\mathbf{V}_2^H\mathbf{x}.
\end{aligned} \tag{1.21}$$

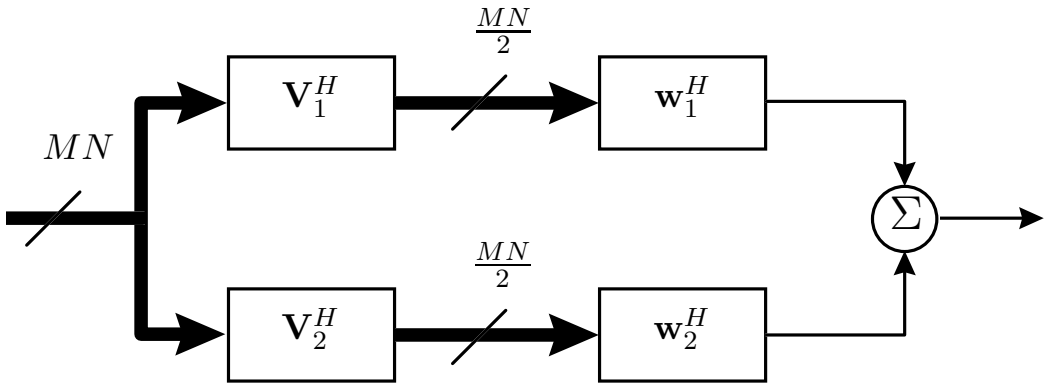


Figure 1.4 Block diagram of the Block STAP processor.

An examination of Eqn. (1.21) leads naturally to the Block STAP processor shown in Fig. 1.4. When the true interference plus noise correlation matrix is unknown, the weight vectors are computed with the SMI method using estimates of the reduced-dimension correlation matrices $\mathbf{V}_1^H\mathbf{R}\mathbf{V}_1$ and $\mathbf{V}_2^H\mathbf{R}\mathbf{V}_2$.

1.4.2 Research Objectives. This research differs from previous STAP research in several aspects. First, the partitioning of the optimum weight vector into two reduced-dimension weight vectors and the implementation structure shown in Fig. 1.4 are based on mathematical principles that preserve optimal SINR performance. In general, with non-data adaptive transformation methods, one expects a loss in SINR performance and designs the transformation to minimize the loss. Typically, the design of non-data adaptive transformations is based on heuristics developed from a detailed understanding of how preprocessing in either the spatial or temporal domains degrades performance. Second, in contrast to data adaptive transformation methods, the Block STAP method does not require the decomposition of the interference plus noise correlation matrix or the proper identification of subspace to achieve optimal performance. Third, we prove that the SMI Block STAP processor has reduced secondary data requirements in comparison to the direct form SMI processor when SINR performance is measured relative to the optimal processor. Ward [41:81] states that SMI reduced-dimension processors may actually out perform a full dimension SMI processor when the secondary data support is limited because they will incur much less estimation loss, but this statement is not supported with any analytical

discussion or references. The original contributors of the CSM GSC [11–14] and OP [37] methods do not provide any analytical discussion on the secondary data requirements of these two data adaptive transformation methods. Lastly, we do not restrict our investigation to the ideal condition that the family of correlation matrices is simultaneously block diagonalizable. We establish a criterion for selecting a transformation to reduce the loss in SINR performance when the family is not simultaneously block diagonalizable.

The key objective of this research is to lay a solid foundation for the Block STAP method to support future research on this method and general STAP issues. To this end, we prove that the secondary data requirements of the Block STAP processor are reduced by approximately 50% in comparison to the direct form processor. We define a family of correlation matrices that is representative of interference plus noise environments typically encountered by airborne surveillance radar systems and investigate its potential for simultaneous block diagonalization. The results of our investigation support the conjecture that the family cannot be simultaneously block diagonalized. Thus, we turn our attention to the problem of selecting a non-block diagonalizing transformation that minimizes the loss in performance. Through a detailed analysis of the Block STAP efficiency relative to the optimal processor, we develop a mathematical criterion for selecting such a transformation. We also present an analog metric to the cross spectral metric for the direct form processor.

1.5 Organization

This dissertation is organized as follows. In Chapter II, the reduced secondary data requirements of the Block STAP processor are proven using a Taylor series expansion of an exact expression for the Block STAP normalized SINR. The focus of Chapter III is on the issue of simultaneously block diagonalizing a family of correlation matrices. We define a family of correlation matrices that is representative of airborne STAP interference plus noise environments, review the mathematical requirements for block diagonalizing a matrix and family of matrices, and then, using existing theorems, provide evidence to support the conjecture that the defined family is not simultaneously block diagonalizable. In Chapter IV, we develop a criterion for selecting a transformation that reduces the loss in SINR performance when the family of correlation matrices is not simultaneously block diagonal-

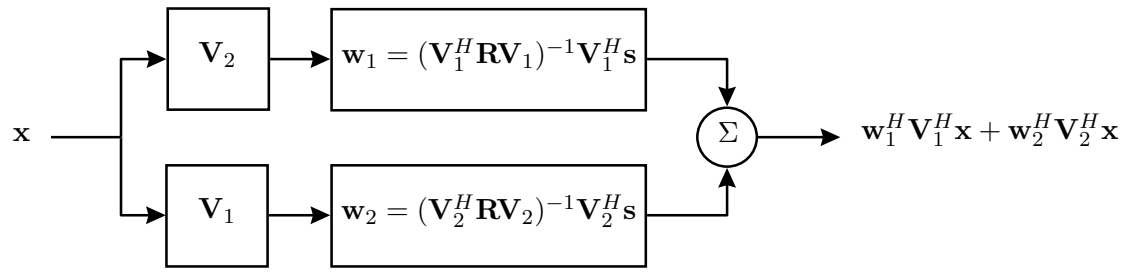
izable. In Chapter V, we propose the SINR metric and a reduced-dimension STAP method based on this metric which are direct form processor analogs to the cross spectral metric (CSM) and reduced-dimension CSM generalized sidelobe canceler. In Chapter VI, we summarize the results and contributions of this dissertation and present recommendations for future research areas.

II. Dual Channel Secondary Data Requirements

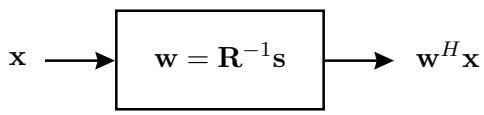
2.1 Introduction

As discussed earlier, the optimum weight vector can be computed as two independent, reduced-dimension weight vectors if the correlation matrix is in a block diagonal form. Thus, when the true correlation matrix $\mathbf{R}$ is known and the unitary matrix $\mathbf{V} = [\mathbf{V}_1 \ \mathbf{V}_2]$ block diagonalizes $\mathbf{R}$, the dual channel system shown in Fig. 2.1(a) is equivalent to the single channel system shown in Fig. 2.1(b) in terms of SINR performance. However, since the true correlation matrix is unknown in practical applications, this equivalence is pointless if the dual channel system cannot deliver the same or nearly the same performance as the single channel system when estimated correlation matrices are used. In this chapter, we demonstrate that the dual channel system can achieve nearly the same performance as the single channel system with only half the secondary data support. Note that the development in this chapter is general and applies to any maximum SINR filtering application. Thus, we use the term dual channel system in place of Block STAP system to avoid the implication that the development in this chapter is restricted to STAP applications. When necessary to avoid confusion, we preface the type of system with SMI to indicate that the weight vector is computed with an estimated correlation matrix.

Reed et. al. [34] analyzed the random SINR performance of the SMI single channel system by defining the conditioned SINR and normalized SINR statistics and then, showed that the normalized SINR, denoted by ρ_{smi}, was distributed as a beta random variable. As a result, the average SINR performance loss of the SMI single channel system relative to the optimum system as a function of secondary data support can easily be determined from the expected value of ρ_{smi}. By setting the expected value of ρ_{smi} equal to 0.5 (an average loss of 3 dB), we get the common rule of thumb that $2N$ secondary data vectors are needed to achieve acceptable performance, where N denotes the dimension of the weight vector. Our objective in this chapter is to perform a similar analysis for the SMI dual channel system. The results of the analysis depend on whether the interference plus noise is a complex or real random process. We present the results for both cases, but our presentation centers on the complex case.



(a) Dual Channel System.



(b) Single Channel System.

Figure 2.1 The dual and single channel systems are equivalent in terms of SINR performance if the matrix $\mathbf{V} = [\mathbf{V}_1 \ \mathbf{V}_2]$ block diagonalizes the true correlation matrix $\mathbf{R}$.

Our analysis begins in Section 2.3 with a derivation of an exact expression for the dual channel normalized SINR, denoted by ρ_{dual}, in terms of random variables with known distributions. After establishing the exact expression for ρ_{dual}, we derive approximate expressions for the mean and variance of ρ_{dual} using a Taylor series expansion of the exact expression. Then, in Section 2.4, we use these approximations to demonstrate the reduced secondary data requirements of the SMI dual channel system relative to the SMI single channel system. Next, we discuss several practical aspects of replacing a single channel system with a dual channel system in Section 2.5, which is followed by a simulation example in Section 2.6. Finally, we summarize the chapter in Section 2.7.

2.2 Notation and Previous Results

Before beginning the analysis, we first comment on the notation and highlight previous results used in the analysis. A complex p-variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ will be denoted as $\tilde{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for the real case.

The symbols $B(\alpha,\beta)$, χ^2_m, and $\gamma(\alpha,\beta)$ will denote a beta random variable with parameters α and β, a chi-square random variable with m degrees of freedom, and a gamma random variable with parameters α and β, respectively. The symbol $=_d$ will denote that two random variables have the same distributions and $\sim$ will denote 'is distributed as.' Let $\{\mathbf{w}_i\}_{i=1}^K$ be a set of independent and identically distributed (i.i.d.) $N \times 1$ random vectors with $\mathbf{w}_i \sim N_N(\mathbf{0},\mathbf{\Sigma})$ and let the $N \times K$ matrix $\mathbf{W} = [\mathbf{w}_1\ \mathbf{w}_2\ \cdots\ \mathbf{w}_K]$. Then, the $N \times N$ matrix $\hat{\mathbf{\Sigma}} = \mathbf{W}\mathbf{W}^T$ has a Wishart distribution with K degrees of freedom which will be denoted by $W_N(K,\mathbf{\Sigma})$ [24]. When each $\mathbf{w}_i \sim \tilde{N}_N(\mathbf{0},\mathbf{\Sigma})$, then $\hat{\mathbf{\Sigma}} = \mathbf{W}\mathbf{W}^H$ has a complex Wishart distribution with K degrees of freedom which will be denoted by $\tilde{W}_N(K,\mathbf{\Sigma})$ [24]. Using Theorem 1 of Khatri and Rao [24], we can derive the following two results. Let $\hat{\mathbf{R}} \sim W_N(K,\mathbf{R})$ and $\mathbf{s}$ be a $N \times 1$ vector, then

$$\xi = \left(\mathbf{s}^T\hat{\mathbf{R}}^{-1}\mathbf{s}\right)^{-1} \sim W_1\left(K-N+1, \left(\mathbf{s}^T\mathbf{R}^{-1}\mathbf{s}\right)^{-1}\right) \tag{2.1}$$

$$\rho = \left(\mathbf{s}^T\mathbf{R}^{-1}\mathbf{s}\right)^{-1}\left(\mathbf{s}^T\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}\right)^{-1}\left(\mathbf{s}^T\hat{\mathbf{R}}^{-1}\mathbf{s}\right)^2 \sim B\left(\frac{K-N+2}{2}, \frac{N-1}{2}\right), \tag{2.2}$$

where ξ and ρ are independently distributed. Let $\hat{\mathbf{R}} \sim \tilde{W}_N(K,\mathbf{R})$ and $\mathbf{s}$ be a $N \times 1$ vector, then

$$\tilde{\xi} = \left(\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}\right)^{-1} \sim \tilde{W}_1\left(K-N+1, \left(\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s}\right)^{-1}\right) \tag{2.3}$$

$$\tilde{\rho} = \left(\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s}\right)^{-1}\left(\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}\right)^{-1}\left(\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}\right)^2 \sim B\left(K-N+2, N-1\right), \tag{2.4}$$

where $\tilde{\xi}$ and $\tilde{\rho}$ are independently distributed. Note that Eqn. (2.4) is the same result derived by Reed et. al [34]. With regard to ξ and $\tilde{\xi}$, one can show that [31:96]

$$\xi\,\mathbf{s}^T\mathbf{R}^{-1}\mathbf{s} = \frac{\mathbf{s}^T\mathbf{R}^{-1}\mathbf{s}}{\mathbf{s}^T\hat{\mathbf{R}}^{-1}\mathbf{s}} \sim \chi^2_{K-N+1} \tag{2.5}$$

and

$$\tilde{\xi}\,\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s} = \frac{\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s}}{\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}} \sim \gamma\left(K-N+1, 1\right). \tag{2.6}$$

We are now ready to begin the analysis.

2.3 Dual Channel Normalized SINR

The first step in deriving an expression for ρ_{dual} in terms of random variables with known distributions is to derive an expression for the dual channel conditioned SINR in terms of random variables with known distributions. Let $\mathbf{n}_i$ and $\mathbf{s}_i$ denote the $N \times 1$ interference plus noise vectors and desired signal vectors for each channel ($i = 1, 2$), respectively, and let $\mathbf{R}_i = \mathrm{E}\left\{\mathbf{n}_i\mathbf{n}_i^H\right\}$ denote the true correlation matrices of the interference plus noise vectors. Further, assume that $\mathbf{n}_1$ and $\mathbf{n}_2$ are uncorrelated (i.e., $\mathrm{E}\left\{\mathbf{n}_1\mathbf{n}_2^H\right\} = \mathbf{0} = \mathrm{E}\left\{\mathbf{n}_2\mathbf{n}_1^H\right\}$). Let $\mathbf{X}_i$ denote the $N \times K$ secondary data matrix for each channel, where the columns of $\mathbf{X}_i$ are i.i.d. $\tilde{N}_N\left(\mathbf{0}, \mathbf{R}_i\right)$. Note to keep the analysis general, we do not explicitly consider the unitary matrix $\mathbf{V}$, since we have assumed that $\mathbf{V}$ block diagonalizes the correlation matrix and the interference plus noise vector is a zero-mean, normal random vector. The matrix $\mathbf{V}$ is implicit in the definition of $\mathbf{n}_i$, $\mathbf{s}_i$ and $\mathbf{R}_i$ (i.e., $\mathbf{n}_i = \mathbf{V}_i^H\mathbf{n}$, $\mathbf{s}_i = \mathbf{V}_i^H\mathbf{s}$, and $\mathbf{R}_i = \mathbf{V}_i^H\mathbf{R}\mathbf{V}_i$, where $\mathbf{n}$ and $\mathbf{s}$ are the full dimension $2N \times 1$ interference plus noise vector and desired signal vector, respectively, and $\mathbf{R} = \mathrm{E}\left\{\mathbf{n}\mathbf{n}^H\right\}$ is the $2N \times 2N$ interference plus noise correlation matrix). The maximum likelihood estimates of the interference plus noise correlation matrices are [34]

$$\hat{\mathbf{R}}_1 = \frac{1}{K}\mathbf{X}_1\mathbf{X}_1^H \tag{2.7}$$

$$\hat{\mathbf{R}}_2 = \frac{1}{K}\mathbf{X}_2\mathbf{X}_2^H. \tag{2.8}$$

We can drop the $1/K$ term in the subsequent analysis, since it appears both in the numerator and denominator of the conditioned SINR. Thus, $\hat{\mathbf{R}}_1 \sim \tilde{W}_N\left(K, \mathbf{R}_1\right)$ and $\hat{\mathbf{R}}_2 \sim \tilde{W}_N\left(K, \mathbf{R}_2\right)$ after dropping the $1/K$ term and note that $\hat{\mathbf{R}}_1$ and $\hat{\mathbf{R}}_2$ are independently distributed [31:92]. Similarly, $\hat{\mathbf{R}}_1 \sim W_N\left(K, \mathbf{R}_1\right)$ and $\hat{\mathbf{R}}_2 \sim W_N\left(K, \mathbf{R}_2\right)$ and are independently distributed for the real case. When the true correlation matrices are unknown, the weight vectors are computed using the SMI method and are given by

$$\hat{\mathbf{w}}_1 = \hat{\mathbf{R}}_1^{-1}\mathbf{s}_1 \tag{2.9}$$

$$\hat{\mathbf{w}}_2 = \hat{\mathbf{R}}_2^{-1}\mathbf{s}_2. \tag{2.10}$$

The output of the dual channel system given the input vector $\mathbf{x}^T = [\mathbf{x}_1^T \ \mathbf{x}_2^T]$ is $y = \hat{\mathbf{w}}_1^H \mathbf{x}_1 + \hat{\mathbf{w}}_2^H \mathbf{x}_2$ and thus, we can write the dual channel conditioned SINR as

$$\mathrm{SINR}_{\mathrm{dual}} | \hat{\mathbf{w}}_1, \hat{\mathbf{w}}_2 = \frac{|\hat{\mathbf{w}}_1^H \mathbf{s}_1 + \hat{\mathbf{w}}_2^H \mathbf{s}_2|^2}{\mathrm{E}\left\{|\hat{\mathbf{w}}_1^H \mathbf{n}_1 + \hat{\mathbf{w}}_2^H \mathbf{n}_2|^2\right\}} = \frac{\left(\mathbf{s}_1^H \hat{\mathbf{R}}_1^{-1} \mathbf{s}_1 + \mathbf{s}_2^H \hat{\mathbf{R}}_2^{-1} \mathbf{s}_2\right)^2}{\mathbf{s}_1^H \hat{\mathbf{R}}_1^{-1} \mathbf{R}_1 \hat{\mathbf{R}}_1^{-1} \mathbf{s}_1 + \mathbf{s}_2^H \hat{\mathbf{R}}_2^{-1} \mathbf{R}_2 \hat{\mathbf{R}}_2^{-1} \mathbf{s}_2}. \tag{2.11}$$

Note that dual channel conditioned SINR for the real case is the same as Eqn. (2.11) with the complex conjugate transpose replaced by transpose. Observe that we can write the denominator of Eqn. (2.11) as

$$\begin{bmatrix} \mathbf{s}_1^H & \mathbf{s}_2^H \end{bmatrix} \begin{bmatrix} \hat{\mathbf{R}}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{R}}_2^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{R}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_2 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{R}}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{R}}_2^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \end{bmatrix} \tag{2.12}$$

and the numerator as

$$\left(\begin{bmatrix} \mathbf{s}_1^H & \mathbf{s}_2^H \end{bmatrix} \begin{bmatrix} \hat{\mathbf{R}}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{R}}_2^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \end{bmatrix} \right)^2. \tag{2.13}$$

Notice that the off-diagonal blocks of the matrices involving the estimated correlation matrices are zero matrices. This represents one of the major differences between the single and dual channel derivations. In a single channel system, only a single correlation matrix is estimated. Although the off-diagonal blocks of the true correlation matrix are zero matrices if $\mathbf{n}_1$ and $\mathbf{n}_2$ are uncorrelated, it does not guarantee that the off-diagonal blocks of the estimated correlation matrix will be zero.

Notice the similarity between the terms in Eqn. (2.11) and the earlier results presented in Eqns. (2.1)-(2.6). To write the dual channel conditioned SINR in terms of random variables with known distribution, we will arrange the terms of Eqn. (2.11) in the numerator to a form similar to Eqn. (2.6) and the terms in the denominator to a form similar to Eqn. (2.4). Each term in the numerator has the form $\mathbf{s}_i^H \hat{\mathbf{R}}_i^{-1} \mathbf{s}_i$ and can be rewritten as

$$\mathbf{s}_i^H \hat{\mathbf{R}}_i^{-1} \mathbf{s}_i = \frac{1}{\left(\mathbf{s}_i^H \hat{\mathbf{R}}_i^{-1} \mathbf{s}_i\right)^{-1}} = \frac{\mathbf{s}_i^H \mathbf{R}_i^{-1} \mathbf{s}_i}{\mathbf{s}_i^H \mathbf{R}_i^{-1} \mathbf{s}_i \left(\mathbf{s}_i^H \hat{\mathbf{R}}_i^{-1} \mathbf{s}_i\right)^{-1}} = \frac{\alpha_i}{u_i}, \tag{2.14}$$

where

$$\alpha_i = \mathbf{s}_i^H \mathbf{R}_i^{-1} \mathbf{s}_i \tag{2.15}$$

$$u_i = \frac{\mathbf{s}_i^H \mathbf{R}_i^{-1} \mathbf{s}_i}{\mathbf{s}_i^H \hat{\mathbf{R}}_i^{-1} \mathbf{s}_i} \sim \begin{cases} \gamma\left(K-N+1,1\right) & \text{Complex Case} \\ \chi^2_{K-N+1} & \text{Real Case} \end{cases} \tag{2.16}$$

with the last result following from Eqns. (2.6) and (2.5). Each term in the denominator has the form $\mathbf{s}_i^H \hat{\mathbf{R}}_i^{-1} \mathbf{R}_i \hat{\mathbf{R}}_i^{-1} \mathbf{s}_i$ and can be rewritten as

$$\begin{aligned} \mathbf{s}_i^H \hat{\mathbf{R}}_i^{-1} \mathbf{R}_i \hat{\mathbf{R}}_i^{-1} \mathbf{s}_i &= \frac{1}{\left(\mathbf{s}_i^H \hat{\mathbf{R}}_i^{-1} \mathbf{R}_i \hat{\mathbf{R}}_i^{-1} \mathbf{s}_i\right)^{-1}} = \frac{\left(\mathbf{s}_i^H \mathbf{R}_i^{-1} \mathbf{s}_i\right)^{-1} \left(\mathbf{s}_i^H \hat{\mathbf{R}}_i^{-1} \mathbf{s}_i\right)^2}{\left(\mathbf{s}_i^H \mathbf{R}_i^{-1} \mathbf{s}_i\right)^{-1} \left(\mathbf{s}_i^H \hat{\mathbf{R}}_i^{-1} \mathbf{R}_i \hat{\mathbf{R}}_i^{-1} \mathbf{s}_i\right)^{-1} \left(\mathbf{s}_i^H \hat{\mathbf{R}}_i^{-1} \mathbf{s}_i\right)^2} \\ &= \mathbf{s}_i^H \mathbf{R}_i^{-1} \mathbf{s}_i \left(\frac{1}{\left(\mathbf{s}_i^H \mathbf{R}_i^{-1} \mathbf{s}_i\right)^{-1} \left(\mathbf{s}_i^H \hat{\mathbf{R}}_i^{-1} \mathbf{R}_i \hat{\mathbf{R}}_i^{-1} \mathbf{s}_i\right)^{-1} \left(\mathbf{s}_i^H \hat{\mathbf{R}}_i^{-1} \mathbf{s}_i\right)^2} \right) \left(\frac{1}{\mathbf{s}_i^H \mathbf{R}_i^{-1} \mathbf{s}_i \left(\mathbf{s}_i^H \hat{\mathbf{R}}_i^{-1} \mathbf{s}_i\right)^{-1}} \right)^2 \\ &= \frac{\alpha_i}{q_i u_i^2}, \end{aligned} \tag{2.17}$$

where

$$\begin{aligned} q_i &= \left(\mathbf{s}_i^H \mathbf{R}_i^{-1} \mathbf{s}_i\right)^{-1} \left(\mathbf{s}_i^H \hat{\mathbf{R}}_i^{-1} \mathbf{R}_i \hat{\mathbf{R}}_i^{-1} \mathbf{s}_i\right)^{-1} \left(\mathbf{s}_i^H \hat{\mathbf{R}}_i^{-1} \mathbf{s}_i\right)^2 \\ &\sim \begin{cases} B\left(K-N+2, N-1\right) & \text{Complex Case} \\ B\left(\frac{K-N+2}{2}, \frac{N-1}{2}\right) & \text{Real Case} \end{cases} \end{aligned} \tag{2.18}$$

with the last result following from Eqns. (2.4) and (2.2). Recall that u_i and q_i are independent for a fixed i and that $\hat{\mathbf{R}}_1$ and $\hat{\mathbf{R}}_2$ are independent and hence, u_1, u_2, q_1 and q_2 are independent. Finally, using these results, we can write the dual channel conditioned SINR in terms of random variables with known distributions as

$$\text{SINR}_{\text{dual}} | \hat{\mathbf{w}}_1, \hat{\mathbf{w}}_2 =_d \frac{\left(\frac{\alpha_1}{u_1} + \frac{\alpha_2}{u_2}\right)^2}{\frac{\alpha_1}{q_1 u_1^2} + \frac{\alpha_2}{q_2 u_2^2}}. \tag{2.19}$$

With the dual channel conditional SINR established, we can achieve our initial objective of expressing the dual channel normalized SINR as a function of random variables with

known distributions by dividing Eqn. (2.19) by the maximum SINR. Under the assumption that $\mathbf{n}_1$ and $\mathbf{n}_2$ are uncorrelated, the maximum SINR is

$$\begin{aligned}\mathrm{SINR}_{\max} &= \mathbf{s}^H\mathbf{R}^{-1}\mathbf{s} = \begin{bmatrix}\mathbf{s}_1^H & \mathbf{s}_2^H\end{bmatrix}\begin{bmatrix}\mathbf{R}_1^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{R}_2^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{s}_1\\ \mathbf{s}_2\end{bmatrix} = \mathbf{s}_1^H\mathbf{R}_1^{-1}\mathbf{s}_1 + \mathbf{s}_2^H\mathbf{R}_2^{-1}\mathbf{s}_2\\ &= \alpha_1 + \alpha_2. \end{aligned} \tag{2.20}$$

Thus, the dual channel normalized SINR in terms of random variables with known distributions is

$$\rho_{\mathrm{dual}} =_d \frac{1}{\alpha_1+\alpha_2}\left[\frac{(\alpha_1 u_2 + \alpha_2 u_1)^2}{\alpha_1 q_2 u_2^2 + \alpha_2 q_1 u_1^2}\right] q_1 q_2, \tag{2.21}$$

which we can rewrite as

$$\rho_{\mathrm{dual}} =_d \left[\frac{(k u_2 + (1-k) u_1)^2}{k q_2 u_2^2 + (1-k) q_1 u_1^2}\right] q_1 q_2 = h\left(u_1, u_2, q_1, q_2\right), \tag{2.22}$$

by letting $k = \alpha_1/(\alpha_1+\alpha_2)$. Note that $0 < k < 1$, since $\mathbf{R}_1$ and $\mathbf{R}_2$ are positive definite and thus, α_1 and α_2 are greater than zero. Ideally, we would like to develop an expression for the probability density function (pdf) of ρ_{dual} to fully characterize its behavior. Although we can express ρ_{dual} as a function of random variables with known pdfs, developing the pdf of ρ_{dual} represents a formidable task, as does developing closed-form expressions for the mean and variance. Thus, we resorted to a Taylor series expansion of $h\left(u_1, u_2, q_1, q_2\right)$ to derive approximate expressions for the mean and variance.

Let $g(x_1, x_2)$ be a function of the continuous random variables x_1 and x_2. Papoulis [32:156–157] provides the following approximation for the expected value of $g(x_1, x_2)$:

$$\mu_g \approx g(x_1,x_2) + \frac{1}{2}\left[\sigma_1^2\frac{\partial^2 g(x_1,x_2)}{\partial x_1^2} + 2\xi\sigma_1\sigma_2\frac{\partial^2 g(x_1,x_2)}{\partial x_1 \partial x_2} + \sigma_2^2\frac{\partial^2 g(x_1,x_2)}{\partial x_2^2}\right], \tag{2.23}$$

where $g(x_1, x_2)$ and the partial derivatives are evaluated at the point (μ_1, μ_2), μ_1 and μ_2 denote the mean of x_1 and x_2, σ_1^2 and σ_2^2 denote the variance of x_1 and x_2, ξ is the correlation coefficient of x_1 and x_2, and $g(x_1, x_2)$ is assumed to be sufficiently smooth

about the point (μ_1, μ_2). Equation (2.23) is derived using the Taylor series of $g(x_1, x_2)$ about the point (μ_1, μ_2) and substituting up to the second order terms into the standard formula of the expected value. The approximation for the variance of $g(x_1, x_2)$ is given by [32:156-157]

$$\sigma_g^2 \approx \left(\frac{\partial g(x_1, x_2)}{\partial x_1}\right)^2 \sigma_1^2 + 2\left(\frac{\partial g(x_1, x_2)}{\partial x_1}\right)\left(\frac{\partial g(x_1, x_2)}{\partial x_2}\right)\xi\sigma_1\sigma_2 + \left(\frac{\partial g(x_1, x_2)}{\partial x_2}\right)^2 \sigma_1^2, \tag{2.24}$$

where the partial derivatives are evaluated at the point (μ_1, μ_2). The extension of these approximations for an arbitrary n is straight forward as follows. Let $g(x_1, \ldots, x_n)$ be a function of the continuous random variables $x_1, \ldots, x_n$. The expected value of $g(x_1, \ldots, x_n)$, denoted as μ_g, is

$$\mu_g = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \ldots, x_n) f(x_1, \ldots, x_n) dx_1 \ldots dx_n, \tag{2.25}$$

and the variance, denoted as σ_g^2, is

$$\sigma_g^2 = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(g(x_1, \ldots, x_n) - \mu_g\right)^2 f(x_1, \ldots, x_n) dx_1 \ldots dx_n, \tag{2.26}$$

where $f(x_1, \ldots, x_n)$ is the joint pdf of $x_1, \ldots, x_n$. Given that the partial derivatives of order three exist, the Taylor polynomial of $g(x_1, \ldots, x_n)$ of degree two about the point $p = (\mu_1, \ldots, \mu_n)$ is [10:154]

$$g(x_1, \ldots, x_n) + \sum_i^n (x_i - \mu_i)\frac{\partial g(x_1, \ldots, x_n)}{\partial x_i} + \frac{1}{2}\sum_i^n \sum_j^n (x_i - \mu_i)(x_j - \mu_j)\frac{\partial^2 g(x_1, \ldots, x_n)}{\partial x_i \partial x_j}, \tag{2.27}$$

where $g(x_1, \ldots, x_n)$ and the partial derivatives are evaluated at point p. Now, observe that if $p = (\mu_1, \ldots, \mu_n)$ where μ_i is the mean of x_i, then

$$\mathrm{E}\left\{g(\mu_1, \ldots, \mu_n)\right\} = g(\mu_1, \ldots, \mu_n) \tag{2.28}$$

$$\mathrm{E}\left\{\sum_{i}^{n}(x_i-\mu_i)\left.\frac{\partial g}{\partial x_i}\right|_p\right\}=\sum_{i}^{n}\left.\frac{\partial g}{\partial x_i}\right|_p(\mathrm{E}\{x_i\}-\mu_i)=0 \tag{2.29}$$

and

$$\mathrm{E}\left\{(x_i-\mu_i)(x_j-\mu_j)\left.\frac{\partial^2 g}{\partial x_i\partial x_j}\right|_p\right\}=\left.\frac{\partial^2 g}{\partial x_i\partial x_j}\right|_p\mathrm{E}\{(x_i-\mu_i)(x_j-\mu_j)\} \tag{2.30}$$

$$=\begin{cases}\sigma_{x_i}^2\left.\frac{\partial^2 g}{\partial x_i^2}\right|_p & \text{if } i=j,\\ \xi_{ij}\sigma_{x_i}\sigma_{x_j}\left.\frac{\partial^2 g}{\partial x_i\partial x_j}\right|_p & \text{if } i\neq j,\end{cases} \tag{2.31}$$

where $\left.\frac{\partial g}{\partial x_i}\right|_p$, $\left.\frac{\partial^2 g}{\partial x_i^2}\right|_p$, and $\left.\frac{\partial^2 g}{\partial x_i\partial x_j}\right|_p$ denote the partial derivatives of $g(x_1,\ldots,x_n)$ evaluated at point p. Thus, substituting Eqn. (2.27) into Eqn. (2.25) yields the following approximation for the mean of $g(x_1,\ldots,x_n)$:

$$\mu_g\approx g(\mu_1,\ldots,\mu_n)+\frac{1}{2}\sum_{i}^{n}\sigma_{x_i}^2\left.\frac{\partial^2 g}{\partial x_i^2}\right|_p+\frac{1}{2}\sum_{i}^{n}\sum_{\substack{j\\ i\neq j}}^{n}\xi_{ij}\sigma_{x_i}\sigma_{x_j}\left.\frac{\partial^2 g}{\partial x_i\partial x_j}\right|_p, \tag{2.32}$$

where μ_i is the mean of x_i. Similarly, by substituting Eqn. (2.27) for $g(x_1,\ldots,x_n)$ and Eqn. (2.32) for μ_g into Eqn. (2.26) and discarding any moments greater than 2, one can show that the variance of $g(x_1,\ldots,x_n)$ is approximated by

$$\sigma_g^2\approx\sum_{i}^{n}\sigma_{x_i}^2\left(\left.\frac{\partial g}{\partial x_i}\right|_p\right)^2+\frac{1}{2}\sum_{i}^{n}\sum_{\substack{j\\ i\neq j}}^{n}\xi_{ij}\sigma_{x_i}\sigma_{x_j}\left(\left.\frac{\partial g}{\partial x_i}\right|_p\right)\left(\left.\frac{\partial g}{\partial x_j}\right|_p\right). \tag{2.33}$$

Using Eqn. (2.32) and (2.33), we can now derive approximate expressions for the mean and variance of the dual channel normalized SINR ρ_{dual}. Recall that u_1 and u_2 are i.i.d. and that q_1 and q_2 are i.i.d. Thus, the mean (μ_u) and variance (σ_u^2) of u_i and the mean (μ_q) and variance (σ_q^2) of q_i for $i=1,2$ are [22]

$$\mu_u=\begin{cases}K-N+1 & \text{Complex Case}\\ K-N+1 & \text{Real Case}\end{cases} \tag{2.34}$$

$$\sigma_u^2 = \begin{cases} K - N + 1 & \text{Complex Case} \\ 2(K - N + 1) & \text{Real Case} \end{cases} \tag{2.35}$$

$$\mu_q = \begin{cases} \frac{K-N+2}{K+1} & \text{Complex Case} \\ \frac{K-N+2}{K+1} & \text{Real Case} \end{cases} \tag{2.36}$$

$$\sigma_q^2 = \begin{cases} \frac{(K-N+2)(N-1)}{(K+2)(K+1)^2} & \text{Complex Case} \\ \frac{2(K-N+2)(N-1)}{(K+3)(K+1)^2} & \text{Real Case.} \end{cases} \tag{2.37}$$

Further, observe that all the partial derivatives of $h(u_1, u_2, q_1, q_2)$ exist at point p, if p does not cause the denominator of Eqn. (2.22) to equal zero, since $h(u_1, u_2, q_1, q_2)$ is a rational function. Thus, the Taylor series of $h(u_1, u_2, q_1, q_2)$ about the point $p = (\mu_u, \mu_u, \mu_q, \mu_q)$ will exist if μ_u and μ_q do not equal zero. Note that μ_u and μ_q are greater than zero if $K > N$. Using the independence of u_1, u_2, q_1, and q_2 (i.e., the correlation coefficients (ξ_{ij}) in Eqn. (2.32) and (2.33) are zero) and with $p = (\mu_u, \mu_u, \mu_q, \mu_q)$, the approximation for the mean $\mathrm{E}\{\rho_{\text{dual}}\}$ and variance $\mathrm{V}\{\rho_{\text{dual}}\}$ are

$$\mathrm{E}\{\rho_{\text{dual}}\} \approx h(\mu_u, \mu_u, \mu_q\mu_q) + \frac{1}{2}\left[\sigma_u^2 \left[\left.\frac{\partial^2 h}{\partial u_1^2}\right|_p + \left.\frac{\partial^2 h}{\partial u_2^2}\right|_p\right] + \sigma_q^2 \left[\left.\frac{\partial^2 h}{\partial q_1^2}\right|_p + \left.\frac{\partial^2 h}{\partial q_2^2}\right|_p\right]\right], \tag{2.38}$$

$$\mathrm{V}\{\rho_{\text{dual}}\} \approx \sigma_u^2 \left[\left(\left.\frac{\partial h}{\partial u_1}\right|_p\right)^2 + \left(\left.\frac{\partial h}{\partial u_2}\right|_p\right)^2\right] + \sigma_q^2 \left[\left(\left.\frac{\partial h}{\partial q_1}\right|_p\right)^2 + \left(\left.\frac{\partial h}{\partial q_2}\right|_p\right)^2\right]. \tag{2.39}$$

Performing the partial derivatives in Eqn. (2.38) and (2.39) and evaluating at the point p yields

$$\mathrm{E}\{\rho_{\text{dual}}\} \approx \mu_q - 2\frac{\mu_q \sigma_u^2}{\mu_u^2} k(1-k) - 2\frac{\sigma_q^2}{\mu_q} k(1-k), \tag{2.40}$$

$$\mathrm{V}\{\rho_{\text{dual}}\} \approx \left(1 - 2k + 2k^2\right)\sigma_q^2. \tag{2.41}$$

Substituting Eqns. (2.34)-(2.37) into Eqns. (2.40) and (2.41) yields

$$\mathrm{E}\{\rho_{\text{dual}}\} \approx \begin{cases} \frac{K-N+2}{K+1} - 2\frac{k(1-k)}{K+1}\left[\frac{K-N+2}{K-N+1} + \frac{N-1}{K+2}\right] & \text{Complex Case} \\ \frac{K-N+2}{K+1} - 4\frac{k(1-k)}{K+1}\left[\frac{K-N+2}{K-N+1} + \frac{N-1}{K+3}\right] & \text{Real Case} \end{cases} \tag{2.42}$$

$$\mathrm{V}\left\{\rho_{\mathrm{dual}}\right\} \approx \begin{cases} \left(1-2k+2k^2\right)\frac{(K-N+2)(N-1)}{(K+2)(K+1)^2} & \text{Complex Case} \\ \left(1-2k+2k^2\right)\frac{2(K-N+2)(N-1)}{(K+3)(K+1)^2} & \text{Real Case,} \end{cases} \tag{2.43}$$

where K is the number of secondary data vectors used to estimate the interference plus noise correlation matrices, N is the dimension of the weight vectors $\hat{\mathbf{w}}_1$ and $\hat{\mathbf{w}}_2$, and $k = \alpha_1/(\alpha_1 + \alpha_2)$, which indicates the relative SINR between the two channels. Observe that the approximations are quadratic functions of k and are a maximum at $k = 0, 1$ and minimum at $k = 0.5$. Thus, an increase in the mean (less SINR loss) comes at the expense of an increase in the variance. This trade-off between the mean and variance will undoubtedly have implications in the probability of detection performance of the dual channel system.

An examination of Eqn. (2.42) reveals that the approximation for the mean has characteristics that match with one's intuition: asymptotically correct, an increasing function of K, and provides an exact answer when $k = 0$ or $k = 1$. As K approaches infinity, the estimated interference plus noise correlation matrix approaches the true interference plus noise correlation matrix, implying that 1) the mean of the conditioned SINR approaches the maximum SINR or equivalently, the mean of normalized SINR approaches one and 2) the mean of normalized SINR is an increasing function of K. In the limit as K goes to infinity, the approximation of the mean for ρ_{dual} given in Eqn. (2.42) is one, indicating that the asymptotic properties of the approximation are correct. An examination of Eqn. (2.42) reveals that the approximation is an increasing function of K when $K > N$ and N and k are held constant. Next, observe that if $k = 0$ (or $k = 1$) the approximation reduces to μ_q and $h(u_1, u_2, q_1, q_2) = q_2$ (or q_1). The mean of q_1 or q_2 is μ_q and thus, the approximation provides an exact answer when $k = 0$ or $k = 1$. Similar observations hold for the approximate expression for the variance (i.e., variance approaches zero as K approaches infinity, decreasing function of K, and exact when $k = 0$ or 1).

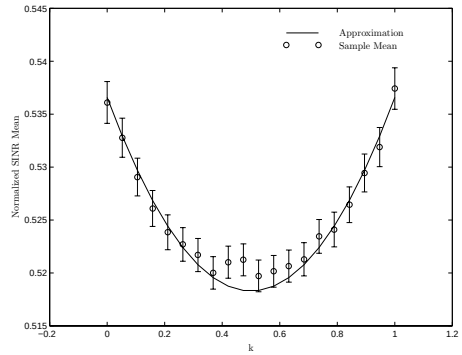
To further verify the validity of the approximations given in Eqns. (2.40) and (2.41), we conducted a series of Monte Carlo simulations with $K = 2N$ for N between 20 and 500 in steps of 20 and for 20 values of k uniformly distributed on the interval $0 \leq k \leq 1$. For each N and k, 10,000 samples of ρ_{dual} were used to compute a sample mean (i.e.,

$\mathrm{E}\{\rho_{\mathrm{dual}}\}$), an approximate 99.5% confidence interval for the sample mean, and a sample variance. The confidence intervals are termed approximate, because the sample standard deviation was used in place of the population (known) standard deviation [32:248]. We present only the results for the complex case, but the results from the real case have the same behavior as the complex case. Figure 2.2 shows the sample mean of ρ_{dual} for $N = 20$, 260, and 500 as k varied between 0 and 1 along with the approximate 99.5% confidence intervals which are indicated by the error bars. Also plotted in Fig. 2.2 is the approximate mean of ρ_{dual} from Eqn. (2.42). An examination of Fig. 2.2 shows excellent agreement between the sample mean and the mean approximation, except when $N = 20$ and near $k = 0.5$. This problem area can be eliminated by including another term in the Taylor series expansion. Plotted in Fig. 2.3 is the sample variance overlaid with the approximate variance from Eqn. (2.43) for the same cases as Fig. 2.2. Again, with exception of $N = 20$ and near $k = 0.5$, there is excellent agreement between the sample variance and variance approximation. This problem area can be eliminated by keeping all the moments (i.e., do not discard moments greater than 2).

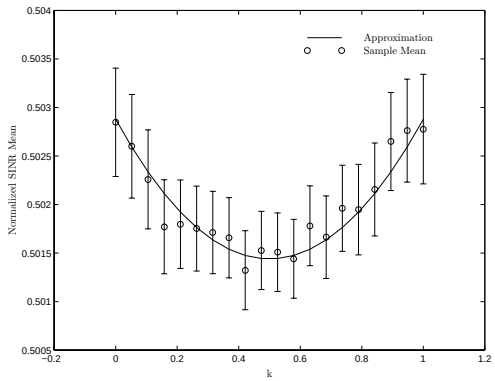
2.4 Reduced Secondary Data Requirements

In this section, we address our earlier claim that the SMI dual channel system requires half the secondary data as the equivalent SMI single channel system to achieve nearly the same normalized SINR performance. We only discuss the complex case, but one can easily show that the results also hold for the real case. This claim is examined under the assumption that the input interference plus noise signals are uncorrelated. That is, if $\mathbf{n}_1$ and $\mathbf{n}_2$ denote the input interference plus noise signal vectors in their respective channels, then $\mathrm{E}\left\{\mathbf{n}_1\mathbf{n}_2^H\right\} = \mathbf{0} = \mathrm{E}\left\{\mathbf{n}_2\mathbf{n}_1^H\right\}$. Let $\mathbf{n}_1$ and $\mathbf{n}_2$ be $N \times 1$ vectors, then the equivalent single channel system must process a $2N \times 1$ input signal vector. Thus, the single channel system requires a $2N \times 1$ weight vector and approximately $4N$ ($K \approx 4N$) secondary data vectors are required to achieve an average normalized SINR of 0.5 (i.e., an average SINR loss of 3 dB). With $K = 4N$, the variance of the single channel normalized SINR is

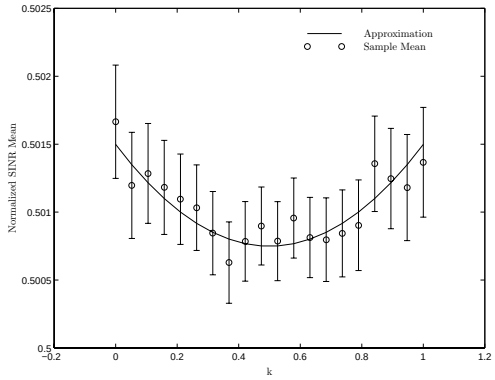
$$\mathrm{V}\{\rho_{\mathrm{smi}}\} = \frac{(4N - 2N + 2)(2N - 1)}{(4N + 2)(4N + 1)^2} \approx \frac{1}{16N}, \tag{2.44}$$



(a) $N = 20$

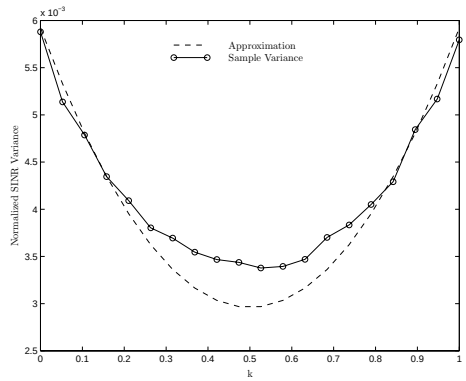


(b) $N = 260$

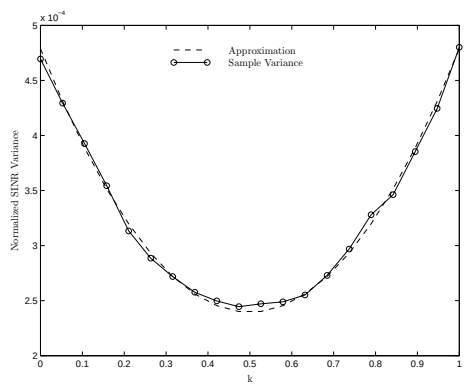


(c) $N = 500$

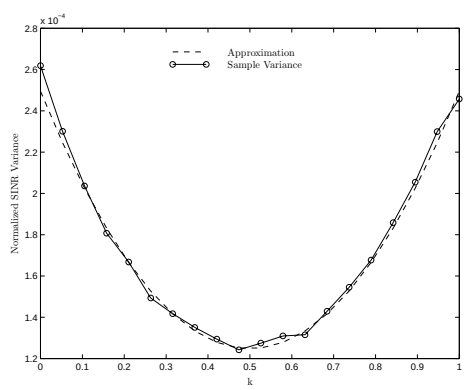
Figure 2.2 Sample mean of ρ_{dual} (i.e., $\mathrm{E}\{\rho_{\text{dual}}\}$) based on 10,000 samples for each N and k overlaid with the approximate $\mathrm{E}\{\rho_{\text{dual}}\}$ computed using Eqn. (2.42) and $K = 2N$. The error bars are approximate 99.5% confidence intervals.



(a) $N = 20$



(b) $N = 260$



(c) $N = 500$

Figure 2.3 Sample Variance of ρ_{dual} (i.e., $\mathrm{V}\{\rho_{\text{dual}}\}$) based on 10,000 samples for each N and k overlaid with the approximate $\mathrm{V}\{\rho_{\text{dual}}\}$ computed using Eqn. (2.43) and $K = 2N$.

for large N. Thus, to support our claim, we must show that the dual channel system has an average normalized SINR of approximately 0.5 ($\mathrm{E}\{\rho_{\mathrm{dual}}\} \approx 0.5$) with a variance approximately equal to Eqn. (2.44), when $K = 2N$.

With the single channel system, we could set the expression for the mean of the normalized SINR equal to 0.5 and solve directly for the number of secondary data vectors needed in terms of the dimension of the weight vector. We can not apply this same approach to Eqn. (2.42), because of its form. Instead, we set $K = 2N$ and $k = 0.5$, since Eqn. (2.42) is quadratic function of k with a minimum at $k = 0.5$, and then, show that Eqn. (2.42) is greater than or equal to 0.5 for all $N \geq 1$. Substituting $k = 0.5$ and $K = 2N$ in Eqn. (2.42) yields

$$\mathrm{E}\{\rho_{\mathrm{dual}}\} \approx \frac{4N+5}{8N+4}. \tag{2.45}$$

Let $N \geq 1$, then Eqn. (2.45) is greater than or equal to 0.5 if

$$4N + 5 \geq 0.5(8N + 4) = 4N + 2, \tag{2.46}$$

which is true for all $N \geq 1$. Thus, the approximate $\mathrm{E}\{\rho_{\mathrm{dual}}\} \geq 0.5$ if $K \geq 2N$. The approximate variance (Eqn. (2.43)) of the dual channel normalized SINR with $k = 0.5$ and $K = 2N$ is

$$\mathrm{V}\{\rho_{\mathrm{dual}}\} \approx \frac{1}{2}\,\frac{(2N - N + 2)(N - 1)}{(2N + 2)(2N + 1)^2} \approx \frac{1}{16N}, \tag{2.47}$$

for large N. These results support our claim that the dual channel system requires half the secondary data vectors as the single channel system to achieve nearly the same normalized SINR performance.

2.5 Practical Aspects

We must address two issues before we can take advantage of the reduced secondary data requirements of the dual channel system to replace a single channel system. First, we must decorrelate the two halves of the interference plus noise vector to meet the hypothesis

of the previous development. That is, we need to find a transformation that block diagonalizes the correlation matrices of interest. Secondly, we need to control the computational cost of the dual channel system. Although the computational cost of computing the weight vectors is less with the dual channel system, one must be concerned with the additional computational cost associated with the transformation (decorrelation preprocessing). The additional matrix-vector multiple introduced by the preprocessing coupled with the fact that every secondary data vector must be preprocessed will significantly reduce any computational savings achieved by reducing the dimension of the weight vector. Clearly, the transformation needs to have an efficient implementation. In general, we can construct a block diagonalizing transformation for any particular correlation matrix from its eigenvectors, but this requires a computationally expensive eigendecomposition and will not, in general, yield an efficient transformation matrix. Thus, we seek a fixed (environment independent) block diagonalizing transformation with an efficient implementation. The possibility of finding such a transformation will depend on the class of correlation matrices of interest.

One class of matrices that can be block diagonalized by a fixed and efficient transformation is the class of centrosymmetric matrices. A $N \times N$ matrix $\mathbf{C}$ is a centrosymmetric matrix if [16]

$$[c]_{N+1-m,N+1-n} = [c]_{m,n} \quad \text{for } m, n = 1, \cdots, N \tag{2.48}$$

where $[c]_{m,n}$ denotes the element of $\mathbf{C}$ in the mth row and n^{th} column. Depending on whether N is even ($N = 2M$) or odd ($N = 2M + 1$), we can write a centrosymmetric matrix $\mathbf{C}$ in one of the following forms [16]:

$$\mathbf{C} = \begin{bmatrix} \mathbf{A} & \mathbf{BJ} \\ \mathbf{JB} & \mathbf{JAJ} \end{bmatrix}_{\text{even}} \quad \text{or} \quad \mathbf{C} = \begin{bmatrix} \mathbf{A} & \mathbf{Jx} & \mathbf{BJ} \\ \mathbf{z}^T\mathbf{J} & \beta & \mathbf{z}^T \\ \mathbf{JB} & \mathbf{x} & \mathbf{JAJ} \end{bmatrix}_{\text{odd}}, \tag{2.49}$$

where $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{J}$ are $M \times M$ matrices, $\mathbf{x}$ and $\mathbf{z}$ are $M \times 1$ vectors, β is a scalar, and $\mathbf{J}$ is the anti-diagonal (reverse diagonal) matrix, i.e.,

$$\mathbf{J} = \begin{bmatrix} 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \iddots & \vdots \\ 0 & 1 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \end{bmatrix}. \tag{2.50}$$

One can easily verify that even ($N = 2M$) and odd ($N = 2M + 1$) centrosymmetric matrices are block diagonalized by the unitary (orthonormal) matrices [8]

$$\frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{I} & \mathbf{J} \\ \mathbf{J} & -\mathbf{I} \end{bmatrix}_{\text{even}} \quad \text{and} \quad \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{I} & 0 & \mathbf{J} \\ 0 & 1 & 0 \\ \mathbf{J} & 0 & -\mathbf{I} \end{bmatrix}_{\text{odd}}, \tag{2.51}$$

where $\mathbf{I}$ and $\mathbf{J}$ are $M \times M$ matrices and $\mathbf{I}$ is the identity matrix. Clearly, the transformation matrices defined in Eqn. (2.51) can be implemented efficiently, requiring only the simple operations of addressing and addition. Observe that real, symmetric Toeplitz matrices and real, symmetric Toeplitz-block-Toeplitz matrices are subclasses of centrosymmetric matrices. Recall that the correlation matrix of a real, stationary random process is a real, symmetric Toeplitz matrix [38:150]. Thus, we can replace a single channel system with a dual channel system, reducing the secondary data requirements and easing the computational cost, in any maximum SINR filtering application involving real, stationary random processes.

In applications where the random process does not yield a centrosymmetric correlation matrix, the problem of block diagonalizing a family of correlation matrices with a fixed, efficient transformation becomes difficult. First, one must answer the question of whether or not a fixed transformation exists for the correlation matrices of interest. Basically, to block diagonalize a family of correlation matrices, we must find two independent subspaces that span the N-dimensional vector space (i.e., the vector space is the direct sum of two subspaces) and the subspaces must be invariant to every member of the fam-

ily. For example, in the case of centrosymmetric matrices, the vector space is the direct sum of the symmetric (i.e., $\mathbf{x} = \mathbf{J}\mathbf{x}$) subspace and the skewed symmetric (i.e., $\mathbf{x} = -\mathbf{J}\mathbf{x}$) subspaces. Mathematical machinery is available for examining the existence issue and is discussed in the next chapter where we address the problem of block diagonalizing STAP correlation matrices. Secondly, assuming that one can find two invariant subspaces, we are still left with the problem of controlling computational cost. That is, can we select basis vectors that span the two subspaces such that the resulting transformation will have an efficient implementation? The answer to this question appears to be an open problem. In the absence of two invariant subspaces, we can attack the problem in an approximate sense. For example, the correlation matrix of a complex, stationary random process is a Hermitian, Toeplitz matrix which is not a subclass of centrosymmetric matrices. However, we can approximately decorrelate (block diagonalize the correlation matrix) a real or complex stationary random process using a filter bank consisting of a high-pass filter and a low-pass filter [33:165]. The use of a filter bank to decorrelate the signal is a central concept in subband image compression and subband adaptive filters where efficiency is also a key issue. The decorrelation properties of a filter bank will depend on the transition regions and stopband attenuation of the filters and the characteristics of the interference plus noise. The price paid for only approximately decorrelating the interference plus noise is a loss in performance which cannot be regained by increased secondary data support, since the dual channel system is no longer equivalent to the optimal system even when the correlation matrix is known.

2.6 Simulation Example

In this section, we present the results of a simulation example to demonstrate the reduced secondary data requirements of the dual channel system. The signal of interest is

$$s(n) = \cos(2\pi 0.13n)(u(n) - u(n-31)) + \cos(2\pi 0.31n)(u(n-32) - u(n-63)) \quad (2.52)$$

for $n = 0, 1, ..., 63$, where $u(n)$ is the unit step function. The interference plus noise environment consists of three uncorrelated interference sources and receiver noise. The

signal from each of the interference sources has the form:

$$t_i(n) = \sqrt{10}\cos(2\pi f_i n + \phi_i), \tag{2.53}$$

where ϕ_i is a random variable uniformly distributed over the interval $[0, 2\pi]$ and $\{f_1 = 0.051, f_2 = 0.23, f_3 = 0.41\}$. The correlation matrix of each interference source is a real, symmetric Toeplitz (centrosymmetric) matrix $\mathbf{T}_i$ with elements given by

$$\mathrm{E}\{t_i(m)t_i(n)\} = [\mathbf{T}_i]_{m,n} = 5\cos(2\pi f_i|m-n|). \tag{2.54}$$

The receiver noise is modeled as white noise with a variance of one and is assumed to be uncorrelated with the interference. Thus, the interference plus noise correlation matrix is simply the sum of the identity matrix (receiver noise) and three real, symmetric Toeplitz matrices and hence, is a real, symmetric Toeplitz matrix. The dual channel transformations are given by

$$\mathbf{V}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix}\mathbf{I}\\ \mathbf{J}\end{bmatrix} \quad \text{and} \quad \mathbf{V}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix}\mathbf{J}\\ -\mathbf{I}\end{bmatrix}, \tag{2.55}$$

in accordance with Eqn. (2.51). The dual channel system only used 64 secondary data vectors to compute the two 32×1 weight vectors. In contrast, the single channel system used 128 secondary data vectors to compute the 64×1 weight vector. Thus, the expected loss in SINR performance was 3 dB for both systems simulated (i.e., $\mathrm{E}\{\rho_{\mathrm{smi}}\} \approx 0.5 \approx \mathrm{E}\{\rho_{\mathrm{dual}}\}$). The simulation results are summarized in Table 2.1 and are based on 20000 runs for each system. The predicted values in Table 2.1 for the single channel system were computed from a beta distribution with parameters 33 and 31.5 from Eqns. (2.42) and (2.43) for the dual channel systems with $K = 64$, $N = 32$, and $k = 0.4934$. Note the good agreement between the predicted and simulation values in Table 2.1, further verifying the utility of the approximations given in Eqns. (2.42) and (2.43). Figure 2.4 shows the normalized SINR cumulative probability distribution curves for each of the systems. The results in Table 2.1 and Fig. 2.4 show that the dual channel system has nearly the same performance as the single channel system with half the secondary data support.

System	Predicted		Simulated	
	Mean	Variance	Mean	Variance
Single Channel	0.5116	0.0038	0.5110	0.0038
Dual Channel	0.5001	0.0037	0.5030	0.0043

Table 2.1 Predicted and simulated normalized SINR performance for the single channel and dual channel systems in a centrosymmetric interference plus noise environment.

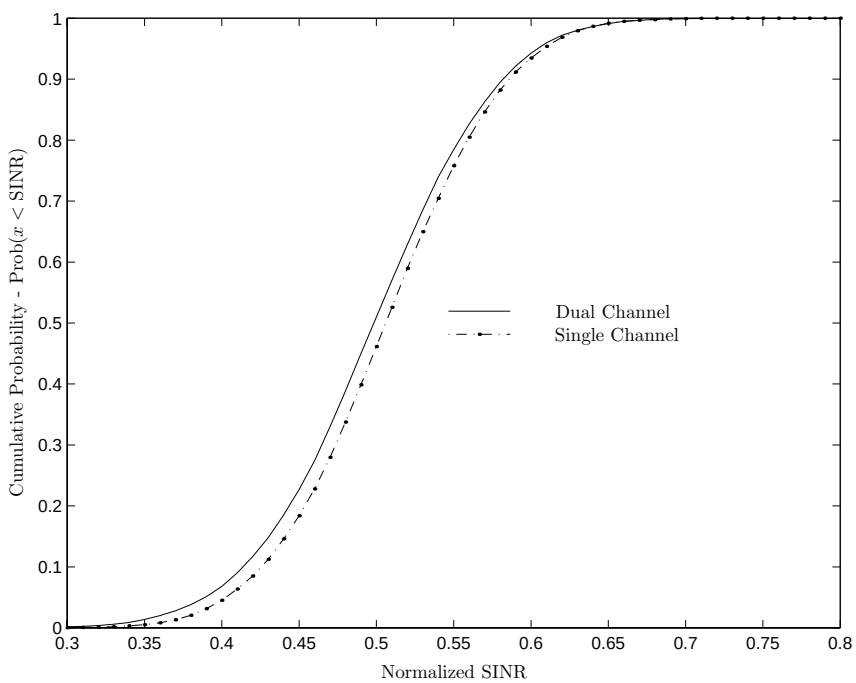


Figure 2.4 Cumulative probability distribution of the normalized SINR for the single channel and dual channel systems in a centrosymmetric interference plus noise environment.

2.7 Summary

In this chapter, we analyzed the SINR performance of a dual channel system assuming that the interference plus noise in one channel was uncorrelated with interference plus noise in the other channel. We derived approximations for mean and variance of the dual channel normalized SINR from an exact expression of the normalized SINR as functions of random variables with known distributions. Using the approximations, we showed that a dual channel system delivers nearly the same normalized SINR performance as a single channel system designed to process both inputs with half the secondary data. These results suggest the possibility of replacing a single channel system with a dual channel system using smaller weight vectors, leading to the reduction in the secondary data support and potentially, a reduction in the computational cost. A key element in replacing a single channel system with a dual channel system is the decorrelation preprocessing, which basically requires the introduction of a transformation that block diagonalizes the correlation matrix. This requirement for preprocessing introduces a new challenge: finding a fixed transformation that block diagonalizes the family of correlation matrices of interest and that has an efficient implementation. Depending on the family of correlation matrices, such a transformation may or may not exist. A family of matrices that can be efficiently block diagonalized with a fixed transformation is the family of centrosymmetric matrices which includes the family of real, symmetric Toeplitz matrices. The correlation matrix of a real, stationary random process is a real, symmetric Toeplitz matrix. Thus, in maximum SINR filtering applications involving real, stationary random processes, we can replace a single channel system with a dual channel system to reduce the secondary data requirements by approximately 50% and ease the computational cost. Unfortunately, STAP correlation matrices are not centrosymmetric matrices and thus, we are faced with the difficult problem of determining if a fixed block diagonalizing transformation exists for STAP correlation matrices. The block diagonalization of STAP correlation matrices is addressed in the next chapter.

III. Block Diagonalizing STAP Correlation Matrices

3.1 Introduction

The key concept of the Block STAP method (dual channel system) is block diagonalizing the interference plus noise correlation matrix. Because a radar system operates in a wide variety of interference environments, the true correlation matrix is not known. However, the configuration of the radar system and the particular types of interference encountered give the correlation matrix, to some degree, a known structure. From this known structure, we can define a family of matrices in which the true correlation matrix is a member. Thus, our objective is to block diagonalize every member of a known family of correlation matrices with some transformation. To avoid the computational cost of searching for a new transformation every time the interference environment changes, the same transformation should block diagonalize every member of the family. Further, the transformation needs to have an efficient implementation to minimize the computational cost associated with transforming (preprocessing) the signal vectors. As noted earlier, this requirement for a fixed, efficient transformation introduces two new challenges: determining whether or not a fixed transformation exists for a particular family and determining if the transformation has an efficient implementation given that it exists.

The focus of this chapter is on the problem of determining if a fixed, block diagonalizing transformation exists for the family of STAP correlation matrices. Recall that the discrete Karhunen-Loève transform (DKLT) is constructed from the eigenvectors of the correlation matrix and is the unique unitary transformation that diagonalizes a correlation matrix [38:176]. One might conjecture the nonexistence of a fixed, block diagonalizing transformation for STAP correlation matrices is a foregone conclusion, based on the uniqueness of the DKLT. However, the uniqueness of the DKLT does not imply that a family of correlation matrices cannot be simultaneously block diagonalized or diagonalized by a fixed transformation. The conditions for simultaneously block diagonalizing a family of matrices are less restrictive than the conditions for simultaneously diagonalizing a family of matrices. Thus, the fact that a family of matrices cannot be simultaneously diagonalized does not imply that the family cannot be simultaneously block diagonalized.

In this chapter, we review the mathematical machinery available to address the problem of simultaneously diagonalizing and block diagonalizing a family of matrices and then, we apply this machinery to a family of STAP correlation matrices. We demonstrate that the defined family of STAP correlation matrices cannot be simultaneously diagonalized and provide evidence to support the conjecture that the family cannot be simultaneously block diagonalized.

The rest of this chapter is organized as follows. In Section 3.2, we define a family of STAP correlation matrices that is representative of the interference plus noise environment typically encountered by an airborne surveillance radar. We review the uniqueness of the DKLT and the conditions for simultaneously diagonalizing a family of correlation matrices and demonstrate that the defined family of STAP correlation matrices is not simultaneously diagonalizable in Section 3.3. We review what block diagonalizing a matrix means in terms of vector spaces and subspaces in Section 3.4. In Section 3.5, we discuss the available theorems on simultaneously block diagonalizing a family of matrices and apply these theorems to the defined family of STAP correlation matrices. We provide evidence to support the conjecture that the defined family cannot be simultaneously block diagonalized. In Section 3.6, we depart from the defined family of STAP correlation matrices and focus on the clutter correlation matrix. We show that the clutter correlation matrix is a centrosymmetric matrix under certain assumptions and discuss some of the potential uses of this result. Finally, we summarize the chapter in Section 3.7.

3.2 STAP Correlation Matrices

In this section, we define the family of STAP correlation matrices considered in this research. The family is representative of the interference plus noise environment typically encountered by a airborne surveillance radar, but we have used certain assumptions to limit the complexity of the family. We restricted the complexity to bound the scope of the research to the basic concepts. However, the family has sufficient complexity to stress the ability of a STAP system to remove both spatially and temporally correlated signals. The unwanted signals considered are receiver noise, barrage noise jamming, and ground clutter. The inclusion of ground clutter ensures the presence of an unwanted signal that is both

spatially and temporally correlated. The barrage noise jamming and ground clutter are referred to as interference and the receiver noise is referred to as noise. We assume that the receiver noise, barrage noise jamming, and ground clutter are zero-mean random processes and are uncorrelated with each other. Thus, the interference plus noise correlation matrix can be written as the sum of three correlation matrices. That is,

$$\mathbf{R} = \mathbf{R}_{\mathrm{R}} + \mathbf{R}_{\mathrm{J}} + \mathbf{R}_{\mathrm{C}}, \tag{3.1}$$

where $\mathbf{R}$, $\mathbf{R}_{\mathrm{R}}$, $\mathbf{R}_{\mathrm{J}}$, and $\mathbf{R}_{\mathrm{C}}$ are the interference plus noise, receiver noise, barrage noise jamming, and ground clutter correlation matrices, respectively. Recall that the radar system under consideration has a uniform linear array of N equally spaced antenna elements and processes M samples (pulses) in a coherent processing interval (CPI). We assume that the spacing between the antenna elements is half a wavelength and that the sampling interval in the temporal domain is the pulse repetition interval which is held constant over the CPI. Using the signal models defined by Ward [41], we now present the structure of the receiver noise, barrage noise jamming, and ground clutter correlation matrices.

The receiver noise is modeled as both spatially and temporally uncorrelated noise (white noise). Thus, the correlation matrix of the receiver noise is

$$\mathbf{R}_{\mathrm{R}} = \sigma^2 \mathbf{I}_{MN}, \tag{3.2}$$

where σ^2 is the variance (power) of the receiver noise and $\mathbf{I}_{MN}$ is a $MN \times MN$ identity matrix. Observe that $\mathbf{R}_{\mathrm{R}}$ is positive definite and Hermitian. Further, note that $\mathbf{R}_{\mathrm{R}}$ can be partitioned into a $M \times M$ block Toeplitz matrix where each diagonal block is a $N \times N$ identity matrix scaled by σ^2 and each off-diagonal block is a $N \times N$ zero matrix. Since the zero matrix and the identity matrix are Toeplitz matrices, $\mathbf{R}_{\mathrm{R}}$ is a Toeplitz-block-Toeplitz matrix.

The interference due to barrage noise jamming consists of one or more uncorrelated jamming sources (jammers). Thus, the jamming correlation matrix can be written as the

sum of the correlation matrices due to the individual jammers. That is,

$$\mathbf{R}_{\mathrm{J}} = \sum_{k=1}^{N_{\mathrm{J}}} \mathbf{R}_{\mathrm{J}}(k), \tag{3.3}$$

where $\mathbf{R}_{\mathrm{J}}(k)$ is the correlation matrix of the kth jammer and N_{J} is the total number of jammers. The signal from each jammer is modeled as a spatially correlated, temporally uncorrelated signal, yielding the following correlation matrix for the kth jammer:

$$\mathbf{R}_{\mathrm{J}}(k) = \sigma^2 \zeta_k \mathbf{I}_M \otimes \mathbf{a}(\upsilon_k)\mathbf{a}^H(\upsilon_k), \tag{3.4}$$

where $\otimes$ denotes the Kronecker product, ζ_k is the jamming to noise ratio (JNR) of the kth jammer, $\mathbf{I}_M$ is a $M \times M$ identity matrix, and $\mathbf{a}(\upsilon_k)$ is the $N \times 1$ spatial steering vector to the kth jammer. The $N \times 1$ spatial steering vector is defined as

$$\mathbf{a}(\upsilon) = \begin{bmatrix} 1 & e^{j2\pi\upsilon} & \dots & e^{j(N-1)2\pi\upsilon} \end{bmatrix}^T, \tag{3.5}$$

where the parameter υ is the spatial frequency (also referred to as the normalized angle) to the source. Now, observe that $\mathbf{I}_M$ and $\mathbf{a}(\upsilon_k)\mathbf{a}^H(\upsilon_k)$ are Toeplitz matrices and the Kronecker product of these two matrices yields a $M \times M$ block Toeplitz matrix, where each diagonal block is a $N \times N$ Toeplitz matrix given by $\mathbf{a}(\upsilon_k)\mathbf{a}^H(\upsilon_k)$ and each off-diagonal block is a $N \times N$ zero matrix. Thus, $\mathbf{R}_{\mathrm{J}}(k)$ is a Toeplitz-block-Toeplitz matrix, and since $\mathbf{R}_{\mathrm{J}}$ is the sum of similarly configured Toeplitz-block-Toeplitz matrices, $\mathbf{R}_{\mathrm{J}}$ is a Toeplitz-block-Toeplitz matrix. Also observe that $\mathbf{R}_{\mathrm{J}}$ is positive semidefinite and Hermitian. Finally, note that the development of Eqn. (3.4) assumes that the bandwidth of the jamming signal is greater than or equal to the bandwidth of the radar receiver and that the propagation time across the array is much less than the inverse of jamming bandwidth.

The interference due to ground clutter is modeled as a series of uncorrelated point scatters, referred to as clutter patches, that surround the radar and are located in the range gate of interest. The radar is assumed to be operating in an unambiguous range scenario (i.e., no second time around clutter) and that the axis of the array is perfectly aligned with the platform's velocity vector. Although the amplitude of the signal from

each clutter patch is random, the amplitude is assumed to be a constant over the CPI (i.e., no intrinsic clutter motion). Under these assumptions, the correlation matrix of the clutter is given by

$$\mathbf{R}_{\mathrm{C}} = \sum_{m=1}^{N_{\mathrm{C}}} \sigma^2 \xi_m \mathbf{b}(\omega_m)\mathbf{b}^H(\omega_m) \otimes \mathbf{a}(\upsilon_m)\mathbf{a}^H(\upsilon_m), \tag{3.6}$$

where N_{C} is the number of clutter patches, ξ_k is the clutter to noise ratio (CNR) of the mth clutter patch, $\mathbf{a}(\upsilon_m)$ is the $N \times 1$ spatial steering to the mth clutter patch, and $\mathbf{b}(\omega_m)$ is the $M \times 1$ Doppler steering vector of the mth clutter patch. The $M \times 1$ Doppler steering vector is defined as

$$\mathbf{b}(\upsilon) = \begin{bmatrix} 1 & e^{j2\pi\omega} & \dots & e^{j(M-1)2\pi\omega} \end{bmatrix}^T, \tag{3.7}$$

where the parameter ω is the normalized Doppler shift. Note that

$$\omega_m = \frac{2v_a T_r}{d}\upsilon_m = \beta\upsilon_m, \tag{3.8}$$

where v_a is the velocity of the platform, T_r is the pulse repetition interval, d is the spacing between antenna elements, and

$$\beta = \frac{2v_a T_r}{d}. \tag{3.9}$$

Now, observe that each clutter patch correlation matrix in Eqn. (3.6) is the Kronecker product of two Toeplitz matrices: $\mathbf{b}(\omega_m)\mathbf{b}^H(\omega_m)$ and $\mathbf{a}(\upsilon_m)\mathbf{a}^H(\upsilon_m)$. Thus, each clutter patch correlation matrix can be partitioned as a $M \times M$ block Toeplitz matrix where each block is a $N \times N$ Toeplitz matrix. That is, each clutter patch correlation matrix is a Toeplitz-block-Toeplitz matrix and therefore, $\mathbf{R}_{\mathrm{C}}$ is a Toeplitz-block-Toeplitz matrix. Also observe that $\mathbf{R}_{\mathrm{C}}$ is positive semidefinite and Hermitian.

With the correlation matrices of the receiver noise, barrage noise jamming, and ground clutter defined, we can now define the family of interference plus noise (STAP) correlation matrices considered in this research. First, observe that all three correlation

matrices are Hermitian and Toeplitz-block-Toeplitz matrices. Second, observe that the sum of a positive definite matrix and a positive semidefinite matrix is positive definite. Thus, the interference plus noise correlation matrices constructed from these three correlation matrices are Hermitian, positive definite, and Toeplitz-block-Toeplitz matrices. The family of correlation of interference plus noise correlation matrices considered in this research is defined as:

Definition 1 (Family of STAP Correlation Matrices). *Assume the radar system under consideration has a linear array of N equally spaced elements and that M samples (pulses) are collected in a coherent processing interval, where $N \geq 2$ and $M \geq 2$. Let $\mathcal{S}$ denote the family of interference plus noise (STAP) correlation matrices and $\mathbf{R}$ denote an arbitrary interference plus noise correlation matrix. Then, $\mathbf{R} \in \mathcal{S}$, if $\mathbf{R}$ is a Hermitian, positive definite, and Toeplitz-block-Toeplitz matrix of the following form:*

$$\begin{aligned}\mathbf{R} = \sigma^2 \mathbf{I}_{MN} \\ &+ \sum_{p=1}^{N_J} \sigma^2 \zeta_p \mathbf{I}_M \otimes \mathbf{a}(\upsilon_p)\mathbf{a}^H(\upsilon_p) \\ &+ \sum_{q=1}^{N_C} \sigma^2 \xi_q \mathbf{b}(\omega_q)\mathbf{b}^H(\omega_q) \otimes \mathbf{a}(\upsilon_q)\mathbf{a}^H(\upsilon_q),\end{aligned} \tag{3.10}$$

where

$$\mathbf{a}(\upsilon) = \begin{bmatrix} 1 & e^{j2\pi\upsilon} & \dots & e^{j(N-1)2\pi\upsilon} \end{bmatrix}^T \tag{3.11}$$

$$\mathbf{b}(\omega) = \begin{bmatrix} 1 & e^{j2\pi\omega} & \dots & e^{j(M-1)2\pi\omega} \end{bmatrix}^T \tag{3.12}$$

$$\omega_q = \frac{2v_a T_r}{d}\upsilon_q, \tag{3.13}$$

$\mathbf{I}_m$ is a $m \times m$ identity matrix, σ^2 is the receiver noise power, ζ_p is the jamming to noise ratio of the pth jammer, υ_p is the normalized angle to the pth jammer, ξ_q is the clutter to noise ratio of the qth clutter patch, υ_q is the normalized angle to the qth clutter patch, v_a is the velocity of the platform, T_r is the pulse repetition interval, and d is the spacing between antenna elements.

In defining the family $\mathcal{S}$, we used the assumptions listed below:

- The antenna is a linear array of equally spaced antenna elements.
- The sampling interval for the output signals from each antenna element is a constant for the coherent processing interval and equal to the pulse repetition interval.
- The receiver noise is both spatially and temporally uncorrelated and zero-mean.
- The jamming sources (jammers) are uncorrelated with each other.
- The jamming signal from any particular jammer is spatially correlated, temporally uncorrelated, zero-mean, has a bandwidth greater than or equal to the bandwidth of the radar receiver, and the propagation time across the array is much less than the inverse of the bandwidth of the jamming signal.
- The radar operates in an unambiguous range scenario (i.e., no second time around clutter).
- The clutter can be modeled as a series of uncorrelated point scatters, referred to as clutter patches, that surround the radar.
- The amplitude of each clutter patch is a zero-mean random variable with the amplitude held constant of the coherent processing interval (i.e., no intrinsic clutter motion).
- The axis of the antenna array is perfectly aligned with the platform's velocity vector (i.e., no velocity misalignment).

In general, without these assumptions, the family of interference plus noise correlation matrices would encompass nearly all Hermitian, positive definite matrices. Some of the assumptions (e.g., a linear array of equally spaced elements) represent a priori knowledge about the radar system that give the interference plus noise correlation matrices structure and bound the family $\mathcal{S}$ to a subset of the family of Hermitian, positive definite matrices. While other assumptions (e.g., no intrinsic clutter motion) represent simplifications designed to further bound the family (problem domain) and focus the research on analyzing the basic concepts of Block STAP. Although the family $\mathcal{S}$ is not the most general family,

because of these assumptions, the family $\mathcal{S}$ is sufficiently representative of typical interference plus noise signal environments encountered by airborne surveillance radars. As such, the members of $\mathcal{S}$ represent interference plus noise environments that will stress the ability of the Block STAP method to effectively remove unwanted signals that are both spatially and temporally correlated.

Observe that the receiver noise and barrage noise jamming correlation matrices have a block diagonal structure. Thus, if the interference environment does not include clutter, then we can use the identity matrix as the Block STAP transformation. That is, we would basically split the CPI into two temporal sub-CPIs and compute a weight vector for each sub-CPI using half the secondary data as a single channel system. However, the problem of block diagonalizing the interference plus noise correlation matrix becomes difficult when clutter is present, since its correlation matrix is not a block diagonal matrix. We cannot restrict the problem to block diagonalizing the clutter correlation matrix, because a transformation that block diagonalizes the clutter correlation matrix may destroy the block diagonal form of the jamming correlation. Notice that if we require the transformation to be unitary, then we can ignore the receiver noise correlation matrix. However, this unitary requirement is overly restrictive, since we can use any non-singular matrix to block diagonalize the correlation matrix and achieve the optimum performance when the correlation matrix is known. We examine the problem of simultaneously block diagonalizing the family $\mathcal{S}$ in the remaining parts of this chapter.

3.3 DKLT and Simultaneous Diagonalization

We can certainly obtain our objective of simultaneously block diagonalizing the family $\mathcal{S}$ if we can simultaneously diagonalize the family. Earlier, we noted that the uniqueness of the discrete Karhunen-Loève transform (DKLT) does not imply that a family of correlation matrices cannot be diagonalized. For completeness, we review the uniqueness of the DKLT in Section 3.3.1. The conditions for diagonalizing a family of matrices are given and applied to the family $\mathcal{S}$ in Section 3.3.2, where we show that the family $\mathcal{S}$ cannot be diagonalized.

3.3.1 Uniqueness of the DKLT. The following discussion on the uniqueness of the DKLT is based on a presentation given by Therrien [38:174–177]. Let $\mathbf{x} = [x_1 \; x_2 \; \cdots \; x_n]^T$ be a zero-mean random vector, then the $n \times n$ correlation matrix of $\mathbf{x}$ is

$$\mathbf{R}_x = \mathrm{E}\left\{\mathbf{x}\mathbf{x}^H\right\} = \begin{bmatrix} \mathrm{E}\left\{x_1 x_1^*\right\} & \mathrm{E}\left\{x_1 x_2^*\right\} & \cdots & \mathrm{E}\left\{x_1 x_n^*\right\} \\ \mathrm{E}\left\{x_2 x_1^*\right\} & \mathrm{E}\left\{x_2 x_2^*\right\} & \cdots & \mathrm{E}\left\{x_2 x_n^*\right\} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}\left\{x_n x_1^*\right\} & \mathrm{E}\left\{x_n x_2^*\right\} & \cdots & \mathrm{E}\left\{x_n x_n^*\right\} \end{bmatrix}. \tag{3.14}$$

The random vector $\mathbf{x}$ is the weighted sum of n basis vectors for the n-dimensional complex vector space, denoted as $\mathbb{C}^n$. Let $\mathbf{e}_i$ denote a $n \times 1$ vector with a one in the ith position and zero in all other positions. The set $\{\mathbf{e}_i\}_{i=1}^n$ is a basis for $\mathbb{C}^n$ and is referred to as the the standard ordered basis. In the standard ordered basis, the elements of $\mathbf{x}$ are the coefficients or coordinates for the basis vectors. That is,

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n. \tag{3.15}$$

Each of the elements (coefficients) of $\mathbf{x}$ is a random variable and the elements of $\mathbf{R}_x$ represent the correlation between the coefficients. Ideally, one would select a basis for representing $\mathbf{x}$ such that the correlation matrix is in a convenient form for reducing the computational cost of computing the optimum weight vector. If the basis is selected such that the coefficients are uncorrelated, then the correlation matrix is a diagonal matrix. A diagonal matrix is desirable since inverting a diagonal matrix only requires the inversion of n scalars. The transformation of $\mathbf{x}$ from one basis to another basis requires a change of basis matrix (transformation). Assume $\{\boldsymbol{\varphi}_i\}_{i=1}^n$ is a set of n orthonormal column vectors (and hence, a basis for $\mathbb{C}^n$) and $\mathbf{x}$ is defined with respect to the standard ordered basis, then the change of basis matrix is

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\varphi}_1 & \boldsymbol{\varphi}_2 & \cdots & \boldsymbol{\varphi}_n \end{bmatrix} \tag{3.16}$$

and the vector of coefficients of $\mathbf{x}$ in the basis $\{\boldsymbol{\varphi}_i\}_{i=1}^n$ is

$$\boldsymbol{\kappa} = \begin{bmatrix} \kappa_1 \\ \kappa_2 \\ \vdots \\ \kappa_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{\varphi}_1^H \mathbf{x} \\ \boldsymbol{\varphi}_2^H \mathbf{x} \\ \vdots \\ \boldsymbol{\varphi}_n^H \mathbf{x} \end{bmatrix} = \boldsymbol{\Phi}^H \mathbf{x}, \tag{3.17}$$

where $\kappa_i = \boldsymbol{\varphi}_i^H \mathbf{x}$. Observe that $\boldsymbol{\Phi}$ is a unitary matrix since $\{\boldsymbol{\varphi}_i\}_{i=1}^n$ is an orthonormal set. The objective is to select $\{\boldsymbol{\varphi}_i\}_{i=1}^n$ or equivalent, the change of basis matrix $\boldsymbol{\Phi}$ such that the resulting correlation matrix is a diagonal matrix. That is,

$$\mathrm{E}\left\{\kappa_p \kappa_q^*\right\} = \mathrm{E}\left\{\boldsymbol{\varphi}_p^H \mathbf{x}\mathbf{x}^H \boldsymbol{\varphi}_q\right\} = \boldsymbol{\varphi}_p^H \mathbf{R}_x \boldsymbol{\varphi}_q = \begin{cases} \varsigma_p^2 & p = q, \\ 0 & p \neq q. \end{cases} \tag{3.18}$$

The DKLT is the unique transformation that achieves the conditions of Eqn. (3.18) and has the property that the columns of the change of basis matrix are the eigenvectors of $\mathbf{R}_x$ (i.e., $\mathbf{R}_x \boldsymbol{\varphi}_i = \lambda_i \boldsymbol{\varphi}_i$, where λ_i is the corresponding eigenvalue). To examine the uniqueness of the DKLT, assume $\{\boldsymbol{\varphi}_i\}_{i=1}^n$ is an orthonormal set which is not necessarily the set of DKLT vectors. From the diagonal matrix objective, we have

$$\boldsymbol{\varphi}_p^H \mathbf{u}_q = \begin{cases} \varsigma_p^2 & p = q, \\ 0 & p \neq q, \end{cases} \tag{3.19}$$

where $\mathbf{u}_q = \mathbf{R}_x \boldsymbol{\varphi}_q$. Assume $\mathbf{u}_q$ is a non-zero vector in $\mathbb{C}^n$ which is not necessarily an eigenvector of $\mathbf{R}_x$, then $\mathbf{u}_q$ must be a linear combination of $\{\boldsymbol{\varphi}_i\}_{i=1}^n$. That is,

$$\mathbf{u}_q = c_1 \boldsymbol{\varphi}_1 + c_2 \boldsymbol{\varphi}_2 + \cdots + c_n \boldsymbol{\varphi}_n, \tag{3.20}$$

where not all of the c_i's are zeros. However, from Eqn. (3.19), we know that $\mathbf{u}_q$ is orthogonal to $\boldsymbol{\varphi}_i$ for $i \neq q$ and thus, $c_i = 0$ for $i \neq q$. Therefore, we must have $\mathbf{u}_q = c_q \boldsymbol{\varphi}_q$ and hence, $\mathbf{u}_q$ must be an eigenvector of $\mathbf{R}_x$ and c_q is the corresponding eigenvalue. Thus, the DKLT is the unique unitary transformation that diagonalizes a correlation matrix.

3.3.2 Diagonalizing the Family $\mathcal{S}$. The uniqueness of the DKLT does not imply that the family of correlation matrices cannot be simultaneously diagonalized with a single unitary transformation. Recall that the members of the family $\mathcal{S}$ are Hermitian matrices. Horn and Johnson [23:172] give the following theorem which provides the necessary and sufficient conditions for the simultaneous diagonalization of a family of Hermitian matrices:

Theorem 1 (Horn and Johnson). *Let $\mathcal{F}$ be a given family of Hermitian matrices. There exists a unitary transform matrix $\mathbf{U}$ such that $\mathbf{UAU}^H$ is diagonal for all $\mathbf{A} \in \mathcal{F}$ if and only if $\mathbf{AB} = \mathbf{BA}$ of all $\mathbf{A}, \mathbf{B} \in \mathcal{F}$.*

Thus, for a family of Hermitian matrices to be simultaneously unitarily similar to a diagonal matrix, the family must be a commuting family. The if and only if structure of Theorem 1 is convenient for showing that the family $\mathcal{S}$ is not simultaneously diagonalizable with a unitary transformation – we simply need to provide one example where two matrices from $\mathcal{S}$ do not commute. Consider two correlation matrices, $\mathbf{R}_1$ and $\mathbf{R}_2$, drawn from $\mathcal{S}$, where the interference plus noise signal consists of receiver noise, one dominate clutter patch (i.e., $N_{\mathrm{C}} = 1$), and no jammers (i.e., $N_{\mathrm{J}} = 0$). That is,

$$\mathbf{R}_i = \mathbf{I} + \mathbf{c}_i \mathbf{c}_i^H, \tag{3.21a}$$

where

$$\mathbf{c}_i = \mathbf{b}(\omega_i) \otimes \mathbf{a}(\upsilon_i), \tag{3.21b}$$

$$\mathbf{b}(\omega_i) = \begin{bmatrix} 1 & e^{j2\pi\omega_i} & e^{j(2)2\pi\omega_i} & \dots & e^{j(M-1)2\pi\omega_i} \end{bmatrix}^T, \tag{3.21c}$$

$$\mathbf{a}(\upsilon_i) = \begin{bmatrix} 1 & e^{j2\pi\upsilon_i} & e^{j(2)2\pi\upsilon_i} & \dots & e^{j(N-1)2\pi\upsilon_i} \end{bmatrix}^T, \tag{3.21d}$$

and $\omega_i = \beta\upsilon_i$. Note that, without loss of generality, we have assumed that the receiver noise power and the clutter power are equal to one. Let $[R]_{p,q}$ denote the element in the pth row and q^{th} column of the matrix $\mathbf{R}$ and $[\mathbf{c}]_p$ denote the pth element of $\mathbf{c}$. Now, notice that $[\mathbf{c}_p \mathbf{c}_q^H]_{0,0} = 1$ for $p, q = 1, 2$ since $[\mathbf{c}_p]_0 = 1$. We will show that $\mathbf{R}_1$ and $\mathbf{R}_2$ do not

commute, in general. If $\mathbf{R}_1$ and $\mathbf{R}_2$ commute, then the following must hold:

$$\begin{aligned}
\mathbf{R}_1\mathbf{R}_2 &= \mathbf{R}_2\mathbf{R}_1 \\
(\mathbf{I}+\mathbf{c}_1\mathbf{c}_1^H)(\mathbf{I}+\mathbf{c}_2\mathbf{c}_2^H) &= (\mathbf{I}+\mathbf{c}_2\mathbf{c}_2^H)(\mathbf{I}+\mathbf{c}_1\mathbf{c}_1^H) \\
\mathbf{c}_1\mathbf{c}_1^H\mathbf{c}_2\mathbf{c}_2^H &= \mathbf{c}_2\mathbf{c}_2^H\mathbf{c}_1\mathbf{c}_1^H. \qquad (3.22)
\end{aligned}$$

Thus, to show that $\mathbf{R}_1$ and $\mathbf{R}_2$ do not commute, we only need show that the clutter correlation matrices, $\mathbf{c}_1\mathbf{c}_1^H$ and $\mathbf{c}_2\mathbf{c}_2^H$, do not commute. First, observe that $\mathbf{c}_1^H\mathbf{c}_2$ and $\mathbf{c}_2^H\mathbf{c}_1$ are scalars and we can rewrite Eqn. (3.22) as

$$(\mathbf{c}_1^H\mathbf{c}_2)\mathbf{c}_1\mathbf{c}_2^H = (\mathbf{c}_2^H\mathbf{c}_1)\mathbf{c}_2\mathbf{c}_1^H. \qquad (3.23)$$

For Eqn. (3.23) to hold, all the elements of the matrix on the left hand side must equal the corresponding elements of the matrix of the right hand side. Since $[\mathbf{c}_1\mathbf{c}_2^H]_{0,0} = 1 = [\mathbf{c}_2\mathbf{c}_1^H]_{0,0}$, Eqn. (3.23) cannot hold if $\mathbf{c}_1^H\mathbf{c}_2$ does not equal $\mathbf{c}_2^H\mathbf{c}_1$. Next, notice that $\mathbf{c}_1^H\mathbf{c}_2 = (\mathbf{c}_2^H\mathbf{c}_1)^*$ which implies that $\mathbf{c}_1^H\mathbf{c}_2$ will equal $\mathbf{c}_2^H\mathbf{c}_1$ only if $\mathbf{c}_1^H\mathbf{c}_2$ is real. Thus, $\mathbf{R}_1$ and $\mathbf{R}_2$ can commute only if $\mathbf{c}_1^H\mathbf{c}_2$ is real. The inner product of $\mathbf{c}_1$ and $\mathbf{c}_2$ is

$$\begin{aligned}
\mathbf{c}_1^H\mathbf{c}_2 &= \Big(\mathbf{b}^H(\omega_1)\otimes\mathbf{a}^H(\upsilon_1)\Big)\Big(\mathbf{b}(\omega_2)\otimes\mathbf{a}(\upsilon_2)\Big) \\
&= \Big(\mathbf{b}^H(\omega_1)\mathbf{b}(\omega_2)\Big)\otimes\Big(\mathbf{a}^H(\upsilon_1)\mathbf{a}(\upsilon_2)\Big) \\
&= \Big(\mathbf{b}^H(\omega_1)\mathbf{b}(\omega_2)\Big)\Big(\mathbf{a}^H(\upsilon_1)\mathbf{a}(\upsilon_2)\Big) \\
&= e^{-j\pi(\upsilon_1-\upsilon_2)(N-1)}e^{-j\pi(\omega_1-\omega_2)(M-1)}\,\frac{\sin(\pi(\upsilon_1-\upsilon_2)N)}{\sin(\pi(\upsilon_1-\upsilon_2))}\,\frac{\sin(\pi(\omega_1-\omega_2)M)}{\sin(\pi(\omega_1-\omega_2))}. \qquad (3.24)
\end{aligned}$$

An examination of Eqn. (3.24) reveals that $\mathbf{c}_1^H\mathbf{c}_2$ is real only if

$$(\upsilon_1-\upsilon_2)(N-1)+(\omega_1-\omega_2)(M-1) = k, \qquad (3.25)$$

where k is an integer. By inspection, Eqn. (3.25) is not true for all values of υ_1 and υ_2 (recall that $\omega_i = \beta\upsilon_i$, where β is a real constant). Therefore, in general, $\mathbf{R}_1$ and $\mathbf{R}_2$ do not commute and we can conclude that the family $\mathcal{S}$ is not simultaneously diagonalizable by a single unitary transformation.

Theorem 1 only addresses the case of simultaneous diagonalization by a unitary transformation and not the more general case of a non-singular transformation. Horn and Johnson [23:239] provide the following theorem for the non-singular transformation case:

Theorem 2 (Horn and Johnson). *Let $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k \in M_n$ (M_n denotes the space of all complex matrices of size $n \times n$) be given Hermitian matrices with $\mathbf{A}_1$ non-singular. There exists a non-singular matrix $\mathbf{T} \in M_n$ such that $\mathbf{T}^H \mathbf{A}_i \mathbf{T}$ is diagonal for all $i = 1, 2, \ldots, k$ if and only if (a) $\mathbf{A}_1^{-1}\mathbf{A}_i$ is similar to a real diagonal matrix for all $i = 1, 2, \ldots, k$, and (b) $\{\mathbf{A}_1^{-1}\mathbf{A}_i : i = 2, \ldots, k\}$ is a commuting family.*

Assume that the family $\mathcal{S}$ can be simultaneously diagonalized by some non-singular transformation, then every subset of $\mathcal{S}$ is diagonalizable with the same non-singular transformation and Theorem 2 must hold. Thus, if we can find a subset of $\mathcal{S}$ such that either condition (a) or (b) of Theorem 2 does not hold, then we can conclude that the family $\mathcal{S}$ is not simultaneously diagonalizable by a non-singular transformation. Let $\mathcal{S}_3 = \{\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3\}$ be a subset of $\mathcal{S}$, where $\mathbf{R}_1 = \mathbf{I}$ and $\mathbf{R}_2$ and $\mathbf{R}_3$ are defined as in Eqn. (3.21). The matrix $\mathbf{R}_1$ represents the receiver noise only case (i.e., no clutter and no jamming). The set $\mathcal{S}_3$ satisfies condition (a) of Theorem 2 which is easily verified as follows. Since $\mathbf{R}_1$ is the identity matrix, condition (a) is simply that each matrix in $\mathcal{S}_3$ is similar to a real diagonal matrix. Each of the matrices in $\mathcal{S}_3$ is Hermitian and thus, is unitarily similar to a real diagonal matrix [23:171]. The set $\mathcal{S}_3$, however, does not satisfy condition (b) of Theorem 2. Since $\mathbf{R}_1$ is the identity matrix, condition (b) is simply that $\{R_i : i = 2, 3\}$ is a commuting family (i.e., $\mathbf{R}_2$ and $\mathbf{R}_3$ commute). Now, notice that the matrices $\mathbf{R}_2$ and $\mathbf{R}_3$ are defined by Eqn. (3.21) and as previously demonstrated, any pair of matrices of this form do not commute, in general. Therefore, the set $\mathcal{S}_3$ cannot be simultaneously diagonalized by a non-singular transformation which implies that $\mathcal{S}$ cannot be simultaneously diagonalized by a non-singular transformation.

The fact that the family $\mathcal{S}$ cannot be simultaneously diagonalized by non-singular (unitary or otherwise) transformation does not imply that $\mathcal{S}$ cannot be simultaneously block diagonalized. The simultaneous block diagonalization of family of matrices is less restrictive as will be discussed in the next sections.

3.4 Invariant Subspaces

To provide a better understanding of the conditions for simultaneously block diagonalizing a family of matrices, we review what block diagonalizing a matrix means in terms of vector spaces and subspaces in this section. Each $\mathbf{R}$ in $\mathcal{S}$ is a matrix representation of a linear operator R that maps a vector from the MN-dimensional vector space over the complex numbers, denoted by $\mathbb{C}^{MN}$, to another vector in $\mathbb{C}^{MN}$. The matrix representation $\mathbf{R}$ of R is defined with respect to some ordered basis for $\mathbb{C}^{MN}$. In the absence of any a priori knowledge about the basis, we will assume the basis is the standard ordered basis for $\mathbb{C}^{MN}$ (i.e., $\{\boldsymbol{e}_i\}_{i=1}^{MN}$, where $\mathbf{e}_i$ is a vector with a 1 in the ith element and zero everywhere else). Given $\mathbf{R}$ and the basis used to define $\mathbf{R}$, the matrix representation of R in another ordered basis is [21:92]

$$\tilde{\mathbf{R}} = \mathbf{P}^{-1}\mathbf{R}\mathbf{P}, \tag{3.26}$$

where the columns of $\mathbf{P}$ are the new basis vectors written in terms of the old basis vectors and $\mathbf{P}$ is referred to as the change of basis matrix. Now, note that a matrix $\mathbf{A}$ is said to be similar to a matrix $\mathbf{B}$ if $\mathbf{A} = \mathbf{U}^{-1}\mathbf{B}\mathbf{U}$ and the transformation $\mathbf{U}^{-1}\mathbf{B}\mathbf{U}$ is referred to as a similarity transformation [23:44]. Thus, the change of basis operation given in Eqn. (3.26) is a similarity transformation. If $\mathbf{R}$ was originally defined with respect to the standard ordered basis, then the columns of $\mathbf{P}$ are simply the new basis vectors. Further, if the basis vectors form an orthonormal set, then the change of basis matrix is unitary (i.e., $\mathbf{P}^H\mathbf{P} = \mathbf{I}$ implying $\mathbf{P}^H = \mathbf{P}^{-1}$) and we can rewrite Eqn. (3.26) as

$$\tilde{\mathbf{R}} = \mathbf{P}^H\mathbf{R}\mathbf{P} \tag{3.27}$$

and $\mathbf{R}$ is said to be unitarily similar to $\tilde{\mathbf{R}}$ through the unitary similarity transformation $\mathbf{P}^H\mathbf{R}\mathbf{P}$.

Recall that our objective is to select a non-singular transformation matrix $\mathbf{V}$ such that for every $\mathbf{R}$ in $\mathcal{S}$ we have

$$\mathbf{V}^H\mathbf{R}\mathbf{V} = \begin{bmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 \end{bmatrix}, \tag{3.28}$$

where $\mathbf{D}_1$ and $\mathbf{D}_2$ are $MN/2 \times MN/2$ matrices. If $\mathbf{V}$ is restricted to unitary, then the block diagonalization of $\mathbf{R}$ is a similarity transformation since $\mathbf{V}^H\mathbf{R}\mathbf{V} = \mathbf{V}^{-1}\mathbf{R}\mathbf{V}$ and as such, involves changing the basis for representing the linear operator R as a matrix. The existence of a block diagonalizing change of basis matrix implies that special subspaces of $\mathbb{C}^{MN}$ exist, which we discuss next. The case where $\mathbf{V}$ is any non-singular matrix will be addressed later.

A subspace is a subset of vectors from a vector space that also forms a vector space. That is, let Z denote a vector space over the complex numbers and let X denote a subset of Z. Then X is a subspace if $cx_1 + x_2 \in X$ for all $x_1, x_2 \in X$, where c is an arbitrary complex number [21:35]. Let $W_1, \ldots, W_k$ be subspaces of the vector space Z, then the subspaces are said to be independent if

$$w_1 + w_2 + \cdots + w_k = 0, \quad \text{for } w_i \in W_i, \tag{3.29}$$

implies that each w_i is the zero vector [21:209]. For the case of two subspaces, W_1 and W_2 are independent if the intersection of W_1 and W_2 is the zero vector. We now narrow our attention to finite-dimensional vector spaces (i.e., a basis for the vector space has a finite number of vectors). The dimension of a finite-dimensional vector space is the number of vectors in a basis for the vector space. The dimension of a finite-dimensional vector space will be denoted as $\dim(\cdot)$. Let $W = W_1 + \cdots + W_k$, where $W_1, \ldots, W_k$ are independent subspaces of the vector space Z and let $\mathcal{B}_i$ denote the set of basis vectors for W_i. Then, the set $\{\mathcal{B}_1, \ldots, \mathcal{B}_k\}$ is a basis for the subspace W and the dimension of W is [21:209]

$$\dim(W) = \dim(W_1) + \cdots + \dim(W_k). \tag{3.30}$$

Now, notice that if the $\dim(W) = \dim(Z)$, then set $\{\mathcal{B}_1, \ldots, \mathcal{B}_k\}$ is a basis for Z and Z is said to be the direct sum of $W_1, \ldots, W_k$ which is denoted as [21:210]

$$Z = W_1 \oplus \cdots \oplus W_k. \tag{3.31}$$

When Z is the direct sum of the subspaces $W_1, \ldots, W_k$, we can write any vector in Z as the sum of vectors from each of the subspaces. That is, let $y \in Z$ and $Z = W_1 \oplus \cdots \oplus W_k$, then

$$y = w_1 + w_2 + \cdots + w_k, \tag{3.32}$$

where $w_i \in W_i$. Thus, the construction of Z as the direct sum of independent subspaces allows us to decompose any vector into several independent components. As we discuss next, this decomposition of the vector space is essential in block diagonalizing a linear operator.

Let Z be a n-dimensional vector space over the complex numbers and let U and W be subspaces of Z such that $Z = U \oplus W$. Let $\mathcal{B}_u = \{\gamma_1, \ldots, \gamma_k\}$ and $\mathcal{B}_u = \{\gamma_{k+1}, \ldots, \gamma_n\}$ be ordered basis for the subspaces U and W, respectively, where $k = n/2$. Hence, the set $\mathcal{B} = \{\mathcal{B}_u, \mathcal{B}_w\} = \{\gamma_1, \ldots, \gamma_k, \gamma_{k+1}, \ldots, \gamma_n\}$ is an ordered basis of Z. Let R be a linear operator on Z. Now, observe that $R\gamma_p \in Z$ and thus, the vector $R\gamma_p$ must be a linear combination of the basis vectors in $\mathcal{B}$. That is,

$$R\gamma_p = c_{1p}\gamma_1 + c_{2p}\gamma_2 + \cdots + c_{np}\gamma_n, \tag{3.33}$$

where the c_{pq}'s are complex numbers and are referred to as the coefficients or coordinates of $R\gamma_p$ with repsect to the basis $\mathcal{B}$. If $\gamma_p \in U$ and $R\gamma_p \in U$ for $1 \leq p \leq k$, then the subspace U is said to be invariant under R or R-invariant [21:199]. If the subspace U is R-invariant and $\gamma_p \in U$, then Eqn. (3.33) reduces to

$$R\gamma_p = c_{1p}\gamma_1 + c_{2p}\gamma_2 + \cdots + c_{kp}\gamma_k. \tag{3.34}$$

Thus, when $\gamma_p \in U$ and U is R-invariant, we only need half the basis vectors of Z to write $R\gamma_p$. The invariance of U under R induces a linear operator R_u on the subspace U (i.e., $Rx = R_u x$ if $x \in U$) [21:199]. If the subspace W is R-invariant, then a linear operator R_w is induced on the subspace W (i.e., $Rx = R_w x$ if $x \in W$). Recall that since $Z = U \oplus W$, any vector $x \in Z$ can be written as $x = w + u$, where $w \in W$ and $u \in U$. Using the invariance of U and W under R, we can write

$$Rx = R(u + w) = R_u u + R_w w. \tag{3.35}$$

Thus, the combination of the direct sum property and the invariance of the subspaces decompose the linear operator R into R_u and R_w. The decomposition of R leads to a block diagonal matrix representation of R with respect to the basis $\mathcal{B} = \{\mathcal{B}_u, \mathcal{B}_w\}$.

The following is based on a discussion presented by Hoffman and Kunze [21:200]. Let $\mathbf{R}$ be the matrix representation of the linear operator R with respect to $\mathcal{B}$ and $[\mathbf{R}]_{pq}$ denote the element of $\mathbf{R}$ in the pth row and q^{th} column. In general, $R\gamma_p$ is a linear combination of the basis vectors in $\mathcal{B}$ as given in Eqn. (3.33). If we let the coordinates of $R\gamma_p$ be the pth column of $\mathbf{R}$, then

$$R\gamma_p = \sum_{q=1}^{n} [\mathbf{R}]_{qp} \gamma_q. \tag{3.36}$$

Now, notice that if the subspace U is R-invariant and $p \leq k$, then $\gamma_p \in U$ and hence, $R\gamma_p \in U$. Thus, $R\gamma_p$ is a linear combination of the basis vectors in $\mathcal{B}_u$ (and not $\mathcal{B}$) and we can rewrite Eqn. (3.36) as

$$R\gamma_p = \sum_{q=1}^{k} [\mathbf{R}]_{qp} \gamma_q, \tag{3.37}$$

which implies the $[\mathbf{R}]_{qp} = 0$ for $k + 1 \leq q \leq n$ and $1 \leq p \leq k$. Similarly, if $p \geq k + 1$, then $\gamma_p \in W$ and $R\gamma_p \in W$. Thus, we can write Eqn. (3.36) as

$$R\gamma_p = \sum_{q=k+1}^{n} [\mathbf{R}]_{qp} \gamma_q, \tag{3.38}$$

which implies the $[\mathbf{R}]_{qp} = 0$ for $1 \leq q \leq k$ and $k+1 \leq p \leq n$. Combining these results, we have

$$[\mathbf{R}]_{qp} = \begin{cases} 0 & \text{if } k+1 \leq q \leq n \text{ and } 1 \leq p \leq k; \\ 0 & \text{if } 1 \leq q \leq k \text{ and } k+1 \leq p \leq n; \\ c_{qp} & \text{otherwise,} \end{cases} \tag{3.39}$$

where c_{qp} is a complex number. An examination of Eqn. (3.39) reveals that the matrix representation $\mathbf{R}$ of R with respect to the basis $\mathcal{B}$ is a block diagonal matrix of the form

$$\mathbf{R} = \begin{bmatrix} \mathbf{D}_u & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_w \end{bmatrix}, \tag{3.40}$$

where $\mathbf{D}_u$ and $\mathbf{D}_w$ are $n/2 \times n/2$ matrices and are the matrix representations of the induced linear operators R_u and R_u, respectively. In general, if $Z = W_1 \oplus \cdots \oplus W_m$ and all the subspace $W_1, \ldots, W_m$ are R-invariant, then

$$\mathbf{R} = \begin{bmatrix} \mathbf{D}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{D}_m \end{bmatrix}, \tag{3.41}$$

where the dimension of the square matrix $\mathbf{D}_i$ is equal to $\dim(W_i)$ [29:371]. Therefore, if we want to diagonalize a linear operator R on a n-dimensional vector space, then n R-invariant, independent subspaces must exist. This is in sharp contrast to the block diagonalization of R with blocks of size $n/2 \times n/2$, which only requires two R-invariant, independent subspaces.

Summarizing, each $\mathbf{R}$ in $\mathcal{S}$ is a matrix representation of a linear operator R on the vector space $\mathbb{C}^{MN}$ with respect to some ordered basis for $\mathbb{C}^{MN}$. The basis is assumed to be the standard ordered basis of $\mathbb{C}^{MN}$. If $\mathbb{C}^{MN}$ is the direct sum of R-invariant subspaces, then the matrix representation of R with respect to the ordered bases for the R-invariant subspaces will be a block diagonal matrix. When $\mathbf{V}$ is restricted to unitary, the block

diagonalization of $\mathbf{R}$ is a unitary similarity transformation involving a unitary change of basis matrix. Thus, to simultaneously block diagonalize $\mathcal{S}$ as in Eqn. (3.28) where $\mathbf{V}$ is unitary, the vector space $\mathbb{C}^{MN}$ must be the direct sum of two subspaces which are invariant to every $\mathbf{R}$ in $\mathcal{S}$. Further, the union of the basis vectors for the two invariant, independent subspaces must be an orthonormal set. The possibility of these conditions occurring is discussed in the next section.

3.5 Simultaneous Block Diagonalization of the Family $\mathcal{S}$

The concept of transforming a set or family of matrices into a particular form (e.g., diagonal and triangular) is a well researched topic in mathematics and continues to be a topic of interest [28]. Watters [42] appears to be one of the first to address the issue of simultaneously block diagonalizing a set of matrices using a similarity transformation. Watters [42] focused on the special case of block diagonalizing a family with a unitary similarity transformation where the blocks on the diagonal are of size 2×2 and possibly one block of size 1×1. Barker et. al. [3] and Shapiro [36] (Laffey [28] provides a survey paper) examined the general case using theorems that pre-date the work of Watters, suggesting that Watters may not have been the first. Regardless of who was first, the end result is a set of theorems for examining the existence of a similarity transformation that simultaneously block diagonalizes a family of matrices. Recall that the block diagonalization of $\mathbf{R} \in \mathcal{S}$ with the transformation $\mathbf{V}^H\mathbf{R}\mathbf{V}$ is a similarity transformation if $\mathbf{V}$ is unitary. Thus, we can use the results of Watters, Barker et. al., Shapiro, and Laffey to examine the existence of a unitary $\mathbf{V}$ that block diagonalizes the family $\mathcal{S}$. If $\mathbf{V}$ is simply non-singular (i.e., not unitary), then the transformation $\mathbf{V}^H\mathbf{R}\mathbf{V}$ is not a similarity transformation and the results from the above cited papers do not apply. When $\mathbf{V}$ is simply non-singular, the transformation $\mathbf{V}^H\mathbf{R}\mathbf{V}$ is referred to a *congruent (star congruent) transformation, where a matrix $\mathbf{A}$ is said to be *congruent to matrix $\mathbf{B}$ if $\mathbf{A} = \mathbf{U}^H\mathbf{B}\mathbf{U}$ for some non-singular matrix $\mathbf{U}$ [23:220]. In contrast to the similarity transformation, a *congruent transformation does not, in general, represent a change of basis and as a result, statements about the relationships between the vector space, subspace, and linear operators do not exist. What can be said about the *congruent transformation is that it is an equivalence relation and it

preserves the inertia of Hermitian matrices (i.e., the ordered triple of the number of positive, negative, and zero eigenvalues) [23:221]. Although research does exists for *congruent diagonalization, research on the *congruent block diagonalization of a family of matrices does not appear to exist. As such, this section will focus on the block diagonalization of $\mathcal{S}$ through a similarity transformation. In particular, we start in Section 3.5.1 with a few observations about block diagonal matrices and then, we present a discussion on the basic concept used by Watters [42], Barker et. al. [3], Shapiro [36], and Laffey [27] [28]. Finally, in Sections 3.5.2 and 3.5.3, we state the results from these papers and apply the results to the family $\mathcal{S}$. Although we are able to show that a unitary $\mathbf{V}$ does not exist in several cases where M and N are small, we cannot provide a conclusive proof for the general case.

3.5.1 Observations and Basic Concepts. Let $\mathrm{BD}(k_1, \ldots, k_n)$ denote a block diagonal matrix with n square blocks (matrices) on the diagonal of sizes $k_1, \ldots, k_n$. Now, observe that the sum and product of two $\mathrm{BD}(k_1, \ldots, k_n)$ matrices are $\mathrm{BD}(k_1, \ldots, k_n)$ matrices which is easily verified as follows. Let $\mathbf{X}$ and $\mathbf{Y}$ be $\mathrm{BD}(k_1, \ldots, k_n)$ matrices:

$$\mathbf{X} = \begin{bmatrix} \mathbf{D}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{D}_n \end{bmatrix} \quad \text{and} \quad \mathbf{Y} = \begin{bmatrix} \mathbf{E}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{E}_n \end{bmatrix},$$

where $\mathbf{D}_i$ and $\mathbf{E}_i$ are $k_i \times k_i$ matrices, and observe that

$$\begin{aligned} \mathbf{X} + \mathbf{Y} &= \begin{bmatrix} \mathbf{D}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{D}_n \end{bmatrix} + \begin{bmatrix} \mathbf{E}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{E}_n \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{D}_1 + \mathbf{E}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 + \mathbf{E}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{D}_n + \mathbf{E}_n \end{bmatrix} \end{aligned}$$

and

$$\mathbf{XY} = \begin{bmatrix} \mathbf{D}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{D}_n \end{bmatrix} \begin{bmatrix} \mathbf{E}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{E}_n \end{bmatrix} = \begin{bmatrix} \mathbf{D}_1\mathbf{E}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2\mathbf{E}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{D}_n\mathbf{E}_n \end{bmatrix}.$$

Using the associative property of matrix addition and multiplication, one can extended these results to the case of an arbitrary number of $\mathrm{BD}(k_1, \ldots, k_n)$ matrices. Next, observe that if $\mathbf{X}$ is a $\mathrm{BD}(k_1, \ldots, k_n)$ matrix as defined above and $p(\cdot)$ denotes a polynomial, then

$$p(\mathbf{X}) = \begin{bmatrix} p(\mathbf{D}_1) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & p(\mathbf{D}_2) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & p(\mathbf{D}_n) \end{bmatrix}.$$

Further, if $p(\mathbf{X}) = \mathbf{0}$, then $p(\mathbf{D}_i) = \mathbf{0}$ for all $i = 1, \ldots, n$. Additionally, observe that if $\mathbf{X}$ is an arbitrary matrix and $\mathbf{U}$ is a non-singular matrix, then $p(\mathbf{U}^{-1}\mathbf{X}\mathbf{U}) = \mathbf{U}^{-1}p(\mathbf{X})\mathbf{U}$. Finally, recall that an algebra is a vector space over a field with an additional operation called vector multiplication which produces another vector in the algebra and is associative, distributive with respect to vector addition, and associative with respect to scalar multiplication [21:117]. Let $\mathcal{T}$ be a set of complex matrices that is simultaneously block diagonalizable with a similarity transformation involving the non-singular matrix $\mathbf{U}$. That is, if $\mathbf{X} \in \mathcal{T}$, then $\mathbf{U}^{-1}\mathbf{X}\mathbf{U}$ is a $\mathrm{BD}(k_1, \ldots, k_n)$ matrix. One can verify that the algebra generated by $\mathcal{T}$, using the standard definitions for matrix addition and multiplication, is block diagonalizable by $\mathbf{U}$ as follows. Let $A_{\mathcal{T}}$ denote the algebra generated by $\mathcal{T}$ and notice that the elements in $A_{\mathcal{T}}$ are generated from the following basic forms:

$$c\mathbf{X}_p, \tag{3.42}$$

$$\mathbf{X}_p + \mathbf{X}_q, \tag{3.43}$$

$$\mathbf{X}_p\mathbf{X}_q, \tag{3.44}$$

where c is in the field and $\mathbf{X}_p$ and $\mathbf{X}_q$ are in $\mathcal{T}$. Now, notice that $\mathbf{D}_p = \mathbf{U}^{-1}\mathbf{X}_p\mathbf{U}$ and $\mathbf{D}_q = \mathbf{U}^{-1}\mathbf{X}_q\mathbf{U}$, where $\mathbf{D}_p$ and $\mathbf{D}_q$ are $\mathrm{BD}(k_1, \ldots, k_n)$ matrices. Next, observe that

$$\mathbf{U}^{-1}(c\mathbf{X}_p)\mathbf{U} = c\mathbf{U}^{-1}\mathbf{X}_p\mathbf{U} = c\mathbf{D}_p, \tag{3.45}$$

$$\mathbf{U}^{-1}(\mathbf{X}_p + \mathbf{X}_q)\mathbf{U} = \mathbf{U}^{-1}\mathbf{X}_p\mathbf{U} + \mathbf{U}^{-1}\mathbf{X}_q\mathbf{U} = \mathbf{D}_p + \mathbf{D}_q, \tag{3.46}$$

$$\mathbf{U}^{-1}(\mathbf{X}_p\mathbf{X}_q)\mathbf{U} = \mathbf{U}^{-1}(\mathbf{U}\mathbf{D}_p\mathbf{U}^{-1})(\mathbf{U}\mathbf{D}_q\mathbf{U}^{-1})\mathbf{U} = \mathbf{D}_p\mathbf{D}_q, \tag{3.47}$$

and as previously noted, the sum and product of block diagonal matrices of the same form are also block diagonal matrices of the same form. Thus, the algebra generated by $\mathcal{T}$ is block diagonalizable if $\mathcal{T}$ is block diagonalizable. Conversely, since $\mathcal{T}$ is a subset of $A_{\mathcal{T}}$, if $A_{\mathcal{T}}$ is block diagonalizable, then $\mathcal{T}$ is block diagonalizable. Using the algebra generated by the set, instead of the set directly, allows one to take full advantage of the vast research available on algebras.

The notion of polynomial identities is the central concept in the works of Watters, Barker et. al., Shapiro, and Laffey. Basically, if the algebra generated by a family (set) of matrices is block diagonalizable, then the algebra must satisfy a polynomial identity. Let $\mathrm{M}_n(F)$ denote the full matrix algebra of $n \times n$ matrices over the field F (e.g., real numbers or complex numbers), then a polynomial identity is defined as [28]

Definition 2. *A non-zero polynomial $p(x_1, \ldots, x_m)$ in the non-commuting indeterminants (variables) $x_1, \ldots, x_n$ is called a polynomial identity (PI) for $M_n(F)$ if*

$$p(\mathbf{A}_1, \ldots, \mathbf{A}_m) = \mathbf{0}$$

for all elements $\mathbf{A}_1, \ldots, \mathbf{A}_m \in M_n(F)$.

An example of a polynomial identity for $\mathrm{M}_2(\mathbb{C})$ is $p_2(x_1, x_2, x_3) = (x_1x_2 - x_2x_1)^2x_3 - x_3(x_1x_2 - x_2x_1)^2$. That is, if we select three arbitrary 2×2 complex matrices, say $\mathbf{X}_1$, $\mathbf{X}_2$,

and $\mathbf{X}_3$, and substitute $\mathbf{X}_i$ for x_i into $p_2(x_1, x_2, x_3)$, then

$$p_2(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3) = (\mathbf{X}_1\mathbf{X}_2 - \mathbf{X}_2\mathbf{X}_1)^2\mathbf{X}_3 - \mathbf{X}_3(\mathbf{X}_1\mathbf{X}_2 - \mathbf{X}_2\mathbf{X}_1)^2 = \mathbf{0}. \tag{3.48}$$

Note that if a polynomial is a polynomial identity for $\mathrm{M}_n(F)$, then we will refer to the polynomial as a polynomial identity of degree n. Additionally, note that if a polynomial is a polynomial identity for $\mathrm{M}_n(F)$, then the polynomial is also a polynomial identity for $\mathrm{M}_k(F)$ where $k < n$ [42].

Now, using the property that a polynomial identity for $\mathrm{M}_n(\mathbb{C})$ is zero for every element in $\mathrm{M}_n(\mathbb{C})$, we can examine the implications on the block diagonalization of an algebra. Let $\mathcal{T}$ be a set of $n \times n$ matrices that are simultaneously block diagonalizable to $\mathrm{BD}(n/2, n/2)$ matrices with a similarity transformation involving the non-singular matrix $\mathbf{U}$. Additionally, let $A_{\mathcal{T}}$ denote the algebra generated by $\mathcal{T}$. Thus, for any element $\mathbf{T}_i$ in $A_{\mathcal{T}}$, we can write

$$\mathbf{U}^{-1}\mathbf{T}_i\mathbf{U} = \mathbf{D}_i = \begin{bmatrix} \mathbf{E}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_i \end{bmatrix}, \tag{3.49}$$

where $\mathbf{E}_i$ and $\mathbf{F}_i$ are $n/2 \times n/2$ matrices. Also notice that $\mathbf{T}_i = \mathbf{U}\mathbf{D}_i\mathbf{U}^{-1}$. Next, let $p(\mathbf{T}_1, \ldots, \mathbf{T}_m)$ be a polynomial, where $\mathbf{T}_1, \ldots, \mathbf{T}_m$ are arbitrary matrices in $A_{\mathcal{T}}$, and observe that

$$p(\mathbf{T}_1, \ldots, \mathbf{T}_m) = \mathbf{U}p(\mathbf{D}_1, \ldots, \mathbf{D}_m)\mathbf{U}^{-1} = \mathbf{U}\begin{bmatrix} p(\mathbf{E}_1, \ldots, \mathbf{E}_m) & \mathbf{0} \\ \mathbf{0} & p(\mathbf{F}_1, \ldots, \mathbf{F}_m) \end{bmatrix}\mathbf{U}^{-1}. \tag{3.50}$$

If the polynomial $p(x_1, \ldots, x_m)$ is a polynomial identity for $\mathrm{M}_{n/2}(\mathbb{C})$, then $p(\mathbf{E}_1, \ldots, \mathbf{E}_m) = \mathbf{0}$ and $p(\mathbf{F}_1, \ldots, \mathbf{F}_m) = \mathbf{0}$ which implies $p(\mathbf{T}_1, \ldots, \mathbf{T}_m) = \mathbf{0}$. That is, the algebra $A_{\mathcal{T}}$ must satisfy a polynomial identity for $\mathrm{M}_{n/2}(\mathbb{C})$. Thus, for a family of matrices to be simultaneously block diagonalizable with a similarity transformation, the algebra generated by the family must satisfy a polynomial identity for the dimension of the largest block. Al-

though the above discussion highlights the basic concept, it does not reflect the complex mathematics used in deriving the results presented next.

3.5.2 Quasi-Diagonalization. Watters [42] investigated the conditions for a family of normal matrices (recall that Hermitian implies normal) to be simultaneously unitarily similar to quasi-diagonal matrices, where a quasi-diagonal matrix is defined as a block diagonal matrix with blocks of size 2×2 and possibly one block of size 1×1 (i.e., $\mathrm{BD}(2, 2, \ldots, 2, \delta)$ where δ is either 1 or 2). Watters [42] proves the following theorem:

Theorem 3 (Watters). *Let $\mathcal{T}$ be a family of Hermitian matrices in $M_n(\mathbb{C})$. The algebra $A_{\mathcal{T}}$ is of unitary type $(2, 2, \ldots, \delta)$, where $\delta = 1$ or 2, depending on the parity of n (i.e., $A_{\mathcal{T}}$ is simultaneously unitarily similar to $BD(2, 2, \ldots, 2, \delta)$ matrices), if and only if*

$$(\mathbf{AP} - \mathbf{PA})^2\mathbf{Q} - \mathbf{Q}(\mathbf{AP} - \mathbf{PA})^2 = \mathbf{0}$$

for all $\mathbf{A} \in A_{\mathcal{T}}$, $\mathbf{P}$ and $\mathbf{Q} \in \mathcal{T}$.

Theorem 3 provides the necessary and sufficient conditions for a family of Hermitian matrices to be simultaneously quasi-diagonalizable with a unitary similarity transformation. Essentially, Theorem 3 states that the algebra generated by a family of Hermitian matrices must satisfy the polynomial identity $p(x_1, x_2, x_3) = (x_1x_2 - x_2x_1)^2x_3 - x_3(x_1x_2 - x_2x_1)^2$ to be simultaneously unitary quasi-diagonalizable. Showing that an algebra satisfies a polynomial identity is a formidable task. However, the true value of Theorem 3 may lie in proving a family of Hermitian matrices is not simultaneously unitarily quasi-diagonalizable, since one only needs to provide a single counter-example. Next, we apply Theorem 3 to the previously defined family $\mathcal{S}$ to conclude that $\mathcal{S}$ is not, in general, quasi-diagonalizable.

Consider the following three matrices in $\mathcal{S}$ and hence, in $A_{\mathcal{S}}$:

$$\begin{aligned}
\mathbf{R}_1 &= \mathbf{I} + \mathbf{c}_1\mathbf{c}_1^H = \mathbf{I} + \mathbf{C}_1, \\
\mathbf{R}_2 &= \mathbf{I} + \mathbf{c}_2\mathbf{c}_2^H = \mathbf{I} + \mathbf{C}_2, \\
\mathbf{R}_3 &= \mathbf{I} + \mathbf{c}_3\mathbf{c}_3^H = \mathbf{I} + \mathbf{C}_3,
\end{aligned}$$

where $\mathbf{C}_i = \mathbf{c}_i\mathbf{c}_i^H$ and represents the clutter correlation matrix from a single dominant scatter (See Eqn. (3.21) for details). To prove $\mathcal{S}$ is not simultaneously unitarily quasi-diagonalizable, we must show that

$$(\mathbf{R}_1\mathbf{R}_2 - \mathbf{R}_2\mathbf{R}_1)^2\mathbf{R}_3 \neq \mathbf{R}_3(\mathbf{R}_1\mathbf{R}_2 - \mathbf{R}_2\mathbf{R}_1)^2. \tag{3.51}$$

We can simplify this condition by first observing that

$$\begin{aligned}
(\mathbf{R}_1\mathbf{R}_2 - \mathbf{R}_2\mathbf{R}_1)^2\mathbf{R}_3 &= \left((\mathbf{I}+\mathbf{C}_1)(\mathbf{I}+\mathbf{C}_2) - (\mathbf{I}+\mathbf{C}_2)(\mathbf{I}+\mathbf{C}_1)\right)^2\mathbf{R}_3 \\
&= \left(\mathbf{I}+\mathbf{C}_1+\mathbf{C}_2+\mathbf{C}_1\mathbf{C}_2 - (\mathbf{I}+\mathbf{C}_1+\mathbf{C}_2+\mathbf{C}_2\mathbf{C}_1)\right)^2\mathbf{R}_3 \\
&= (\mathbf{C}_1\mathbf{C}_2 - \mathbf{C}_2\mathbf{C}_1)^2(\mathbf{I}+\mathbf{C}_3) \\
&= (\mathbf{C}_1\mathbf{C}_2 - \mathbf{C}_2\mathbf{C}_1)^2 + (\mathbf{C}_1\mathbf{C}_2 - \mathbf{C}_2\mathbf{C}_1)^2\mathbf{C}_3 \qquad (3.52) \\
\mathbf{R}_3(\mathbf{R}_1\mathbf{R}_2 - \mathbf{R}_2\mathbf{R}_1)^2 &= (\mathbf{C}_1\mathbf{C}_2 - \mathbf{C}_2\mathbf{C}_1)^2 + \mathbf{C}_3(\mathbf{C}_1\mathbf{C}_2 - \mathbf{C}_2\mathbf{C}_1)^2. \qquad (3.53)
\end{aligned}$$

Then, since Eqns. (3.52) and (3.53) share the common term $(\mathbf{C}_1\mathbf{C}_2 - \mathbf{C}_2\mathbf{C}_1)^2$, we only need to show that

$$(\mathbf{C}_1\mathbf{C}_2 - \mathbf{C}_2\mathbf{C}_1)^2\mathbf{C}_3 \neq \mathbf{C}_3(\mathbf{C}_1\mathbf{C}_2 - \mathbf{C}_2\mathbf{C}_1)^2. \tag{3.54}$$

Proceeding, we expand $(\mathbf{C}_1\mathbf{C}_2 - \mathbf{C}_2\mathbf{C}_1)^2$ yielding

$$(\mathbf{C}_1\mathbf{C}_2 - \mathbf{C}_2\mathbf{C}_1)^2 = \mathbf{C}_1\mathbf{C}_2\mathbf{C}_1\mathbf{C}_2 - \mathbf{C}_1\mathbf{C}_2\mathbf{C}_2\mathbf{C}_1 - \mathbf{C}_2\mathbf{C}_1\mathbf{C}_1\mathbf{C}_2 + \mathbf{C}_2\mathbf{C}_1\mathbf{C}_2\mathbf{C}_1. \tag{3.55}$$

and then, expand each of the components on the right-hand side of Eqn. (3.55) to yield

$$\mathbf{C}_1\mathbf{C}_2\mathbf{C}_1\mathbf{C}_2 = \mathbf{c}_1\mathbf{c}_1^H\mathbf{c}_2\mathbf{c}_2^H\mathbf{c}_1\mathbf{c}_1^H\mathbf{c}_2\mathbf{c}_2^H, \tag{3.56}$$

$$\mathbf{C}_1\mathbf{C}_2\mathbf{C}_2\mathbf{C}_1 = \mathbf{c}_1\mathbf{c}_1^H\mathbf{c}_2\mathbf{c}_2^H\mathbf{c}_2\mathbf{c}_2^H\mathbf{c}_1\mathbf{c}_1^H, \tag{3.57}$$

$$\mathbf{C}_2\mathbf{C}_1\mathbf{C}_1\mathbf{C}_2 = \mathbf{c}_2\mathbf{c}_2^H\mathbf{c}_1\mathbf{c}_1^H\mathbf{c}_1\mathbf{c}_1^H\mathbf{c}_2\mathbf{c}_2^H, \tag{3.58}$$

$$\mathbf{C}_2\mathbf{C}_1\mathbf{C}_2\mathbf{C}_1 = \mathbf{c}_2\mathbf{c}_2^H\mathbf{c}_1\mathbf{c}_1^H\mathbf{c}_2\mathbf{c}_2^H\mathbf{c}_1\mathbf{c}_1^H. \tag{3.59}$$

Now, notice that $\mathbf{c}_m^H \mathbf{c}_n$ is a scalar which will be denoted as k_{mn}, where

$$
\begin{aligned}
k_{mn} &= \mathbf{c}_m^H \mathbf{c}_n \\
&= \begin{cases} e^{-j\pi(\upsilon_m-\upsilon_n)(N-1)} e^{-j\pi(\omega_m-\omega_n)(M-1)} \frac{\sin(\pi(\upsilon_m-\upsilon_n)N)}{\sin(\pi(\upsilon_m-\upsilon_n))} \frac{\sin(\pi(\omega_m-\omega_n)M)}{\sin(\pi(\omega_m-\omega_n))} & \text{if } m \neq n, \\ MN & \text{if } m = n. \end{cases}
\end{aligned} \tag{3.60}
$$

Further, notice that $k_{nm} = k_{mn}^*$. Using this observation and the fact that $\mathbf{c}_m^H \mathbf{c}_n$ is scalar, we can rewrite Eqns. (3.56)-(3.59) as

$$\mathbf{C}_1\mathbf{C}_2\mathbf{C}_1\mathbf{C}_2 = k_{12}|k_{12}|^2 \mathbf{c}_1\mathbf{c}_2^H, \tag{3.61}$$

$$\mathbf{C}_1\mathbf{C}_2\mathbf{C}_2\mathbf{C}_1 = MN|k_{12}|^2 \mathbf{c}_1\mathbf{c}_1^H, \tag{3.62}$$

$$\mathbf{C}_2\mathbf{C}_1\mathbf{C}_1\mathbf{C}_2 = MN|k_{12}|^2 \mathbf{c}_2\mathbf{c}_2^H, \tag{3.63}$$

$$\mathbf{C}_2\mathbf{C}_1\mathbf{C}_2\mathbf{C}_1 = k_{12}^*|k_{12}|^2 \mathbf{c}_2\mathbf{c}_1^H, \tag{3.64}$$

and substituting Eqns. (3.61)-(3.64) into Eqn. (3.55) yields

$$(\mathbf{C}_1\mathbf{C}_2 - \mathbf{C}_2\mathbf{C}_1)^2 = |k_{12}|^2(k_{12}\mathbf{c}_1\mathbf{c}_2^H - MN\mathbf{c}_1\mathbf{c}_1^H - MN\mathbf{c}_2\mathbf{c}_2^H + k_{12}^*\mathbf{c}_2\mathbf{c}_1^H). \tag{3.65}$$

Thus,

$$
\begin{aligned}
&(\mathbf{C}_1\mathbf{C}_2 - \mathbf{C}_2\mathbf{C}_1)^2\mathbf{C}_3 \\
&\quad = |k_{12}|^2(k_{12}\mathbf{c}_1\mathbf{c}_2^H\mathbf{c}_3\mathbf{c}_3^H - MN\mathbf{c}_1\mathbf{c}_1^H\mathbf{c}_3\mathbf{c}_3^H - MN\mathbf{c}_2\mathbf{c}_2^H\mathbf{c}_3\mathbf{c}_3^H + k_{12}^*\mathbf{c}_2\mathbf{c}_1^H\mathbf{c}_3\mathbf{c}_3^H) \\
&\quad = |k_{12}|^2(k_{12}k_{23}\mathbf{c}_1\mathbf{c}_3^H - MNk_{13}\mathbf{c}_1\mathbf{c}_3^H - MNk_{23}\mathbf{c}_2\mathbf{c}_3^H + k_{12}^*k_{13}\mathbf{c}_2\mathbf{c}_3^H), \qquad (3.66) \\
&\mathbf{C}_3(\mathbf{C}_1\mathbf{C}_2 - \mathbf{C}_2\mathbf{C}_1)^2 \\
&\quad = |k_{12}|^2(k_{12}\mathbf{c}_3\mathbf{c}_3^H\mathbf{c}_1\mathbf{c}_2^H - MN\mathbf{c}_3\mathbf{c}_3^H\mathbf{c}_1\mathbf{c}_1^H - MN\mathbf{c}_3\mathbf{c}_3^H\mathbf{c}_2\mathbf{c}_2^H + k_{12}^*\mathbf{c}_3\mathbf{c}_3^H\mathbf{c}_2\mathbf{c}_1^H) \\
&\quad = |k_{12}|^2(k_{12}k_{13}^*\mathbf{c}_3\mathbf{c}_2^H - MNk_{13}^*\mathbf{c}_3\mathbf{c}_1^H - MNk_{23}^*\mathbf{c}_3\mathbf{c}_2^H + k_{12}^*k_{23}^*\mathbf{c}_3\mathbf{c}_1^H). \qquad (3.67)
\end{aligned}
$$

Proving that Eqn. (3.66) does not equal Eqn. (3.67) is a difficult problem. However, we can gain further insight by examining the first element in the first column of the matrices

in these equations. First, observe that the first element of the column vector $\mathbf{c}_i$ is 1 for $i = 1, 2, 3$ which implies that the first element of the first column of the matrix $\mathbf{c}_m \mathbf{c}_n^H$ is 1 for $m, n = 1, 2, 3$. That is, $[\mathbf{c}_m \mathbf{c}_n^H]_{00} = 1$ and thus,

$$[(\mathbf{C}_1\mathbf{C}_2 - \mathbf{C}_2\mathbf{C}_1)^2\mathbf{C}_3]_{00} = |k_{12}|^2(k_{12}k_{23} - MNk_{13} - MNk_{23} + k_{12}^* k_{13}) \tag{3.68}$$

$$[\mathbf{C}_3(\mathbf{C}_1\mathbf{C}_2 - \mathbf{C}_2\mathbf{C}_1)^2]_{00} = |k_{12}|^2(k_{12}k_{13}^* - MNk_{13}^* - MNk_{22}^* + k_{13}^* k_{23}^*). \tag{3.69}$$

Now, observe that $[(\mathbf{C}_1\mathbf{C}_2 - \mathbf{C}_2\mathbf{C}_1)^2\mathbf{C}_3]_{00} = [\mathbf{C}_3(\mathbf{C}_1\mathbf{C}_2 - \mathbf{C}_2\mathbf{C}_1)^2]_{00}^*$, and thus, in general, $[(\mathbf{C}_1\mathbf{C}_2 - \mathbf{C}_2\mathbf{C}_1)^2\mathbf{C}_3]_{00}$ does not equal $[\mathbf{C}_3(\mathbf{C}_1\mathbf{C}_2 - \mathbf{C}_2\mathbf{C}_1)^2]_{00}$ unless $[(\mathbf{C}_1\mathbf{C}_2 - \mathbf{C}_2\mathbf{C}_1)^2\mathbf{C}_3]_{00}$ is real. Although at this time we cannot provide a rigorous proof that $[(\mathbf{C}_1\mathbf{C}_2 - \mathbf{C}_2\mathbf{C}_1)^2\mathbf{C}_3]_{00}$ is not real for all values of υ_i for $i = 1, 2, 3$, M, N, and β (recall that $\omega_i = \beta\upsilon_i$), it is not difficult to find values for these parameters such that $[(\mathbf{C}_1\mathbf{C}_2 - \mathbf{C}_2\mathbf{C}_1)^2\mathbf{C}_3]_{00}$ is not real. Hence, since $[(\mathbf{C}_1\mathbf{C}_2 - \mathbf{C}_2\mathbf{C}_1)^2\mathbf{C}_3]_{00}$ does not equal $[\mathbf{C}_3(\mathbf{C}_1\mathbf{C}_2 - \mathbf{C}_2\mathbf{C}_1)^2]_{00}$ in general, then

$$(\mathbf{C}_1\mathbf{C}_2 - \mathbf{C}_2\mathbf{C}_1)^2\mathbf{C}_3 \neq \mathbf{C}_3(\mathbf{C}_1\mathbf{C}_2 - \mathbf{C}_2\mathbf{C}_1)^2 \tag{3.70}$$

which implies that

$$(\mathbf{R}_1\mathbf{R}_2 - \mathbf{R}_2\mathbf{R}_1)^2\mathbf{R}_3 \neq \mathbf{R}_3(\mathbf{R}_1\mathbf{R}_2 - \mathbf{R}_2\mathbf{R}_1)^2 \tag{3.71}$$

and we conclude the $\mathcal{S}$ is not simultaneously unitarily similar to quasi-diagonal matrices. Note that Laffey [27] also investigated the issue of quasi-diagonalization, but restricted the family to a pair of Hermitian matrices. Essentially, the main result of Watters requires a family to satisfy a near-infinite number of conditions to be simultaneously block diagonalizable. In contrast, Laffey [27] provides several theorems with a finite set of conditions for a pair of Hermitian matrices to be simultaneously quasi-diagonalizable. Although we could use Laffey's theorems to prove that the family $\mathcal{S}$ is not quasi-diagonalizable, these theorems do not lead to simpler expressions than those developed using Watters' theorem.

3.5.3 General Case. Barker et. al. [3] and Shapiro [36] investigated the general case (i.e., block size not restricted to 2×2) of block diagonalizing a family of matrices with a similarity transformation. Both Barker et. al. and Shapiro provide theorems or corollaries linking the dimension of the largest block on the diagonal of a block diagonalizable family to a polynomial identity. Recall that our objective is the block diagonalization of $\mathbf{R} \in \mathcal{S}$ with the transformation $\mathbf{V}^H\mathbf{R}\mathbf{V}$, where $\mathbf{V}$ is non-singular. If $\mathbf{V}$ is unitary, then the transformation $\mathbf{V}^H\mathbf{R}\mathbf{V}$ is a unitary similarity transformation since $\mathbf{V}^H\mathbf{R}\mathbf{V} = \mathbf{V}^{-1}\mathbf{R}\mathbf{V}$. Thus, we are concerned with the simultaneous block diagonalization of $\mathcal{S}$ with a unitary similarity transformation. However, the unitary distinction is not required, since Barker et. al., Shapiro, and Laffey [27] all note that if a family is block diagonalizable with a similarity transformation, then the family is block diagonalizable by a unitary similarity transformation. Before presenting the main results of Barker et. al. and Shapiro, we use the following two observations to establish two properties of the algebra $A_{\mathcal{S}}$ generated by the family $\mathcal{S}$:

1. Let $\mathcal{T}$ be a set of normal matrices and $A_{\mathcal{T}}$ denote algebra generated by $\mathcal{T}$, then $A_{\mathcal{T}} = A_{\mathcal{T}}^H$ [36],
2. If $A_{\mathcal{T}}$ is an algebra of complex matrices with $A_{\mathcal{T}} = A_{\mathcal{T}}^H$, then $A_{\mathcal{T}}$ is semi-simple [27].

Recall that the members of $\mathcal{S}$ are Hermitian and that Hermitian implies normal. Thus, the algebra $A_{\mathcal{S}}$ generated by $\mathcal{S}$ is semi-simple and $A_{\mathcal{S}} = A_{\mathcal{S}}^H$.

Barker et. al. [3] present the following special case of the Wedderburn-Artin theorem which serves as the foundation for their main result:

Theorem 4 (Barker). *Let $\mathcal{A} \neq \{0\}$ be a finite dimensional algebra over $\mathbb{C}$. If $\mathcal{A}$ is semi-simple, then $\mathcal{A}$ is algebra isomorphic to $M_{p_1} \oplus \cdots \oplus M_{p_n}$.*

Thus, any finite dimensional semi-simple algebra can be decomposed into full matrix algebras of size $p_i \times p_i$ for $i = 1, \ldots, n$ which suggests the potential for block diagonalizing a semi-simple algebra of matrices. In fact, based on Theorem 4, Barker et. al. derive the following theorem on the block diagonalization of a semi-simple algebra:

Theorem 5 (Barker). *Let $\mathcal{A}$ be a semi-simple subalgebra of $L(V)$, where V is a finite dimensional inner product space over $\mathbb{C}$. Then V has a basis $\mathcal{U}$, and there are integers*

$p_1, k_1, \ldots, p_n, k_n, r$ such that for each $A \in \mathcal{A}$ we have

$$\left[A\right]_{\mathcal{U}} = diag(B_1, \ldots, B_n, 0_r),$$

where $B_i \in M_{p_i}^{(k_i)}$ for $i = 1, \ldots, n$, and such that

$$\dim(V) = r + \sum_{i=1}^{n} p_i k_i.$$

Furthermore, if $\mathcal{A}$ is a *-subalgebra of $L(V)$, then the basis $\mathcal{U}$ can be chosen to be orthonormal.

In Theorem 5, $L(V)$ denotes the algebra of linear transformations on V, $[A]_{\mathcal{U}}$ denotes the matrix representation of the linear operator A with respect to the basis $\mathcal{U}$, $\mathrm{diag}(B_1, \ldots, B_n, 0_r)$ denotes a block diagonal matrix with blocks of $B_1, \ldots, B_n, 0_r$ where 0_r is a zero matrix of size $r \times r$, $\mathrm{M}_p^{(k)}$ denotes the algebra of $pk \times pk$ matrices of the form $\mathrm{diag}(B, \ldots, B)$ where $B \in \mathrm{M}_p$ and there are k blocks, and a subalgebra is a *-subalgebra if and only if $A \in \mathcal{A}$ implies $A^H \in \mathcal{A}$ (i.e., $\mathcal{A} = \mathcal{A}^H$). For discussion purposes, assume the dimension of the inner product space V is q, then matrix representation of the algebra $L(V)$ is the full matrix algebra of $q \times q$ complex matrices (denote by $\mathrm{M}_q(\mathbb{C})$) which is not simultaneously block diagonalizable [28]. Let $\mathcal{T}$ denote a set of $q \times q$ normal matrices, and let $A_{\mathcal{T}}$ denote the algebra generated by $\mathcal{T}$. Since $\mathcal{T}$ is a set of normal matrices, $A_{\mathcal{T}}$ is semi-simple. Thus, from Theorem 5, if $A_{\mathcal{T}}$ is a subalgebra of $L(V)$, then $A_{\mathcal{T}}$ is simultaneously block diagonalizable by a similarity transformation. Satisfying the hypothesis of Theorem 5 represents one of the key challenges in using the theorem. One must first show that $A_{\mathcal{T}}$ is semi-simple and then, show that $A_{\mathcal{T}} \neq \mathrm{M}_q(\mathbb{C})$ which, according to Laffey [28], is a difficult problem. Even if the hypothesis of Theorem 5 is satisfied, the theorem merely states that the algebra can be block diagonalized, but does not specify the basis, the number of blocks, or the size of the blocks. Thus, one is left with the problem of determining the basis, the number of blocks, and the size of the blocks. However, Barker et. al. do provide the following corollary linking the size of the largest block to a polynomial identity:

Corollary 1 (Barker). *Let $\mathcal{A}$ be a semi-simple subalgebra of $L(V)$ which satisfies a polynomial identity which is also satisfied by M_p but not by M_{p+1}. Then there is a basis $\mathcal{U}$ for V such that for each $A \in \mathcal{A}$, $[A]_{\mathcal{U}} = diag(B_1, \ldots, B_n)$, where each B_i is a $p_i \times p_i$ matrix and $p_i \leq p$. Furthermore, if $\mathcal{A}$ is a *-subalgebra, then $\mathcal{U}$ can be chosen orthonormal.*

Shapiro [36] proves the following theorem which also links the size of the largest block to a polynomial:

Theorem 6 (Shapiro). *Let Ω be a non-empty set of complex $n \times n$ matrices, and let $\mathcal{A}$ be the algebra generated by Ω over $\mathbb{C}$. Assume $\mathcal{A} = \mathcal{A}^H$. Let $P(x_1, \ldots, x_r)$ be a polynomial in the non-commutative variables $x_1, \ldots, x_r$ with coefficients in an algebraically closed field F. Suppose the equation $P(x_1, \ldots, x_r) = 0$ is satisfied by every r-tuple of $k \times k$ matrices over F, but there exists an r-tuple of $(k+1) \times (k+1)$ matrices over F which does not satisfy the equation. Then the following are equivalent:*

1. *There is a unitary matrix U such that for all $A \in \mathcal{A}$, the matrix UAU^H is block diagonal with blocks of sizes $n_1, \ldots, n_t$ and $\max\{n_1, \ldots, n_t\} \leq k$.*
2. *$P(A_1, \ldots, A_r) = 0$ for all $A_1, \ldots, A_r \in \mathcal{A}$.*

Shapiro notes that polynomials that satisfy the hypothesis of Theorem 6 do exist and gives the standard polynomial as one such polynomial. The standard polynomial is defined as [36]

Definition 3. *The standard polynomial in m variables is*

$$S_m(x_1, \ldots, x_m) = \sum_{\sigma} \pm x_{\sigma(1)} x_{\sigma(2)} \cdots x_{\sigma(m)},$$

where the sum is over all permutations σ of the integers $1, \ldots, m$, and the coefficient of the term $x_{\sigma(1)} x_{\sigma(2)} \cdots x_{\sigma(m)}$ is $+1$ if σ is an even permutation and -1 if σ is an odd permutation.

Shapiro refers to the equation $S_m(x_1, \ldots, x_m) = 0$ as the standard identity. Shapiro cites results of Amitsur and Levitzi and of Levitzi to conclude that the standard identity $S_{2m}(x_1, \ldots, x_{2m}) = 0$ is the polynomial identity of minimal degree for $\mathrm{M}_m(F)$ which satisfies the hypothesis of Theorem 6. Note that Laffey [27] refers to the standard polynomial

defined above as the standard identity of degree m and reaches a similar conclusion as Shapiro. Thus, from Theorem 6, if a family of $n \times n$ normal matrices is simultaneously similar to block diagonal matrices with the largest block of size $k \times k$, then the algebra generated by the family must satisfy the standard identity of degree $2k$. Conversely, if the algebra generated by a family of $n \times n$ normal matrices satisfies the standard identity of degree $2k$, then the algebra is simultaneously similar to block diagonal matrices, where the size of the largest block is less than or equal to k.

We now apply the results of Barker et. al. and Shapiro to examine the existence of a unitary $\mathbf{V}$ that simultaneously block diagonalizes the family $\mathcal{S}$ through a similarity transformation. Ideally, we would use Theorem 3 to prove or disprove the existence of $\mathbf{V}$. However, at this time, we can only establish that the algebra $A_{\mathcal{S}}$ generated by the family $\mathcal{S}$ is semi-simple and not that $A_{\mathcal{S}}$ is a subalgebra of the algebra of $\mathrm{M}_{MN}(\mathbb{C})$. As noted earlier, that in addition to being semi-simple, the algebra $A_{\mathcal{S}}$ is equal to $A_{\mathcal{S}}^H$. Thus, the hypothesis of Theorem 6 is satisfied and can be used to examine the existence of $\mathbf{V}$. Recall that of our objective is to simultaneously block diagonalize every member of $\mathcal{S}$ such that the resulting block diagonal matrices have two blocks of size $MN/2 \times MN/2$. Therefore, by Theorem 6, if we assume a $\mathbf{V}$ exists that block diagonalizes $\mathcal{S}$ as desired, then the algebra $A_{\mathcal{S}}$ must satisfy the standard identity of degree MN. That is, $A_{\mathcal{S}}$ must satisfy

$$S_{MN}(\mathbf{R}_1, \ldots, \mathbf{R}_{MN}) = \sum_{\sigma} (\operatorname{sgn} \sigma) \mathbf{R}_{\sigma(1)} \cdots \mathbf{R}_{\sigma(MN)} = \mathbf{0}, \quad \text{for all } \mathbf{R}_1, \ldots, \mathbf{R}_{MN} \in A_{\mathcal{S}}, \tag{3.72}$$

where $(\operatorname{sgn} \sigma)$ is $+1$ if σ is an even permutation and -1 is odd permutation. Obviously, proving that Eqn. (3.72) holds is an extremely difficult problem. However, if we can provide a single counter-example, then the assumption is contradicted and we can conclude that $\mathbf{V}$ does not exist. Our intuition lead us to believe that $\mathbf{V}$ does not exist and thus, we attempt to provide a counter example.

Consider the following set of matrices in $\mathcal{S}$:

$$\mathbf{R}_i = \mathbf{I} + \mathbf{c}_i \mathbf{c}_i^H = \mathbf{I} + \mathbf{C}_i, \quad \text{for } i = 1, \ldots, MN, \tag{3.73}$$

where $\mathbf{C}_i = \mathbf{c}_i\mathbf{c}_i^H$ and represents the clutter correlation matrix from a single dominant scatter (See Eqn. (3.21) for details). Now, observe that if $\mathbf{V}$ is unitary, then

$$\mathbf{V}^H\mathbf{R}_i\mathbf{V} = \mathbf{V}^H(\mathbf{I}+\mathbf{C}_i)\mathbf{V} = \mathbf{I}+\mathbf{V}^H\mathbf{C}_i\mathbf{V}. \tag{3.74}$$

Thus, we only need to show that the set $\{\mathbf{C}_1, \ldots, \mathbf{C}_{MN}\}$ does not satisfy the standard identity of degree MN. That is,

$$S_{MN}(\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_{MN}) = \sum_{\sigma}(\operatorname{sgn}\sigma)\mathbf{C}_{\sigma(1)}\mathbf{C}_{\sigma(2)}\cdots\mathbf{C}_{\sigma(MN)} \neq \mathbf{0}. \tag{3.75}$$

Observe that a single term of the summation is given by

$$\mathbf{C}_{\sigma(1)}\mathbf{C}_{\sigma(2)}\cdots\mathbf{C}_{\sigma(MN)} = \mathbf{c}_{\sigma(1)}\mathbf{c}_{\sigma(1)}^H\mathbf{c}_{\sigma(2)}\mathbf{c}_{\sigma(2)}^H\cdots\mathbf{c}_{\sigma(MN)}\mathbf{c}_{\sigma(MN)}^H \tag{3.76}$$

$$= \left[\prod_{k=1}^{MN-1}\mathbf{c}_{\sigma(k)}^H\mathbf{c}_{\sigma(k-1)}\right]\mathbf{c}_{\sigma(1)}\mathbf{c}_{\sigma(MN)}^H. \tag{3.77}$$

Recall that

$$\mathbf{c}_m^H\mathbf{c}_n = e^{-j\pi(\upsilon_m-\upsilon_n)(N-1)}e^{-j\pi(\omega_m-\omega_n)(M-1)}\,\frac{\sin(\pi(\upsilon_m-\upsilon_n)N)}{\sin(\pi(\upsilon_m-\upsilon_n))}\,\frac{\sin(\pi(\omega_m-\omega_n)M)}{\sin(\pi(\omega_m-\omega_n))} \tag{3.78}$$

and thus,

$$\left[\prod_{k=1}^{MN-1}\mathbf{c}_{\sigma(k)}^H\mathbf{c}_{\sigma(k-1)}\right] = \left[\prod_{k=1}^{MN-1}e^{-j\pi(\upsilon_{\sigma(k)}-\upsilon_{\sigma(k-1)})(N-1)}\right]\cdot\left[\prod_{k=1}^{MN-1}e^{-j\pi(\omega_{\sigma(k)}-\omega_{\sigma(k-1)})(M-1)}\right]$$
$$\cdot\left[\prod_{k=1}^{MN-1}\frac{\sin(\pi(\upsilon_{\sigma(k)}-\upsilon_{\sigma(k-1)})N)}{\sin(\pi(\upsilon_{\sigma(k)}-\upsilon_{\sigma(k-1)}))}\right]\cdot\left[\prod_{k=1}^{MN-1}\frac{\sin(\pi(\omega_{\sigma(k)}-\omega_{\sigma(k-1)})M)}{\sin(\pi(\omega_{\sigma(k)}-\omega_{\sigma(k-1)}))}\right] \tag{3.79}$$

which reduces to

$$\left[\prod_{k=1}^{MN-1} \mathbf{c}_{\sigma(k)}^H \mathbf{c}_{\sigma(k-1)}\right] = e^{-j\pi(\upsilon_{\sigma(1)}-\upsilon_{\sigma(MN)})(N-1)} \left[\prod_{k=1}^{MN-1} \frac{\sin(\pi(\upsilon_{\sigma(k)}-\upsilon_{\sigma(k-1)})N)}{\sin(\pi(\upsilon_{\sigma(k)}-\upsilon_{\sigma(k-1)}))}\right]$$
$$\cdot\, e^{-j\pi(\omega_{\sigma(1)}-\omega_{\sigma(MN)})(M-1)} \left[\prod_{k=1}^{MN-1} \frac{\sin(\pi(\omega_{\sigma(k)}-\omega_{\sigma(k-1)})M)}{\sin(\pi(\omega_{\sigma(k)}-\omega_{\sigma(k-1)}))}\right]. \quad (3.80)$$

To show that Eqn. (3.75) does not hold, it is sufficient to show that the first element of the first column of the summation is not zero. Recall that $[\mathbf{c}_m \mathbf{c}_n^H]_{00} = 1$. Thus, we only need to show that

$$\sum_{\sigma} (\operatorname{sgn}\sigma) \left[\prod_{k=1}^{MN-1} \mathbf{c}_{\sigma(k)}^H \mathbf{c}_{\sigma(k-1)}\right] \neq \mathbf{0}, \quad (3.81)$$

and by substituting Eqn (3.80), we can rewrite this condition as

$$\sum_{\sigma} (\operatorname{sgn}\sigma) e^{-j\pi(\upsilon_{\sigma(1)}-\upsilon_{\sigma(MN)})(N-1)} \left[\prod_{k=1}^{MN-1} \frac{\sin(\pi(\upsilon_{\sigma(k)}-\upsilon_{\sigma(k-1)})N)}{\sin(\pi(\upsilon_{\sigma(k)}-\upsilon_{\sigma(k-1)}))}\right]$$
$$\cdot\, e^{-j\pi(\omega_{\sigma(1)}-\omega_{\sigma(MN)})(M-1)} \left[\prod_{k=1}^{MN-1} \frac{\sin(\pi(\omega_{\sigma(k)}-\omega_{\sigma(k-1)})M)}{\sin(\pi(\omega_{\sigma(k)}-\omega_{\sigma(k-1)}))}\right] \neq \mathbf{0}. \quad (3.82)$$

At this time, we cannot prove Eqn. (3.82) holds in general. However, we can provide numerical results for several cases when M and N are small in which Eqn. (3.82) does not hold. Although the numerical results support our intuition that $\mathbf{V}$ does not exist, numerical results are not a proof for the general case. Additionally, observe that the summation in Eqn. (3.82) is over $MN!$ terms where each term involves two products of $(MN-1)$ terms and thus, the numerical evaluation of the summation is computationally intensive (approximately on the order of $(MN)^2(MN!)$) and of limited value. Because of the computational burden, we only evaluated the summation for small values of M and N. We present the results from three cases: Case 1 ($M = 2, N = 2$), Case 2 ($M = 2, N = 3$), and Case 3 ($M = 2, N = 4$). Recall that $\omega_i = \beta\upsilon_i$ and notice that the numerical evaluation of the summation requires MN values of υ_i. The values of υ_i for $i = 1, \ldots, MN$ were randomly selected from a uniform distribution on the interval $(0, 1)$. For each case, we

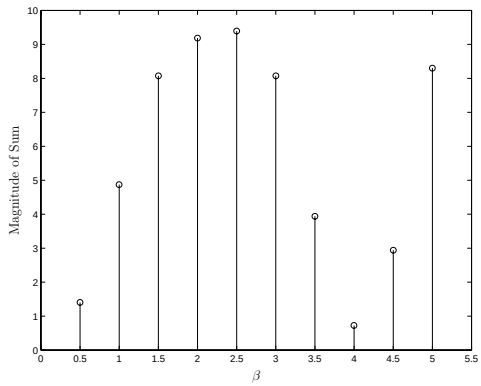
	Case 1	Case 2	Case 3
υ_1	0.8214	0.4103	0.9501
υ_2	0.4447	0.8936	0.2311
υ_3	0.6154	0.0579	0.6068
υ_4	0.7919	0.3529	0.4860
υ_5		0.8132	0.8913
υ_6		0.0099	0.7621
υ_7			0.4565
υ_8			0.0185

Table 3.1 Spatial frequencies used in the numerical evaluation of the standard identity of degree MN (See Eqn. (3.82)).

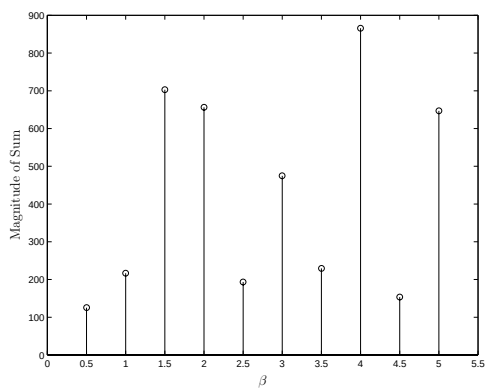
present the results for one instantiation of υ_i for $i = 1, \ldots, MN$ and 10 values of β uniformly distributed on the interval $(0.5, 5.0)$. The values of υ_i for the instantiation presented here are listed in Table 3.1. The magnitude of the summation for the three cases as function of β are plotted in Fig. 3.1. An examination of Fig. 3.1 clearly reveals that the summation in Eqn. (3.82) is not zero from the parameters selected which implies that the family $\mathcal{S}$ cannot be simultaneously block diagonalized with blocks of size $MN/2 \times MN/2$ when $M = 2$ and $N = 2, 3, 4$. Based on these limited numerical results, we conjecture that the family $\mathcal{S}$ cannot, in general, be simultaneously block diagonalized with blocks of size $MN/2 \times MN/2$ through a similarity transformation. That is, we conjecture that a unitary $\mathbf{V}$ does not exist.

3.6 Centrosymmetric Clutter

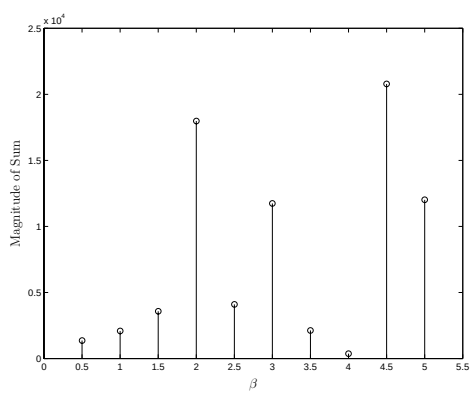
In the previous section, we provided evidence to support the conjecture that the family of STAP correlation matrices $\mathcal{S}$ cannot be simultaneously diagonalized by non-singular transformation. In this section, we show that the clutter correlation matrix is a centrosymmetric matrix if we introduce additional assumptions and thus, we can block diagonalize a family of clutter correlation matrices under these assumptions. We will show this centrosymmetric property for two different sets of assumptions. In Section 3.6.1, we add the assumptions that number of clutter patches is infinite and that each clutter patch has the same power level to the previous assumptions used in defining the family $\mathcal{S}$. In Section 3.6.2, we add the assumption that the power level is symmetric about the



(a) Case 1: $M = 2$ and $N = 2$



(b) Case 2: $M = 2$ and $N = 3$



(c) Case 3: $M = 2$ and $N = 4$

Figure 3.1 Numerical evaluation of the standard identity of degree MN (first element of the first column only, see Eqn. (3.82)) as a function of β with the parameters listed in Table 3.1.

radar platform to the previous assumptions used in defining the family $\mathcal{S}$. Although these assumptions are not realistic, the centrosymmetric property could be potentially useful for approximate or data adaptive methods as briefly discussed in Section 3.6.3.

3.6.1 Infinite Number of Clutter Patches and Constant Power. Recall from Section 3.2 that the clutter was modeled as a series of point scatters, referred to as clutter patches, that surround the radar and are located in the range gate of interest. Further, in development of the family $\mathcal{S}$, we assumed that the clutter patches were uncorrelated, the radar was operating in an unambiguous range scenario (i.e., no second time around clutter) and there was no intrinsic clutter motion or velocity misalignment. Under these assumptions, the clutter correlation matrix is given by

$$\mathbf{R} = \sum_{m=0}^{N_c-1} \xi_m \mathbf{C}(\upsilon_m), \tag{3.83}$$

where N_c is the number of clutter patches and ξ_m, $\mathbf{C}(\upsilon_m)$, and υ_m are the power, correlation matrix, and spatial frequency of the mth clutter patch, respectively (Note the slight notation change from Eqn. (3.6) and the assumption that the receiver noise variance σ^2 is one). The correlation matrix of the mth clutter patch is given by

$$\mathbf{C}(\upsilon_m) = \mathbf{b}(\omega_m)\mathbf{b}^H(\omega_m) \otimes \mathbf{a}(\upsilon_m)\mathbf{a}^H(\upsilon_m) = \mathbf{B}(\omega_m) \otimes \mathbf{A}(\upsilon_m), \tag{3.84}$$

where

$$\mathbf{B}(\omega_m) = \mathbf{b}(\omega_m)\mathbf{b}^H(\omega_m) \tag{3.85}$$

$$\mathbf{A}(\upsilon_m) = \mathbf{a}(\upsilon_m)\mathbf{a}^H(\upsilon_m) \tag{3.86}$$

$$\mathbf{b}(\omega_m) = \begin{bmatrix} 1 & e^{j2\pi\omega_m} & \dots & e^{j2\pi(M-1)\omega_m} \end{bmatrix}^T \tag{3.87}$$

$$\mathbf{a}(\upsilon_m) = \begin{bmatrix} 1 & e^{j2\pi\upsilon_m} & \dots & e^{j2\pi(N-1)\upsilon_m} \end{bmatrix}^T \tag{3.88}$$

$$\omega_m = \beta\upsilon_m. \tag{3.89}$$

Using the properties of the Kronecker product, we can express Eqn. (3.84) as

$$\mathbf{C}(\upsilon_m) = \begin{bmatrix} \mathbf{A}(\upsilon_m) & e^{-j2\pi\beta\upsilon_m}\mathbf{A}(\upsilon_m) & \cdots & e^{-j2\pi(M-1)\beta\upsilon_m}\mathbf{A}(\upsilon_m) \\ e^{j2\pi\beta\upsilon_m}\mathbf{A}(\upsilon_m) & \mathbf{A}(\upsilon_m) & \cdots & e^{-j2\pi(M-2)\beta\upsilon_m}\mathbf{A}(\upsilon_m) \\ \vdots & \vdots & \ddots & \vdots \\ e^{j2\pi(M-1)\beta\upsilon_m}\mathbf{A}(\upsilon_m) & e^{j2\pi(M-2)\beta\upsilon_m}\mathbf{A}(\upsilon_m) & \cdots & \mathbf{A}(\upsilon_m) \end{bmatrix}, \tag{3.90}$$

where

$$\mathbf{A}(\upsilon_m) = \begin{bmatrix} 1 & e^{-j2\pi\upsilon_m} & \cdots & e^{-j2\pi(N-1)\upsilon_m} \\ e^{j2\pi\upsilon_m} & 1 & \cdots & e^{-j2\pi(N-2)\upsilon_m} \\ \vdots & \vdots & \ddots & \vdots \\ e^{j2\pi(N-1)\upsilon_m} & e^{j2\pi(N-2)\upsilon_m} & \cdots & 1 \end{bmatrix}. \tag{3.91}$$

An examination of Eqn. (3.90) reveals that $\mathbf{C}(\upsilon_m)$ is an $M \times M$ block matrix where each block is a $N \times N$ matrix. Note that $\mathbf{R}$ will also have this block structure. The blocks of Eqn. (3.90) will serve as the basic building blocks for examining the sum given in Eqn. (3.83). Before examining the sum, we need to discuss the clutter patch power parameter ξ_m.

The power of the mth clutter patch is

$$\xi_m = \frac{P_t G_t(\phi_m, \theta_m) g(\phi_m, \theta_m) \lambda_o \sigma_m}{(4\pi)^3 R^4}, \tag{3.92}$$

where P_t is the peak power of the transmitter, $G_t(\phi_m, \theta_m)$ is the transmit antenna gain, $g(\phi_m, \theta_m)$ is the received element gain, λ_o is the wavelength of the transmitted signal, σ_m is the radar cross section (RCS) of the mth clutter patch, R is the range to the clutter, and ϕ_m and θ_m are the azimuth and elevation angles to the mth clutter patch. The RCS of the mth clutter patch is

$$\sigma_m = \sigma_o(\phi_m, \theta_m) R \Delta\phi \Delta R \sec\psi, \tag{3.93}$$

where $\sigma_o(\phi_m, \theta_m)$ is the area reflectivity of the ground, ΔR is the range covered by a single range gate, ψ is the grazing angle, and $\Delta\phi = 2\pi/N_c$. The parameter $\Delta\phi$ defines the azimuth extent of the clutter patches and is the only parameter directly effected by increasing or decreasing the number of the clutter patches. Observe that the power of mth clutter patch is inversely proportional to N_c and thus, we can write clutter patch power as

$$\xi_m = \frac{\kappa_m}{N_c}, \tag{3.94}$$

where

$$\kappa_m = \frac{P_t G_t(\phi_m, \theta_m) g(\phi_m, \theta_m) \lambda_o \sigma_o(\phi_m, \theta_m) 2\pi R \Delta R \sec\psi}{(4\pi)^3 R^4}. \tag{3.95}$$

In examining the clutter correlation matrix as the number of clutter patches approaches infinity, we assume that the transmit gain, receiver element gain, and area reflectivity of the ground are constant for all angles. Under these assumptions, the parameter κ_m is a constant κ for all m. Thus, we can write the equation for the clutter correlation matrix for fixed N_c as

$$\mathbf{R} = \frac{\kappa}{N_c} \sum_{m=0}^{N_c-1} \mathbf{C}(\upsilon_m) \tag{3.96}$$

and for the case when N_c approaches infinity as

$$\mathbf{R}^{\infty} = \lim_{N_c \to \infty} \frac{\kappa}{N_c} \sum_{m=0}^{N_c-1} \mathbf{C}(\upsilon_m). \tag{3.97}$$

Without loss of generality, the parameter κ is assumed to be one. For a given β, the clutter patch correlation matrices are only a function of the spatial frequency υ_m which is defined as

$$\upsilon_m = \frac{d}{\lambda_o} \sin\phi_m, \tag{3.98}$$

assuming a constant elevation angle of zero. Recall that ϕ_m is the azimuth angle to the mth clutter patch and the clutter patches are assumed to be uniformly distributed around the radar. Therefore, we define ϕ_m as

$$\phi_m = \frac{2\pi}{N_c} m \quad \text{for } m = 0, 1, \ldots, N_c - 1. \tag{3.99}$$

Assuming $d = \lambda_o/2$, we can write the spatial frequency υ_m as

$$\upsilon_m = \frac{1}{2} \sin\left(\frac{2\pi}{N_c} m\right). \tag{3.100}$$

We now proceed by first expressing the blocks in $\mathbf{R}$ as a summation of the individual $N \times N$ blocks of the clutter patch correlation matrices and then, by examining the individual elements in each block. Let $\mathbf{C}(\upsilon_m)_{p,q}$ denote the $N \times N$ matrix (block) in the pth row and q^{th} column of $\mathbf{C}(\upsilon_m)$ and $[\mathbf{C}(\upsilon_m)_{p,q}]_{k,n}$ denote the element in the kth row and n^{th} column of $\mathbf{C}(\upsilon_m)_{p,q}$, where $0 \leq p, q \leq M - 1$ and $0 \leq k, n \leq N - 1$. The block in the pth row and q^{th} column of $\mathbf{R}$ is given as

$$\mathbf{R}_{p,q} = \frac{1}{N_c} \sum_{m=0}^{N_c-1} \mathbf{C}(\upsilon_m)_{p,q}. \tag{3.101}$$

From Eqns. (3.90) and (3.91), we can write $\mathbf{C}(\upsilon_m)_{p,q}$ as

$$\mathbf{C}(\upsilon_m)_{p,q} = e^{j2\pi(p-q)\beta\upsilon_m} \begin{bmatrix} 1 & e^{-j2\pi\upsilon_m} & \cdots & e^{-j2\pi(N-1)\upsilon_m} \\ e^{j2\pi\upsilon_m} & 1 & \cdots & e^{-j2\pi(N-2)\upsilon_m} \\ \vdots & \vdots & \ddots & \vdots \\ e^{j2\pi(N-1)\upsilon_m} & e^{j2\pi(N-2)\upsilon_m} & \cdots & 1 \end{bmatrix} \tag{3.102}$$

and an arbitrary element of $\mathbf{C}(\upsilon_m)_{p,q}$ as

$$\begin{aligned} [\mathbf{C}(\upsilon_m)_{p,q}]_{k,n} &= e^{j2\pi(p-q)\beta\upsilon_m} \, e^{j2\pi(k-n)\upsilon_m} \\ &= e^{j2\pi[(p-q)\beta+(k-n)]\upsilon_m} \\ &= e^{j2\pi z\upsilon_m} \end{aligned} \tag{3.103}$$

where we have defined $z = (p-q)\beta + (k-n)$ to ease notation. From Eqns. (3.101) and (3.103), we can write an arbitrary element of $\mathbf{R}_{p,q}$ as

$$[\mathbf{R}_{p,q}]_{k,n} = \frac{1}{N_c} \sum_{m=0}^{N_c-1} e^{j2\pi z \upsilon_m}. \tag{3.104}$$

Substituting Eqn. (3.100) into Eqn.(3.104) yields

$$[\mathbf{R}_{p,q}]_{k,n} = \frac{1}{N_c} \sum_{m=0}^{N_c-1} e^{j\pi z \sin\left(\frac{2\pi}{N_c} m\right)} = \frac{1}{N_c} \sum_{m=0}^{N_c-1} e^{j\pi z \sin \phi_m}. \tag{3.105}$$

Now, we analyze the summation in Eqn. (3.105) using Bessel functions. The moment generating function of the Bessel function is [1:361]

$$e^{(x/2)(t-1/t)} = \sum_{l=-\infty}^{\infty} J_l(x) t^l \quad t \neq 0, \tag{3.106}$$

where $J_l(x)$ is the Bessel function of order l. Substituting $t = e^{j\phi_m}$ and $x = \pi z$ into Eqn. (3.106) yields

$$e^{j\pi z \sin \phi_m} = \sum_{l=-\infty}^{\infty} J_l(\pi z) e^{j\phi_m l}. \tag{3.107}$$

Thus, by substituting Eqn. (3.107) into Eqn. (3.105), we can write $[\mathbf{R}_{p,q}]_{k,n}$ as

$$[\mathbf{R}_{p,q}]_{k,n} = \frac{1}{N_c - 1} \sum_{m=0}^{N_c-1} \sum_{l=-\infty}^{\infty} J_l(\pi z) e^{j\phi_m l} \tag{3.108}$$

Observe that the only non-zeros terms in Eqn. (3.108) occur when l is a multiple of N_c, since for a fixed $l \neq N_c k$ (k is an integer) we have

$$\frac{J_l(\pi z)}{N_c} \sum_{m=0}^{N_c-1} e^{j\frac{2\pi}{N_c} l m} = 0.$$

Further, observe that $e^{j\phi_m l} = 1$ when $l = N_c k$ and the summation for a fixed $l = N_c k$ yields $N_c J_l(\pi z)$. Thus, we can rewrite Eqn. (3.108) as

$$
\begin{aligned}
[\mathbf{R}_{p,q}]_{k,n} &= \sum_{l=-\infty}^{\infty} J_{N_c l}(\pi z) \\
&= \begin{cases} J_0(\pi z) + 2\sum_{l=1}^{\infty} J_{2N_c l}(\pi z) & \text{if } N_c \text{ is odd;} \\ J_0(\pi z) + 2\sum_{l=1}^{\infty} J_{N_c l}(\pi z) & \text{if } N_c \text{ is even,} \end{cases}
\end{aligned}
\tag{3.109}
$$

where we have used the fact that $J_{-n}(x) = (-1)^n J(x)$ when n is an integer. For a fixed x, notice that $J_n(x)$ approaches zero as n approaches infinity [1]. Thus, as N_c approaches infinity, we would expect $[\mathbf{R}_{p,q}]_{k,n}$ to approach $J_0(\pi z)$. In fact, as N_c approaches infinity, the summation in Eqn. (3.105) becomes a definite integral which equals $J_0(\pi z)$ [20:221]. That is,

$$
[\mathbf{R}^{\infty}_{p,q}]_{k,n} = \lim_{N_c \to \infty} \frac{1}{N_c - 1} \sum_{m=0}^{N_c - 1} e^{j\pi z \sin \phi_m} = \frac{1}{2\pi} \int_0^{2\pi} e^{j\pi z \sin \phi} d\phi = J_0(\pi z). \tag{3.110}
$$

Now, substituting back in $z = (p-q)\beta + (k-n)$ into Eqn. (3.110) yields

$$
[\mathbf{R}^{\infty}_{p,q}]_{k,n} = J_0([(p-q)\beta + (k-n)]\pi) \tag{3.111}
$$

Note that Ward [41:28] gives a similar result without derivation or further comment. Equation (3.111) reveals that each block of $\mathbf{R}^{\infty}$ is a real Toeplitz matrix, since each element only depends on the difference between k and n when p and q are fixed and $J_0(x)$ is real for all real x. Note that, in general, $\mathbf{R}^{\infty}_{0,0}$ is the only block which is a symmetric Toeplitz matrix. Using the results from Eqn. (3.111), we can write $\mathbf{R}^{\infty}_{p,q}$ as

$$
\mathbf{R}^{\infty}_{p,q} = \begin{bmatrix} J_0(k_1\beta\pi) & J_0([k_1\beta - 1]\pi) & \cdots & J_0([k_1\beta - (N-1)]\pi) \\ J_0([k_1\beta + 1]\pi) & J_0(k_1\beta\pi) & \cdots & J_0([k_1\beta - (N-2)]\pi) \\ \vdots & \vdots & \ddots & \vdots \\ J_0([k_1\beta + (N-1)]\pi) & J_0([k_1\beta + (N-2)]\pi) & \cdots & J_0(k_1\beta\pi) \end{bmatrix} \tag{3.112}
$$

where $k_1 = p - q$. A further examination of Eqn. (3.111) along with the fact that $J_0(x) = J_0(-x)$ reveals that

$$\mathbf{R}^{\infty}_{q,p} = \mathbf{R}^{\infty\,T}_{p,q}, \tag{3.113}$$

which easily follows since

$$\begin{aligned}[\mathbf{R}^{\infty\,T}_{p,q}]_{k,n} &= [\mathbf{R}^{\infty}_{p,q}]_{n,k} \\ &= J_0([(p-q)\beta + (n-k)]\pi)\end{aligned}$$

and

$$\begin{aligned}[\mathbf{R}^{\infty}_{q,p}]_{k,n} &= J_0([(q-p)\beta + (k-n)]\pi) \\ &= J_0(-[(p-q)\beta + (n-k)]\pi) \\ &= J_0([(p-q)\beta + (n-k)]\pi).\end{aligned}$$

Using Eqn. (3.113), we can write $\mathbf{R}^{\infty}$ as

$$\mathbf{R}^{\infty} = \begin{bmatrix} \mathbf{R}^{\infty}_{0,0} & \mathbf{R}^{\infty}_{0,1} & \cdots & \mathbf{R}^{\infty}_{0,M-1} \\ \mathbf{R}^{\infty\,T}_{0,1} & \mathbf{R}^{\infty}_{0,0} & \cdots & \mathbf{R}^{\infty}_{0,M-2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}^{\infty\,T}_{0,M-1} & \mathbf{R}^{\infty\,T}_{0,M-2} & \cdots & \mathbf{R}^{\infty}_{0,0} \end{bmatrix}. \tag{3.114}$$

Thus, $\mathbf{R}^{\infty}$ is a real symmetric, Toeplitz-block-Toeplitz matrix. Next, we will show that $\mathbf{R}^{\infty}$ is also centrosymmetric.

Recall from Chapter II that a $n \times n$ matrix $\mathbf{Q}$ is centrosymmetric if

$$[\mathbf{Q}]_{p,q} = [\mathbf{Q}]_{n-1-p,n-1-q} \quad \text{for } p, q = 0, \cdots, n-1, \tag{3.115}$$

or equivalently, $\mathbf{Q} = \mathbf{JQJ}$, where $\mathbf{J}$ is the anti-diagonal (reverse diagonal) matrix [2]. Thus, to prove that $\mathbf{R}^{\infty}$ is centrosymmetric, we must show that $\mathbf{R}^{\infty} = \mathbf{JR}^{\infty}\mathbf{J}$. We start by pre-

and post-multiplying $\mathbf{R}^{\infty}$ by $\mathbf{J}$ which yields

$$\mathbf{J}\mathbf{R}^{\infty}\mathbf{J} = \begin{bmatrix} \mathbf{J}\mathbf{R}_{0,0}^{\infty}\mathbf{J} & \mathbf{J}\mathbf{R}_{0,1}^{\infty}{}^{T}\mathbf{J} & \cdots & \mathbf{J}\mathbf{R}_{0,M-1}^{\infty}{}^{T}\mathbf{J} \\ \mathbf{J}\mathbf{R}_{0,1}^{\infty}\mathbf{J} & \mathbf{J}\mathbf{R}_{0,0}^{\infty}\mathbf{J} & \cdots & \mathbf{J}\mathbf{R}_{0,M-2}^{\infty}{}^{T}\mathbf{J} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{J}\mathbf{R}_{0,M-1}^{\infty}\mathbf{J} & \mathbf{J}\mathbf{R}_{0,M-2}^{\infty}\mathbf{J} & \cdots & \mathbf{J}\mathbf{R}_{0,0}^{\infty}\mathbf{J} \end{bmatrix}. \tag{3.116}$$

Thus, we only need to show that $\mathbf{R}_{0,p}^{\infty} = \mathbf{J}\mathbf{R}_{0,p}^{\infty}{}^{T}\mathbf{J}$ to prove that $\mathbf{R}^{\infty}$ is centrosymmetric, since $\mathbf{J}^2 = \mathbf{I}$. It is easily shown for any Toeplitz matrix $\mathbf{Q}$ that $\mathbf{Q} = \mathbf{J}\mathbf{Q}^T\mathbf{J}$ by noting that an equivalent condition is $\mathbf{J}\mathbf{Q} = \mathbf{Q}^T\mathbf{J}$ and examining the elements of the resulting matrices. Since $\mathbf{R}_{p,q}^{\infty}$ is a real Toeplitz matrix, we have $\mathbf{R}_{0,p}^{\infty} = \mathbf{J}\mathbf{R}_{0,p}^{\infty}{}^{T}\mathbf{J}$. Therefore, $\mathbf{R}^{\infty}$ is a real symmetric, centrosymmetric (Toeplitz-block-Toeplitz) matrix. As noted earlier, we can easily block diagonalize any centrosymmetric matrix with a fixed, efficient transformation (See Chapter II, Section 2.5).

3.6.2 Finite Number of Clutter Patches and Symmetric Power. In this section, we ease the assumption that the transmit gain, receiver element gain, and area reflectivity of the ground are constant for all angles (i.e., each clutter patch has the same power level) and we examine the clutter correlation matrix when the number of the clutter patches is fixed and even. The development for the odd case is similar. We now assume that the transmit gain, receiver element gain, and area reflectivity of the ground are symmetric about the zero angle in the azimuth plane. That is,

$$\begin{aligned} G_t(\phi,\theta) &= G_t(-\phi,\theta) \quad \phi \neq 0,\pi \\ g(\phi,\theta) &= g(-\phi,\theta) \quad \phi \neq 0,\pi \\ \sigma_o(\phi,\theta) &= \sigma_o(-\phi,\theta) \end{aligned}$$

Under these assumptions, the clutter patch power is no longer constant for all m. Thus, we must write the clutter correlation matrix as

$$\mathbf{R} = \frac{1}{N_c}\sum_{m=0}^{N_c-1} \kappa_m \mathbf{C}(\upsilon_m). \tag{3.117}$$

Folding the clutter patch power κ_m back into Eqns. (3.101) and (3.105) yields

$$\mathbf{R}_{p,q} = \frac{1}{N_c} \sum_{m=0}^{N_c-1} \kappa_m \mathbf{C}(\upsilon_m)_{p,q} \tag{3.118}$$

and

$$[\mathbf{R}_{p,q}]_{k,n} = \frac{1}{N_c} \sum_{m=0}^{N_c-1} \kappa_m e^{j\pi z \sin\phi_m}, \tag{3.119}$$

where, as before, $z = (p-q)\beta + (k-n)$ and $\phi_m = 2\pi m/N_c$. Now, observe that when N_c is even, $\phi_m = -\phi_{N_c-m}$ for all $m \notin \{0, N_c/2\}$. In conjunction with the above symmetric assumptions, this observation implies that $\kappa_m = \kappa_{N_c-m}$. Additionally, recall that $\sin(-x) = -\sin(x)$. Using these observations, we can rewrite Eqn. (3.119) as

$$\begin{aligned}
[\mathbf{R}_{p,q}]_{k,n} &= \frac{1}{N_c}\left(\kappa_0 - \kappa_{N_c/2} + 2\sum_{m=1}^{(N_c/2)-1} \kappa_m \cos\big(\pi z \sin(\phi_m)\big)\right) \\
&= \frac{1}{N_c}\left(\kappa_0 - \kappa_{N_c/2} + 2\sum_{m=1}^{(N_c/2)-1} \kappa_m \cos\big(\pi[(p-q)\beta + (k-n)]\sin(\phi_m)\big)\right),
\end{aligned} \tag{3.120}$$

where we have used the fact that $e^{j\pi z \sin\phi_0} = 1$ and $e^{j\pi z \sin\phi_{N_c/2}} = -1$. An examination of Eqn. (3.120) reveals that $\mathbf{R}_{p,q}$ is a real Toeplitz matrix and that $\mathbf{R}_{q,p} = \mathbf{R}_{p,q}^T$. Using these results, we can write the clutter correlation matrix as

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{0,0} & \mathbf{R}_{0,1} & \cdots & \mathbf{R}_{0,M-1} \\ \mathbf{R}_{0,1}^T & \mathbf{R}_{0,0} & \cdots & \mathbf{R}_{0,M-2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{0,M-1}^T & \mathbf{R}_{0,M-2}^T & \cdots & \mathbf{R}_{0,0} \end{bmatrix}. \tag{3.121}$$

An examination of Eqn. (3.121) reveals that $\mathbf{R}$ is a real symmetric, Toeplitz-block-Toeplitz matrix. In the previous section, we only used the fact that $\mathbf{R}^\infty$ is a symmetric matrix with real Toeplitz blocks to prove its centrosymmetric property (i.e., the individual elements were not considered). Since $\mathbf{R}$ is a symmetric matrix with real Toeplitz blocks, we can conclude that $\mathbf{R}$ is a real symmetric centrosymmetric (Toeplitz-block-Toeplitz) matrix.

Recall that we defined the family $\mathcal{S}$ assuming that the amplitude for each clutter patch was random, but constant for the CPI (i.e., no intrinsic clutter motion). We can ease this assumption some what and retain the centrosymmetric property of the clutter correlation matrix under the symmetric power assumption. A clutter patch is said to have intrinsic clutter motion when the individual scatters in a clutter patch move (e.g., wind blown trees). When a clutter patch has intrinsic clutter motion, the amplitude of the clutter patch fluctuates during the CPI. Ward [41:39–49] models the temporal fluctuations as a wide-sense stationary random process which has a real, symmetric Toeplitz correlation matrix. Let $\mathbf{\Gamma}(\upsilon_m)$ denote $M \times M$ correlation matrix of the fluctuations for the mth clutter patch. Following Ward, the individual elements of $\mathbf{\Gamma}(\upsilon_m)$ are given as

$$[\mathbf{\Gamma}(\upsilon_m)]_{p,q} = \xi_m \gamma_m(p-q) = \xi_m \exp\left\{ -\left(\frac{4\pi\sigma_v T_r}{\sqrt{2}\lambda_o}\right)^2 (p-q)^2 \right\}, \tag{3.122}$$

where σ_v is the velocity standard deviation of the scatters in the clutter patch. With the intrinsic clutter motion, the clutter correlation matrix given in Eqn. (3.83) is rewritten as [41:47]

$$\mathbf{R} = \sum_{m=0}^{N_c-1} (\mathbf{\Gamma}(\upsilon_m) \odot \mathbf{B}(\omega_m)) \otimes \mathbf{A}(\upsilon_m), \tag{3.123}$$

where $\odot$ denotes the Hadamard matrix product. Now, if we assume that all the clutter patches have the same intrinsic clutter motion, then we can rewrite Eqn. (3.118) for the individual blocks of $\mathbf{R}$ as

$$\mathbf{R}_{p,q} = \frac{\gamma(p-q)}{N_c} \sum_{m=0}^{N_c-1} \kappa_m \mathbf{C}(\upsilon_m)_{p,q} \tag{3.124}$$

and the block equation (Eqn. (3.121)) for $\mathbf{R}$ as

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{0,0} & \gamma(1)\mathbf{R}_{0,1} & \cdots & \gamma(M-1)\mathbf{R}_{0,M-1} \\ \gamma(1)\mathbf{R}_{0,1}^T & \mathbf{R}_{0,0} & \cdots & \gamma(M-2)\mathbf{R}_{0,M-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(M-1)\mathbf{R}_{0,M-1}^T & \gamma(M-2)\mathbf{R}_{0,M-2}^T & \cdots & \mathbf{R}_{0,0} \end{bmatrix}. \tag{3.125}$$

An examination of Eqn. (3.125) reveals that $\mathbf{R}$ is a real symmetric Toeplitz-block-Toeplitz matrix and thus, is also a centrosymmetric matrix.

3.6.3 Potential Uses. Assuming the conjecture that the family $\mathcal{S}$ cannot be simultaneously block diagonalized with a fixed transformation is true, then we can attack the problem of replacing a single channel system with a dual channel system in two ways. One, we can retain the concept of having a fixed transformation to save computational cost and select the transformation to minimize the loss in SINR performance. Two, we could abandon the fixed transformation concept and move to a more computationally expensive data adaptive transformation method. We believe that the centrosymmetric results presented in the previous two sections could be useful in the above two approaches and other applications. We briefly discuss some the potential uses of the centrosymmetric results in this section. These potential uses represent future areas of research.

The assumptions used in Section 3.6.1 and 3.6.2 force the clutter environment to a symmetric structure that results in a centrosymmetric clutter correlation matrix. In general, the clutter environment will not have a symmetric structure, but one could divide the clutter into symmetric and anti-symmetric parts. If the symmetric clutter dominates the the anti-symmetric clutter, then a transformation that block diagonalizes a centrosymmetric matrix could be a good choice for reducing the loss in SINR performance with the additional benefit of having an efficient implementation. We experiment with this concept in the next chapter.

Recall that the block diagonalization of a matrix requires two independent subspaces that are invariant to the matrix. In the case of centrosymmetric clutter, the two subspaces are the symmetric subspace and the skewed symmetric subspaces. If the clutter has an anti-symmetric part, then these two subspaces are no longer invariant to the clutter correlation matrix and thus, we must find two new invariant subspaces. The symmetric and skewed symmetric subspaces may represent good initial guesses in a data adaptive search for the new invariant subspaces. One could envision an approach similar in concept to computing the eigenvectors with rank one updates, where we start with the symmetric and skewed symmetric subspaces and computed new subspaces as data becomes available.

One final area of potential use involves the clutter correlation matrix form Section 3.6.1, referred to as the constant clutter correlation matrix, where we assumed an infinite number of clutter patches and that every clutter patch had the same power level. Under these assumptions, the clutter correlation matrix is parameterized by only two parameters: β and the clutter to noise ratio (CNR). Without these assumptions, the clutter correlation matrix is parameterized by approximately $2N_c$ parameters, where N_c is the number of clutter patches. Since the constant clutter correlation matrix is parameterized by β and the CNR, one could use it in modeling and simulation applications requiring the quick generation of interference environments, but not requiring high fidelity. With the proper selection of the CNR, this constant clutter correlation matrix could represent a worst-case clutter scenario. The constant clutter correlation matrix could be beneficial in worst-case optimization applications, because there are only two parameters associated with the clutter instead of approximately $2N_c$.

3.7 Summary

In this chapter, we examined one of the main issues with the Block STAP method (dual channel system) - the block diagonalization of a family of STAP correlation matrices with a fixed transformation. We first defined a family of STAP correlation matrices, denoted by $\mathcal{S}$, that is representative of the interference environments typically encountered by an airborne surveillance radar and has sufficient complexity to stress the ability of STAP algorithms to remove both spatially and temporally correlated signals. Then, we addressed the problems of simultaneously diagonalizing and block diagonalizing the family $\mathcal{S}$ with a fixed transformation.

The simultaneous diagonalization of a family of matrices essentially requires the family to be a commuting family. We demonstrated that the family $\mathcal{S}$ is not a commuting family and hence, cannot be simultaneously diagonalized by a fixed transformation. The fact that a family cannot be simultaneously diagonalized does not imply that the family cannot be simultaneously block diagonalized by a fixed transformation. We reviewed the conditions for block diagonalizing a matrix in terms of vector spaces and subspaces, noting that the simultaneous block diagonalization of a family requires two independent

subspaces that are invariant to every matrix in the family. A transformation (change of basis matrix) constructed from the basis vectors for the two subspaces will block diagonalize the family through a similarity transformation. A review of the literature provided theorems for determining if a family of matrices can be block diagonalized or equivalently, if a block diagonalizing transformation exists. Basically, a family of matrices must satisfy a polynomial identity of the appropriate degree to be simultaneously block diagonalizable. Unfortunately, showing that a family satisfies a polynomial identity is extremely difficult, requiring the satisfaction of a near infinite number of conditions. For small M and N, we demonstrated with numerical examples that the family $\mathcal{S}$ did not satisfy the appropriate polynomial identity. Thus, we conjecture that the family of STAP correlation matrices cannot be simultaneously block diagonalized with a fixed transformation.

In Section 3.6, we departed from the family $\mathcal{S}$ and only considered the clutter correlation matrix under two additional sets of assumptions. Basically, these additional assumptions force the clutter environment to have symmetry in the azimuth angle plane. Under these assumptions, we derived the new result that the clutter correlation matrix is a centrosymmetric matrix. Thus, when these assumptions hold, the clutter correlation matrix can be easily block diagonalized with a efficient, fixed transformation. This centrosymmetric result provides additional insight into the characteristics of the clutter and has potential uses in selecting a suboptimal transformation for the block STAP processor that minimizes or reduces the SINR loss. Further, under the assumptions that all the clutter patches have the same power level and the number of clutter patches is infinite, the clutter correlation matrix is not only a centrosymmetric matrix, but is also parameterized by two parameters: the clutter to noise ratio and β. This simple two parameter characterization of the clutter correlation matrix has potential uses in modeling and simulation and optimization applications.

We proceeded with our research under the assumption that the conjecture that STAP correlation matrices cannot be simultaneously block diagonalized is true and with the objective using a fixed non-block diagonalizing transformation that reduces the loss in SINR performance. We examine the problem of selecting such a transformation in the next chapter.

IV. Transformation Selection

4.1 Introduction

In the previous chapter, we provided evidence to support the conjecture that a block diagonalizing transformation $\mathbf{V}$ does not exist for the family of interference plus noise correlation matrices $\mathcal{S}$. Assuming this conjecture is true, then the Block STAP processor (dual system channel) is not equivalent to the optimal process and will experience a loss in SINR performance. We can proceed with the development of a suboptimal Block STAP processor (dual channel system) in one of two ways. One, we could retain the concept of having a fixed transformation to save computational cost and select a non-block diagonalizing transformation that minimizes or reduces the loss in SINR performance. Two, we could abandon the fixed transformation concept and move to a more computationally expensive data adaptive transformation method which could potentially offer better SINR performance. In either case, we need a criterion for selecting the transformation that minimizes or reduces the loss in SINR performance. We examine the problem of how to select such a transformation in this chapter.

We start in Section 4.2 with a top-level analysis of the SINR performance of a Block STAP processor where the transformation $\mathbf{V}$ equals the identity matrix. Observe that with a non-block diagonalizing transformation, the potential exists for the Block STAP processor to have SINR performance worse than one of the channels, since one of the channels could act as a noise source for the other channel. The identity matrix configured Block STAP processor has the desirable property of providing SINR performance greater than or equal to either channel of the Block STAP processor and thus, will serve as our baseline system. Basically, this identity matrix configured Block STAP processor assumes that the off-diagonal blocks of the correlation are zero matrices which is not true, in general. In Section 4.3, we examine the notion of approximately block diagonalizing the correlation matrix in the sense of reducing the norm of the off-diagonal blocks. One might heuristically reason that a Block STAP processor based on a transformation that approximately block diagonalizes the correlation matrix (i.e., reducing the magnitude of the elements in off-diagonal blocks) would perform better than the identity matrix configuration, since the

identity matrix configuration simply assumes that off-diagonal blocks are zero. We demonstrate through a simulation example that reducing the l_2-norm of the off-diagonal blocks is not a sufficient criterion for improving the SINR performance of the baseline system. One might also hypothesize that similar weight vectors provide similar performance. We examine this hypothesis in Section 4.4, where we discuss the transformation selection problem from the perspective of a perturbation problem, leading to the notion of minimizing the difference between the optimum weight vector and the equivalent full-dimension Block STAP weight vector. The development of a selection criterion based on minimizing the difference between weight vectors does not directly consider the issue of SINR performance. Thus, in Section 4.5, we develop a transformation selection criterion through an in-depth analysis of the SINR efficiency (loss) of the Block STAP processor relative to the optimum STAP processor and discuss the utility of this criterion. In Section 4.6, we summarize the chapter.

4.2 Identity Transformation

One approach to implementing a suboptimal Block STAP processor would be to simply assume that the off-diagonal blocks of the correlation matrices are zero, i.e., assume the correlation matrices have a block diagonal form. Under this block diagonal assumption, a natural choice for the transformation $\mathbf{V}$ is the identity matrix. With $\mathbf{V} = \mathbf{I}_{MN}$, the rank reduction transformations $\mathbf{V}_1$ and $\mathbf{V}_2$ simply perform a selection process (which is essentially computationally free) that segments the signal vectors into two temporal periods (sub-CPIs): one associated with the antenna element samples from the first $M/2$ pulses and another associated with the remaining $M/2$ pulses. With $\mathbf{V} = \mathbf{I}_{MN}$, the Block STAP processing equates to summing the outputs from two sub-CPI optimum weight vectors. In this configuration, the cross-correlation information between the sub-CPIs, which is contained in the off-diagonal blocks, is not used in computing the weight vectors leading to a reduction in SINR performance. As noted earlier, the potential exists for the Block STAP processor to have SINR performance worse than one or both of the channels when a non-block diagonalizing transformation is used. In this section, we show that the SINR

performance of the Block STAP process with $\mathbf{V} = \mathbf{I}_{MN}$ is equal to or greater than the SINR performance of either channel of the Block STAP processor.

Recall that the SINR of the Block STAP process is given by

$$\mathrm{SINR}_{\mathrm{blk}} = \frac{P_{sblk}}{P_{nblk}} = \frac{\left(\mathbf{s}^H \mathbf{V} \tilde{\mathbf{Q}}^{-1} \mathbf{V}^H \mathbf{s}\right)^2}{\mathbf{s}^H \mathbf{V} \tilde{\mathbf{Q}}^{-1} \tilde{\mathbf{R}} \tilde{\mathbf{Q}}^{-1} \mathbf{V}^H \mathbf{s}}, \tag{4.1}$$

where $\mathbf{s}$ is the steering vector (desired signal),

$$\tilde{\mathbf{R}} = \begin{bmatrix} \mathbf{V}_1^H \mathbf{R} \mathbf{V}_1 & \mathbf{V}_1^H \mathbf{R} \mathbf{V}_2 \\ \mathbf{V}_2^H \mathbf{R} \mathbf{V}_1 & \mathbf{V}_2^H \mathbf{R} \mathbf{V}_2 \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{A}} & \tilde{\mathbf{B}} \\ \tilde{\mathbf{B}}^H & \tilde{\mathbf{C}} \end{bmatrix}, \tag{4.2}$$

$$\tilde{\mathbf{Q}} = \begin{bmatrix} \mathbf{V}_1^H \mathbf{R} \mathbf{V}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2^H \mathbf{R} \mathbf{V}_2 \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{A}} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{C}} \end{bmatrix}, \tag{4.3}$$

and $\mathbf{R}$ is the interference plus noise correlation matrix which is partitioned as

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^H & \mathbf{C} \end{bmatrix}. \tag{4.4}$$

Note that the blocks in the above equations are square matrices. Now, notice that since $\mathbf{s} = \mathbf{b}(\omega) \otimes \mathbf{a}(\upsilon)$ and $\mathbf{b}(\omega) = [1 \;\; e^{j2\pi\omega} \;\; \cdots \;\; e^{j2\pi(M-1)\omega}]^T$, we can write the partitioned steering vector as

$$\mathbf{s} = \begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{s}_1 \\ e^{j\theta} \mathbf{s}_1 \end{bmatrix}, \tag{4.5}$$

where $\theta = 2\pi\omega M/2$. The transformed steering vector $\mathbf{z} = \mathbf{V}^H \mathbf{s}$ is given as

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{V}_1^H \mathbf{s} \\ \mathbf{V}_2^H \mathbf{s} \end{bmatrix}. \tag{4.6}$$

Using the above observations, we can write the output signal power of the Block STAP processor as

$$P_{sblk} = \left(\mathbf{z}_1^H \tilde{\mathbf{A}}^{-1}\mathbf{z}_1 + \mathbf{z}_2^H \tilde{\mathbf{C}}^{-1}\mathbf{z}_2\right)^2, \tag{4.7}$$

and the average output interference plus noise power as

$$\begin{aligned} P_{nblk} &= \mathbf{z}_1^H \tilde{\mathbf{A}}^{-1}\mathbf{z}_1 + \mathbf{z}_2^H \tilde{\mathbf{C}}^{-1}\mathbf{z}_2 + \mathbf{z}_1^H \tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}}^H \tilde{\mathbf{C}}^{-1}\mathbf{z}_2 + \mathbf{z}_2^H \tilde{\mathbf{C}}^{-1}\tilde{\mathbf{B}}\tilde{\mathbf{A}}^{-1}\mathbf{z}_1 \\ &= \mathbf{z}_1^H \tilde{\mathbf{A}}^{-1}\mathbf{z}_1 + \mathbf{z}_2^H \tilde{\mathbf{C}}^{-1}\mathbf{z}_2 + 2\mathrm{Re}\{\mathbf{z}_1^H \tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}}\tilde{\mathbf{C}}^{-1}\mathbf{z}_2\}, \end{aligned} \tag{4.8}$$

where $\mathrm{Re}\{x\}$ denotes the real part of the complex scalar x. Each channel of the Block STAP processor acts as a reduced rank STAP processor and thus, the upper channel output SINR is [19]

$$\mathrm{SINR}_1 = \mathbf{z}_1^H \tilde{\mathbf{A}}^{-1}\mathbf{z}_1 \tag{4.9}$$

and the lower channel output SINR is

$$\mathrm{SINR}_2 = \mathbf{z}_2^H \tilde{\mathbf{C}}^{-1}\mathbf{z}_2. \tag{4.10}$$

Our objective is to show that

$$\mathrm{SINR}_{\mathrm{blk}} \geq \mathrm{SINR}_1 \quad \text{and} \quad \mathrm{SINR}_{\mathrm{blk}} \geq \mathrm{SINR}_2 \quad \text{if } \mathbf{V} = \mathbf{I}_{MN}. \tag{4.11}$$

First, observe that

$$\mathrm{Re}\{\mathbf{z}_1^H \tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}}\tilde{\mathbf{C}}^{-1}\mathbf{z}_2\} \leq |\mathbf{z}_1^H \tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}}\tilde{\mathbf{C}}^{-1}\mathbf{z}_2\ |, \tag{4.12}$$

and since $\mathbf{R}$ ($\tilde{\mathbf{R}}$) is positive definite, that [23:473]

$$2|\mathbf{z}_1^H \tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}}\tilde{\mathbf{C}}^{-1}\mathbf{z}_2| \leq \mathbf{z}_1^H \tilde{\mathbf{A}}^{-1}\mathbf{z}_1 + \mathbf{z}_2^H \tilde{\mathbf{C}}^{-1}\mathbf{z}_2. \tag{4.13}$$

Thus, we can place an upper bound on P_{nblk} as

$$\begin{aligned} P_{nblk} &= \mathbf{z}_1^H \tilde{\mathbf{A}}^{-1}\mathbf{z}_1 + \mathbf{z}_2^H \tilde{\mathbf{C}}^{-1}\mathbf{z}_2 + 2\mathrm{Re}\{\mathbf{z}_1^H \tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}}\tilde{\mathbf{C}}^{-1}\mathbf{z}_2\} \\ &\leq \mathbf{z}_1^H \tilde{\mathbf{A}}^{-1}\mathbf{z}_1 + \mathbf{z}_2^H \tilde{\mathbf{C}}^{-1}\mathbf{z}_2 + 2|\mathbf{z}_1^H \tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}}\tilde{\mathbf{C}}^{-1}\mathbf{z}_2| \\ &\leq 2\big(\mathbf{z}_1^H \tilde{\mathbf{A}}^{-1}\mathbf{z}_1 + \mathbf{z}_2^H \tilde{\mathbf{C}}^{-1}\mathbf{z}_2\big), \end{aligned} \tag{4.14}$$

which, in turn, places a lower bound on $\mathrm{SINR}_{\mathrm{blk}}$ of

$$\mathrm{SINR}_{\mathrm{blk}} \geq \frac{1}{2}\big(\mathbf{z}_1^H \tilde{\mathbf{A}}^{-1}\mathbf{z}_1 + \mathbf{z}_2^H \tilde{\mathbf{C}}^{-1}\mathbf{z}_2\big). \tag{4.15}$$

When $\mathbf{V} = \mathbf{I}_{MN}$, observe that $\tilde{\mathbf{R}} = \mathbf{R}$ and $\mathbf{z} = \mathbf{s}$ ($\mathbf{z}_1 = \mathbf{s}_1$ and $\mathbf{z}_2 = e^{j\theta}\mathbf{s}_1$). Thus, we can rewrite the bound in Eqn. (4.15) as

$$\mathrm{SINR}_{\mathrm{blk}} \geq \frac{1}{2}\big(\mathbf{s}_1^H \mathbf{A}^{-1}\mathbf{s}_1 + \mathbf{s}_1^H \mathbf{C}^{-1}\mathbf{s}_1\big), \tag{4.16}$$

and the upper and lower channel SINRs as

$$\mathrm{SINR}_1 = \mathbf{s}_1^H \mathbf{A}^{-1}\mathbf{s}_1 \tag{4.17}$$

$$\mathrm{SINR}_2 = \mathbf{s}_1^H \mathbf{C}^{-1}\mathbf{s}_1. \tag{4.18}$$

Finally, since the members of the family of interference plus noise correlation matrices $\mathcal{S}$ have a Toeplitz-block-Toeplitz structure, we know that $\mathbf{A} = \mathbf{C}$ which reduces Eqn. (4.16) to

$$\mathrm{SINR}_{\mathrm{blk}} \geq \mathbf{s}_1^H \mathbf{A}^{-1}\mathbf{s}_1 = \mathrm{SINR}_1 = \mathrm{SINR}_2. \tag{4.19}$$

Therefore, the SINR performance of Block SINR processor is greater than or equal to the SINR performance of either channel when $\mathbf{V} = \mathbf{I}_{MN}$, i.e., Eqn. (4.11) holds. As shown below, Eqn. (4.11) also holds when $\mathbf{V}$ is a unitary block diagonal matrix since the unitary block diagonal matrix configuration is identical to the identity matrix configuration. From Eqn. (4.15), we cannot conclusively state that Eqn. (4.11) does or does not hold for an arbitrary unitary transformation. However, through computer simulations, we have found

some configurations of interference plus noise correlation matrices, unitary transformations, and steering vectors where the Block STAP SINR performance is below one or both of the channels. That is, the potential exist for the Block STAP processor to perform worse than a single channel reduced rank STAP processor based on either $\mathbf{V}_1$ or $\mathbf{V}_2$ if the transformation $\mathbf{V}$ is not properly selected. The SINR performance of the identity matrix configured Block STAP processor will serve as the baseline since its performance meets or exceeds the performance of either channel. In the next three sections, we examine the problem of developing a criterion for properly selecting the transformation $\mathbf{V}$ to minimize the loss in the SINR performance.

Before moving onto the next section, we first establish our claim that a unitary block diagonal matrix configuration and the identity matrix configuration are equivalent. As a result, we can eliminate unitary block diagonal matrices from the list of potential transformation candidates for improving SINR performance of the baseline system. Observe that $\tilde{\mathbf{R}} = \mathbf{R}$, $\tilde{\mathbf{Q}} = \mathbf{Q}$, and $\mathbf{V}^H\mathbf{s} = \mathbf{s}$ when $\mathbf{V} = \mathbf{I}_{MN}$ and thus, we can rewrite Eqn. (4.1) as

$$\mathrm{SINR}_{\mathrm{blk}} = \frac{\left(\mathbf{s}^H\mathbf{Q}^{-1}\mathbf{s}\right)^2}{\mathbf{s}^H\mathbf{Q}^{-1}\mathbf{R}\mathbf{Q}^{-1}\mathbf{s}} \quad \text{if } \mathbf{V} = \mathbf{I}_{MN}, \tag{4.20}$$

where

$$\mathbf{Q} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \end{bmatrix}. \tag{4.21}$$

To prove our claim, we must show that the SINR of the unitary block diagonal matrix configuration is equal to Eqn. (4.20). Let $\mathbf{V}$ be a unitary block diagonal matrix, i.e.,

$$\mathbf{V} = \begin{bmatrix} \mathbf{U}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_2 \end{bmatrix}, \tag{4.22}$$

where $\mathbf{U}_1$ and $\mathbf{U}_2$ are unitary matrices. Now, observe that

$$\tilde{\mathbf{Q}} = \begin{bmatrix} \mathbf{V}_1^H\mathbf{R}\mathbf{V}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2^H\mathbf{R}\mathbf{V}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{U}_1^H\mathbf{A}\mathbf{U}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_2^H\mathbf{C}\mathbf{U}_2 \end{bmatrix} = \mathbf{V}^H\mathbf{Q}\mathbf{V}, \tag{4.23}$$

since

$$\mathbf{V}_1 = \begin{bmatrix} \mathbf{U}_1 \\ \mathbf{0} \end{bmatrix} \quad \text{and} \quad \mathbf{V}_2 = \begin{bmatrix} \mathbf{0} \\ \mathbf{U}_2 \end{bmatrix}.$$

Recalling that $\tilde{\mathbf{R}} = \mathbf{V}^H \mathbf{R} \mathbf{V}$ and substituting Eqn. (4.23) into Eqn. (4.1) yields

$$\text{SINR}_{\text{blk}} = \frac{\left(\mathbf{s}^H \mathbf{V}\mathbf{V}^H \mathbf{Q}^{-1} \mathbf{V}\mathbf{V}^H \mathbf{s}\right)^2}{\mathbf{s}^H \mathbf{V}\mathbf{V}^H \mathbf{Q}^{-1} \mathbf{V}\mathbf{V}^H \mathbf{R} \mathbf{V}\mathbf{V}^H \mathbf{Q}^{-1} \mathbf{V}\mathbf{V}^H \mathbf{s}}, \tag{4.24}$$

which reduces to

$$\text{SINR}_{\text{blk}} = \frac{\left(\mathbf{s}^H \mathbf{Q}^{-1} \mathbf{s}\right)^2}{\mathbf{s}^H \mathbf{Q}^{-1} \mathbf{R} \mathbf{Q}^{-1} \mathbf{s}} \tag{4.25}$$

since $\mathbf{V}\mathbf{V}^H = \mathbf{I}$. Thus, our claim is proven.

4.3 Approximate Block Diagonalization

As noted earlier, the identity matrix configured Block STAP processor does not use the cross-correlation information contained in the off-diagonal blocks since these blocks are assumed to be zero. In general, the off-diagonal blocks of the correlation matrix are not zero. Heuristically, one might reason that selecting a transformation $\mathbf{V}$ that drives the off-diagonal blocks towards zero would improve the SINR performance of the Block STAP processor. This suggests the concept of approximately block diagonalizing the correlation matrix in the sense of reducing the total magnitude of the elements in the off-diagonal blocks. That is, instead of simply assuming that the off-diagonal blocks are zero matrices, we select a transformation to make the off-diagonal blocks look more like zero matrices in a l_2-norm sense. The l_2-norm (also called the Frobenius norm) of a $n \times n$ matrix $\mathbf{X}$ will be denote by $\|\mathbf{X}\|_2$ and is defined as [23:291]

$$\|\mathbf{X}\|_2 = \left(\sum_{p,q=1}^{n} |[\mathbf{X}]_{p,q}|^2 \right)^{1/2}, \tag{4.26}$$

where $[\mathbf{X}]_{p,q}$ denotes the elements of $\mathbf{X}$. To isolate the off-diagonal blocks of the transformed correlation matrix $\tilde{\mathbf{R}}$, we define an error matrix $\mathbf{E}$ as

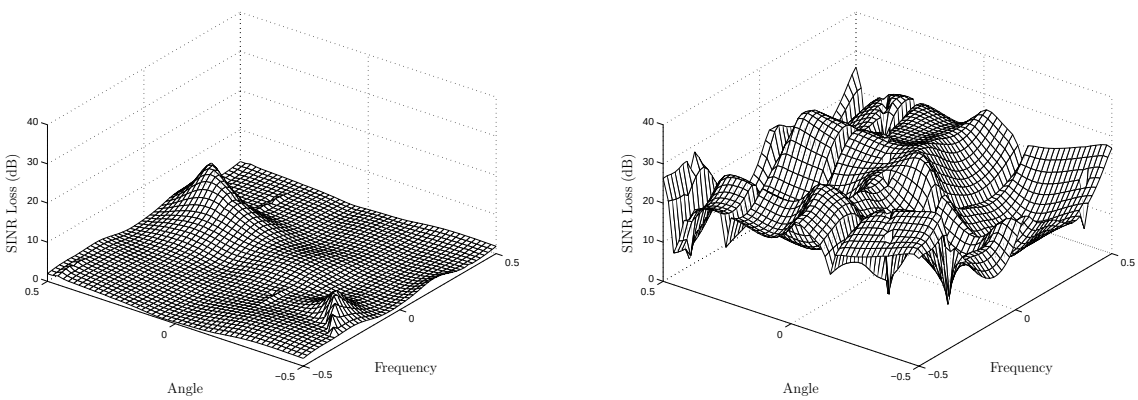
$$\mathbf{E} = \tilde{\mathbf{R}} - \tilde{\mathbf{Q}} = \begin{bmatrix} \mathbf{0} & \tilde{\mathbf{B}} \\ \tilde{\mathbf{B}}^H & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{V}_1^H \mathbf{R} \mathbf{V}_2 \\ \mathbf{V}_2^H \mathbf{R} \mathbf{V}_1 & \mathbf{0} \end{bmatrix}. \tag{4.27}$$

Following this heuristic line of reasoning, we would select the transformation $\mathbf{V}$ to reduce $\|\mathbf{E}\|_2$ and as a result, reduce the loss in SINR performance. Certainly, if we reduce $\|\mathbf{E}\|_2$ to zero, then $\tilde{\mathbf{R}}$ would be a block diagonal matrix and we would have no loss in SINR performance. However, as we show next, by way of a simulation example, reducing the l_2-norm of the error matrix $\mathbf{E}$ does not necessarily improve the overall SINR performance.

The simulated interference plus noise environment consisted of receiver noise, clutter, and three barrage noise jammers with the relevant simulation parameters listed in Table 4.1. The selected transform $\mathbf{V}$ was an 80 point discrete Fourier transform (DFT) matrix with $\mathbf{V}_1$ and $\mathbf{V}_2$ equal to first and last 40 columns of $\mathbf{V}$, respectively. The DFT matrix configuration reduced the off-diagonal norm ($\|\mathbf{E}\|_2$) from 3.67×10^6 to 2.53×10^6. Figure 4.1 shows the SINR performance (loss) of the identity matrix and DFT matrix configured Block STAP processors relative to the optimum processor over the entire angle-Doppler plane. Figure 4.1 clearly shows that the DFT matrix configuration has significantly worse SINR performance even though the l_2-norm of the off-diagonal blocks was reduced. The in-depth reasons for this decreased performance are discussed in Section 4.5.2. On a general level, the problem with the notion of approximately block diagonalizing the correlation matrix lies in the fact that the inverse of the correlation matrix is used to compute the weight vectors. The inverse of a 2×2 block matrix is not simply the inverse of the blocks. Each block of the inverted matrix is a function of the four blocks in the 2×2 block matrix. Thus, approximately block diagonalizing the correlation matrix by reducing the off-diagonal norm does not necessarily imply that the inverted correlation matrix will have a desirable block diagonal form. In the next section, we examine another heuristic approach for developing a criterion for selecting the transformation $\mathbf{V}$ which is based on the hypothesis that similar weight vectors should provided similar performance. This heuristic

Parameters	Values
Normalized Jammer Angles	0.25, 0.433, -0.15
Jammer to Noise Ratio (dB)	40, 30, 30
Clutter to Noise Ratio (dB)	55
Clutter Beta	1.3
Pulses per CPI	10
Number of Antenna Elements	8

Table 4.1 Simulation parameters.



(a) Identity matrix configuration

(b) DFT matrix configuration

Figure 4.1 SINR performance of the identity matrix and DFT matrix configurations relative to the optimum processor.

approach considers the inverse of the correlation matrix and can be discussed from the perspective of a perturbation problem.

4.4 Similar Weight Vectors

Under the hypothesis that similar weight vectors provide similar performance (similarity hypothesis), one would reason that the transformation $\mathbf{V}$ should be selected to reduce the difference between the optimum weight vector and the Block STAP weight vectors. The dual channel processing of the Block STAP method can be expressed in terms of an equivalent full dimensional weight vector as

$$\mathbf{w}_{\text{blk}} = \tilde{\mathbf{Q}}^{-1}\mathbf{z}, \tag{4.28}$$

with the optimum weight vector given by

$$\mathbf{w}_{\text{opt}} = \tilde{\mathbf{R}}^{-1}\mathbf{z}, \tag{4.29}$$

where $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{Q}}$ are defined in Eqns. (4.2) and (4.3) and $\mathbf{z} = \mathbf{V}^H\mathbf{s}$ is the transformed steering vector. To examine the validity of the similarity hypothesis, we need some measure of the similarity between the optimum weight vector and the equivalent full dimension Block STAP weight vector. The l_2-norm of the difference between two vectors is a common criterion for measuring the similarity between two vectors and is convenient mathematically. Thus, we define an error vector $\mathbf{e}$ between the optimum and Block STAP weight vectors as $\mathbf{e} = \mathbf{w}_{\text{opt}} - \mathbf{w}_{\text{blk}}$. We can bound the l_2-norm of $\mathbf{e}$ as

$$\|\mathbf{e}\|_2 = \|\tilde{\mathbf{R}}^{-1}\mathbf{z} - \tilde{\mathbf{Q}}^{-1}\mathbf{z}\|_2 \leq \|\tilde{\mathbf{R}}^{-1} - \tilde{\mathbf{Q}}^{-1}\|_2\|\mathbf{z}\|_2. \tag{4.30}$$

Assuming the similarity hypothesis is true, then we would select the transformation $\mathbf{V}$ to reduce $\|\tilde{\mathbf{R}}^{-1} - \tilde{\mathbf{Q}}^{-1}\|_2$. Note that reducing the norm of the matrix $\mathbf{E}$, defined in Eqn. (4.27), does not necessarily imply a reduction in $\|\tilde{\mathbf{R}}^{-1} - \tilde{\mathbf{Q}}^{-1}\|_2$. From the previous simulation example, we find that $\|\tilde{\mathbf{R}}^{-1} - \tilde{\mathbf{Q}}^{-1}\|_2$ equals 2.47 for the identity matrix configuration and 5.07 for the DFT matrix configuration. Thus, in general, the weight vector of the DFT matrix configuration is less similar to the optimum weight vector than the identity matrix configuration, providing some rationale for its decreased SINR performance assuming the similarity hypothesis is true. Before discussing the validity of the similarity hypothesis further, we will review some additional results available on the error vector. The error analysis between $\mathbf{w}_{\text{opt}}$ and $\mathbf{w}_{\text{blk}}$ can be viewed as a perturbation problem, which is a well researched problem in matrix analysis.

Suppose we reverse the roles of $\mathbf{w}_{\text{opt}}$ and $\mathbf{w}_{\text{blk}}$. Then, $\tilde{\mathbf{Q}}$ represents the unperturbed system and $\tilde{\mathbf{R}}$ represents the perturbed system since $\tilde{\mathbf{R}} = \tilde{\mathbf{Q}} + \mathbf{E}$. The weight vectors $\mathbf{w}_{\text{blk}}$ and $\mathbf{w}_{\text{opt}}$ are the solutions to the unperturbed and perturbed systems, respectively. Horn

and Johnson [23:337] provide the following bound for the norm between $\mathbf{w}_{\text{blk}}$ and $\mathbf{w}_{\text{opt}}$:

$$\frac{\|\mathbf{w}_{\text{blk}} - \mathbf{w}_{\text{opt}}\|}{\|\mathbf{w}_{\text{blk}}\|} \leq \frac{\|\tilde{\mathbf{Q}}^{-1}\mathbf{E}\|}{1 - \|\tilde{\mathbf{Q}}^{-1}\mathbf{E}\|} \quad \text{if } \|\tilde{\mathbf{Q}}^{-1}\mathbf{E}\| < 1. \tag{4.31}$$

Note that Eqn.(4.31) was derived under the condition that $\rho(\tilde{\mathbf{Q}}^{-1}\mathbf{E}) < 1$, where $\rho(\tilde{\mathbf{Q}}^{-1}\mathbf{E})$ denotes the spectral radius of $\tilde{\mathbf{Q}}^{-1}\mathbf{E}$ (i.e., $\rho(\mathbf{X}) = \max\{|\lambda_i|\colon \lambda_i \text{ are the eigenvalues of } \mathbf{X}\}$) [23:296]. Also note that $\|\tilde{\mathbf{Q}}^{-1}\mathbf{E}\| < 1$ implies $\rho(\tilde{\mathbf{Q}}^{-1}\mathbf{E}) < 1$, but the reverse is not true [23:297]. Finally, note that in Eqns. (4.31) the matrix and vector norms must be compatible (i.e., $\|\mathbf{U}\mathbf{x}\| \leq \|\mathbf{U}\|\,\|\mathbf{x}\|$) [23:293]. Now, if we assume that $\|\mathbf{E}\| < 1/\|\tilde{\mathbf{Q}}^{-1}\|$, then the bound in Eqn. (4.31) can be rewritten as [23:337]

$$\frac{\|\mathbf{w}_{\text{blk}} - \mathbf{w}_{\text{opt}}\|}{\|\mathbf{w}_{\text{blk}}\|} \leq \frac{\kappa(\tilde{\mathbf{Q}})}{1 - \kappa(\tilde{\mathbf{Q}})(\|\mathbf{E}\|/\|\tilde{\mathbf{Q}}\|)} \frac{\|\mathbf{E}\|}{\|\tilde{\mathbf{Q}}\|} \quad \text{if } \|\tilde{\mathbf{Q}}^{-1}\|\,\|\mathbf{E}\| < 1, \tag{4.32}$$

where

$$\kappa(\tilde{\mathbf{Q}}) = \begin{cases} \|\tilde{\mathbf{Q}}^{-1}\|\,\|\tilde{\mathbf{Q}}\| & \text{if } \tilde{\mathbf{Q}} \text{ is nonsingular} \\ \infty & \text{if } \tilde{\mathbf{Q}} \text{ is singular.} \end{cases} \tag{4.33}$$

With regard to the similarity between $\mathbf{w}_{\text{blk}}$ and $\mathbf{w}_{\text{opt}}$, we observe from Eqn. (4.32) that the earlier notation of reducing the l_2-norm of the off-diagonal blocks ($\|\mathbf{E}\|_2$) is not without merit if $\|\tilde{\mathbf{Q}}^{-1}\|\|\mathbf{E}\| < 1$. The right-hand side of Eqns. (4.31) represents upper bound for $\|\tilde{\mathbf{Q}}^{-1} - \tilde{\mathbf{R}}^{-1}\|$ relative to $\tilde{\mathbf{Q}}^{-1}$ when $\|\tilde{\mathbf{Q}}^{-1}\mathbf{E}\| < 1$. These results suggest that the transformation $\mathbf{V}$ should be selected to reduce the norm of $\tilde{\mathbf{Q}}^{-1}\mathbf{E}$ to minimize the difference between $\mathbf{w}_{\text{opt}}$ and $\mathbf{w}_{\text{blk}}$ and hence, minimize the loss in SINR performance if the similarity hypothesis is true. However, we will demonstrate that the similarity hypothesis does not hold, in general, with a simulation example.

In this simulation example, we use the same interference plus noise environment as the simulation example in Section 4.3, but we change the transformation matrix $\mathbf{V}$. Instead of a DFT matrix, we used a transformation that block diagonalizes a centrosymmetric

matrix, i.e.,

$$\mathbf{V}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{I} \\ \mathbf{J} \end{bmatrix} \quad \text{and} \quad \mathbf{V}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{J} \\ -\mathbf{I} \end{bmatrix}. \tag{4.34}$$

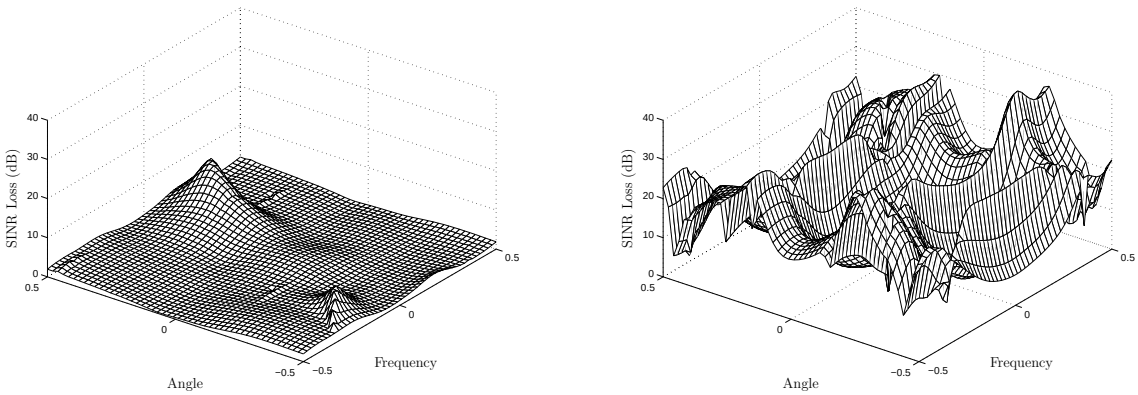
Figure 4.2 shows the SINR performance (loss) of the identity and centrosymmetric matrix configured Block STAP processors relative to the optimum processor over the entire angle-Doppler plane. Figure 4.2 shows that the identity matrix configuration has significantly better performance than the centrosymmetric matrix configuration. The difference in SINR performance is discussed in Section 4.5.2 (Note, for the centrosymmetric matrix configuration that $\|\mathbf{E}\|_2 = 3.60 \times 10^5$ and $\|\tilde{\mathbf{R}}^{-1} - \tilde{\mathbf{Q}}^{-1}\|_2 = 2.51$). To demonstrate that the similarity hypothesis does not hold, we simply need show that the SINR of the identity matrix configuration is greater than the centrosymmetric matrix configuration even through the l_2-norm of the identity matrix configuration error vector $\mathbf{e}$ is greater than the centrosymmetric matrix configuration. Let the $\mathrm{SINR}_{id}(\upsilon, \omega)$ and $\mathrm{SINR}_c(\upsilon, \omega)$ denote the SINR of the identity matrix and centrosymmetric matrix configurations, respectively, and let $\mathrm{e}_{id}(\upsilon, \omega)$ and $\mathrm{e}_c(\upsilon, \omega)$ denote the l_2-norm of the error vector for the respective systems. Note the SINR and error norm are functions of υ and ω since they depend on the pointing direction (normalized angle υ and normalized Doppler ω) of the steering vector. Now, define an indicator function $\chi(x)$ such that

$$\chi(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{4.35}$$

The similarity hypothesis does not hold if the function

$$f(\upsilon, \omega) = \chi(\mathrm{SINR}_{id}(\upsilon, \omega) - \mathrm{SINR}_c(\upsilon, \omega))\, \chi(\mathrm{e}_{id}(\upsilon, \omega) - \mathrm{e}_c(\upsilon, \omega)) \tag{4.36}$$

is not zero for all υ and ω. Figure 4.3(a) shows a plot of the function $f(\upsilon, \omega)$ and clearly, shows that the similarity hypothesis does not hold. Figures 4.3(b) and 4.3(c) show the difference between the SINR and the norm of the error vector for the two configuration



(a) Identity matrix configuration

(b) Centrosymmetric matrix configuration

Figure 4.2 SINR performance of the identity matrix and centrosymmetric matrix configurations relative to the optimum processor.

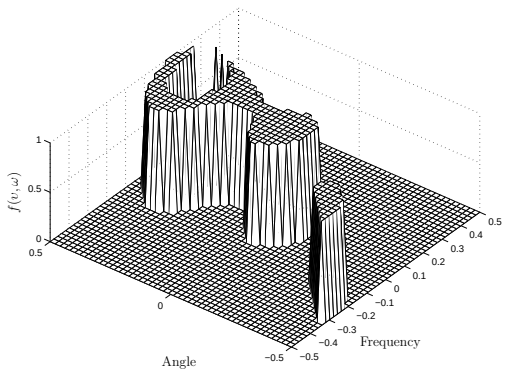
in the region where $f(\upsilon, \omega)$ is one, respectively. The differences shown in Figs. 4.3(b) and 4.3(c) are non-trivial (i.e., cannot be attributed round-off error).

Although the hypothesis that similar weight vectors should provide similar performance is not true in a l_2-norm sense, the notion of reducing the norm of $\mathbf{e}$ with regard to the condition in Eqn. (4.31) does influence the worst-case SINR performance of the Block STAP processor as discussed in the next section. In the next section, we explicitly consider the issue of SINR performance in developing a criterion for selecting the transformation $\mathbf{V}$.

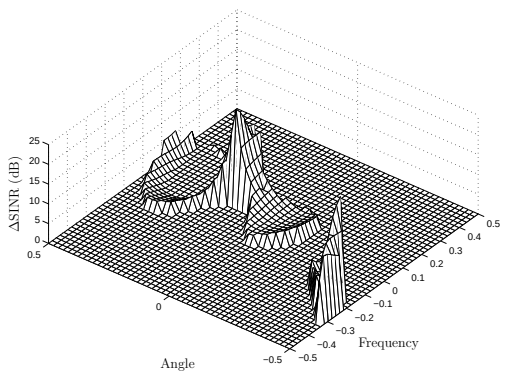
4.5 Block STAP SINR Efficiency

In the previous two sections, we discussed the problem of selecting the transformation $\mathbf{V}$ from primarily a heuristic perspective without directly considering the SINR performance of the Block STAP processor. In this section, we directly consider SINR performance by defining and analyzing the SINR efficiency of the Block STAP processor relative to the optimum processor. Recall that the output SINR of the optimum process is

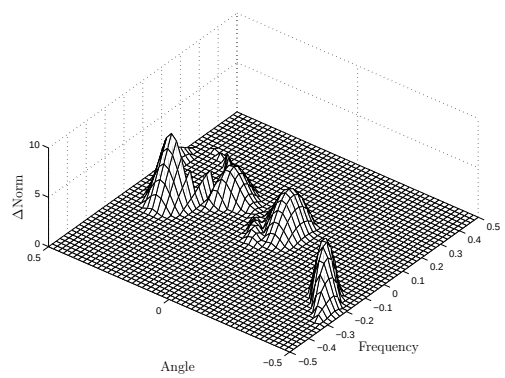
$$\mathrm{SINR}_{\max} = \mathbf{s}^H \mathbf{R}^{-1} \mathbf{s} = \mathbf{z}^H \tilde{\mathbf{R}}^{-1} \mathbf{z}, \tag{4.37}$$



(a) Similarity hypothesis fails when $f(\upsilon,\omega)$=1



(b) Difference in SINR



(c) Difference in l_2-norm of $\mathbf{e}$

Figure 4.3 Plots showing the regions of the angle-Doppler plane where the SINR and $\|\mathbf{e}\|_2$ of the identity matrix configuration simultaneously exceed those of the centrosymmetric matrix configuration.

where $\mathbf{z} = \mathbf{V}^H\mathbf{s}$, $\tilde{\mathbf{R}} = \mathbf{V}^H\mathbf{R}\mathbf{V}$, and $\mathbf{V}$ is a unitary matrix. We define the SINR efficiency (loss) of the Block STAP processor relative to the optimum STAP process as

$$\varepsilon = \frac{\mathrm{SINR}_{\mathrm{blk}}}{\mathrm{SINR}_{\max}} = \frac{\left(\mathbf{z}^H\tilde{\mathbf{Q}}^{-1}\mathbf{z}\right)^2}{\mathbf{z}^H\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{R}}\tilde{\mathbf{Q}}^{-1}\mathbf{z}}\,\frac{1}{\mathbf{z}^H\tilde{\mathbf{R}}^{-1}\mathbf{z}}. \tag{4.38}$$

Observe that the efficiency ε lies between 0 and 1 and that larger values of ε indicate a Block STAP processor with better SINR performance. The SINR efficiency in Eqn. (4.38) can also be viewed as the loss in SINR performance of the Block STAP processor relative to the optimum STAP processor in the sense that $\mathrm{SINR}_{\mathrm{blk}} = \varepsilon\mathrm{SINR}_{\max}$, i.e., the SINR loss is zero when $\varepsilon = 1$. In the next section, we analyze Eqn. (4.38) and develop a lower bound for the efficiency ε. From this bound, we can define a criterion for selecting the transformation $\mathbf{V}$ for improving the worst-case SINR performance of the Block STAP processor. Because the bound is not sharp in the sense that the actual SINR performance of the Block STAP processor may not approach the bound, we have to temper the criterion as discussed in Section 4.5.2.

4.5.1 Efficiency Analysis. Our basic approach is to introduce a transformation so that we can rewrite Eqn. (4.38) in a form where we can apply existing results to bound the efficiency. We start with the following known result. Suppose $\mathbf{X}$ is positive definite and Hermitian and $\mathbf{Y}$ is Hermitian, then there exists a non-singular matrix $\mathbf{U}$ such that $\mathbf{U}\mathbf{X}\mathbf{U}^H = \mathbf{I}$ and $\mathbf{U}\mathbf{Y}\mathbf{U}^H = \mathbf{\Lambda}$, where $\mathbf{\Lambda}$ is a diagonal matrix of the real eigenvalues of $\mathbf{X}^{-1}\mathbf{Y}$ [23:250]. Recall that $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{Q}}$ are positive definite and Hermitian and thus, there exists a nonsingular matrix $\mathbf{T}$ such that

$$\mathbf{T}^H\tilde{\mathbf{Q}}\mathbf{T} = \mathbf{I} \tag{4.39}$$

$$\mathbf{T}^H\tilde{\mathbf{R}}\mathbf{T} = \mathbf{D}, \tag{4.40}$$

where $\mathbf{D}$ is a diagonal matrix of the real eigenvalues of $\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{R}}$. Since $\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{R}}$ is the product of two positive definite and Hermitian matrices, all of the eigenvalues of $\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{R}}$ are not only real, but also greater than zero [23:465]. Now, observe from Eqns. (4.39) and (4.40)

that

$$\tilde{\mathbf{Q}} = \mathbf{T}^{-H}\mathbf{I}\mathbf{T}^{-1} = \mathbf{T}^{-H}\mathbf{T}^{-1} \tag{4.41}$$

$$\tilde{\mathbf{R}} = \mathbf{T}^{-H}\mathbf{D}\mathbf{T}^{-1} \tag{4.42}$$

which implies that

$$\tilde{\mathbf{Q}}^{-1} = \mathbf{T}\mathbf{T}^{H} \tag{4.43}$$

$$\tilde{\mathbf{R}}^{-1} = \mathbf{T}\mathbf{D}^{-1}\mathbf{T}^{H}. \tag{4.44}$$

Using the above equations, we can write $\mathrm{SINR}_{\mathrm{blk}}$ as

$$\mathrm{SINR}_{\mathrm{blk}} = \frac{\left(\mathbf{z}^{H}\mathbf{T}\mathbf{T}^{H}\mathbf{z}\right)^{2}}{\mathbf{z}^{H}\mathbf{T}\mathbf{T}^{H}\mathbf{T}^{-H}\mathbf{D}\mathbf{T}^{-1}\mathbf{T}\mathbf{T}^{H}\mathbf{z}} = \frac{\left(\mathbf{z}^{H}\mathbf{T}\mathbf{T}^{H}\mathbf{z}\right)^{2}}{\mathbf{z}^{H}\mathbf{T}\mathbf{D}\mathbf{T}^{H}\mathbf{z}} \tag{4.45}$$

and $\mathrm{SINR}_{\mathrm{max}}$ as

$$\mathrm{SINR}_{\mathrm{max}} = \mathbf{z}^{H}\mathbf{T}\mathbf{D}^{-1}\mathbf{T}^{H}\mathbf{z}. \tag{4.46}$$

Substituting Eqns. (4.45) and (4.46) into Eqn. (4.38) yields

$$\varepsilon = \frac{\left(\mathbf{z}^{H}\mathbf{T}\mathbf{T}^{H}\mathbf{z}\right)^{2}}{\mathbf{z}^{H}\mathbf{T}\mathbf{D}\mathbf{T}^{H}\mathbf{z}}\,\frac{1}{\mathbf{z}^{H}\mathbf{T}\mathbf{D}^{-1}\mathbf{T}^{H}\mathbf{z}} \tag{4.47}$$

and by letting $\mathbf{x} = \mathbf{T}^{H}\mathbf{z}$, we have

$$\varepsilon = \frac{\left(\mathbf{x}^{H}\mathbf{x}\right)^{2}}{\left(\mathbf{x}^{H}\mathbf{D}\mathbf{x}\right)\left(\mathbf{x}^{H}\mathbf{D}^{-1}\mathbf{x}\right)}. \tag{4.48}$$

Thus, the SINR efficiency of the Block STAP processor relative to the optimum processor is depended on the eigenvalues of $\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{R}}$ which are the elements of $\mathbf{D}$. Using Kantorovich's

inequality, we can bound the SINR efficiency as [23:444]

$$\varepsilon = \frac{\left(\mathbf{x}^H\mathbf{x}\right)^2}{\left(\mathbf{x}^H\mathbf{D}\mathbf{x}\right)\left(\mathbf{x}^H\mathbf{D}^{-1}\mathbf{x}\right)} \geq \frac{4\lambda_{min}\lambda_{max}}{(\lambda_{min}+\lambda_{max})^2}. \tag{4.49}$$

where λ_{min} and λ_{max} are the minimum and maximum eigenvalues of $\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{R}}$, respectively. We can simplify this bound by examining the behavior of the eigenvalues of $\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{R}}$.

We first observe that the product $\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{R}}$ can be written as

$$\begin{aligned}\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{R}} &= \begin{bmatrix}\tilde{\mathbf{A}}^{-1} & \mathbf{0}\\ \mathbf{0} & \tilde{\mathbf{C}}^{-1}\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{A}} & \tilde{\mathbf{B}}\\ \tilde{\mathbf{B}}^H & \tilde{\mathbf{C}}\end{bmatrix} = \begin{bmatrix}\mathbf{I} & \tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}}\\ \tilde{\mathbf{C}}^{-1}\tilde{\mathbf{B}}^H & \mathbf{I}\end{bmatrix} = \mathbf{I} + \begin{bmatrix}\mathbf{0} & \tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}}\\ \tilde{\mathbf{C}}^{-1}\tilde{\mathbf{B}}^H & \mathbf{0}\end{bmatrix}\\ &= \mathbf{I} + \mathbf{G}.\end{aligned} \tag{4.50}$$

One can show that $\mathbf{G}$ is similar to a Hermitian matrix and hence, is diagonalizable with real eigenvalues. Let $\{\lambda_i\}_{i=1}^{MN}$ and $\{\sigma_i\}_{i=1}^{MN}$ be the eigenvalues of $\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{R}}$ and $\mathbf{G}$, respectively. As noted earlier, all of the eigenvalues of $\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{R}}$ are real and greater than zero, i.e., $\lambda_i > 0$ for all i. Now, since $\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{R}}$ is the sum of an identity matrix and a diagonalizable matrix, we observe that $\lambda_i = 1 + \sigma_i$ and when combined with the fact that $\lambda_i > 0$ for all i, it implies

$$\sigma_i > -1 \quad \text{for all } i. \tag{4.51}$$

We can also place an upper bound on $\{\sigma_i\}_{i=1}^{MN}$ as follows. Let $\mathbf{y}^T = [\mathbf{y}_1^T\ \mathbf{y}_2^T]$ and σ_y form an eigenpair for $\mathbf{G}$, i.e., $\mathbf{G}\mathbf{y} = \sigma_y\mathbf{y}$. Then, observe that

$$\mathbf{G}\mathbf{y} = \begin{bmatrix}\mathbf{0} & \tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}}\\ \tilde{\mathbf{C}}^{-1}\tilde{\mathbf{B}}^H & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} = \begin{bmatrix}\tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}}\mathbf{y}_2\\ \tilde{\mathbf{C}}^{-1}\tilde{\mathbf{B}}^H\mathbf{y}_1\end{bmatrix} = \sigma_y\mathbf{y} = \begin{bmatrix}\sigma_y\mathbf{y}_1\\ \sigma_y\mathbf{y}_2\end{bmatrix} \tag{4.52}$$

which implies that

$$\tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}}\mathbf{y}_2 = \sigma_y\mathbf{y}_1 \tag{4.53}$$

$$\tilde{\mathbf{C}}^{-1}\tilde{\mathbf{B}}^H\mathbf{y}_1 = \sigma_y\mathbf{y}_2. \tag{4.54}$$

Solving Eqn. (4.53) for $\mathbf{y}_1$ and substituting into Eqn. (4.54) yields

$$\tilde{\mathbf{C}}^{-1}\tilde{\mathbf{B}}^H\tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}}\mathbf{y}_2 = \sigma_y^2\mathbf{y}_2. \tag{4.55}$$

Thus, the nonzero eigenvalues of $\mathbf{G}$ occur in positive-negative pairs. Next, we show that the spectral radius of $\mathbf{G}$ is less than or equal to one (i.e., $\rho(\mathbf{G}) \leq 1$). Since $\tilde{\mathbf{R}}$ is positive definite, we know that [23:473]

$$\rho(\tilde{\mathbf{B}}^H\tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}}\tilde{\mathbf{C}}^{-1}) \leq 1. \tag{4.56}$$

Recall that the eigenvalues of the matrix $\mathbf{U}^H$ are the complex conjugate of the eigenvalues of $\mathbf{U}$ and thus, $\rho(\mathbf{U}) = \rho(\mathbf{U}^H)$. Now, notice that $\tilde{\mathbf{C}}^{-1}\tilde{\mathbf{B}}^H\tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}} = (\tilde{\mathbf{B}}^H\tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}}\tilde{\mathbf{C}}^{-1})^H$ and hence,

$$\rho(\tilde{\mathbf{C}}^{-1}\tilde{\mathbf{B}}^H\tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}}) \leq 1 \tag{4.57}$$

which implies that

$$\rho(\mathbf{G}) \leq 1. \tag{4.58}$$

Combining the results of Eqns. (4.51) and (4.58) and the observation that the eigenvalues of $\mathbf{G}$ occur in positive-negative pairs, we can bound the eigenvalues of $\mathbf{G}$ as

$$-1 < \sigma_i < 1 \quad \text{for all } i, \tag{4.59}$$

which implies that $\rho(\mathbf{G}) < 1$. Using this result, we can bound the eigenvalues of $\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{R}}$ as

$$0 < \lambda_i < 2 \quad \text{for all } i. \tag{4.60}$$

Using the observation that eigenvalues of $\mathbf{G}$ occur in positive-negative pairs and recalling that $\rho(\mathbf{G}) = \max\{|\sigma_i|\}$, we can write the largest and smallest eigenvalues of $\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{R}}$ as

$$\lambda_{max} = 1 + \rho(\mathbf{G}) \tag{4.61}$$

$$\lambda_{min} = 1 - \rho(\mathbf{G}) \tag{4.62}$$

Substituting Eqns. (4.61) and (4.62) into Eqn. (4.49) yields

$$\varepsilon \geq \frac{4\,(1-\rho(\mathbf{G}))\,(1+\rho(\mathbf{G}))}{(1-\rho(\mathbf{G})+1+\rho(\mathbf{G}))^2} = 1 - \rho(\mathbf{G})^2. \tag{4.63}$$

Recall the SINR efficiency of the Block STAP processor is defined relative to the optimum STAP processor and thus, larger values of ε indicate better performance with maximum efficiency (no SINR loss) occurring when $\varepsilon = 1$. From Eqn. (4.63), we observe that ε approaches one as $\rho(\mathbf{G})$ approaches zero. Thus, we could define a transformation selection criterion based on the spectral radius of $\mathbf{G}$ as follows: to maximize the lower bound of the Block STAP SINR efficiency relative to the optimum SINR, the transformation $\mathbf{V}$ should be selected to minimize the spectral radius of $\mathbf{G}$. However, a transformation selection criterion based strictly on the spectral radius of $\mathbf{G}$ has limited utility.

The problem with using Eqn. (4.63) to a establish a transformation selection criterion lies with Kantorovich's inequality which was used to set the bound in Eqn. (4.63). Referring to Eqn. (4.49), Kantorovich's inequality assumes the vector $\mathbf{x}$ can be any vector in the vector space and only achieves the bound with equality if $\mathbf{x} = c(\boldsymbol{\varphi}_{max} \pm \boldsymbol{\varphi}_{min})$, where $\boldsymbol{\varphi}_{max}$ and $\boldsymbol{\varphi}_{min}$ are the eigenvectors associated with the maximum and minimum eigenvalues, respectively, and c is an arbitrary constant [17]. Essentially , the bound in Eqn. (4.63) represents a worst-case analysis which only occurs if $\mathbf{x} = c(\boldsymbol{\varphi}_{max} \pm \boldsymbol{\varphi}_{min})$. If $\mathbf{x}$ never takes on this somewhat unique form, then the SINR efficiency will never reach the bound and may not even approach the bound as $\mathbf{x}$ varies. From Eqn. (4.49), observe that $\mathbf{D}$ is a diagonal matrix and thus, the standard ordered basis vectors ($\{\mathbf{e}_i\}_{i=1}^{MN}$) are eigenvectors of $\mathbf{D}$. Suppose the diagonal elements of $\mathbf{D}$ are ordered from largest to smallest, then $\mathbf{x}$ must equal $c(\mathbf{e}_1 \pm \mathbf{e}_{MN})$ to achieve the bound in Eqn. (4.49). The possibility of $\mathbf{x}$ having form $[c\ 0\ \cdots\ 0\ \pm c]^T$ appears to be extremely small (most likely, impossible)

given that $\mathbf{x} = \mathbf{T}^H\mathbf{V}^H\mathbf{s}$. Thus, the bound in Eqn. (4.63) is not useful for comparing the SINR performance of two suboptimal Block STAP processors. For example, consider the identity and DFT matrix configurations from the simulation example in Section 4.3. From Eqn. (4.63), the lower bound on the SINR efficiency for the identity matrix configuration is 3.05×10^{-6} versus 7.16×10^{-6} for the DFT matrix configuration, suggesting that the former configuration will have the worst performance. However, the opposite occurs - the lowest efficiency from the simulation is 5.76×10^{-2} for the identity matrix configuration and 3.86×10^{-4} for the DFT matrix configuration. In the next section, we discuss modifications to the spectral radius criterion to provide a better indication of actual performance.

Before we move into the next section, we address our earlier statement that reducing the norm of the error vector $\mathbf{e} = \mathbf{w}_{\text{blk}} - \mathbf{w}_{\text{opt}}$ has an influence on worst-case SINR performance. Observe from Eqn. (4.31) that the norm of $\mathbf{e}$ is minimized by minimizing $\|\tilde{\mathbf{Q}}^{-1}\mathbf{E}\|$, assuming that $\|\tilde{\mathbf{Q}}^{-1}\mathbf{E}\| < 1$. Now, observe that

$$\tilde{\mathbf{Q}}^{-1}\mathbf{E} = \begin{bmatrix} \tilde{\mathbf{A}}^{-1} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{C}}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{0} & \tilde{\mathbf{B}} \\ \tilde{\mathbf{B}}^H & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}} \\ \tilde{\mathbf{C}}^{-1}\tilde{\mathbf{B}}^H & \mathbf{0} \end{bmatrix} = \mathbf{G} \tag{4.64}$$

and recall that $\rho(\mathbf{G}) \leq \|\mathbf{G}\|$. Thus, the notion of reducing the norm of $\mathbf{e}$ by reducing $\|\tilde{\mathbf{Q}}^{-1}\mathbf{E}\|$ $(= \|\mathbf{G}\|)$ is consistent with reducing $\rho(\mathbf{G})$ to improve the worst-case SINR performance of the Block STAP processor given by Eqn. (4.63).

4.5.2 Criterion Discussion. In the previous section, we noted that a transformation selection criterion based solely on the spectral radius of $\mathbf{G}$ was not particularly useful, because the bound in Eqn. (4.63) represented worst-case SINR performance which would, in general, not occur. In this section, we discuss modifications to the spectral radius criterion to improve its utility. First, observe that since $\mathbf{D}$ is a diagonal matrix of the eigenvalues of $\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{R}}$, we can rewrite the Block STAP SINR efficiency given in Eqn. (4.48)

as

$$\varepsilon = \frac{\left(\sum_{p=1}^{MN} |x_p|^2\right)^2}{\left(\sum_{p=1}^{MN} \frac{|x_p|^2}{\lambda_p}\right)\left(\sum_{q=1}^{MN} \lambda_q |x_q|^2\right)} = \frac{\left(\sum_{p=1}^{MN} |x_p|^2\right)^2}{\sum_{p=1}^{MN}\sum_{q=1}^{MN} |x_p|^2 |x_q|^2 \frac{\lambda_q}{\lambda_p}}, \tag{4.65}$$

where $\{\lambda_p\}_{p=1}^{MN}$ are the eigenvalues of $\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{R}}$ and recall that $0 < \lambda_p < 2$ for all p. Now, observe that the efficiency depends on the ratio λ_q/λ_p and the elements of $\mathbf{x}$. If the transformation $\mathbf{V}$ block diagonalizes the correlation matrix $\mathbf{R}$, then $\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{R}} = \mathbf{I}$ and $\lambda_q/\lambda_p = 1$ for all p, q which yields an efficiency of one. If $\mathbf{V}$ does not block diagonalize $\mathbf{R}$, then $\lambda_q/\lambda_p \neq 1$ when $p \neq q$ and the efficiency will be less than one. In general, the efficiency will be low when the ratio λ_q/λ_p is large and $|x_p|^2|x_q|^2$ is not small, for at least one set of p and q. The ratio λ_q/λ_p is large only if λ_p is small since $0 < \lambda_q < 2$ for all q. The efficiency also depends on the component of $\mathbf{x}$ in the direction of the eigenvectors associated with small eigenvalues. Suppose λ_p is a small eigenvalue, then the efficiency will be low if x_p (i.e., p^{th} element of $\mathbf{x}$) is not small. Conversely, if x_p is small, then the small eigenvalue λ_p will not have a significant impact on the efficiency. Obviously, not all the components of $\mathbf{x}$ can be small and thus, if a large number of the eigenvalues of $\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{R}}$ are small, then the efficiency will be poor over a large portion of the angle-Doppler plane. Therefore, to keep the efficiency high, we want to minimize the number of eigenvalues of $\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{R}}$ that are small.

Recall that $\lambda_p = 1 + \sigma_p$, where $\{\sigma_p\}_{p=1}^{MN}$ are the eigenvalues of the matrix $\mathbf{G}$, and $-1 < \sigma_p < 1$ for all p. Also recall that eigenvalues of $\mathbf{G}$ occur in positive-negative pairs and hence, we can write $\sigma_p = -\sigma_{MN+1-p}$. Thus, at most, half the eigenvalues of $\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{R}}$ will be small. Now, observe that if $\rho(\mathbf{G}) \approx 1$, then at least one of the eigenvalues of $\mathbf{G}$ is close to one which implies that at least one of the eigenvalues of $\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{R}}$ is small. When the spectral radius of $\mathbf{G}$ is close to one, we essentially only have information about two eigenvalues of $\mathbf{G}$ and of $\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{R}}$. That is, when $\rho(\mathbf{G}) \approx 1$, it is possible that several of the eigenvalues of $\mathbf{G}$ are close to one and in which case, several of the eigenvalues of $\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{R}}$ are small, leading to poor efficiency. On the other hand, if $\rho(\mathbf{G}) \approx 0$, then we know that all

the eigenvalues of $\mathbf{G}$ are small and thus, all the eigenvalues of $\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{R}}$ are approximately one, leading to high efficiency. Now, suppose that the $MN \times MN$ matrix $\mathbf{G}$ has a rank of $r < MN$, then $MN-r$ eigenvalues of $\mathbf{G}$ are zero and $MN-r$ eigenvalues of $\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{R}}$ are one. Thus, one way to ensure that $\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{R}}$ only has a few small eigenvalues is to keep the rank of $\mathbf{G}$ low. The rank of $\mathbf{G}$ is determined by the rank of $\tilde{\mathbf{B}} = \mathbf{V}_1^H\mathbf{R}\mathbf{V}_2$, since $\tilde{\mathbf{A}}^{-1}$ and $\tilde{\mathbf{C}}^{-1}$ are full rank and square matrices. This suggests we should select the transformation $\mathbf{V}$ such that $\tilde{\mathbf{B}}$ has low rank (ideally, zero which implies that $\rho(\mathbf{G}) = 0$ and $\mathbf{V}$ block diagonalizes $\mathbf{R}$). However, restricting our attention to the rank of $\mathbf{G}$ ignores the fact that some of the non-zero eigenvalues of $\mathbf{G}$ could be approximately one and in which case, some of the eigenvalues of $\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{R}}$ would be small. When $\mathbf{x}$ has components in the directions of these small eigenvalues, the efficiency will be low. Clearly, our transformation selection criterion must consider both the spectral radius and rank of $\mathbf{G}$. Before proposing a transformation selection criterion, we present simulation examples to demonstrate the concepts discussed above.

Earlier, we compared the SINR efficiency of the identity, DFT, and centrosymmetric matrix configurations in an interference plus noise environment consisting of receiver noise, clutter, and three barrage noise jammers (See Table 4.1 and Figs. 4.1 and 4.2). We noted that the identity matrix configuration had significantly better performance than the other two configurations (See Table 4.2 for the average efficiency over the angle-Doppler plane for the three configurations). The spectral radius of $\mathbf{G}$ was approximately one of all three configurations yielding a lower bound for the efficiency of approximately zero (See Table 4.2). Thus, the lower bound (or equivalently, $\rho(\mathbf{G})$) is not useful for comparing the performance of these three configurations. As noted previously, we need to consider the rank of $\mathbf{G}$ in addition to the spectral radius of $\mathbf{G}$ to get a more complete picture of SINR performance. Plotted in Fig. 4.4 are the eigenvalues of $\mathbf{G}$ in ascending order for all three configurations. Referring to Fig. 4.4, we observe that the ordering of the configurations according to rank is the identity matrix configuration, then the centrosymmetric matrix configuration, and finally, the DFT matrix configuration. An examination of Table 4.2 reveals the same ordering among configurations in terms of the highest average efficiency (lowest average SINR loss). Basically, the identity matrix configuration performs the best

Parameter	Identity	DFT	Centrosymmetric
Average Efficiency	0.69	0.04	0.11
Average SINR Loss (dB)	1.6	13.6	9.8
Lower Bound of ε	3.05×10^{-6}	7.16×10^{-6}	5.56×10^{-5}
Minimum ε	5.76×10^{-2}	3.86×10^{-4}	1.04×10^{-4}
Maximum ε	0.96	0.89	0.95
$\|\mathbf{E}\|_2$	3.67×10^{6}	2.53×10^{6}	3.61×10^{5}
$\|\tilde{\mathbf{R}}^{-1} - \tilde{\mathbf{Q}}^{-1}\|_2$	2.47	5.07	2.51
Average of $\|\sigma_p\|$	0.2275	0.7995	0.6733

Table 4.2 Summary of simulation results for the identity, DFT, and centrosymmetric matrix configurations, where the interference plus noise environment consisted of receiver noise, clutter, and three barrage noise jammers.

because the possibility of $\mathbf{x}$ having a component associated with a small eigenvalue of $\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{R}}$ is less since there are fewer small eigenvalues.

Recall that $\mathbf{G}$ has r non-zero eigenvalues, where r is the rank of $\mathbf{G}$, and these non-zero eigenvalues of $\mathbf{G}$ could result in small eigenvalues of $\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{R}}$. We can gain insight into the rank of $\mathbf{G}$ and the rank ordering of the configurations by examining the transformed interference plus noise correlation matrix. Recall that the interference plus noise correlation matrix is the sum of the receiver noise, barrage noise jamming, and clutter correlation matrices. Also recall that the receiver noise correlation matrix is a scaled identity matrix and the jamming correlation matrix has a block diagonal form. Note that we do not have to consider the receiver noise correlation matrix since the transformations are unitary matrices. The approximate rank of the clutter correlation matrix is 19 based on Brennan's rule [41:29] and the rank of the jamming correlation matrix is 30 (note that the rank is invariant to a unitary transformation). As noted earlier, the rank of $\mathbf{G}$ depends on the rank of $\tilde{\mathbf{B}} = \mathbf{V}_1^H \mathbf{R} \mathbf{V}_2$ which is the off-diagonal block of the transformed interference plus noise correlation matrix. For the identity matrix configuration, we have $\tilde{\mathbf{B}} = \mathbf{B}$ which is just the off-diagonal block of the clutter correlation matrix since the receiver noise and jamming correlation matrices are block diagonal matrices before and after transformation. Thus, the rank of $\tilde{\mathbf{B}}$ is less than 19 which implies that the rank of $\mathbf{G}$ is less than 38. The clutter has constant power for all clutter patches and hence, its a centrosymmetric matrix. For the centrosymmetric matrix configuration, the transformed clutter correlation matrix

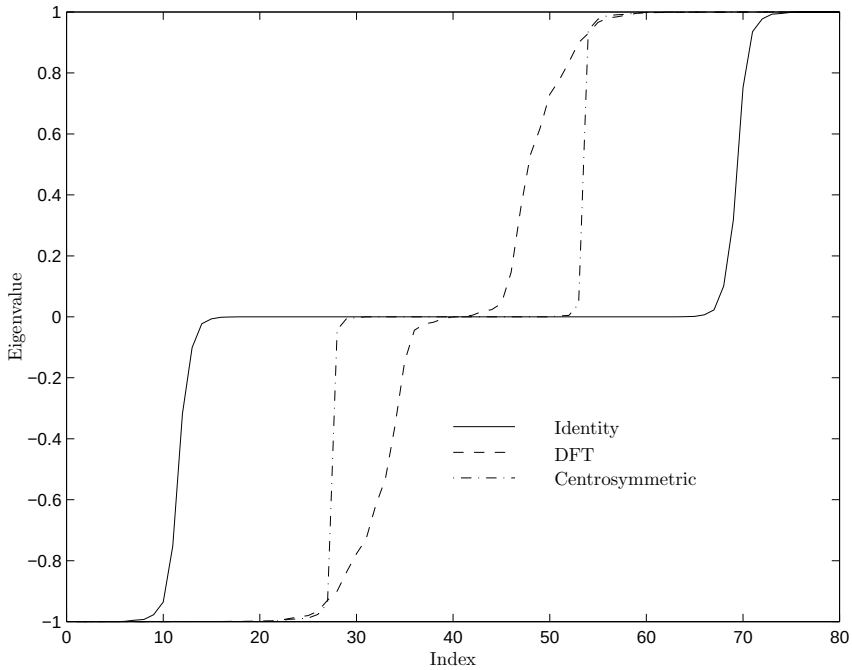


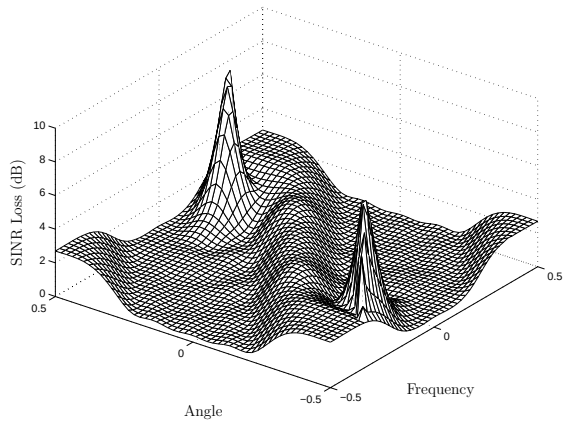
Figure 4.4 Eigenvalues of $\mathbf{G}$ sorted in ascending order for the identity, DFT, and centrosymmetric configurations, where the interference plus noise environment consisted of receiver noise, clutter, and three barrage noise jammers.

has a block diagonal form, but the transformation destroys the block diagonal form of the jamming correlation matrix. Thus, $\tilde{\mathbf{B}}$ is only a function of the jamming correlation matrix which has a rank of 30 and therefore, the rank of $\mathbf{G}$ is less than 60. For DFT matrix configuration, the block diagonal form of the jamming correlation matrix is destroyed and thus, $\tilde{\mathbf{B}}$ becomes a function of both the clutter and jamming. This mixing of the clutter and jamming by the DFT transformation accounts for the rank of $\mathbf{G}$ being higher than the other two configurations. Clearly, the transformation should not increase the rank of $\mathbf{G}$ unless it also significantly reduces the spectral radius of $\mathbf{G}$.

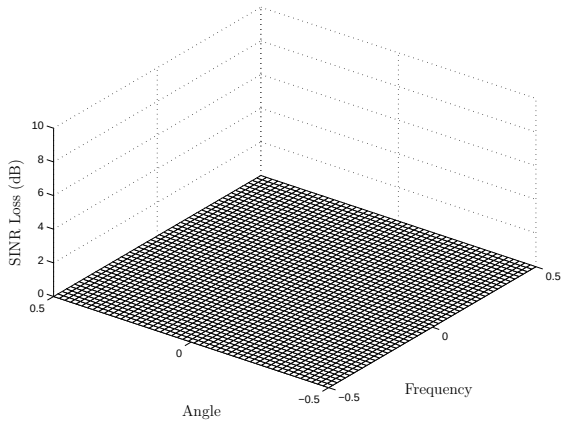
In this simulation example, the interference plus noise environment consists of clutter and receiver noise and we only consider the identity and centrosymmetric matrix configurations (note that the DFT matrix configuration increased the rank without a reduction in the spectral radius). The clutter is divided into two parts: symmetric and anti-symmetric. Recall that the clutter from the previous simulation example was constructed with all the clutter patches having the same power level, giving it a symmetric structure and a

centrosymmetric correlation matrix. Thus, the symmetric part is the same clutter from previous simulation example. The anti-symmetric part consisted of 11 clutter patches uniformly distributed between normalized angle of 0.13 and 0.21. We considered two cases for the anti-symmetric clutter: one with the anti-symmetric clutter patches having a 0 dB gain relative to the symmetric clutter patch power level and the other with the anti-symmetric clutter patches having a 20 dB gain relative to the symmetric clutter patch power level. The simulation results are summarized in Table 4.3. Figure 4.5 shows the SINR (loss) performance of the identity and centrosymmetric matrix configurations relative to the optimum processor. Table 4.3 and Fig. 4.5 show that the centrosymmetric matrix configuration had better performance than identity matrix configuration. An examination of Table 4.3 reveals that the lower bound on efficiency does, in this example, indicate which configuration will perform better, but only in the 0 dB case with the centrosymmetric matrix configuration does the lower bound approach the simulated efficiency. Plotted in Fig. 4.6 are the eigenvalues of $\mathbf{G}$, showing that $\mathbf{G}$ has a lower rank for the centrosymmetric matrix configuration. The better performance of the centrosymmetric matrix configuration is attributed to this lower rank. The centrosymmetric matrix configuration has a lower rank because the transformation block diagonalizes the symmetric clutter correlation matrix and thus, $\tilde{\mathbf{B}}$ is only function of the anti-symmetric clutter which only has a few clutter patches. In the case of the identity matrix configuration, $\tilde{\mathbf{B}}$ is function both the symmetric and anti-symmetric clutter. Again, this simulation example highlights the importance of selecting a transformation that decreases the rank of the matrix $\mathbf{G}$. This simulation example also suggest that the centrosymmetric block diagonalizing transformation is good candidate in non-jamming environments as noted earlier.

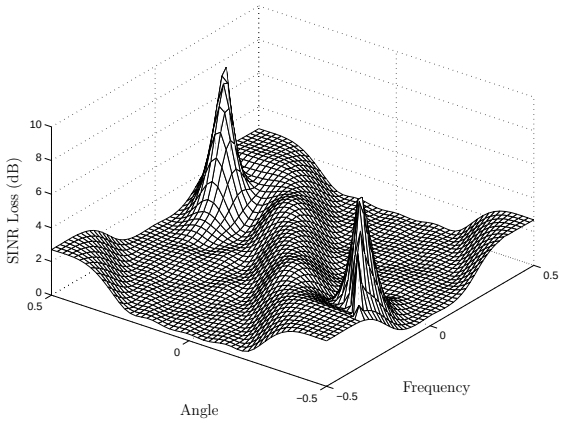
The previous discussion and simulation examples have highlighted that the Block STAP SINR efficiency depends both on the spectral radius and rank of $\mathbf{G}$. The efficiency is, in general, higher when these two factors are small and is one when either the spectral radius or rank is zero (one implies the other). Thus, the transformation selection criterion should consider the spectral radius and rank of $\mathbf{G}$ with the objective of minimizing both. One way to simultaneously consider both the spectral radius and rank of $\mathbf{G}$ is by averaging the magnitude of the eigenvalues of $\mathbf{G}$. This average is zero only if the spectral radius or



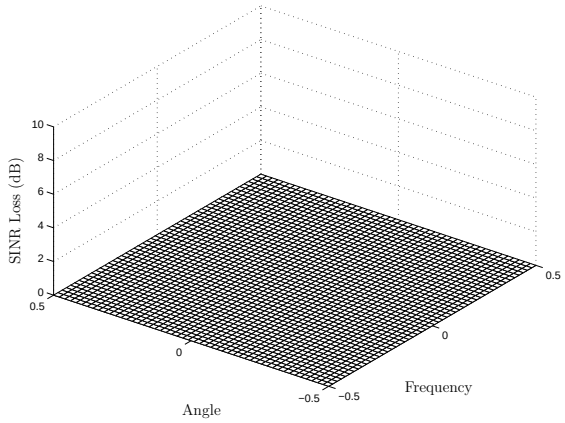
(a) Identity matrix configuration - 0 dB case



(b) Centrosymmetric matrix configuration - 0 dB case

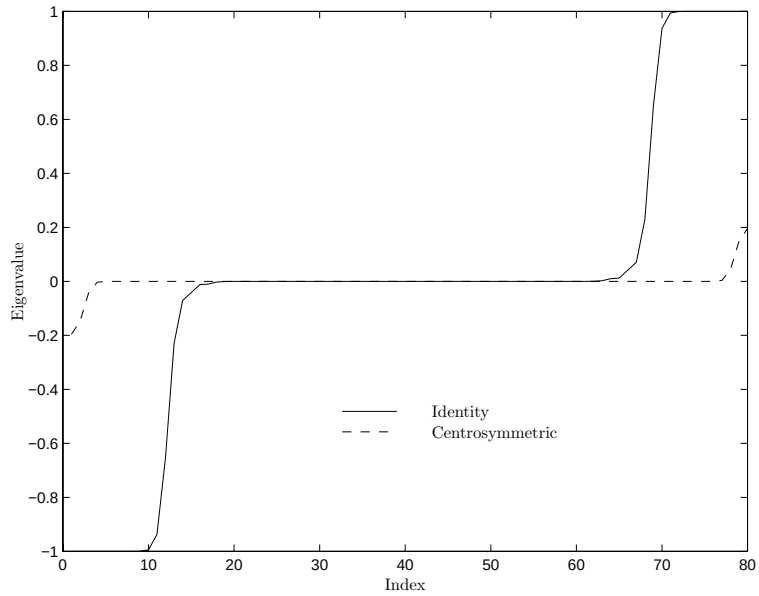


(c) Identity matrix configuration - 20 dB case

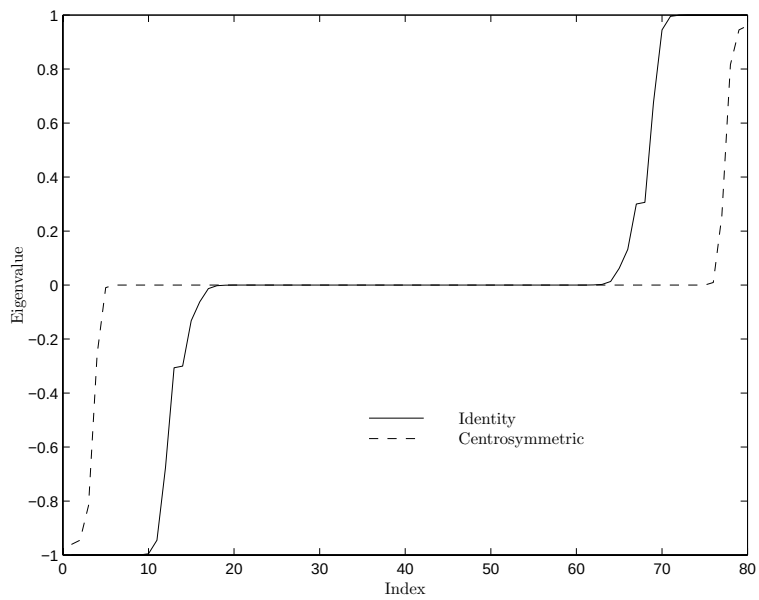


(d) Centrosymmetric matrix configuration - 20 dB case

Figure 4.5 SINR performance of the identity matrix and centrosymmetric matrix configurations relative to the optimum processor, where the interference plus noise environment consisted of receiver noise, symmetric clutter, and anti-symmetric clutter with a gain of either 0 dB or 20 dB above the symmetric clutter patches.



(a) 0 dB Case



(b) 20 dB Case

Figure 4.6 Eigenvalues of **G** sorted in ascending order for the identity and centrosymmetric configurations, where the interference plus noise environment consisted of receiver noise, symmetric clutter, and anti-symmetric clutter with a gain of either 0 dB or 20 dB above the symmetric clutter patches.

Parameter	Identity		Centrosymmetric	
Case	0 dB	20 dB	0 dB	20 dB
Average Efficiency	0.7490	0.7487	1	0.9993
Lower Bound of ε	1.4967×10^{-6}	1.3638×10^{-7}	9.6167×10^{-1}	7.7436×10^{-2}
Minimum ε	0.1345	0.1311	0.9807	0.9469
Maximum ε	0.9341	0.9338	1	1
Average of $\lvert\sigma_p\rvert$	2.9878×10^{-1}	3.1080×10^{-1}	9.773×10^{-3}	7.4690×10^{-2}

Table 4.3 Summary of simulation results for the identity and centrosymmetric matrix configurations, where the interference plus noise environment consisted of receiver noise, symmetric clutter, and anti-symmetric clutter with a gain of either 0 dB or 20 dB above the symmetric clutter patches.

rank of $\mathbf{G}$ is zero. In general, the average is small if the rank is low or if the spectral radius is small. The average of the magnitude of the eigenvalues of $\mathbf{G}$ from the simulation examples are listed on the last lines of Tables 4.2 and 4.3. Observe that the average provides a reliable indication of the configuration with the highest average efficiency. Thus, we propose the following transformation selection criterion:

Definition 4 (Proposed Block STAP Transformation Selection Criterion). *Let* $\mathbf{R}$ *be the positive definite and Hermitian interference plus noise correlation matrix,* $\mathbf{V}$ *be a unitary matrix, and* $\mathbf{s}$ *be the steering vector (desired signal). Define the SINR efficiency of the Block STAP processor relative to the optimum STAP processor as*

$$\varepsilon = \frac{\mathrm{SINR}_{\mathrm{blk}}}{\mathrm{SINR}_{\max}} = \frac{\left(\mathbf{z}^H \tilde{\mathbf{Q}}^{-1}\mathbf{z}\right)^2}{\mathbf{z}^H \tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{R}}\tilde{\mathbf{Q}}^{-1}\mathbf{z}} \; \frac{1}{\mathbf{z}^H \tilde{\mathbf{R}}^{-1}\mathbf{z}}, .$$

where

$$\tilde{\mathbf{R}} = \begin{bmatrix} \mathbf{V}_1^H\mathbf{R}\mathbf{V}_1 & \mathbf{V}_1^H\mathbf{R}\mathbf{V}_2 \\ \mathbf{V}_2^H\mathbf{R}\mathbf{V}_1 & \mathbf{V}_2^H\mathbf{R}\mathbf{V}_2 \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{A}} & \tilde{\mathbf{B}} \\ \tilde{\mathbf{B}}^H & \tilde{\mathbf{C}} \end{bmatrix},$$

$$\tilde{\mathbf{Q}} = \begin{bmatrix} \mathbf{V}_1^H\mathbf{R}\mathbf{V}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2^H\mathbf{R}\mathbf{V}_2 \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{A}} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{C}} \end{bmatrix},$$

and $\mathbf{z} = \mathbf{V}^H\mathbf{s}$. *Let*

$$\mathbf{G} = \begin{bmatrix} \mathbf{0} & \tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}} \\ \tilde{\mathbf{C}}^{-1}\tilde{\mathbf{B}}^H & \mathbf{0} \end{bmatrix}.$$

To maximize the average SINR efficiency (minimize the SINR loss) of the Block STAP processor relative to the optimum STAP processor, the transformation $\mathbf{V}$ *should be selected to minimize the average of the magnitude of the eigenvalues of* $\mathbf{G}$.

Although our analysis and simulation examples support the proposed transformation selection criterion, a rigorous analysis and proof this criterion remains as an open research area.

4.6 Summary

Without a block diagonalizing transformation, the Block STAP processor is not equivalent to the optimum STAP processor and as a result, a Block STAP processor based on non-block diagonalizing transformation will experience a loss in SINR performance. In this chapter, we addressed the problem of how to select a non-block diagonalizing transformation to minimize or reduce the loss SINR performance. We first noted that a suboptimal Block STAP processor has the potential to perform worse than one of its channels if the transformation is not selected properly. We showed that a suboptimal Block STAP process based on the identity matrix has the desirable property of providing SINR performance greater than or equal to either channel. Then, we discussed two heuristic approaches to developing a transformation selection criterion. The first approach was to approximately block diagonalize the correlation matrix by reducing the l_2-norm of the off-diagonal blocks. The second approach was to reduce the l_2-norm of the difference between the optimum and the equivalent full dimension Block STAP weight vectors. Using simulation examples, we showed that these two approaches did not provide a reliable indication of SINR performance. Next, we analyzed the SINR efficiency of the Block STAP processor relative to the optimum processor and provided a lower bound for the efficiency. The utility of a transformation selection criterion based solely on this lower bound is limited, because the bound

is not obtained, in general. We discussed and demonstrated through simulation examples that the transformation selection criterion should combine the condition for improving the lower bound with additional information (i.e., spectral radius and rank of $\mathbf{G}$). Finally, we proposed a transformation selection criterion based on this combined information that provides a reliable indication of SINR performance for the simulation examples presented. The rigorous analysis and proof of this criterion is a future area of research.

V. Reduced-Rank Direct Form Transformation Selection

5.1 Introduction

In this chapter, we depart from the dual channel concept and address the problem of selecting the optimal rank reduction transformation for a data adaptive, reduced-rank direct form processor. We introduce the SINR metric and propose a reduced-rank direct form processor based on this SINR metric, which we refer to as the SINR metric method. The SINR metric is a quantifier that identifies the eigenvectors of the interference plus noise correlation matrix having the greatest impact on the output SINR. If the rank reduction transformation is constructed from r eigenvectors of the correlation matrix, then the r eigenvectors with the largest SINR metric are the optimal set of r eigenvectors in terms of SINR performance. That is, any other set of r eigenvectors will yield an output SINR less than the r eigenvectors with the largest SINR metric. The SINR metric and the reduced-rank SINR metric method are the direct form analogs to the cross-spectral metric (CSM) and the reduced-rank CSM generalized sidelobe canceler introduced by Goldstein and Reed [11–14]. Both the SINR metric direct form processor and the CSM generalized sidelobe canceler exhibit a graceful degradation in SINR performance as the transformation rank is reduced below full dimension. In contrast, principal component methods, such as the principal component inverse (PCI) [25] and minimum norm eigencanceler [18], can exhibit a sharp decrease in performance if the number of principal components is underestimated [14]. Although ordering the eigenvectors according to the largest eigenvalues (principal components) yields the best low-rank approximation to the full dimension correlation matrix [35:45-46], the principal component ordering does not consider the output SINR and as such, does not yield the maximum output SINR. Basically, the SINR metric and CSM methods use the output SINR as a cost function in the transformation selection process which yields a smooth decrease in SINR performance as the transformation rank is reduced.

The remainder of this chapter is organized as follows. In Section 5.2, we briefly review the cross-spectral metric generalized sidelobe canceler. Note that in Appendix A, we show that the CSM for each of the noise subspace eigenvectors is zero which implies the

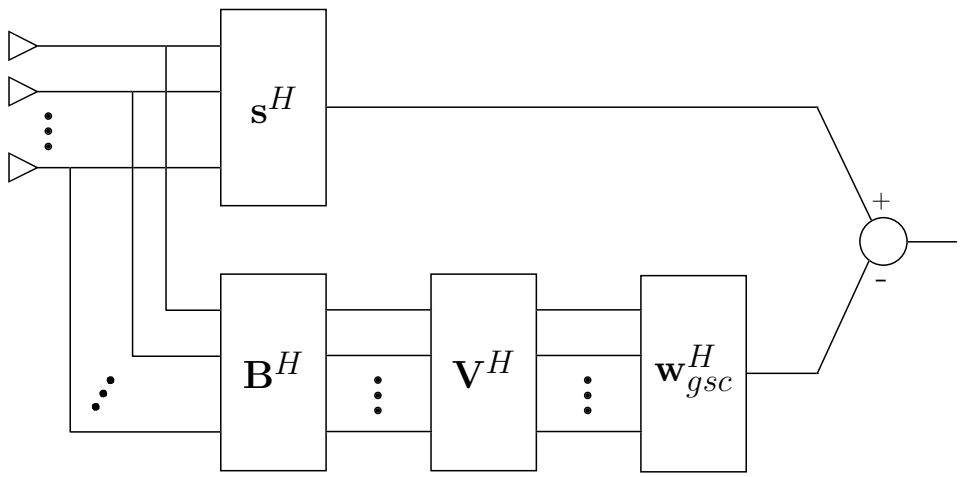


Figure 5.1 Block diagram of a reduced-dimension (rank) generalized sidelobe canceler.

optimum generalized sidelobe canceler weight vector lies in the interference subspace. The development of the SINR metric method is presented in Section 5.3. In Section 5.4, we present simulation results that highlight the importance of incorporating a cost function in the transformation selection process and demonstrate that the performance of the SINR metric and CSM methods depend on the interference plus noise environment and available resources. In Section 5.5, we discuss the limitations of the SINR metric and CSM methods in practical applications and the relevance of these limitations to the Block STAP method. We summarize the chapter in Section 5.6.

5.2 Cross-Spectral Metric Generalized Sidelobe Canceler

The basic implementation structure of the reduced-rank generalized sidelobe canceler (GSC) is shown in Fig. 5.1. The upper branch forces the GSC to have a response in the spatial and Doppler directions defined by the steering vector, $\mathbf{s}$. The lower branch provides an estimate of the noise in the upper branch and the final processing step of subtracting the lower branch from the upper branch reduces the output noise level. In the lower branch, the $MN - 1 \times MN$ matrix $\mathbf{B}$, referred to as the blocking matrix, annihilates the steering vector (i.e., $\mathbf{Bs} = \mathbf{0}$) and has full rank. The blocking matrix prevents the cancellation of signals received in the spatial and Doppler directions defined by the steering vector. The $MN - 1 \times r$ matrix $\mathbf{V}$ is the rank reduction transformation and the weight vector $\mathbf{w}_{gsc}$ is

an unconstrained Wiener filter. Without the rank reduction transformation matrix $\mathbf{V}$, the weight vector for the Wiener filter is given as [14]

$$\mathbf{w}_{gsc} = \mathbf{R}_b^{-1}\mathbf{r}_{bd} \tag{5.1}$$

where $\mathbf{r}_{bd} = \mathbf{BRs}$, $\mathbf{R}_b = \mathbf{BRB}^H$, and $\mathbf{R}$ is the interference plus noise correlation matrix. The output SINR from the GSC shown in Fig. 5.1 (ignoring $\mathbf{V}$) is [14]

$$\mathrm{SINR}_{\mathrm{gsc}} = \frac{|\alpha|^2}{\sigma_d^2 - \mathbf{r}_{bd}^H\mathbf{R}_b^{-1}\mathbf{r}_{bd}} = \frac{|\alpha|^2}{\sigma_d^2 - \sum_{i=1}^{L}\frac{|\mathbf{u}_i^H\mathbf{r}_{bd}|^2}{\lambda_i}}, \tag{5.2}$$

where $\sigma_d^2 = \mathbf{s}^H\mathbf{Rs}$, $|\alpha|^2$ is the power of the desired signal, $\{\mathbf{u}_i\}_{i=1}^L$ are the eigenvectors of $\mathbf{R}_b$, $\{\lambda_i\}_{i=1}^L$ are the associated eigenvalues, and $L = MN - 1$. In Eqn. (5.2), the term

$$\frac{|\mathbf{u}_i^H\mathbf{r}_{bd}|^2}{\lambda_i} \tag{5.3}$$

is the so-called cross-spectral metric (CSM) [14]. Since the interference plus noise correlation matrix is positive definite, the cross-spectral metric is non-negative.

Recall the interference plus noise correlation matrix can be written as the sum of the interference (correlated noise) correlation matrix and the receiver noise (uncorrelated noise) correlation matrix. The interference correlation matrix is the sum of the clutter and barrage noise jamming correlation matrices and in general, is not full rank. As noted earlier, the eigenvectors of the interference correlation matrix are also the eigenvectors of the interference plus noise correlation matrix, since the correlation matrix of the receiver noise is a scaled identity matrix. The eigenvectors of the interference correlation associated with the non-zero eigenvalues form the interference subspace and the remaining eigenvectors form the noise subspace. Note that these two subspaces are defined relative to the interference plus noise correlation matrix $\mathbf{R}_b$ in the case of the GSC. One can show that the CSM for each of the noise subspace eigenvectors is zero which implies the optimum GSC weight vector lies in the interference subspace (See Appendix A for a proof). Thus, we can exploit the low rank nature of the interference correlation matrix to reduce the dimension

of the weight vector $\mathbf{w}_{gsc}$. That is, if the rank reduction transformation $\mathbf{V}$ spans the interference subspace, then the reduced-rank GSC provides the same SINR performance as the full dimension GSC. Now, since the optimum GSC weight vector lies in the interference subspace, we can rewrite Eqn. (5.2) as

$$\mathrm{SINR}_{\mathrm{gsc}} = \frac{|\alpha|^2}{\sigma_d^2 - \sum_{i=1}^{P} \frac{|\mathbf{u}_i^H \mathbf{r}_{bd}|^2}{\lambda_i}}, \tag{5.4}$$

where P is the rank of the interference subspace, $\{\mathbf{u}_i\}_{i=1}^{P}$ are the eigenvectors that span the interference subspace, and $\{\lambda_i\}_{i=1}^{P}$ are the associated eigenvalues.

Suppose we want to reduce the rank of the transformation $\mathbf{V}$ (or equivalently, the dimension of the weight vector) below the rank of the interference subspace, say to rank r. We are now faced with the problem of how to select the r columns of $\mathbf{V}$ to minimize the loss in SINR performance. Observe that if the r columns of $\mathbf{V}$ are a subset of the interference subspace eigenvectors, then the summation in Eqn.(5.4) will only include the terms associated with the r eigenvectors used in $\mathbf{V}$. Thus, the output SINR of an eigen-based reduced-rank GSC is given as

$$\mathrm{SINR}_{\mathrm{RRgsc}} = \frac{|\alpha|^2}{\sigma_d^2 - \sum_{i=1}^{r} \frac{|\mathbf{u}_i^H \mathbf{r}_{bd}|^2}{\lambda_i}}, \tag{5.5}$$

where $\{\mathbf{u}_i\}_{i=1}^{r}$ are the columns of $\mathbf{V}$ and $\{\lambda_i\}_{i=1}^{r}$ are the associated eigenvalues. The partial sum given in Eqn. (5.5) will obviously be less than the summation over all P interference subspace eigenvectors and thus, the output SINR of the reduced-rank GSC with $r < P$ will be less than the full dimension GSC. The objective is to select the eigenvectors which minimize the reduction in SINR performance. Clearly, the partial sum in Eqn. (5.5) is maximized by selecting the r eigenvectors with the largest CSM which, in turn, minimizes the loss in SINR performance. Thus, when the rank reduction transformation $\mathbf{V}$ of a GSC is constructed from r eigenvectors of the interference plus noise correlation matrix, the SINR is maximized by selecting the r eigenvectors with the largest CSM.

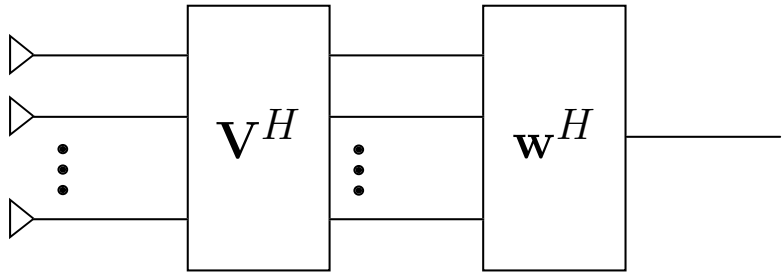


Figure 5.2 Block diagram of a reduced-dimension (rank) direct form processor.

5.3 SINR Metric Direct Form Processor

The basic implementation structure for the reduced-rank direct form processor is shown in Fig. 5.2, where the $MN \times r$ matrix $\mathbf{V}$ is the rank reduction transformation and $\mathbf{w}$ is a $r \times 1$ weight vector designed to maximize the SINR. The full dimension (i.e., without $\mathbf{V}$) weight vector of the direct form processor is given by

$$\mathbf{w} = \mathbf{w}_{\text{dfp}} = \mathbf{R}^{-1}\mathbf{s}, \tag{5.6}$$

and the reduced-rank weight vector by

$$\mathbf{w} = \mathbf{w}_{\text{RRdfp}} = (\mathbf{V}^H\mathbf{R}\mathbf{V})^{-1}\mathbf{V}^H\mathbf{s}, \tag{5.7}$$

where $\mathbf{R}$ is the interference plus noise correlation matrix and $\mathbf{s}$ is the steering vector. As with the reduced-rank GSC, the SINR performance of the reduced-rank direct form processor will be less than the full dimension direct form processor. Thus, the objective is to select the r columns of $\mathbf{V}$ such that the loss in SINR performance is minimized. Inspired by Goldstein and Reed's CSM method, we propose a method where the r columns of $\mathbf{V}$ are eigenvectors of $\mathbf{R}$ and are selected based on their impact on the output SINR from a direct form perspective. That is, we are using the output SINR as a cost function to identify the r eigenvectors (columns of $\mathbf{V}$) that minimize the loss in SINR performance.

Without the rank reduction transformations, the GSC shown in Fig. 5.1 provides the same output SINR as the direct form processor shown in Fig. 5.2 when the weight vectors $\mathbf{w}_{gsc}$ and $\mathbf{w}_{\text{dfp}}$ are defined by Eqns. (5.1) and (5.6), respectively. However, the

SINR equations for the full dimension GSC and direct form processor are different. The SINR for the full dimension direct form processor is

$$\mathrm{SINR}_{\mathrm{dfp}} = |\alpha|^2 \mathbf{s}^H \mathbf{R}^{-1} \mathbf{s} = |\alpha|^2 \sum_{i=1}^{MN} \frac{|\mathbf{f}_i^H \mathbf{s}|^2}{\tilde{\lambda}_i}, \tag{5.8}$$

where $\{\mathbf{f}_i\}_{i=1}^{MN}$ are the eigenvectors of $\mathbf{R}$ and $\{\tilde{\lambda}_i\}_{i=1}^{MN}$ are the associated eigenvalues. In Eqn. (5.8), the term

$$\frac{|\mathbf{f}_i^H \mathbf{s}|^2}{\tilde{\lambda}_i}, \tag{5.9}$$

is referred to as the SINR metric. Since $\mathbf{R}$ is positive definite, the SINR metric is always greater than zero. Thus, unlike the GSC, we cannot state that the optimum weight vector lies in the interference subspace (or in the noise subspace).

With the reduced-rank weight vector defined as in Eqn. (5.7), the output SINR for the reduced-rank direct form processor is given by [19]

$$\mathrm{SINR}_{\mathrm{RRdfp}} = |\alpha|^2 \mathbf{s}^H \mathbf{V} \left(\mathbf{V}^H \mathbf{R} \mathbf{V}\right)^{-1} \mathbf{V}^H \mathbf{s}. \tag{5.10}$$

Now, if we restricted the r columns of $\mathbf{V}$ to be a unique subset of the eigenvectors of $\mathbf{R}$, then we can rewrite Eqn. (5.10) as

$$\mathrm{SINR}_{\mathrm{RRdfp}} = |\alpha|^2 \mathbf{s}^H \mathbf{V} \tilde{\mathbf{\Lambda}}^{-1} \mathbf{V}^H \mathbf{s} = |\alpha|^2 \sum_{i=1}^{r} \frac{|\mathbf{f}_i^H \mathbf{s}|^2}{\tilde{\lambda}_i}, \tag{5.11}$$

where $\tilde{\mathbf{\Lambda}}$ is a diagonal matrix of the eigenvalues associated with the r eigenvectors ($\{\mathbf{f}_i\}_{i=1}^{r}$) that compose the columns of $\mathbf{V}$. The output SINR of the reduced-rank direct form processor given by Eqn. (5.11) is only a partial sum of the output SINR for the fully adaptive processor given by Eqn. (5.8). Thus, in general, the reduced-rank direct form processor will incur a loss in SINR performance. The objective is to select the columns of $\mathbf{V}$ as the eigenvectors of $\mathbf{R}$ that minimize the loss in SINR performance, which is equivalent to maximizing the partial sum given in Eqn. (5.11). Clearly, the partial sum is maximized by selecting the r columns of $\mathbf{V}$ to be the eigenvectors with the largest SINR metric. Thus,

with the SINR metric method, the r columns of $\mathbf{V}$ are the r eigenvectors of $\mathbf{R}$ with the highest SINR metric and as a result, yields the optimum eigen-based rank r direct form processor in terms of SINR performance.

The eigenvectors $\{\mathbf{f}_i\}_{i=1}^{MN}$ form an orthonormal basis for the signal space and thus, the steering vector and the interference plus noise vector can be written as a linear combination of the eigenvectors. The SINR metric given by Eqn. (5.9) basically represents the SINR along each of the basis vectors. We denote the subspace spanned by the r eigenvectors with the highest SINR metric as the high SINR subspace and the subspace spanned by the remaining eigenvectors as the low SINR subspace. Now, note that the high SINR subspace is orthogonal to the low SINR subspace. If we select the r columns of $\mathbf{V}$ to be the eigenvectors with the largest SINR metric, then the weight vector defined by Eqn. (5.7) lies in the high SINR subspace and thus, cancels any signal components (eigenvectors) in the low SINR subspace.

In summary, both the CSM and SINR metric methods introduce the output SINR as a cost function into the process of selecting the rank reduction transformations. The CSM and SINR metric provide an ordering of the eigenvectors of $\mathbf{BRB}^H$ and $\mathbf{R}$, respectively, based on their relative impact on the output SINR. In contrast, ordering the eigenvectors according to the principal components (i.e largest eigenvalues) is not directly related to the output SINR. The principal component (PC) ordering provides the best low rank approximation to the full dimension matrix as noted earlier. One can show that the CSM GSC and PC GSC provide the same SINR performance when the rank of the transformation exceeds the dimension of the interference subspace. However, as Goldstein and Reed have shown, the best low rank approximation does not translate into maximizing the output SINR when the rank is below the dimension of the interference subspace. One can also show that the full dimension direct form processor and GSC have the same SINR performance, but as the rank is reduced below full dimension, a direct analytical comparison of the CSM and SINR metric methods becomes a difficult task. The SINR performance of the CSM and SINR metric methods as a function of the transformation rank r (i.e., the number of columns in the rank reduction transformation) is determined by their respective SINR

equations:

$$\mathrm{SINR}_{\mathrm{gsc}}(r) = \frac{|\alpha|^2}{\sigma_d^2 - \sum_{i=1}^{r} \frac{|\mathbf{u}_i^H \mathbf{r}_{bd}|^2}{\lambda_i}} \qquad \text{for } 1 \leq r \leq MN-1, \tag{5.12}$$

$$\mathrm{SINR}_{\mathrm{dfp}}(r) = |\alpha|^2 \sum_{i=1}^{r} \frac{|\mathbf{f}_i^H \mathbf{s}|^2}{\tilde{\lambda}_i} \qquad \text{for } 1 \leq r \leq MN, \tag{5.13}$$

where r is the number of columns in the rank-reduction transformation. Basically, the relative performance of the CSM and SINR metric methods will depend on the rate that Eqns. (5.12) and (5.13) increase as a function of r. The direct relationship between Eqns. (5.12) and (5.13) as well as the behavior of $\mathbf{u}_i^H \mathbf{r}_{bd}$ and $\mathbf{f}_i^H \mathbf{s}$ as a function of the steering vector and the interference plus noise environment remain as open research areas. In the next section, we present simulation results that show the SINR performance of the CSM GSC and SINR metric direct form processor in several scenarios. Our intent in showing the simulation results is not to suggest that one method is better than the other, but to show the importance of incorporating a cost function in the process of selecting the rank reduction transformation. Additionally, the simulation results highlight that the SINR performance of each method is dependent on the interference plus noise environment and the available resources (i.e., transformation rank).

5.4 Simulation Results

In this section, we examine the SINR performance as a function of the number of eigenvectors used in the rank reduction transformation for the SINR metric method, the CSM GSC method, and a principal component (PC) version of the GSC processor. With the PC GSC, the columns of $\mathbf{V}$ are filled with the eigenvectors associated with the largest eigenvalues of $\mathbf{BRB}^H$. We also present simulation results from a hypothetical direct form processor where the columns of $\mathbf{V}$ are filled with the eigenvectors associated with the smallest eigenvalues, which we refer to as the small method. Recall that each eigenvalue gives an indication of the interference plus noise power level along its associated eigenvector. The small method represents the heuristic reasoning that one should select

Case	Target			Jammer 1	Jammer 2	Jammer 3	Clutter
	Angle	Doppler	SNR	JNR	JNR	JNR	CNR
1	0.0	0.0750	15 dB	40 dB	30 dB	30 dB	55 dB
2	0.0	0.0375	15 dB	40 dB	30 dB	30 dB	55 dB
3	0.0	0.0125	15 dB	40 dB	30 dB	30 dB	55 dB
4	0.0	0.0750	15 dB	0 dB	0 dB	0 dB	20 dB
5	0.0	0.0375	15 dB	0 dB	0 dB	0 dB	20 dB
6	0.0	0.0125	15 dB	0 dB	0 dB	0 dB	20 dB

Table 5.1 Simulation parameters, where SNR, JNR, and CNR denote the signal-to-receiver noise, jammer-to-receiver noise, and total clutter-to-receiver noise ratios, respectively.

the eigenvectors with the smallest eigenvalues because the interference plus noise power level along these eigenvectors is the smallest. Basically, the small method forces the weight vector $\mathbf{w}$ to lie in a subspace orthogonal to the eigenvectors with higher interference plus noise power levels and thus, leading to their cancellation. The simulated radar had a linear array of 14 antenna elements spaced at half a wavelength with 16 pulses in a coherent processing interval. The interference environment consisted of three barrage noise jammers and clutter. The three jammers were at normalized angles of 0.25, 0.433, and -0.433, and are referred to as Jammer 1, Jammer 2, and Jammer 3, respectively. The clutter was simulated by 360 point scatters evenly distributed in azimuth and with the radar parameters selected such that $\beta = 1$, where β defines the slope of the clutter ridge across the normalized angle-Doppler plane (See Ref. [41:24-28] for a complete discussion of β). The simulation results from six cases (scenarios) are presented with the relevant parameters listed in Table 5.1. Cases 1-3 represent a high clutter and jamming environment and Cases 4-6 represent a low clutter and jamming environment.

Figures 5.3 and 5.4 show the SINR performance of the four methods as a function of the number of columns (transformation rank) in the rank reduction transformation for Cases 1 and 4, respectively. Plotted in Figs. 5.3 and 5.4 are Eqn. (5.12) with the summation ordered by the largest CSM and by the largest (PC) eigenvalues and Eqn. (5.13) with the summation ordered by the largest SINR metric and the smallest eigenvalues. An examination of the plots in Figs. 5.3 and 5.4 reveals that the SINR metric method outperforms the small method and the CSM GSC method outperforms the PC GSC method

as the transformation rank is reduced below full dimension, attesting to the importance of incorporating a cost function into the process of selecting the rank reduction transformation. Similar plots for the other cases show the same characteristics in regards to the difference in SINR performance of the SINR metric method and CSM GSC method with their respective counterparts. Note that additional comments on the small method are provided at the end of this section.

A further examination of Figs. 5.3 and 5.4 reveals that above a certain rank the CSM GSC method outperforms the SINR metric method and below this rank, the converse is true. As shown in Figs. 5.5 and 5.6, the SINR performance curve crossover trend holds for all the cases in Table 5.1. Figure 5.5 shows the performance curves (SINR vs. transformation rank) for SINR metric and CSM GSC methods in the high clutter and jamming environment (Case 1-3). Figure 5.6 shows the performance curves for the low clutter and jamming environment (Cases 4-6). Observe from Figs. 5.5 and 5.6 that the crossover rank for the SINR performance curves for the SINR metric and CSM GSC methods varies as the interference plus noise environment and steering vector change. The determination of the crossover rank remains an open research area and its resolution will depend on resolving the relationship between Eqns. (5.12) and (5.13). In general, we expect the crossover rank to move to the left as the dimension of the interference subspace decreases, since the reduced-rank GSC achieves optimal SINR performance when the rank of the transformation matches the dimension of the interference subspace. However, by the same token, we expect the crossover rank to move to the right as the dimension of the interference subspace increases. Figures 5.5 and 5.6 highlight that the SINR performance of the two given methods, and most likely any STAP method, is dependent on the scenario as well as the transformation rank. Thus, one can not in general claim that one method is 'better' than another without specifying the scenario and the transformation rank. We now examine Cases 1 and 4 in more detail to gain additional insight into the difference between the SINR metric and CSM GSC methods, starting with Case 1.

First, observe that the interference plus noise correlation matrix has a rank of approximately 77 and as noted previously, the PC method displays a sharp decrease in performance as the transformation rank decreases below 77 (See Fig.5.3). Both the CSM

and SINR metric methods display a more graceful degradation in performance as the transformation rank decreased below 77. The SINR metric method has better performance as the transformation is decreased below 57. The eigenvalues of $\mathbf{R}$ and the SINR metric for Case 1 are plotted in Fig. 5.7. Figure 5.7 reveals that the largest SINR metrics occur in the noise subspace (i.e., the eigenvectors associated with the smallest eigenvalues) which is orthogonal to the interference subspace. Thus, even as the rank of the transformation approaches 1, the SINR metric method will provide a weight vector that lies in a subspace orthogonal to the interference subspace which effectively cancels all the interference. The SINR performance of CSM and PC GSC is significantly less than the SINR metric method at a transformation rank of 1 because the GSC can only cancel a single interference component leaving approximately 76 interference components uncancelled. The eigenvalues of $\mathbf{R}_b$ and the CSM are plotted in Fig. 5.8. Figure 5.8 shows that the largest eigenvalues do not necessarily correspond to the largest CSM. The SINR performance of the PC GSC is less than the CSM GSC because the sum in Eqn. (5.12) will not be maximized when the eigenvectors associated with largest eigenvalues are used. The interference plus noise power after upper branch processing, σ_d^2, is also plotted in Fig. 5.8.

The difference between Case 4 and Case 1 is the reduction of the CNR from 55 dB to 20 dB and JNR for each jammer 0 dB. A comparison of Figs. 5.3 and 5.4 reveals that the SINR metric method has the same basic performance in both cases, while the performance of the CSM and PC GSC is better in Case 4 than in Case 1. An examination of the SINR and CSM metric plots for Case 4 (See Figs. 5.9 and 5.10) provides insight into explaining these observations. As Fig. 5.9 shows, a large percentage of the highest SINR metrics occur in the noise subspace and in fact, the largest SINR metric occurs in the noise subspace. Thus, as the transformation rank approaches 1, the weight vector $\mathbf{w}$ lies in a subspace orthogonal to the interference subspace providing complete cancellation of the interference. Since the power in the noise subspace did not change from Case 1 to Case 4 and most of the largest SINR metrics occur in the noise subspace, the performance of the SINR metric method is nearly identical in both cases. The eigenvalues of $\mathbf{R}_b$, the CSM, and σ_d^2 are plotted in Fig. 5.10. A comparison of Figs. 5.8 and 5.10 reveals that σ_d^2 and the CSM are approximately 35 dB less in Case 4 than in Case 1. The improved

performance of the GSC methods can be attributed to this 35 dB difference. Since the SINR at a transformation rank of 1 is starting approximately 35 dB higher in Case 4, the addition of each CSM to the summation in Eqn. 5.12 causes a greater change in the SINR performance in Case 4 than in Case 1.

The nearly identical SINR performance of the SINR metric method between Case 1 and Case 4 is also present between Cases 2 and 5 and Cases 3 and 6 (See Figs. 5.5 and 5.6). This similarity in SINR performance suggest that the SINR metric method is nearly invariant to changes in the interference power level, which is not true for the CSM GSC method. We also expect the SINR performance of the SINR metric method to be relatively invariant to changes in the dimension of the interference subspace. That is, the basic shape of the SINR metric method performance curves will remain constant as the dimension of the interference subspace changes, but will shift up or down by an amount consistent with the change in the full dimension SINR performance. As the interference environment changes, the dimension of the interference and noise subspaces will change. However, the low rank nature of the interference plus noise correlation matrix should ensure that the eigenvalues of the noise subspace remain relatively constant, since these eigenvalues are essentially determined by the receiver noise power which should be constant for a given radar system. Recall that the numerator of the SINR metric is the projection of the steering vector onto a particular eigenvector. Thus, as long as the projections of the steering vector on the noise subspace do not change significantly as the interference environment changes, then the SINR metric method should be relatively invariant to the interference changes. The above argument assumes that the largest SINR metric occurs in the noise subspace which is not guaranteed.

Recall that as the steering (target) vector approaches the interference, the SINR performance of a STAP process will decrease. Now, notice that in Cases 1-3 and Cases 4-6 the steering (target) vector is moving towards the clutter ridge which passes through the zero Doppler and zero angle point of the normalized angle-Doppler plane since $\beta = 1$. Thus, Figs. 5.5 and 5.6 provide an indication of the SINR performance of the SINR metric and CSM GSC methods as the steering vector approaches the interference. An examination of Figs. 5.5 and 5.6 reveals the expected decrease in SINR performance, but also reveals

that the basic shape of the SINR performance curves of the two methods do not change significantly as the steering vector approaches the interference.

In Case 1 (high clutter and jamming), we observe that the small method provides better performance than the CSM method below a rank of approximately 53 and approximately 3 to 4 dB less than the SINR metric method (See Fig. 5.3). As noted earlier, when the rank decreases below 77, the CSM method does not have sufficient degrees of freedom to cancel all the interference components. In contrast, the small method forces the weight vector to lie in the noise subspace effectively canceling all the interference components. An examination of Fig. 5.7 reveals a wide separation in the SINR metric values associated with the noise and interference subspaces. Thus, the SINR is primarily determined by the noise subspace SINR metric values and leads the SINR metric method to compute a weight vector that also lies in the noise subspace. Therefore, the small and SINR metric methods provide similar performance, with the difference attributed to the ordering of the eigenvectors. Note that this similarity does not carry over to Case 4 (lower clutter and jamming, See Fig. 5.4), since the separation in the SINR metric values associated with the noise and interference subspace is decreased (See Fig. 5.9). Thus, the relative weight of each noise subspace SINR metric is decreased and the SINR builds up at a slower rate as a function of rank for the small method.

5.5 Limitations and Relevance to Block STAP

Although the CSM and SINR metric methods provide a graceful degradation in performance as the rank is reduced, the computational cost of both of these methods (and in general, any eigen-based method) is an issue due to the high computational cost associated with the eigendecomposition. Further, one must consider the cost of performing the rank reduction transformation on the signals, evaluating the cost function, and sorting/selecting the eigenvectors based on the cost function. In the end, these additional computational costs may be counter productive to the goal of reducing the high computational costs associated with STAP.

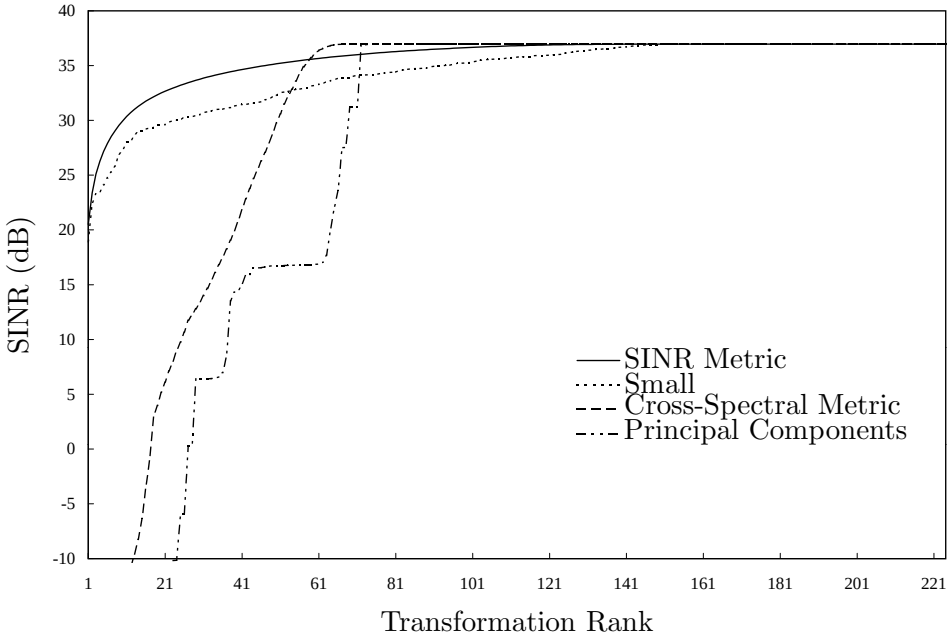


Figure 5.3 SINR Performance of the SINR metric, Small, CSM GSC, and PC GSC methods as a function of the transformation rank for Case 1: SNR=15 dB, Jammers at normalized angles and JNRs of (0.25, 40 dB) and (±0.433, 30 dB), and CNR=55 dB.

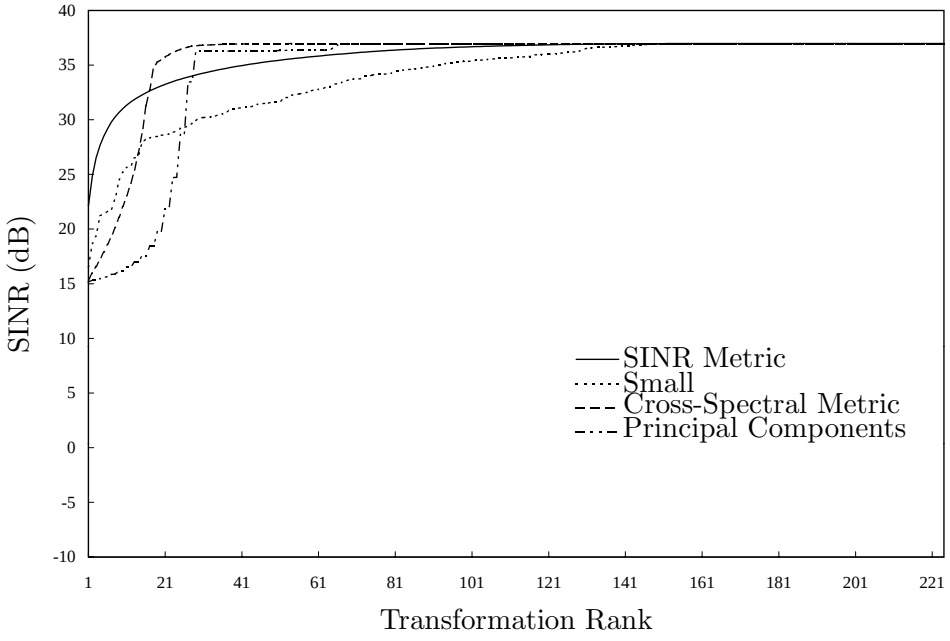


Figure 5.4 SINR Performance of the SINR metric, Small, CSM GSC, and PC GSC methods as a function of the transformation rank for Case 4: SNR=15 dB, Jammers at normalized angles and JNRs of (0.25, 0 dB) and (±0.433, 0 dB), and CNR=20 dB.

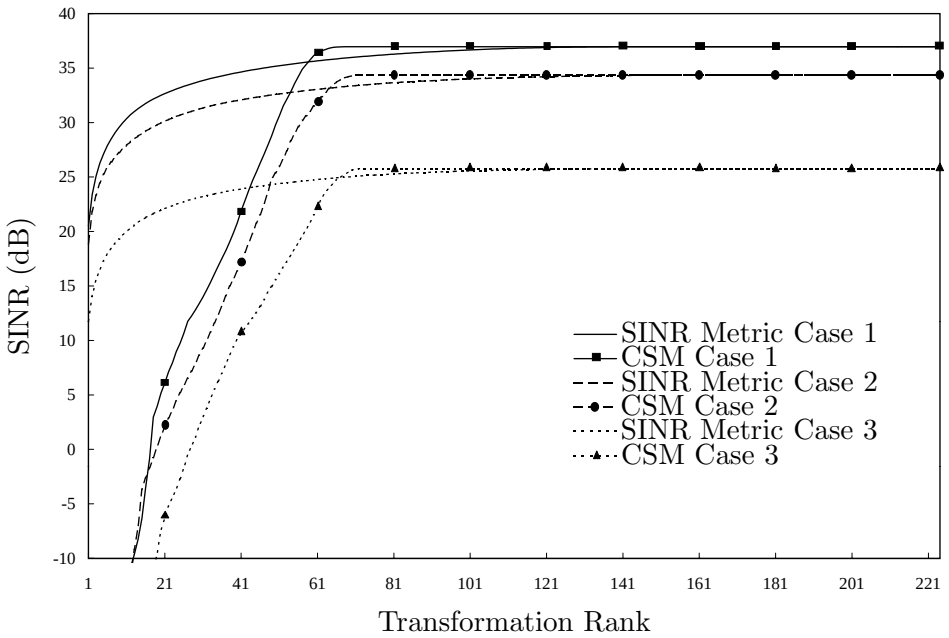


Figure 5.5 SINR Performance of the SINR metric and CSM GSC methods as a function of the transformation rank for the high clutter and jamming environment (Cases 1-3).

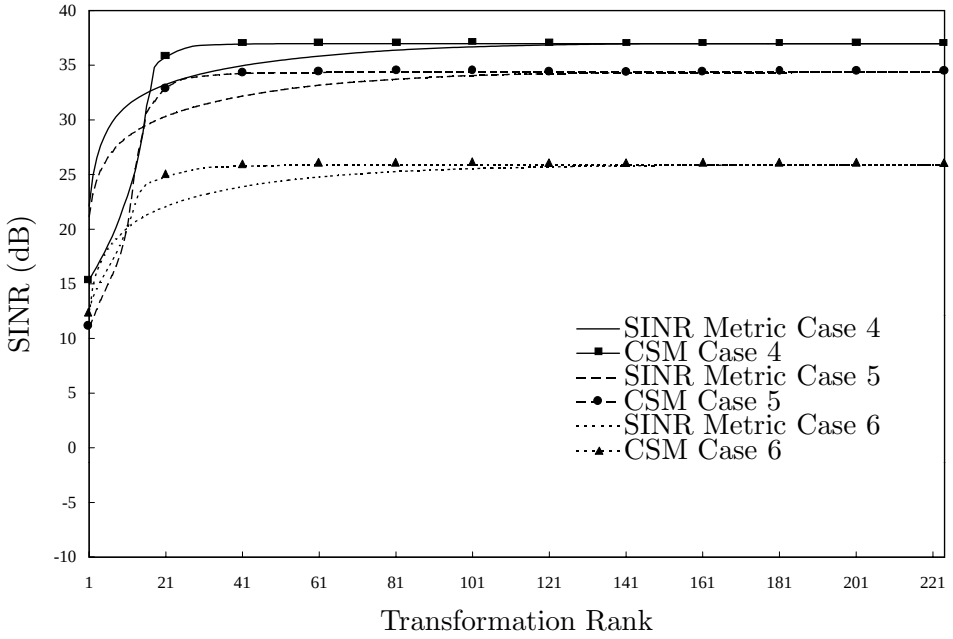


Figure 5.6 SINR Performance of the SINR metric and CSM GSC methods as a function of the transformation rank for the low clutter and jamming environment (Cases 4-6).

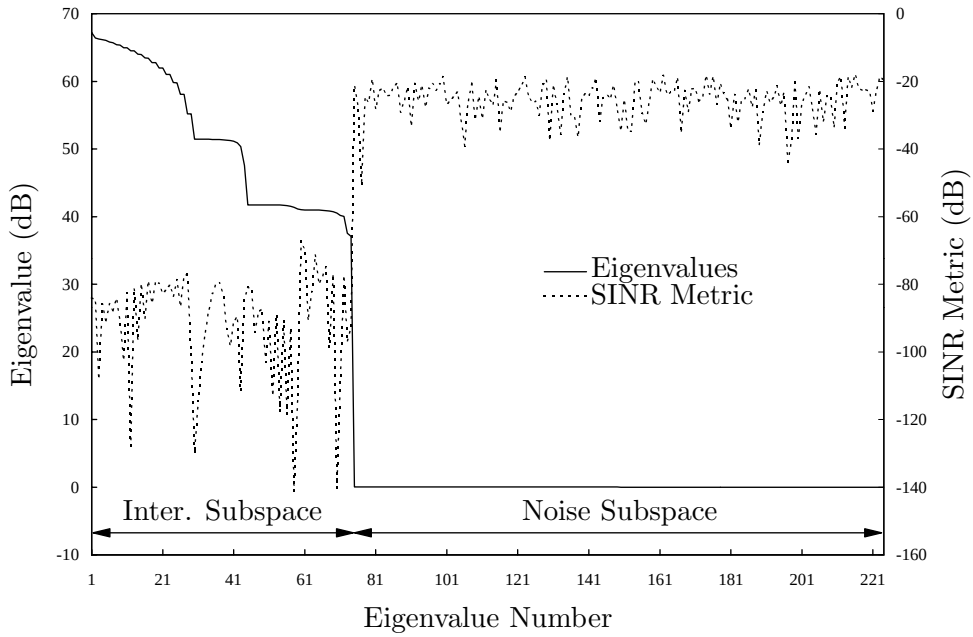


Figure 5.7 Eigenvalues of $\mathbf{R}$ and SINR metric for each eigenvector of $\mathbf{R}$ for Case 1: SNR=15 dB, Jammers at normalized angles and JNRs of (0.25, 40 dB) and ($\pm$0.433, 30 dB), and CNR=55 dB.

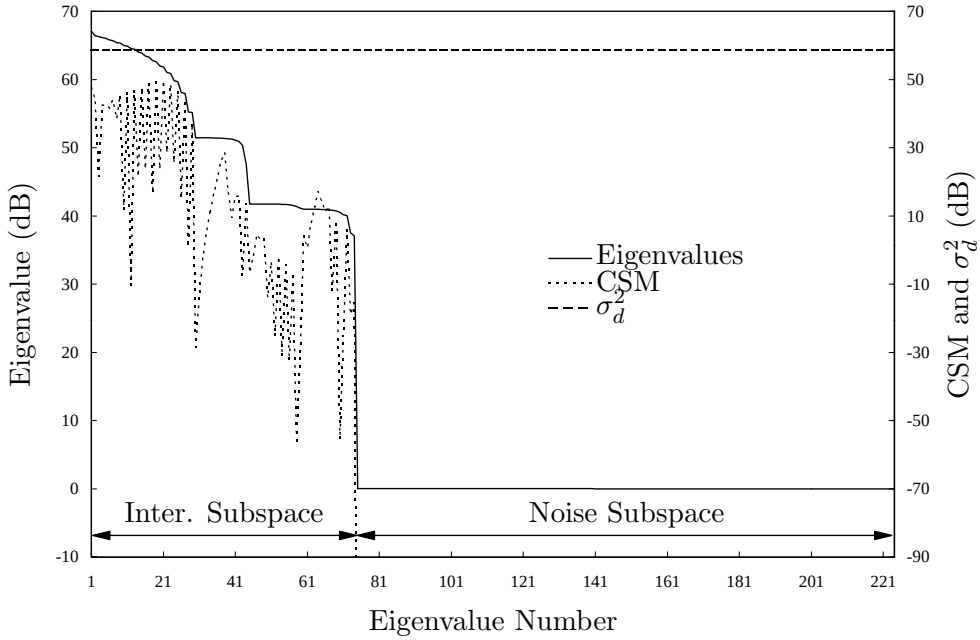


Figure 5.8 Eigenvalues of $\mathbf{R}_b$ and CSM for each eigenvector of $\mathbf{R}_b$ for Case 1: SNR=15 dB, Jammers at normalized angles and JNRs of (0.25, 40 dB) and ($\pm$0.433, 30 dB), and CNR=55 dB.

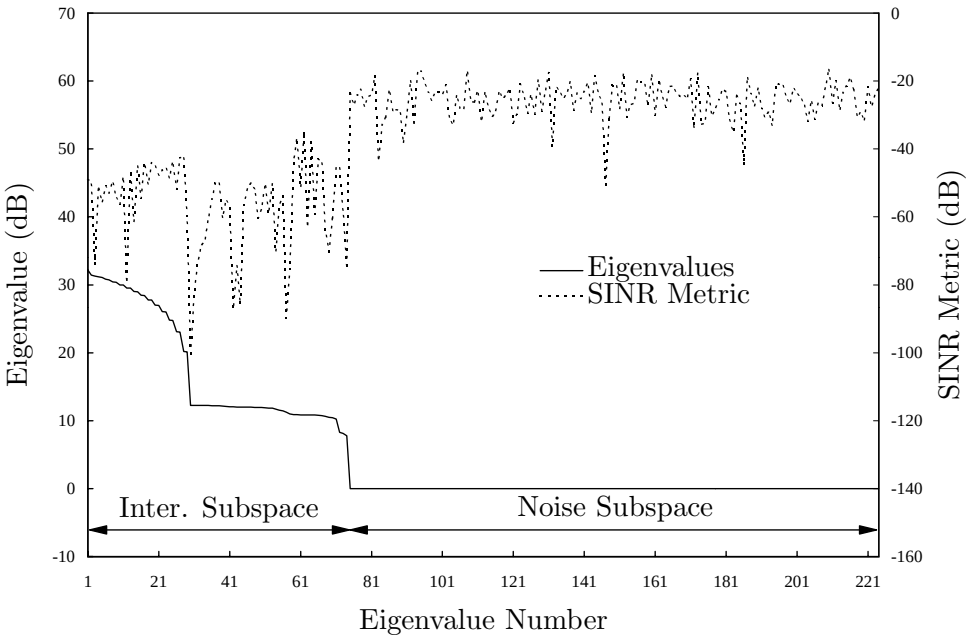


Figure 5.9 Eigenvalues of $\mathbf{R}$ and SINR metric for each eigenvector of $\mathbf{R}$ for Case 4: SNR=15 dB, Jammers at normalized angles and JNRs of (0.25, 0 dB) and ($\pm$0.433, 0 dB), and CNR=20 dB.

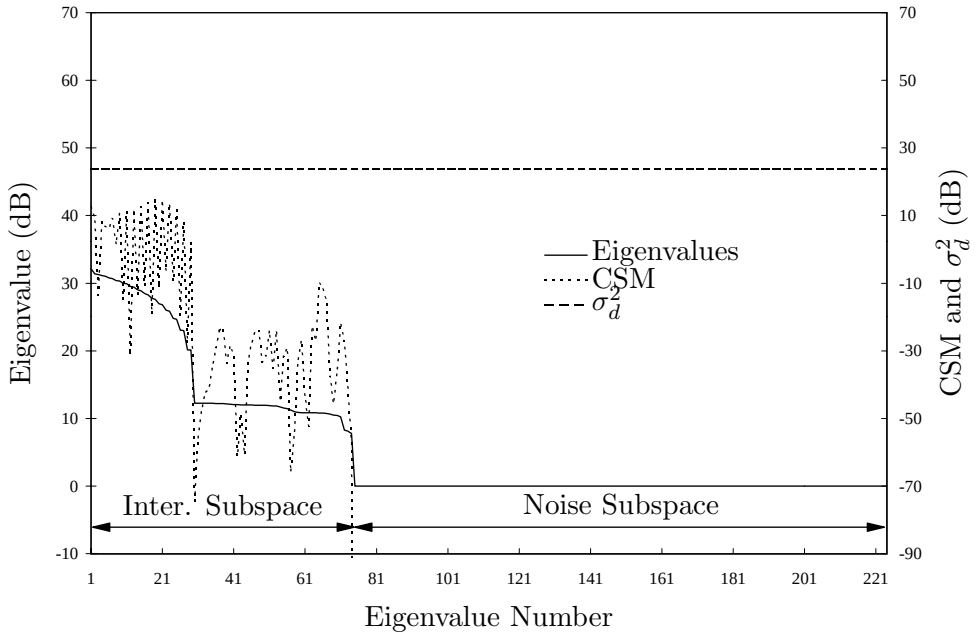


Figure 5.10 Eigenvalues of $\mathbf{R}_b$ and CSM for each eigenvector of $\mathbf{R}_b$ for Case 4: SNR=15 dB, Jammers at normalized angles and JNRs of (0.25, 0 dB) and ($\pm$0.433, 0 dB), and CNR=20 dB.

To implement the SINR metric CSM methods in practice, we must overcome the high computational burden of these methods. As with the Block STAP method, this requirement to reduce the computational burden suggests the need for a fixed, efficient transformation. The key to the SINR metric and CSM methods is that the metric provides a direct indication of the basis vectors that have the greatest impact of the output SINR. Thus, the fixed, efficiency transformation should provide a similar capability. That is, we need a metric that is similar to or is an approximation to the SINR metric or CSM so that we can express the output SINR as a sum similar to Eqns (5.12) and (5.13). This suggests the concept of approximate eigenvectors that approximately diagonalize the correlation matrices such that diagonal elements are approximate eigenvalues. This concept of reducing the computation burden with a fixed, efficiency transformation constructed from approximate eigenvectors is not new. In the applications involving wide-sense stationary random processes, the DFT or other sinusoidal transformations (e.g., the discrete cosine transformation) are commonly used to approximately diagonalize the correlation matrix, since the basis vectors of these transformations are approximate eigenvectors of Hermitian, Toeplitz matrices and can be implemented efficiently [33,38,39]. Unfortunately, STAP correlation matrices are not Hermitian, Toeplitz matrices which implies the vector form of the interference plus noise is not stationary. Thus, the basis vectors of these sinusoidal transformations are not approximate eigenvectors of STAP correlation matrices. The development of approximate eigenvectors for STAP correlation matrices is an open research area.

The resolution of approximate eigenvectors for STAP correlation matrices would also be beneficial to the Block STAP methods. Recall that the overall efficiency of the Block STAP method improves when the rank of the matrix $\mathbf{G}$ is low and that the rank of $\mathbf{G}$ is determined by the rank of the off-diagonal blocks of the correlation matrix. Suppose that we had approximate eigenvectors for STAP correlation matrices, then we could approximately diagonalize a STAP correlation matrix in the sense that the resulting matrix would be a banded diagonal matrix with bandwidth p. Note that a banded diagonal matrix $\mathbf{X}$ with elements $[\mathbf{X}]_{m,n}$ has bandwidth p if $[\mathbf{X}]_{m,n} = 0$ when $m > n+p$ and $n > m+p$ [15:149]. Now, observe that the off-diagonal blocks of such a banded diagonal matrix will have at most

p non-zero rows (columns) and thus, will have a rank less than or equal to p. Therefore, if we can approximately diagonalize a STAP correlation matrix such that the bandwidth is narrow (small), then the rank of $\mathbf{G}$ will be low and we would expect the Block STAP processor to have a high efficiency over a large portion of the angle-Doppler plane.

5.6 Summary

In this chapter, we have extended the cost function concept used by Goldstein and Reed with the GSC to the direct form processor. The result is a new reduced-rank direct form processor, where the columns of the rank reduction transformation are selected as the eigenvectors of the interference plus noise correlation matrix based on the SINR metric. The SINR metric is used to identify the eigenvectors which minimize the loss in output SINR. We presented simulation results that demonstrate the potential of the SINR metric method under ideal conditions and highlight that the SINR performance of the SINR metric and CSM methods are dependent on the interference plus noise environment and transformation rank. For a given interference plus noise environment, the simulation results show that the CSM method outperforms the SINR metric method above a certain rank, while the converse is true below this rank. These results suggest the potential need for more than one implementation structure (method) in a STAP system, where the method is selected based on the scenario and available resources (e.g., secondary data support and computational power). Tools, such as the CSM and SINR metric, should prove to be invaluable in assessing candidate methods for a given environment and resource level.

We noted that the high computational burden of the SINR metric and CSM methods limit their utility in practice and that like the Block STAP processor, we need a fixed, efficiency transformation to overcome the high computational burden. For the SINR metric and CSM methods, this requirement for a fixed, efficiency transformation leads to the concept of approximate eigenvectors for STAP correlation matrices, which is an open research area. Finally, we noted that the Block STAP processor would benefit from the discovery of approximate eigenvectors for STAP correlation matrices, since one could then approximately diagonalize the correlation matrix which would have low rank off-diagonal blocks.

VI. Conclusions and Recommendations

We summarize the major results and contributions of this research effort and present several areas of future research in this chapter.

6.1 Results and Contributions

We have proposed a dual channel matched filtering system that addresses two key challenges in the practical implementation of a single channel matched filter (maximum SINR filter) in an unknown environment: secondary data support and computational cost. When the correlation matrix of the interference plus noise is a block diagonal matrix or can be efficiency transformed to a block diagonal matrix, the proposed dual channel system requires 50% less secondary data to achieve nearly the same level of SINR performance as an equivalent single channel system with a reduction in the computational cost of approximately 75%. We derived an exact expression for the normalized SINR in terms of random variables with known distribution and approximate expressions for the mean and variance of the normalized SINR as a function of weight vector dimension and secondary data support that characterize the SINR performance of the dual channel system in an unknown environment. Using these approximate expressions, we demonstrated the reduced secondary data requirements of the dual channel system. The correlation matrix from any real, wide-sense stationary random process is a member of the centrosymmetric family of matrices which can be efficiently block diagonalized with a fixed transformation. Thus, in any matched filtering applications involving real, wide-sense stationary random processes, a dual channel system can be used in place of a single channel system to realize a reduction in the secondary data support and computation cost.

We investigated the potential of applying a dual channel system to the problem domain of space-time adaptive processing (STAP) for airborne surveillance radars, referring to the system as Block STAP. We defined a family of STAP correlation matrices that was representative of the interference plus noise environments typically encountered by an airborne surveillance radar. We provided evidence to support the conjecture that STAP correlation matrices cannot be block diagonalized by a fixed transformation. Based on this

conjecture, optimum STAP processing cannot be divided into two independent processing steps and thus, any implementation of the Block STAP method with a fixed transformation will be suboptimal. Numerous suboptimal STAP methods have been proposed based on a fixed transformation, but often signal processing heuristics are used to select the transformation. In contrast, we have provided an in-depth analysis of the Block STAP efficiency relative to the optimal processor, yielding a worst-case lower bound for the efficiency and the identification of the factors that control the efficiency. From this analysis, we proposed a mathematical-based criterion for selecting the Block STAP transformation.

Finally, we addressed the problem of selecting the optimum eigen-based rank reduction transformation for a direct form processor. We introduced the SINR metric for the direct form processor which is an extension the cross-spectral metric (CSM) introduced by Goldstein and Reed [11–14] for the generalized sidelobe canceler. The result is a new reduced-rank direct form processor, referred to as the SINR metric method, where the columns of the rank reduction transformation are selected as the eigenvectors of the interference plus correlation matrix based on the SINR metric. The SINR metric identifies the eigenvectors that have the greatest impact on the SINR performance of a direct form processor. If the rank reduction transformation is constructed from r eigenvectors of the correlation matrix, then the r eigenvectors with the largest SINR are the optimal set of r eigenvectors in terms of minimizing the loss in SINR performance of an eigen-based reduced-rank direct form processor. Via simulations, we demonstrated the importance of including a cost function (output SINR) in the transformation selection process and that the best implementation structure (SINR metric director form processor versus CSM generalized sidelobe canceler) depends on the interference plus noise environment and available resources (i.e., transformation rank). Tools, such as the SINR metric and CSM, should prove to be invaluable in identifying and assessing methods for a given environment and resource level. With regard to the CSM, we proved that the CSM for each of the noise subspace eigenvectors is zero, a result that does not appear to be widely known. Our proof also provides a clear proof of the more widely known result that the optimum generalized sidelobe canceler weight vector lies in the interference subspace.

6.2 Directions for Future Research

Although the normalized SINR is a useful performance metric, we are often interested in how a particular detection system performs in terms of probability of detection and probability of false alarms. Thus, one area of future research would be to develop expressions for the dual channel probability of detection and probability of false alarms. Additionally, since the ultimate objective is to declare the presence or absence of a target in an unknown environment, we need to develop a constant false alarm rate test statistic for the dual channel system or method for selecting the threshold to achieve a given level of performance. Along these lines, the exact expression for the normalized SINR in terms of random variables with known distributions should be revisited with the goal of establishing its probability density function. Finally, in regards to future research related to the general dual channel concept, we note that the concept of block diagonalizing the correlation matrix is not restricted to maximum SINR (matched) filtering applications. The block diagonalization concept could be extended to Wiener filtering, subband adaptive filtering, and compression applications.

With regard to the Block STAP method, we proposed a transformation selection criterion based on a lower bound for the efficiency and an analysis of the efficiency expression. The criterion provided a consistent indication of the system with the highest average efficiency for the systems and environments simulated. However, a rigorous analysis and proof of the proposed criterion is needed. This represents a challenging problem, since it requires the development of a relationship between the steering vector and the eigenvectors of the correlation matrix. As mentioned earlier, we could abandon the fixed transformation concept and move to a data adaptive transformation approach. In which case, one would need to develop a method for efficiently implementing the proposed transformation selection criterion with only knowledge of the estimated diagonal blocks of the correlation matrix. Along these lines, a data adaptive method that starts with the symmetric and skewed symmetric subspaces (i.e., subspaces invariant to centrosymmetric matrices) and updates the subspaces as secondary data becomes available is one option. In general, modern surveillance radar platforms have additional sources of information that could be used to predict the anti-symmetric part of the interference environment and aid in the subspace

update cycle. For example, an electronic support measures system could provide estimates of the spatial location and power level of the barrage noise jammers in the environment. Additionally, using a map/terrain database, one could also predict the spatial location of strong clutter discretes such as cities. To effectively incorporate this predicted information into subspace update cycle, one would need to develop an understanding of how the subspaces change with the introduction of anti-symmetric interference.

As noted earlier, the STAP community would certainly benefit from the development of approximate eigenvectors of STAP correlation matrices. As evident by the fact the STAP correlation matrices are not Hermitian, Toeplitz matrices, one could consider the vector form of the interference plus noise as a nonstationary random process. Both Kozek [26] and Mallet et. al. [30] have reported that functions that are localized in time and frequency are approximate eigenvectors for certain classes of non-stationary random processes. Kozek provides an in-depth treatment of the general problem of defining and analyzing the time-varying power spectrum of nonstationary random processes while the treatment of Mallet et. al. is more specialized to local cosine basis functions and adaptive best basis selection. We believe the works of Kozek and Mallet et. al. (and references contained there in) represent a good starting point for gaining additional insight into the characteristics of the STAP correlation matrices and the development of approximate eigenvectors.

Appendix A. CSM of the Noise Subspace Eigenvectors

In this appendix, we show that the cross-spectral metric (CSM) for the eigenvectors that span the noise subspace is zero. This result does not appear to be widely known, although one could deduce this result from Van Veen's [40] observation that the GSC weight vector lies in the interference subspace. In addition to showing that the CSM associated with each noise subspace eigenvector is zero, the analysis provides a clear proof that the GSC weight vector lies in the interference subspace. As a result, one can conclude that a reduced-rank generalized sidelobe canceler will not experience a SINR loss if the span of rank reduction transformation contains the interference subspace. The following analysis is based on the single interference source plus receiver noise analysis given by Van Veen [40].

Recall that the CSM is defined as

$$\frac{|\mathbf{u}_i^H \mathbf{r}_{bd}|^2}{\lambda_i}, \tag{A.1}$$

where $\{\mathbf{u}_i\}_{i=1}^L$ and $\{\lambda_i\}_{i=1}^L$ are the eigenvectors and eigenvalues of $\mathbf{R}_b = \mathbf{BRB}^H$, respectively, $\mathbf{r}_{bd} = \mathbf{BRs}$, $\mathbf{R}$ is the $MN \times MN$ interference plus noise correlation matrix, $\mathbf{s}$ is the $MN \times 1$ steering vector, $\mathbf{B}$ is the $L \times MN$ blocking matrix with the property that $\mathbf{Bs} = \mathbf{0}$, and $L = MN - 1$. Now, observe that the interference plus noise correlation matrix can be written as the sum of the interference correlation matrix and the receiver (white) noise correlation matrix. That is,

$$\mathbf{R} = \mathbf{R}_I + \sigma_w^2 \mathbf{I}_{MN}, \tag{A.2}$$

where $\mathbf{R}_I$ is the interference correlation matrix, $\sigma_w^2 \mathbf{I}_{MN}$ is the receiver noise correlation matrix, $\mathbf{I}_{MN}$ is a $MN \times MN$ identity matrix, and σ_w^2 is the variance of the receiver noise. We can express $\mathbf{R}_I$ in terms of its eigendecomposition as

$$\mathbf{R}_I = \mathbf{A\Gamma A}^H, \tag{A.3}$$

where $\mathbf{A}$ is a unitary matrix composed of the eigenvectors and $\mathbf{\Gamma}$ is a diagonal matrix of the associated eigenvalues. Now, if we assume $\mathbf{R}_I$ has a rank of $P < MN$, then only P of

the eigenvalues are non-zero (i.e. positive) and the remaining eigenvalues are zero. Using this low rank assumption, we can rewrite Eqn. (A.3) as

$$\mathbf{R}_I = \mathbf{A}_1 \mathbf{\Gamma}_1 \mathbf{A}_1^H, \tag{A.4}$$

where $\mathbf{\Gamma}_1$ is an $P \times P$ diagonal matrix of the positive eigenvalues and $\mathbf{A}_1$ is a $MN \times P$ matrix composed of eigenvectors associated with the positive eigenvalues. Using Eqn. (A.4) and by assuming that $\mathbf{B}\mathbf{B}^H = \mathbf{I}_L$, we can write $\mathbf{R}_b$ as

$$\begin{aligned}
\mathbf{R}_b &= \mathbf{B}\mathbf{R}\mathbf{B}^H \\
&= \mathbf{B}\mathbf{R}_I\mathbf{B}^H + \sigma_w^2 \mathbf{I}_L \\
&= \mathbf{B}\mathbf{A}_1\mathbf{\Gamma}_1\mathbf{A}_1^H\mathbf{B}^H + \sigma_w^2 \mathbf{I}_L \\
&= \mathbf{B}\mathbf{A}_1\mathbf{\Gamma}_1^{1/2}\mathbf{\Gamma}_1^{1/2}\mathbf{A}_1^H\mathbf{B}^H + \sigma_w^2 \mathbf{I}_L,
\end{aligned} \tag{A.5}$$

where we used the fact that since the diagonal elements of $\mathbf{\Gamma}_1$ are positive and real, we can write $\mathbf{\Gamma}_1 = \mathbf{\Gamma}_1^{1/2}\mathbf{\Gamma}_1^{1/2}$. Now, let

$$\mathbf{Q} = \mathbf{B}\mathbf{A}_1\mathbf{\Gamma}_1^{1/2} \tag{A.6}$$

and substitute into Eqn. (A.5) to yield

$$\mathbf{R}_b = \mathbf{B}\mathbf{R}_I\mathbf{B}^H + \sigma_w^2 \mathbf{I}_L = \mathbf{Q}\mathbf{Q}^H + \sigma_w^2 \mathbf{I}_L. \tag{A.7}$$

Note that the rank of $\mathbf{B}\mathbf{R}_I\mathbf{B}^H$, and hence, the rank of $\mathbf{Q}$, is less than or equal to P. We can write the singular value decomposition of $\mathbf{Q}$ as

$$\mathbf{Q} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H, \tag{A.8}$$

where $\mathbf{U}$ is the $L \times L$ unitary matrix of the left singular vectors, $\mathbf{V}$ is the $P \times P$ unitary matrix of the right singular vectors, and $\mathbf{\Sigma}$ is the $L \times P$ matrix of the non-negative real

singular values. Observe that $\mathbf{\Sigma}$ can be written as

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_1 \\ \mathbf{0} \end{bmatrix}, \tag{A.9}$$

where $\mathbf{\Sigma}_1$ is an $P \times P$ diagonal matrix of the singular values. Then, by partitioning $\mathbf{U}$ as $\mathbf{U} = [\mathbf{U}_1 \ \mathbf{U}_2]$, where $\mathbf{U}_1$ is a $L \times P$ matrix and $\mathbf{U}_2$ is a $L \times L - P$ matrix, we can rewrite Eqn. (A.8) as

$$\mathbf{Q} = \mathbf{U}_1 \mathbf{\Sigma}_1 \mathbf{V}^H, \tag{A.10}$$

and thus,

$$\mathbf{B}\mathbf{R}_I\mathbf{B}^H = \mathbf{Q}\mathbf{Q}^H = \mathbf{U}_1\mathbf{\Sigma}_1\mathbf{V}^H\mathbf{V}\mathbf{\Sigma}_1\mathbf{U}_1^H = \mathbf{U}_1\mathbf{\Sigma}_1^2\mathbf{U}_1^H. \tag{A.11}$$

The right-hand side of Eqn. (A.11) represents the eigendecomposition of $\mathbf{B}\mathbf{R}_I\mathbf{B}^H$, where the columns of $\mathbf{U}_1$ are the eigenvectors of the interference subspace and the diagonal elements of $\mathbf{\Sigma}_1^2$ are the associated eigenvalues. Now, observe that

$$\mathbf{B}\mathbf{R}_I\mathbf{B}^H = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{U}_1^H \\ \mathbf{U}_2^H \end{bmatrix}, \tag{A.12}$$

and thus, we can express the eigendecomposition of $\mathbf{R}_b$ as

$$\begin{aligned} \mathbf{R}_b &= \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{U}_1^H \\ \mathbf{U}_2^H \end{bmatrix} + \begin{bmatrix} \sigma_w^2\mathbf{I}_P & \mathbf{0} \\ \mathbf{0} & \sigma_w^2\mathbf{I}_{L-P} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma}_1 + \sigma_w^2\mathbf{I}_P & \mathbf{0} \\ \mathbf{0} & \sigma_w^2\mathbf{I}_{L-P} \end{bmatrix} \begin{bmatrix} \mathbf{U}_1^H \\ \mathbf{U}_2^H \end{bmatrix} \\ &= \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H, \end{aligned} \tag{A.13}$$

where

$$\mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Sigma}_1 + \sigma_w^2 \mathbf{I}_P & \mathbf{0} \\ \mathbf{0} & \sigma_w^2 \mathbf{I}_{L-P} \end{bmatrix}. \tag{A.14}$$

The columns of $\mathbf{U}$ are the eigenvectors $\{\mathbf{u}_i\}_{i=1}^L$ of $\mathbf{R}_b$ and the diagonal elements of Λ are the associated eigenvalues $\{\lambda_i\}_{i=1}^L$. Note that the columns of $\mathbf{U}_2$ are the eigenvectors that span the noise subspace and are orthonormal to the columns of $\mathbf{U}_1$ since the eigenvectors $\{\mathbf{u}_i\}_{i=1}^L$ form an orthonormal set.

With these building blocks in hand, we can now examine the properties of the CSM and weight vector $\mathbf{w}_{gsc}$ as function of the eigenvectors of $\mathbf{R}_b$. Recall that the eigenvectors $\{\mathbf{u}_i\}_{i=1}^L$ form an orthonormal basis for the L dimensional vector space and thus, we can write the weight vector as a linear combination of the eigenvectors. That is,

$$\mathbf{w}_{gsc} = \sum_{i=1}^{L} \frac{c_i \mathbf{u}_i}{\sqrt{\lambda_i}}, \tag{A.15}$$

where the c_i's are scalars to be determined and the reason for the $\sqrt{\lambda_i}$ term will become obvious in a few steps. Recall that $\mathbf{w}_{gsc} = \mathbf{R}_b^{-1}\mathbf{r}_{bd}$ which can be rewritten as $\mathbf{R}_b\mathbf{w}_{gsc} = \mathbf{r}_{bd}$ and after substituting Eqns. (A.13) and (A.15), we have

$$\mathbf{U}\mathbf{\Lambda}\mathbf{U}^H \sum_{i=1}^{L} \frac{c_i \mathbf{u}_i}{\sqrt{\lambda_i}} = \mathbf{r}_{bd}. \tag{A.16}$$

To solve for the c_i's, we premultiply both sides of Eqn. (A.16) by $\mathbf{u}_k^H$ for some k yielding

$$\mathbf{u}_k^H \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H \sum_{i=1}^{L} \frac{c_i \mathbf{u}_i}{\sqrt{\lambda_i}} = \mathbf{u}_k^H \mathbf{r}_{bd}, \tag{A.17}$$

which reduces to

$$\sqrt{\lambda_k} c_k = \mathbf{u}_k^H \mathbf{r}_{bd}, \tag{A.18}$$

where we used the fact that $\mathbf{u}_k^H \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H = \lambda_k \mathbf{u}_k^H$ and the orthonormal property of the eigenvectors. Solving Eqn. (A.18) for c_k yields

$$c_k = \frac{\mathbf{u}_k^H \mathbf{r}_{bd}}{\sqrt{\lambda_k}} \quad \text{for each } k. \tag{A.19}$$

Notice that the magnitude squared of the right side of Eqn. (A.19) is the CSM of the kth eigenvector. Next, using Eqns. (A.2), (A.4), (A.6), (A.10) and $\mathbf{Bs} = \mathbf{0}$, we can write

$$\begin{aligned}
\mathbf{r}_{bd} &= \mathbf{BRs} \\
&= \mathbf{BR}_I \mathbf{s} \\
&= \mathbf{B}\mathbf{A}_1 \mathbf{\Gamma}_1^{1/2} \mathbf{\Gamma}_1^{1/2} \mathbf{A}_1^H \mathbf{s} \\
&= \mathbf{Q} \mathbf{\Gamma}_1^{1/2} \mathbf{A}_1^H \mathbf{s} \\
&= \mathbf{U}_1 \mathbf{\Sigma}_1 \mathbf{V}^H \mathbf{\Gamma}_1^{1/2} \mathbf{A}_1^H \mathbf{s}.
\end{aligned} \tag{A.20}$$

Substituting Eqn. (A.20) into Eqn. (A.19) yields

$$c_k = \frac{\mathbf{u}_k^H \mathbf{U}_1 \mathbf{\Sigma}_1 \mathbf{V}^H \mathbf{\Gamma}_1^{1/2} \mathbf{A}_1^H \mathbf{s}}{\sqrt{\lambda_k}}. \tag{A.21}$$

Using the orthonormal property of the eigenvectors, we observe that

$$\mathbf{u}_k^H \mathbf{U}_1 = \begin{cases} \mathbf{e}_k & \text{if } \mathbf{u}_k \in \text{span}(\mathbf{U}_1), \\ \mathbf{0} & \text{otherwise,} \end{cases} \tag{A.22}$$

where $\mathbf{e}_k$ is a $1 \times P$ vector with a 1 in the kth position and zeros in all other positions and $\mathbf{0}$ is a $1 \times P$ vector of zeros. Finally, using Eqn. (A.22), we can write

$$c_k = \begin{cases} \dfrac{\mathbf{u}_k^H \mathbf{r}_{bd}}{\sqrt{\lambda_k}} & \text{if } \mathbf{u}_k \in \text{span}(\mathbf{U}_1), \\ 0 & \text{otherwise.} \end{cases} \tag{A.23}$$

From Eqn. (A.23), we can conclude that the CSM is zero for any eigenvector of $\mathbf{R}_b$ that is not in the interference subspace, since the kth CSM is equal to the magnitude square of

c_k. Using Eqn. (A.23), we can rewrite Eqn. (A.15) as

$$\mathbf{w}_{gsc} = \sum_{i=1}^{P} \frac{c_i \mathbf{u}_i}{\sqrt{\lambda_i}}, \quad \text{where } \mathbf{u}_i \in \text{span}(\mathbf{U}_1). \tag{A.24}$$

Thus, the weight vector $\mathbf{w}_{gsc}$ is a linear combination of the eigenvectors $\{\mathbf{u}_i\}_{i=1}^{P}$ which span the interference subspace defined by $\mathbf{B}\mathbf{R}_I\mathbf{B}^H$. That is, $\mathbf{w}_{gsc}$ lies in the interference subspace. The fact that the weight vector of the GSC lies in the interference subspace is not surprising. If the weight vector was not confined to the interference subspace, then the lower (blocking) branch of the GSC would emit uncorrelated noise in addition to the estimate of the interference (correlated) noise in the upper branch. Any uncorrelated noise emitted by the lower branch would combine with the uncorrelated noise in the upper branch, effectively increasing the receiver noise floor.

Bibliography

1. Abramowitz, Milton and Irene A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Washington, D.C.: National Bureau of Standards, 1964.

2. Andrew, A.L. "Eigenvectors of Certain Matrices," *Linear Algebra and Its Applications*, *7*:151–162 (1973).

3. Barker, G.P., et al. "A Non-Commutative Spectral Theorem," *Linear Algebra and Its Applications*, *20*:95–100 (1978).

4. Berger, Scott D. and Byron M. Welsh. "Selecting a Reduced-Rank Transformation for STAP - A Direct Form Perspective." Accepted for publication in the IEEE Transactions on Aerospace and Electronic Systems.

5. Boroson, Don M. "Sample Size Considerations for Adaptive Arrays," *IEEE Transactions on Aerospace and Electronic Systems*, *16*(4):446–451 (July 1980).

6. Brennan, L.E. and I.S. Reed. "Theory of Adaptive Radar," *IEEE Transactions on Aerospace and Electronic Systems*, *9*(2):237–252 (March 1973).

7. Brooks, Lowell W. and Irving S. Reed. "Equivalence of the Likelihood Ratio Processor, the Maximum Signal-to-Noise Ratio Filter, and the Wiener Filter," *IEEE Transactions on Aerospace and Electronic Systems*, *8*(5):690–692 (September 1972).

8. Cantoni, A. and P. Butler. "Eigenvalues and Eigenvectors of Symmetric Centrosymmetric Matrices," *Linear Algebra and Its Applications*, *13*:275–288 (1976).

9. Chang, Lena and Chien-Chung Yeh. "Performance of DMI and Eigenspace-Based Beamformers," *IEEE Transactions on Antennas and Propagation*, *40*(11):1336–1347 (November 1992).

10. Corwin, Lawrence J. and Robert H. Szczarba. *Multivariable Calculus*. New York, NY: Marcel Dekker, 1982.

11. Goldstein, J. Scott and Irving S. Reed. "Performance Measures for Optimal Constrained Beamformers," *IEEE Transactions on Antennas and Propagation*, *45*(1):11–14 (January 1997).

12. Goldstein, J. Scott and Irving S. Reed. "Reduced-Rank Adaptive Filtering," *IEEE Transactions of Signal Processing*, *45*(2):492–496 (February 1997).

13. Goldstein, J. Scott and Irving S. Reed. "Subspace Selection of Partially Adaptive Sensor Array Processing," *IEEE Transactions on Aerospace and Electronic Systems*, *33*(2):539–543 (April 1997).

14. Goldstein, J. Scott and Irving S. Reed. "Theory of Partially Adaptive Radar," *IEEE Transactions on Aerospace and Electronic Systems*, *33*(4):1309–1325 (October 1997).

15. Golub, Gene E. and Charles F. Van Loan. *Matrix Computations* (second Edition). Baltimore, MD: John Hopkins University Press, 1989.

16. Good, I.J. "The Inverse of a Centrosymmetric Matrix," *Technometrics*, *12*(4):925–928 (November 1970).

17. Gui Wang, Song and Jun Shao. "Constrained Kantorovich Inequalities and Relative Efficiency of Least Squares," *Journal of Multivariate Analysis*, *42*:284–298 (1992).

18. Haimovich, Alexander. "The Eigencanceler: Adaptive Radar by Eigenanalysis Methods," *IEEE Transactions on Aerospace and Electronic Systems*, *32*(2):532–542 (April 1996).

19. Haimovich, A.M., et al. "Performance Analysis of Reduced-Rank STAP." *Proceedings of the 1997 IEEE National Radar Conference*. 42–47. New York, NY: The Institute of the Electrical and Electronics Engineers, 1997.

20. Henrici, Peter. *Applied and Computational Complex Analysis*. New York, NY: John Wiley and Sons, 1974.

21. Hoffman, Kenneth and Ray Kunze. *Linear Algerbra* (Second Edition). Englewood Cliff, NJ: Prentice-Hall, 1971.

22. Hogg, Robert V. and Allen T. Craig. *Introduction to Mathematical Statistics* (Fifth Edition). Englewood Cliffs, NJ: Prentice-Hall, 1995.

23. Horn, Roger A. and Charles R. Johnson. *Matrix Analysis*. New York, NY: Cambridge, 1996.

24. Khatri, C.G. and C. Radhakrishna Rao. "Effects of Estimated Noise Correlation Matrix in Optimal Signal Detection," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *35*(5):671–679 (May 1987).

25. Kirsteins, Ivars P. and Donald W. Tufts. "Adaptive Detection Using Low Rank Approximation to a Data Matrix," *IEEE Transactions on Aerospace and Electronic Systems*, *30*(1):55–67 (January 1994).

26. Kozek, Werner. *Matched Weyl-Heisenberg Expansions of Nonstationary Environments*. PhD dissertation, Vienna University of Technology, 1996.

27. Laffey, Thomas J. "Simultaneous Quasidiagonalization of Complex Matrices," *Linear Algebra and Its Applications*, *16*:189–201 (1977).

28. Laffey, Thomas J. "Simultaneous Reduction of Sets of Matrices under Similarity," *Linear Algebra and Its Applications*, *84*:123–138 (1986).

29. Lipschutz, Seymour. *Theory and Problems of Linear Algebra* (second Edition). Schaum's Outline Series, New York, NY: McGraw-Hill, 1991.

30. Mallat, Stéphane, et al. "Adaptive Covariance Estimation of Locally Stationary Processes," *Annales of Statistics*, *26*(1):1–47 (February 1998).

31. Muirhead, Robb J. *Aspects of Multivariate Statistical Theory*. New York, NY: John Wiley and Sons, 1982.

32. Papoulis, Athanasios. *Probability, Random Variables, and Stochastic Processes* (Third Edition). New York, NY: McGraw-Hill, 1991.

33. Ramstad, T.A., et al. *Subband Compression of Images: Principles and Examples*. Amsterdam, The Netherlands: Elsevier, 1995.

34. Reed, I. S., et al. "Rapid Convergence Rate in Adaptive Arrays," *IEEE Transactions on Aerospace and Electronic Systems*, *10*(6):853–863 (November 1974).

35. Scharf, Louis L. *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Reading, MA: Addison-Wesley, 1991.

36. Shapiro, H. "Simultaneous Block Triangularization and Block Diagonalization of Sets of Matrices," *Linear Algebra and Its Applications*, *25*:129–137 (1979).

37. Subbaram, Harish and Ken Abend. "Interference Suppression Via Orthogonal Projections: A Performance Analysis," *IEEE Transactions on Antennas and Propagation*, *41*(9):1187–1194 (September 1993).

38. Therrien, Charles W. *Discrete Random Signals and Statistical Signal Processing*. Englewood Cliff, NJ: Prentice-Hall, 1992.

39. Unser, Michael. "On the Approximation of the Discrete Karhunen-Loeve Transform for Stationary Processes," *Signal Processing*, *7*:231–249 (1984).

40. Veen, Barry D. Van. "An Analysis of Several Partially Adaptive Beamformer Designs," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *37*(2):192–203 (February 1989).

41. Ward, James. *Space-Time Adaptive Processing for Airborne Radar*. Technical Report TR-1015, Lincoln Laboratory, 1994.

42. Watters, J.F. "Simultaneous Quasi-Diagonalization of Normal Matrices," *Linear Algebra and Its Applications*, *9*:103–117 (1974).

Vita

Major Scott D. Berger was born on 11 August 1963 in Henderson, Nevada and graduated from Cooper City High School in 1981. He received a Bachelor of Science in Electrical Engineering from Florida Atlantic University in 1985 and a Masters of Science in Electrical Engineering from the Air Force Institute of Technology in 1989. In 1985, he received his commission in the USAF through the Officer Training School at Lackland AFB, Texas. From 1985 to 1989, he served as an Electronic Warfare Engineer at the Air Force Electronic Warfare Center, Kelly AFB, Texas. He then attended the Air Force Institute of Technology and after graduation in 1989, was assigned to National Air Intelligent Center at Wright-Patterson AFB, Ohio as a Senior Radar Engineer. From 1993 to 1996, he was assigned to the 84th Test Squadron, Tyndall AFB, Florida serving as the Lead F-15 Engineer, Lead Electronic Countermeasures Engineer, Chief Engineer and later, as the Flight Commander for the Radar Test Facility. He entered the electrical engineering Ph.D. program at the Air Force Institute of Technology in 1996. After completing his studies, he will be assigned to the Air Force Research Laboratory at Kirtland AFB, New Mexico.

Permanent address: 5045 S.W. 87th Ter.
Cooper City, FL 33328