GROWTH CURVE MODELS
IN SIGNAL PROCESSING APPLICATIONS

By

LUZHOU XU

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2006

To my wife, my son, and my parents

ACKNOWLEDGMENTS

I feel extremely lucky to have met all the people I have and to have received their help. Foremost, I would like to express my most sincere gratitude to my advisor, Dr. Jian Li, for her constant support, encouragement, and guidance. I am specially grateful for the opportunity she has offered me to pursue my research under her supervision. She gives me freedom to try new ideas, while providing the needed assistance at the right time. She is always willing to share her knowledge and career experience as an advisor, collaborator, and friend. The experience working with her has proven to be invaluable and memorable.

I would also like to thank Dr. Jenshen Lin, Dr. Clint Slatton, and Dr. Rongling Wu in the department of statistics for serving on my supervisory committee and for their valuable advice. Their wonderful teaching has widened my horizon to their intriguing research fields, and was and will always be helpful to my research work.

My special gratitude is due to Dr. Petre Stoica at Uppsala University, Sweden, for his guidance in many interesting topics. His wide knowledge, strong analytical skill, and keen insight have never ceased to amaze me. I was so fortunate to have the opportunity to work with him and to benefit from his insightful ideas and constructive advice. I am grateful to Dr. Mingzhou Ding in the department of biomedical engineering for his support and guidance in my research on neural data analysis.

I also want to thank my groupmates in Spectral Analysis Lab: Bin Guo, Yi Jiang, Jianhua Liu, Zhipeng Liu, Guoqing Liu, William Roherts, Yijun Sun, Yanwei Wang, Zhisong Wang, Yao Xie, Hong Xiong, Changjiang Xu, Xiayu Zheng, Xumin Zhu, and others. Their friendship, support, and help have made my life easier and more enjoyable.

Last but not least, I wish to thank my wife and my son for all the happiness they have brought to my life. I am also deeply thankful to my grandparents, parents and sisters for their constant love and support.

TABLE OF CONTENTS

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

GROWTH CURVE MODELS
IN SIGNAL PROCESSING APPLICATIONS

By

Luzhou Xu

August 2006

Chair: Jian Li
Major Department: Electrical and Computer Engineering

As a powerful statistical tool, the growth-curve (GC) model is attracting increasing attentions in various areas. In this dissertation, we study several variations of the growth-curve model, and discuss their applications to the emerging multiple-input multiple-output (MIMO) radar system.

We first study the statistical properties of two estimators for the regression coefficient matrix in the GC model, i.e., the maximum likelihood (ML) and Capon methods. We derive the closed-form expression of the Cramér-Rao bound (CRB) for the unknown regression coefficient matrix, and then analyze the bias properties and mean-squared errors (MSEs) of the two estimators. We show that the multivariate ML estimator is unbiased whereas the multivariate Capon estimator is biased downward for finite data samples. Both estimators are asymptotically statistically efficient when the number of data samples is large.

Next, we consider a variation of the GC model, referred to as the diagonal growth-curve (DGC) model, where the regression matrix is constrained to be diagonal. A closed-form approximate maximum likelihood (AML) estimator for this model is derived based on the maximum likelihood principle. We analyze

the statistical properties of this method theoretically and show that the AML estimate is unbiased and asymptotically statistically efficient for a large number of data samples. Via several numerical examples in array signal processing and spectral analysis, we also show that the proposed AML estimator can achieve better estimation accuracy and exhibit greater robustness than the best existing methods.

Then we consider a general growth-curve model, referred to as the block diagonal growth-curve (BDGC) model, where the unknown regression coefficient matrix is constrained to be block-diagonal, and which can unify the GC and DGC models. We proposed a closed-form approximate maximum likelihood (AML) estimator for the block-diagonal constrained matrix, which is proved to be unbiased and asymptotically statistically efficient for a large data sample number. Several applications of this model in signal processing are then presented.

Finally, we consider a multiple-input multiple-output (MIMO) radar system with a general antenna configuration, i.e., both the transmitter and receiver have multiple well-separated subarrays with each subarray containing closely-spaced antennas. Hence, both the coherent processing gain and the spatial diversity gain can be achieved by the system simultaneously. We introduce several spatial spectral estimators, including Capon and APES, for target detection and parameter estimation. We also provide a generalized likelihood ratio test (GLRT) and a conditional generalized likelihood ratio test (cGLRT) for the system. Based on GLRT and iGLRT, we then propose an iterative GLRT (iGLRT) procedure for target detection and parameter estimation. Via several numerical examples, we show that iGLRT can provide excellent detection and estimation performance at a low computational cost.

CHAPTER 1
INTRODUCTION

1.1 Growth-Curve Model and Its Variations

The growth-curve (GC) model is a generalized multivariate analysis of variance (GMANOVA) model, which was first formulated by Potthoff and Roy in 1964 [1] for investigating growth curve problems in statistical applications. Since then it has been studied by many authors, including Rao [2] - [4], Khatri [5], Gleser and Olkin [6], Geisser [7], von Rosen [8] - [10], Verbyla and Venables [11], and Srivastava [12]. It is one of the main tools for dealing with longitudinal data, especially for serial correlations [13] as well as repeated measurements [14] - [18], and is attracting increasing attentions in various areas, such as economics, biology, medical research and epidemiology. Recently, this model was extended to the complex-valued field and was adopted in the signal processing literature [19] - [22].

Consider an observed data matrix $\mathbf{X} \in C^{M\times L}$, which can be written as

$$\mathbf{X} = \mathbf{ABS} + \mathbf{Z}. \tag{1–1}$$

In (1–1), $\mathbf{A} \in C^{M\times N}$ and $\mathbf{S} \in C^{K\times L}$ are both known matrices, $\mathbf{B} \in C^{N\times K}$ is an unknown regression matrix, and $\mathbf{Z} \in C^{M\times L}$ is the error matrix whose columns are independently and identically distributed (i.i.d.) zero-mean Gaussian random vectors with an unknown covariance matrix. The problem of interest is to estimate $\mathbf{B}$ from the observed data matrix $\mathbf{X}$.

A special case of (1–1), when $N = K = 1$, has been studied widely in signal processing, such as high-resolution spectral analysis [23] [24] and array signal processing [25] - [35]. In this case, the GC model reduces to a univariate GC

(UGC) model

$$\mathbf{X} = \mathbf{a}\beta\mathbf{s}^T + \mathbf{Z}, \tag{1–2}$$

with $\mathbf{a}$ and $\mathbf{s}^T$ being column and row vectors, respectively, and β being an unknown scalar variable. The performance of the maximum-likelihood (ML) and Capon estimators for the UGC model has been thoroughly studied in [25]. It was shown theoretically that ML is unbiased whereas Capon is biased downward, and both estimators are asymptotically statistically efficient for a large number of data samples (i.e., $L \gg M$). In this dissertation, we will extend this result to the GC model.

In many practical applications, the observed signal consists of multiple components. For this scenario, the UGC model in (1–2) can be extended as

$$\mathbf{X} = \sum_{k=1}^{K} \mathbf{a}_k\beta_k\mathbf{s}_k^T + \mathbf{Z}, \tag{1–3}$$

with $\{\mathbf{a}_k\}$ and $\{\mathbf{s}_k^T\}$ being the known column and row vectors, and $\{\beta_k\}$ being unknown scalar variables. Obviously, the model in (1–3) can be rewritten as

$$\mathbf{X} = \mathbf{ABS} + \mathbf{Z} \quad \text{with} \quad \mathbf{B} = \text{diag}(\boldsymbol{\beta}), \tag{1–4}$$

where $\text{diag}(\boldsymbol{\beta})$ denotes a diagonal matrix with its diagonal formed by the elements of the vector $\boldsymbol{\beta}$, $\mathbf{A} = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_K]$, $\mathbf{S} = [\mathbf{s}_1 \ \cdots \ \mathbf{s}_K]^T$, and $\boldsymbol{\beta} = [\beta_1 \ \cdots \beta_K]^T$. We note that the model in (1–4) is similar to the GC model in (1–1) except that the unknown regression coefficient matrix $\mathbf{B}$ is constrained to be diagonal. Hence, it is referred to as a diagonal growth-curve (DGC) model [36]. Despite the seemingly minor difference between the GC and DGC models, the ML estimator [21] [22] for the GC model is invalid for the DGC model. In fact, to our knowledge, no closed-form ML estimator for $\boldsymbol{\beta}$ in (1–4) exists in the literature. In this dissertation, we propose an approximate maximum likelihood (AML) estimator for $\boldsymbol{\beta}$ in (1–4). We also investigate its statistical properties via theoretical analysis and numerical

simulations, and show that the AML estimator for the DGC model is unbiased and asymptotically statistically efficient for a large number of data samples.

A more general variation of the GC model was studied by Verbyla [11], Rosen [8] and Srivastava [12], where the authors consider an estimation problem of unknown regression matrices $\{\mathbf{B}_k\}$ from the observed data matrix $\mathbf{X}$ in the equation

$$\mathbf{X} = \sum_{k=1}^{K} \mathbf{A}_k \mathbf{B}_k \mathbf{S}_k + \mathbf{Z}. \tag{1–5}$$

Again, in (1–5), $\{\mathbf{A}_k\}$ and $\{\mathbf{S}_k\}$ $(k = 1,\ 2,\ ,\cdots K)$ are all known matrices, and $\mathbf{Z}$ is defined as in the GC model. Note that (1–5) can be rewritten in the form of the GC model in (1–1) by constraining the unknown regression coefficient matrix $\mathbf{B}$ to be a block-diagonal matrix; hence (1–5) is referred to as a block diagonal growth-curve (BDGC) model in this dissertation. An iterative numerical approach for the estimation of the unknown regression coefficient matrices $\mathbf{B}_k$ in (1–5) was proposed by Verbyla [11] by using the canonical reduction method. However, this approach is both conceptually and practically complicated. Moreover, being an iterative method, it may suffer from convergence problems. Two nested variations of the BDGC model were studied by Rosen [8] and Srivastava [12], independently, where explicit forms of the ML estimators were presented. However, some additional assumptions have be imposed in [8] and [12]. In [8], the rows of $\mathbf{S}_k$ $(k = 1,\ 2,\ \cdots K)$ are assumed to be nested, i.e., $\mathcal{R}(\mathbf{S}_K^T) \subseteq \mathcal{R}(\mathbf{S}_{K-1}^T) \subseteq \cdots \subseteq \mathcal{R}(\mathbf{S}_1^T)$, with $\mathcal{R}(\cdot)$ denoting the range space; and in [12] the columns of $\mathbf{A}_k$ $(k = 1,\ 2,\ \cdots K)$ are assumed to be nested, i.e., $\mathcal{R}(\mathbf{A}_K) \subseteq \mathcal{R}(\mathbf{A}_{K-1}) \subseteq \cdots \subseteq \mathcal{R}(\mathbf{A}_1)$. Neither of these two nested subspace conditions can be satisfied in signal processing applications. In this dissertation, we will consider a general BDGC model in (1–5), and propose an approximate maximum likelihood (AML) estimator [37] for the unknown regression coefficient matrices $\mathbf{B}_k$, which will be shown both theoretically and numerically to be unbiased and asymptotically statistically efficient.

1.2 Multiple-Input Multiple-Output Radar

A multiple-input multiple-output (MIMO) radar uses multiple antennas to simultaneously transmit several (possibly linearly independent) waveforms and it also uses multiple antennas to receive the reflected signals [38] [39]. It has been shown that by exploiting this waveform diversity, MIMO radar can overcome performance degradations caused the radar cross-section (RCS) fluctuations [40] - [43], achieve flexible spatial transmit beampattern design [44] [45], provide high-resolution spatial spectral estimates [46] - [57], and significantly improve the parameter identifiability [58].

The statistical MIMO radar, studied in [40] - [43], aims at resisting the "scintillation" effect encountered in radar systems. It is well-known that the RCS of a target, which represents the amount of energy reflected from the target toward the receiver, changes rapidly as a function of the target aspect [59] and the locations of the transmitting and receiving antennas. The target "scintillation" causes severe degradations in the target detection and parameter estimation performance of the radar. By spacing the transmit antennas, which transmit linearly independent signals, far away from each other, a spatial diversity gain can be obtained as in the MIMO wireless communications to this "scintillation" effect [40] - [43].

Flexible transmit beampattern designs are investigated in [44] and [45]. Different from the "statistical" MIMO radar above, the transmitting antennas are closely spaced. The authors in [44] and [45] show that the waveforms transmitted via its antennas can be optimized to obtain several transmit beampattern designs with superior performance. For example, the covariance matrix of the waveforms can be optimized to maximize the power around the locations of interest and also to minimize the cross-correlation of the signals reflected back to the radar by these targets, thereby significantly improving the performance of the adaptive MIMO

radar techniques. Due to the significantly larger number of degrees of freedom of a MIMO system, a much better transmit beampattern with a MIMO radar can be achieved than with its phased-array counterpart.

In [48], a MIMO radar technique is suggested to improve the radar resolution. The idea is to transmit N ($N > 1$) orthogonal coded waveforms by N antennas and to receive the reflected signals by M ($M > 1$) antennas. At each receiving antenna output, the signal is matched-filtered using each of the transmitted waveforms to obtain NM channels, to which the data-adaptive Capon beamformer [60] is applied. It is proved in [48] that the beampattern of the proposed MIMO radar is obtained by the multiplication of the transmitting and receiving beampatterns, which gives high resolution. However, only the single-target case is considered in [48].

A MIMO radar scheme is considered in [55] - [57] that can deal with the presence of multiple targets. Similar to some of the aforementioned MIMO radar approaches, linearly independent waveforms are transmitted simultaneously via multiple antennas. Due to the different phase shifts associated with different propagation paths from transmitting antennas to targets, these independent waveforms are linearly combined at the targets with different phase factors. As a result, the signal waveforms reflected from different targets are linearly independent of each other, which allows the direct application of many adaptive techniques to achieve high resolution and excellent interference rejection capability. Several adaptive nonparametric algorithms in the presence or absence of steering vector errors are presented in [55] - [57].

Note that the MIMO radars discussed in the aforementioned literature can be grouped into two classes according to their antenna configurations. One is the conventional radar array, in which both transmitting and receiving antennas are closely spaced for coherent transmission and detection [44] - [57]. The other is the

diverse antenna configuration, where the antennas are separated far away from each other to achieve spatial diversity gain [40] - [43]. To reap the benefits of both schemes, in this dissertation, we consider a general antenna configuration, i.e., both the transmitting and receiving antenna arrays consist of several well-separated subarrays with each subarray containing closely-spaced antennas [61]. By using some results of the growth-curve models, we provide a generalized likelihood ratio test (GLRT) and a conditional generalized likelihood ratio test (cGLRT) for the system. Based on GLRT and iGLRT, we then propose an iterative GLRT (iGLRT) procedure for target detection and parameter estimation. Via several numerical examples, we show that iGLRT can provide excellent detection and estimation performance at a low computational cost.

1.3 Study Overview

In Chapter 2, we consider estimating the unknown regression coefficient matrix $\mathbf{B}$ in the GC model in (1–1). Two multivariate approaches, Maximum Likelihood (ML) and Capon, are provided. We derive the closed-form expression of the Cramér-Rao bound (CRB) for the unknown complex amplitudes. We also analyze the bias properties and Mean Squared Errors (MSE) of the two estimators. A comparative study shows that the multivariate ML estimator is unbiased whereas the multivariate Capon estimator is biased downward for finite data samples. Both estimators are asymptotically statistically efficient when the number of data samples is large.

In Chapter 3, we consider a variation of the GC model, referred to as the diagonal growth-curve (DGC) model, where the matrices $\mathbf{A}$ and $\mathbf{S}$ in (1–4) are both known and the regression coefficient matrix $\mathbf{B}$ is constrained to be diagonal. A closed-form approximate maximum likelihood (AML) estimator for this model is derived based on the maximum likelihood principle. We analyze the statistical properties of this method theoretically and show that the AML

estimate is unbiased and asymptotically statistically efficient for a large number of data samples. Via several numerical examples in array signal processing and spectral analysis, we also show that the proposed AML estimator can achieve better estimation accuracy and exhibit greater robustness than the best existing methods.

In Chapter 4, we consider a general variation of the growth-curve (GC) model, referred to as the block diagonal growth-curve (BDGC) model, which can unify the GC and DGC models in Chapters 2 and 3. In BDGC, the unknown regression coefficient matrix is constrained to be block-diagonal. A closed-form approximate maximum likelihood (AML) estimator for this model is then derived, which is shown to be unbiased and asymptotically statistically efficient for a large number of data samples.

In Chapter 5, we consider a multiple-input multiple-output (MIMO) radar system with a general antenna configuration, i.e., both the transmitter and receiver have multiple well-separated subarrays with each subarray containing closely-spaced antennas. We introduce several spatial spectral estimators, including Capon and APES, for target detection and parameter estimation. We also provide a generalized likelihood ratio test (GLRT) and a conditional generalized likelihood ratio test (cGLRT) for the system. Based on GLRT and iGLRT, we then propose an iterative GLRT (iGLRT) procedure for target detection and parameter estimation.

Finally, we summarize the dissertation and point out future research directions in Chapter 6.

CHAPTER 2
GROWTH-CURVE MODEL

2.1 Introduction and Problem Formulation

In this chapter, we consider the following multivariate complex amplitude estimation problem

$$\mathbf{X} = \mathbf{ABS} + \mathbf{Z}. \tag{2–1}$$

In (2–1), $\mathbf{X} \in C^{M\times L}$ denotes the observed snapshots with L being the number of snapshots. The columns in $\mathbf{A} \in C^{M\times N}$ are the known linearly independent spatial vectors, e.g., steering vectors. The rows in $\mathbf{S} \in C^{K\times L}$ are the known temporal vectors, e.g., waveforms, assumed to be linearly independent of each other or not completely correlated with each other. The matrix $\mathbf{B} \in C^{N\times K}$ contains the multivariate unknown complex amplitudes. Throughout this chapter, we assume that $M \geq N$ and $L \geq K + M$. The columns of the interference and noise matrix $\mathbf{Z} \in C^{M\times L}$ are statistically independent circularly symmetric complex Gaussian random vectors with zero-mean and unknown covariance matrix $\mathbf{Q}$. The problem of interest is to estimate the unknown matrix $\mathbf{B}$.

We note that the data model of (2–1) has general applications. Its real-valued counterpart, called growth-curve (GC) Model, has been studied and used widely for investigating growth problems in the statistics field [62] [63] [18]. This real-valued growth-curve model was extended and introduced to the signal processing field in [21]. Using the extended model, the authors in [21] unified many existing algorithms proposed for radar array processing [64] [65], spectral analysis [23] [66] and wireless communication [67] - [70] applications.

The focus of this chapter is on the performance analysis of the multivariate Maximum Likelihood (ML) and Capon estimators for the data model in (2–1). We derive the closed-form expression of the Cramér-Rao Bound (CRB) of the unknown complex amplitude parameters. We also analyze the bias properties and Mean-Squared-Errors (MSE) of the two estimators. A comparative study shows that the multivariate ML estimator is unbiased whereas the multivariate Capon estimator is biased downward for finite snapshots. Yet in finite data samples and at low SNR, Capon can provide a smaller MSE than ML. Both estimators are asymptotically statistically efficient when the number of snapshots is large.

The remainder of the chapter is organized as follows. Section 2.2 provides the multivariate Capon and ML estimators. Section 2.3 gives the performance analysis of the two estimators and the CRB of the unknown complex amplitudes. Numerical examples are provided in Section 2.4. Finally, we present our conclusions in Section 2.5.

2.2 Multivariate Parameter Estimation

Based on the data model in (2–1), we describe the multivariate Capon and ML estimators in this section.

2.2.1 Multivariate Capon Estimation

The multivariate Capon estimator consists of two main steps. The first is the Capon beamforming [60] [71] [72]. The other is the Least-Squares (LS) estimation [16] [73], which is basically the matched filtering.

We first consider the Capon beamforming. Let

$$\mathbf{R} = \mathbf{X}\mathbf{X}^H. \tag{2–2}$$

Then, the Capon beamformer can be formulated as

$$\hat{\mathbf{W}} = \arg\min_{\mathbf{W}} \operatorname{tr}(\mathbf{W}^H\mathbf{R}\mathbf{W}) \qquad \text{subject to} \quad \mathbf{W}^H\mathbf{A} = \mathbf{I}, \tag{2–3}$$

where $\hat{\mathbf{W}}$ is a multivariate weighting matrix for noise and interference suppression while keeping the desired signals undistorted. Solving the above optimization problem yields

$$\hat{\mathbf{W}} = \mathbf{R}^{-1}\mathbf{A}(\mathbf{A}^H\mathbf{R}^{-1}\mathbf{A})^{-1}. \tag{2–4}$$

Note that since $M \geq N$ and the columns in $\mathbf{A}$ are linearly independent of each other, $\mathbf{A}^H\mathbf{R}^{-1}\mathbf{A}$ has full rank N with probability one. The beamforming output, denoted by $\mathbf{Y}$, is

$$\mathbf{Y} = \hat{\mathbf{W}}^H\mathbf{X} = (\mathbf{A}^H\mathbf{R}^{-1}\mathbf{A})^{-1}\mathbf{A}^H\mathbf{R}^{-1}\mathbf{X}. \tag{2–5}$$

Now we consider the LS estimation. Substituting (2–1) into (2–5) yields

$$\mathbf{Y} = \mathbf{BS} + (\mathbf{A}^H\mathbf{R}^{-1}\mathbf{A})^{-1}\mathbf{A}^H\mathbf{R}^{-1}\mathbf{Z}. \tag{2–6}$$

Estimating $\mathbf{B}$ from $\mathbf{Y}$ based on (2–6) is a standard Multivariate Analysis of Variance (MANOVA) problem [62] [63] [21]. Note that after spatial beamforming, the noise vectors remain temporally white, and hence the LS estimator gives the best performance. Using the LS algorithm yields

$$\hat{\mathbf{B}}_{\text{Capon}} = \mathbf{YS}^H(\mathbf{SS}^H)^{-1}. \tag{2–7}$$

Substituting (2–5) into (2–7), the multivariate Capon estimator has the form

$$\hat{\mathbf{B}}_{\text{Capon}} = (\mathbf{A}^H\mathbf{R}^{-1}\mathbf{A})^{-1}\mathbf{A}^H\mathbf{R}^{-1}\mathbf{XS}^H(\mathbf{SS}^H)^{-1}. \tag{2–8}$$

Note that the Capon estimator for the univariate case in [25] is a special case of (2–8).

2.2.2 Multivariate Maximum Likelihood Estimation

A general derivation of the multivariate ML estimator has been given in [21]. In this chapter, we assume that both $\mathbf{A}$ and $\mathbf{S}$ are known and the multivariate ML estimator can be briefly derived as follows to make this chapter self-contained.

Based on the data model in (2–1), the negative log-likelihood function is proportional to

$$\mathcal{L}(\mathbf{B},\mathbf{Q}) = L\ln|\mathbf{Q}| + \mathrm{tr}\big[\mathbf{Q}^{-1}(\mathbf{X}-\mathbf{ABS})(\mathbf{X}-\mathbf{ABS})^H\big], \tag{2–9}$$

where $|\cdot|$, $\mathrm{tr}(\cdot)$ and $(\cdot)^H$ denote the determinant, trace and conjugate transpose of a matrix, respectively.

Minimizing the negative log-likelihood function with respect to $\mathbf{Q}$ yields

$$\hat{\mathbf{Q}} = \frac{1}{L}(\mathbf{X}-\mathbf{ABS})(\mathbf{X}-\mathbf{ABS})^H. \tag{2–10}$$

Inserting (2–10) into (2–9), the ML estimator of $\mathbf{B}$ can be formulated as

$$\hat{\mathbf{B}}_{\mathrm{ML}} = \arg\min_{\mathbf{B}} |(\mathbf{X}-\mathbf{ABS})(\mathbf{X}-\mathbf{ABS})^H|. \tag{2–11}$$

Note that

$$\begin{aligned}
&|(\mathbf{X}-\mathbf{ABS})(\mathbf{X}-\mathbf{ABS})^H| \\
&= \Big|\big[\mathbf{AB}-\mathbf{XS}^H(\mathbf{SS}^H)^{-1}\big](\mathbf{SS}^H)\big[\mathbf{AB}-\mathbf{XS}^H(\mathbf{SS}^H)^{-1}\big]^H + \mathbf{T}\Big| \\
&= |\mathbf{T}|\,\Big|\mathbf{I} + \mathbf{T}^{-\frac{1}{2}}\big[\mathbf{AB}-\mathbf{XS}^H(\mathbf{SS}^H)^{-1}\big](\mathbf{SS}^H)\big[\mathbf{AB}-\mathbf{XS}^H(\mathbf{SS}^H)^{-1}\big]^H\mathbf{T}^{-\frac{1}{2}}\Big| \\
&= |\mathbf{T}|\,\Big|\mathbf{I} + (\mathbf{SS}^H)^{\frac{1}{2}}\big[\mathbf{AB}-\mathbf{XS}^H(\mathbf{SS}^H)^{-1}\big]^H\mathbf{T}^{-1}\big[\mathbf{AB}-\mathbf{XS}^H(\mathbf{SS}^H)^{-1}\big](\mathbf{SS}^H)^{\frac{1}{2}}\Big| \\
&= |\mathbf{T}|\,\Big|\mathbf{I} + (\mathbf{SS}^H)^{-\frac{1}{2}}\mathbf{SX}^H\big[\mathbf{T}^{-1}-\mathbf{T}^{-1}\mathbf{A}(\mathbf{A}^H\mathbf{T}^{-1}\mathbf{A})^{-1}\mathbf{A}^H\mathbf{T}^{-1}\big]\mathbf{XS}^H(\mathbf{SS}^H)^{-\frac{1}{2}} + \\
&\qquad (\mathbf{SS}^H)^{\frac{1}{2}}\big[\mathbf{B}-(\mathbf{A}^H\mathbf{T}^{-1}\mathbf{A})^{-1}\mathbf{A}^H\mathbf{T}^{-1}\mathbf{XS}^H(\mathbf{SS}^H)^{-1}\big]^H(\mathbf{A}^H\mathbf{T}^{-1}\mathbf{A}) \\
&\qquad \big[\mathbf{B}-(\mathbf{A}^H\mathbf{T}^{-1}\mathbf{A})^{-1}\mathbf{A}^H\mathbf{T}^{-1}\mathbf{XS}^H(\mathbf{SS}^H)^{-1}\big](\mathbf{SS}^H)^{\frac{1}{2}}\Big| \\
&\geq |\mathbf{T}|\,\Big|\mathbf{I} + (\mathbf{SS}^H)^{-\frac{1}{2}}\mathbf{SX}^H\big[\mathbf{T}^{-1}-\mathbf{T}^{-1}\mathbf{A}(\mathbf{A}^H\mathbf{T}^{-1}\mathbf{A})^{-1}\mathbf{A}^H\mathbf{T}^{-1}\big]\mathbf{XS}^H(\mathbf{SS}^H)^{-\frac{1}{2}}\Big|,
\end{aligned} \tag{2–12}$$

where

$$\mathbf{T} = \mathbf{X}[\mathbf{I}-\mathbf{S}^H(\mathbf{SS}^H)^{-1}\mathbf{S}]\mathbf{X}^H, \tag{2–13}$$

and $\mathbf{I}$ is an identity matrix. In the above derivation we have used the fact that $|\mathbf{I}+\mathbf{XY}| = |\mathbf{I}+\mathbf{YX}|$ [74]. Since the rows in $\mathbf{S}$ are linearly independent of each other and $L \geq K+M$, the rank of $\mathbf{I}-\mathbf{S}^H(\mathbf{SS}^H)^{-1}\mathbf{S}$, which is $L-K$, is greater than or equal to M. Hence, $\mathbf{T}$ and $\mathbf{A}^H\mathbf{T}^{-1}\mathbf{A}$ in (2–12) have full ranks M and N, respectively, with probability one.

From (2–12), the ML estimator of $\mathbf{B}$ is written as

$$\hat{\mathbf{B}}_{\text{ML}} = (\mathbf{A}^H\mathbf{T}^{-1}\mathbf{A})^{-1}\mathbf{A}^H\mathbf{T}^{-1}\mathbf{XS}^H(\mathbf{SS}^H)^{-1}. \tag{2–14}$$

Note again that the ML estimator for the univariate case in [25] is a special case of (2–14).

To better understand the above ML estimator intuitively, we insert (2–1) into (2–13) and get

$$\mathbf{T} = \mathbf{Z}[\mathbf{I}-\mathbf{S}^H(\mathbf{SS}^H)^{-1}\mathbf{S}]\mathbf{Z}^H. \tag{2–15}$$

It shows that $\frac{1}{L}\mathbf{T}$ is an estimate of the unknown noise covariance $\mathbf{Q}$. We also note that the estimator in (2–14) can be divided into two steps, including the ML beamforming spatially corresponding to the left-multiplication matrix $(\mathbf{A}^H\mathbf{T}^{-1}\mathbf{A})^{-1}\mathbf{A}^H\mathbf{T}^{-1}$ and the LS estimation temporally corresponding to the right-multiplication matrix $\mathbf{S}^H(\mathbf{SS}^H)^{-1}$.

Note that like the univariate case in [25], the only difference between the Capon and ML estimators is that the matrix $\mathbf{R}$ in (2–8) is replaced by $\mathbf{T}$ in (2–14). However, as we will show in the following analysis, this seemingly minor difference in fact leads to significant and interesting performance differences between the two estimators.

2.3 Performance Analysis

2.3.1 Performance Analysis of the Multivariate ML Estimator

We consider below the statistical performance analysis of the multivariate ML estimator. We show that the conclusions in [25] can be extended to the

multivariate case, i.e., the multivariate ML estimator is unbiased and it is asymptotically statistically efficient when the number of snapshots is large. In the following derivations, several techniques presented in [25], [62] and [18] are employed.

2.3.1.1 Bias analysis

For convenience, we denote

$$\mathbf{X}_S = \mathbf{X}\mathbf{S}^H, \qquad \mathbf{X}_S^{\perp} = \mathbf{X}[\mathbf{I} - \mathbf{S}^H(\mathbf{S}\mathbf{S}^H)^{-1}\mathbf{S}], \tag{2–16}$$

$$\mathbf{Z}_S = \mathbf{Z}\mathbf{S}^H, \qquad \mathbf{Z}_S^{\perp} = \mathbf{Z}[\mathbf{I} - \mathbf{S}^H(\mathbf{S}\mathbf{S}^H)^{-1}\mathbf{S}]. \tag{2–17}$$

Clearly, we have

$$\mathbf{X}_S = \mathbf{A}\mathbf{B}\mathbf{S}\mathbf{S}^H + \mathbf{Z}_S, \qquad \mathbf{T} = \mathbf{X}_S^{\perp}(\mathbf{X}_S^{\perp})^H = \mathbf{Z}_S^{\perp}(\mathbf{Z}_S^{\perp})^H. \tag{2–18}$$

Note that the columns of $\mathbf{Z}$ are independent zero-mean Gaussian random vectors. Note also that the columns of $\mathbf{S}^H$ are orthogonal to those of $\mathbf{I} - \mathbf{S}^H(\mathbf{S}\mathbf{S}^H)^{-1}\mathbf{S}$. By the property of joint Gaussian distribution, $\mathbf{Z}_S$ and $\mathbf{Z}_S^{\perp}$ are two independent Gaussian random matrices [75]. Hence, $\mathbf{X}_S$ and $\mathbf{T}$ are also independent of each other. Utilizing this conclusion, we can readily show that

$$E(\hat{\mathbf{B}}_{\text{ML}}) = E_{\mathbf{T}}\left[(\mathbf{A}^H\mathbf{T}^{-1}\mathbf{A})^{-1}\mathbf{A}^H\mathbf{T}^{-1}E_{\mathbf{Z}_S}(\mathbf{A}\mathbf{B}\mathbf{S}\mathbf{S}^H + \mathbf{Z}_{\mathbf{S}})(\mathbf{S}\mathbf{S}^H)^{-1}\right] = \mathbf{B}, \tag{2–19}$$

where $E_{\mathbf{C}}[.]$ denotes calculating the expectation with respect to the random matrix $\mathbf{C}$.

Hence, the multivariate ML estimator is, like its univariate counterpart in [25], unbiased.

2.3.1.2 Mean-squared-error analysis

First, we consider the best possible performance bound for any unbiased estimator of $\mathbf{B}$, i.e., the CRB. Appendix A.1 shows that the CRB of $\mathbf{B}$ based on

the data model in (2–1) has the following form

$$\mathrm{CRB}(\mathbf{B}) \triangleq \mathrm{CRB}(\mathrm{vec}(\mathbf{B})) = (\mathbf{S}^*\mathbf{S}^T)^{-1} \otimes (\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{A})^{-1}, \tag{2–20}$$

where $\mathrm{vec}(\cdot)$ denotes stacking the columns of a matrix on top of each other, $(\cdot)^*$ denotes the complex conjugate and $\otimes$ denotes the Kronecker matrix product [74].

Before calculating the MSE of the multivariate ML estimator, we introduce the following three lemmas which will be used in our derivation. The counterparts of the three lemmas for real-valued variables have been proved and used in the statistics literature [62] [63] [18] and part of Lemma 1 has been proved in [75].

Lemma 2.1. *Suppose $\boldsymbol{\Upsilon}$ is an $m \times m$ random matrix with the complex Wishart distribution with covariance matrix $\boldsymbol{\Sigma}_{m\times m}$ and l ($l \geq m$) degrees of freedom, denoted by $\boldsymbol{\Upsilon} \sim CW(l, m; \boldsymbol{\Sigma})$. Let $\boldsymbol{\Upsilon}$ and $\boldsymbol{\Sigma}$ be partitioned as*

$$\boldsymbol{\Upsilon} = \begin{pmatrix} \boldsymbol{\Upsilon}_{11} & \boldsymbol{\Upsilon}_{12} \\ \boldsymbol{\Upsilon}_{21} & \boldsymbol{\Upsilon}_{22} \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \tag{2–21}$$

where $\boldsymbol{\Upsilon}_{11}$ and $\boldsymbol{\Sigma}_{11}$ are $m_1 \times m_1$ matrices and $\boldsymbol{\Upsilon}_{22}$ and $\boldsymbol{\Sigma}_{22}$ are $m_2 \times m_2$ matrices with $m_1 + m_2 = m$. Let $\boldsymbol{\Upsilon}_{11.2} = \boldsymbol{\Upsilon}_{11} - \boldsymbol{\Upsilon}_{12}\boldsymbol{\Upsilon}_{22}^{-1}\boldsymbol{\Upsilon}_{21}$ and $\boldsymbol{\Sigma}_{11.2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$. Then the following properties hold.

(i). $\boldsymbol{\Upsilon}_{11.2}$ is independent of $\boldsymbol{\Upsilon}_{22}$ and $\boldsymbol{\Upsilon}_{12}$ and $\boldsymbol{\Upsilon}_{11.2} \sim CW(l - m_2, m_1; \boldsymbol{\Sigma}_{11.2})$;

(ii). $\boldsymbol{\Upsilon}_{22} \sim CW(l, m_2; \boldsymbol{\Sigma}_{22})$;

(iii). The conditional distribution of $\boldsymbol{\Upsilon}_{12}$ given $\boldsymbol{\Upsilon}_{22}$ is the matrix-variate complex Gaussian distribution $CN(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Upsilon}_{22}; \boldsymbol{\Sigma}_{11.2}, \boldsymbol{\Upsilon}_{22})$ [74] [75], whose probability density function (pdf) is given by

$$\begin{aligned} f_{\boldsymbol{\Upsilon}_{12}|\boldsymbol{\Upsilon}_{22}} =& (2\pi)^{-m_1 m_2}|\boldsymbol{\Sigma}_{11.2}|^{-m_2}|\boldsymbol{\Upsilon}_{22}|^{-m_1} \\ & exp\Big\{ - tr\big[\boldsymbol{\Sigma}_{11.2}^{-1}(\boldsymbol{\Upsilon}_{12} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Upsilon}_{22})\boldsymbol{\Upsilon}_{22}^{-1}(\boldsymbol{\Upsilon}_{12} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Upsilon}_{22})^H\big]\Big\}. \end{aligned} \tag{2–22}$$

Proof: Appendix A.2.

Lemma 2.2. *Let* $\mathbf{C}$ *be a* $p\times p$ *constant matrix. Suppose* $\boldsymbol{\Upsilon}_{n\times p} \sim CN(\mathbf{H}_{n\times p}; \boldsymbol{\Sigma}_{n\times n}, \boldsymbol{\Omega}_{p\times p})$, *i.e.,*

$$f(\boldsymbol{\Upsilon}) = (2\pi)^{-np}|\boldsymbol{\Sigma}|^{-p}|\boldsymbol{\Omega}|^{-n} exp\Big\{ - tr\big[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Upsilon} - \mathbf{H})\boldsymbol{\Omega}^{-1}(\boldsymbol{\Upsilon} - \mathbf{H})^H\big]\Big\}, \tag{2–23}$$

then

$$E(\boldsymbol{\Upsilon}\mathbf{C}\boldsymbol{\Upsilon}^H) = tr(\boldsymbol{\Omega}\mathbf{C})\boldsymbol{\Sigma} + \mathbf{H}\mathbf{C}\mathbf{H}^H. \tag{2–24}$$

Proof: Appendix A.3.

Lemma 2.3. *If* $\boldsymbol{\Upsilon} \sim CW(l, m; \boldsymbol{\Sigma})$, *then*

$$E(\boldsymbol{\Upsilon}^{-1}) = \frac{\boldsymbol{\Sigma}^{-1}}{(l-m)}. \tag{2–25}$$

Proof: Appendix A.4.

Now we consider the MSE of $\hat{\mathbf{B}}_{\text{ML}}$. The error of the multivariate ML estimate of $\mathbf{B}$ is

$$\begin{aligned} \Delta\mathbf{B} &\triangleq \hat{\mathbf{B}}_{\text{ML}} - \mathbf{B} \\ &= (\mathbf{A}^H\mathbf{T}^{-1}\mathbf{A})^{-1}\mathbf{A}^H\mathbf{T}^{-1}(\mathbf{ABS} + \mathbf{Z})\mathbf{S}^H(\mathbf{SS}^H)^{-1} - \mathbf{B} \\ &= (\mathbf{A}^H\mathbf{T}^{-1}\mathbf{A})^{-1}\mathbf{A}^H\mathbf{T}^{-1}\mathbf{Z}_S(\mathbf{SS}^H)^{-1}. \end{aligned} \tag{2–26}$$

Since $[\mathbf{S}^H(\mathbf{SS}^H)^{-\frac{1}{2}}]^H[\mathbf{S}^H(\mathbf{SS}^H)^{-\frac{1}{2}}] = \mathbf{I}$, i.e., $\mathbf{S}^H(\mathbf{SS}^H)^{-\frac{1}{2}}$ is an $L \times K$ semi-unitary matrix, we can construct an $L \times L$ unitary matrix

$$\mathbf{U} = [\mathbf{U}_1 \quad \mathbf{U}_2], \qquad \mathbf{U}_1 = \mathbf{S}^H(\mathbf{SS}^H)^{-\frac{1}{2}}. \tag{2–27}$$

Thus,

$$\mathbf{T} = \mathbf{Z}[\mathbf{U}\mathbf{U}^H - \mathbf{U}_1\mathbf{U}_1^H]\mathbf{Z}^H = \mathbf{Z}\mathbf{U}_2(\mathbf{Z}\mathbf{U}_2)^H. \tag{2–28}$$

Since the column random vectors of $\mathbf{Z}$ are statistically independent of each other and the columns of $\mathbf{U}_2$ are orthogonal to each other, $\mathbf{Z}\mathbf{U}_2 \sim CN(\mathbf{0}; \mathbf{Q}_{M\times M}, \mathbf{I}_{L-K})$ and $\mathbf{Q}^{-\frac{1}{2}}\mathbf{Z}\mathbf{U}_2 \sim CN(\mathbf{0}; \mathbf{I}_M, \mathbf{I}_{L-K})$ [75]. According to the definition of the complex

Wishart distribution, we have

$$\tilde{\mathbf{T}} \triangleq \mathbf{Q}^{-\frac{1}{2}}\mathbf{T}\mathbf{Q}^{-\frac{1}{2}} = (\mathbf{Q}^{-\frac{1}{2}}\mathbf{Z}\mathbf{U}_2)(\mathbf{Q}^{-\frac{1}{2}}\mathbf{Z}\mathbf{U}_2)^H \sim \mathrm{CW}(L-K, M; \mathbf{I}). \tag{2–29}$$

Let

$$\tilde{\mathbf{Z}}_S = \mathbf{Q}^{-\frac{1}{2}}\mathbf{Z}_S(\mathbf{S}\mathbf{S}^H)^{-\frac{1}{2}}, \tag{2–30}$$

which has the $CN(\mathbf{0}; \mathbf{I}_M, \mathbf{I}_K)$ distribution [75]. Denote

$$\tilde{\mathbf{A}} = \mathbf{Q}^{-\frac{1}{2}}\mathbf{A}(\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{A})^{-\frac{1}{2}}. \tag{2–31}$$

Then, inserting (2–29), (2–30) and (2–31) into (2–26) gives

$$\Delta\mathbf{B} = (\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{A})^{-\frac{1}{2}}(\tilde{\mathbf{A}}^H\tilde{\mathbf{T}}^{-1}\tilde{\mathbf{A}})^{-1}\tilde{\mathbf{A}}^H\tilde{\mathbf{T}}^{-1}\tilde{\mathbf{Z}}_S(\mathbf{S}\mathbf{S}^H)^{-\frac{1}{2}}. \tag{2–32}$$

Since $\tilde{\mathbf{A}}^H\tilde{\mathbf{A}} = \mathbf{I}$, we can decompose $\tilde{\mathbf{A}}$ as

$$\tilde{\mathbf{A}} = \tilde{\mathbf{U}}\mathbf{P}, \tag{2–33}$$

where $\tilde{\mathbf{U}}$ is an $M \times M$ unitary matrix with its first N columns being $\tilde{\mathbf{A}}$; $\mathbf{P}_{M\times N} = [\mathbf{I}_N \quad \mathbf{0}]^T$. Since $\tilde{\mathbf{U}}$ is unitary, like $\tilde{\mathbf{T}}$, $\tilde{\mathbf{U}}^H\tilde{\mathbf{T}}\tilde{\mathbf{U}}$ remains to be the complex Wishart distribution [75], i.e.,

$$\mathbf{\Gamma} \triangleq \tilde{\mathbf{U}}^H\tilde{\mathbf{T}}\tilde{\mathbf{U}} \sim \mathrm{CW}(L-K, M; \mathbf{I}). \tag{2–34}$$

Let

$$\mathbf{\Xi} = \tilde{\mathbf{U}}^H\tilde{\mathbf{Z}}_S, \tag{2–35}$$

which obviously has the $CN(0; \mathbf{I}_M, \mathbf{I}_K)$ distribution.

We next partition $\mathbf{\Xi}_{M\times K}$, $\mathbf{\Gamma}_{M\times M}$, and $\mathbf{\Gamma}^{-1}_{M\times M}$, respectively, as follows.

$$\mathbf{\Xi} = \begin{bmatrix} \mathbf{\Xi}_1 \\ \mathbf{\Xi}_2 \end{bmatrix}, \qquad \mathbf{\Gamma} = \begin{bmatrix} \mathbf{\Gamma}_{11} & \mathbf{\Gamma}_{12} \\ \mathbf{\Gamma}_{21} & \mathbf{\Gamma}_{22} \end{bmatrix}, \qquad \text{and} \quad \mathbf{\Gamma}^{-1} = \begin{bmatrix} \mathbf{\Gamma}^{11} & \mathbf{\Gamma}^{12} \\ \mathbf{\Gamma}^{21} & \mathbf{\Gamma}^{22} \end{bmatrix}, \tag{2–36}$$

where $\mathbf{\Xi}_1$ and $\mathbf{\Xi}_2$ are $N \times K$ and $(M-N) \times K$, respectively, and both $\mathbf{\Gamma}_{11}$ and $\mathbf{\Gamma}^{11}$ are $N \times N$ matrices.

Inserting (2–33) to (2–36) into (2–32) gives

$$\begin{aligned}\Delta\mathbf{B} &= (\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{A})^{-\frac{1}{2}}(\mathbf{P}^H\mathbf{\Gamma}^{-1}\mathbf{P})^{-1}\mathbf{P}^H\mathbf{\Gamma}^{-1}\mathbf{\Xi}(\mathbf{S}\mathbf{S}^H)^{-\frac{1}{2}} \\ &= (\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{A})^{-\frac{1}{2}}(\mathbf{\Xi}_1 - \mathbf{\Gamma}_{12}\mathbf{\Gamma}_{22}^{-1}\mathbf{\Xi}_2)(\mathbf{S}\mathbf{S}^H)^{-\frac{1}{2}}.\end{aligned} \tag{2–37}$$

To obtain (2–37), we have used the inversion lemma of partitioned matrices.

Hence, using the lemma that $\mathrm{vec}(\mathbf{XYZ}) = (\mathbf{Z}^T \otimes \mathbf{X})\mathrm{vec}(\mathbf{Y})$ and the fact that $\mathbf{\Xi}_1$, $\mathbf{\Xi}_2$ are two independent random matrices with the distributions $CN(\mathbf{0}; \mathbf{I}_N, \mathbf{I}_K)$ and $CN(\mathbf{0}; \mathbf{I}_{M-N}, \mathbf{I}_K)$, respectively, as well as (2–20), we have

$$\begin{aligned}\mathrm{MSE}(\hat{\mathbf{B}}_{\mathrm{ML}}) \triangleq& E\{\mathrm{vec}(\Delta\mathbf{B})\,\mathrm{vec}(\Delta\mathbf{B})^H\} \\ =&[(\mathbf{S}^*\mathbf{S}^T)^{-\frac{1}{2}} \otimes (\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{A})^{-\frac{1}{2}}]E\{\mathrm{vec}(\mathbf{\Xi}_1 - \mathbf{\Gamma}_{12}\mathbf{\Gamma}_{22}^{-1}\mathbf{\Xi}_2) \\ &\mathrm{vec}(\mathbf{\Xi}_1 - \mathbf{\Gamma}_{12}\mathbf{\Gamma}_{22}^{-1}\mathbf{\Xi}_2)^H\}[(\mathbf{S}^*\mathbf{S}^T)^{-\frac{1}{2}} \otimes (\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{A})^{-\frac{1}{2}}] \\ =&[\mathrm{CRB}(\mathbf{B})]^{\frac{1}{2}}\,\{\mathbf{I} + E[\mathrm{vec}(\mathbf{\Gamma}_{12}\mathbf{\Gamma}_{22}^{-1}\mathbf{\Xi}_2)\mathrm{vec}(\mathbf{\Gamma}_{12}\mathbf{\Gamma}_{22}^{-1}\mathbf{\Xi}_2)^H]\}\,[\mathrm{CRB}(\mathbf{B})]^{\frac{1}{2}}.\end{aligned} \tag{2–38}$$

Using the facts that $\mathrm{vec}(\mathbf{XY}) = (\mathbf{I} \otimes \mathbf{X})\mathrm{vec}(\mathbf{Y})$ and $\mathbf{\Xi}_2 \sim CN(\mathbf{0}; \mathbf{I}, \mathbf{I})$ yields

$$\begin{aligned}&E[\mathrm{vec}(\mathbf{\Gamma}_{12}\mathbf{\Gamma}_{22}^{-1}\mathbf{\Xi}_2)\,\mathrm{vec}(\mathbf{\Gamma}_{12}\mathbf{\Gamma}_{22}^{-1}\mathbf{\Xi}_2)^H] \\ =\ & E\{[\mathbf{I} \otimes (\mathbf{\Gamma}_{12}\mathbf{\Gamma}_{22}^{-1})]E_{\mathbf{\Xi}_2|\mathbf{\Gamma}_{12},\mathbf{\Gamma}_{22}}[\mathrm{vec}(\mathbf{\Xi}_2)\,\mathrm{vec}(\mathbf{\Xi}_2)^H][\mathbf{I} \otimes (\mathbf{\Gamma}_{12}\mathbf{\Gamma}_{22}^{-1})]^H\} \\ =\ & \mathbf{I} \otimes E(\mathbf{\Gamma}_{12}\mathbf{\Gamma}_{22}^{-1}\mathbf{\Gamma}_{22}^{-1}\mathbf{\Gamma}_{12}^H),\end{aligned} \tag{2–39}$$

where $E_{\mathbf{A}|\mathbf{B}}(\cdot)$ denotes the expectation with respect to random matrix $\mathbf{A}$ given $\mathbf{B}$.

By Lemma 2.1 and (2–34), we know that $\mathbf{\Gamma}_{12}$ given $\mathbf{\Gamma}_{22}$ has the $CN(\mathbf{0}; \mathbf{I}, \mathbf{\Gamma}_{22})$ distribution. Hence, applying Lemma 2.2 gives

$$\begin{aligned}E(\mathbf{\Gamma}_{12}\mathbf{\Gamma}_{22}^{-1}\mathbf{\Gamma}_{22}^{-1}\mathbf{\Gamma}_{12}^H) &= E_{\mathbf{\Gamma}_{22}}\left[E_{\mathbf{\Gamma}_{12}|\mathbf{\Gamma}_{22}}(\mathbf{\Gamma}_{12}\mathbf{\Gamma}_{22}^{-2}\mathbf{\Gamma}_{12}^H)\right] \\ &= \left\{E_{\mathbf{\Gamma}_{22}}[\mathrm{tr}(\mathbf{\Gamma}_{22}^{-1})]\right\}\,\mathbf{I} \\ &= \left\{\mathrm{tr}[E_{\mathbf{\Gamma}_{22}}(\mathbf{\Gamma}_{22}^{-1})]\right\}\,\mathbf{I}.\end{aligned} \tag{2–40}$$

Furthermore, by Lemma 2.1 we know that $\mathbf{\Gamma}_{22} \sim \mathrm{CW}(L-K, M-N; \mathbf{I}_{M-N})$. Hence, by Lemma 3, we have

$$E_{\mathbf{\Gamma}_{22}}(\mathbf{\Gamma}_{22}^{-1}) = \frac{1}{L-K-M+N}\,\mathbf{I}_{M-N}. \tag{2–41}$$

Thus, it follows from (2–38), (2–39), (2–40) and (2–41) that

$$\mathrm{MSE}(\hat{\mathbf{B}}_{\mathrm{ML}}) = \frac{L-K}{L-K-M+N}\mathrm{CRB}(\mathbf{B}). \tag{2–42}$$

From (2–42), we note that $\mathrm{MSE}(\hat{\mathbf{B}}_{\mathrm{ML}})$ approaches $\mathrm{CRB}(\mathbf{B})$ for large L, which means that the multivariate ML estimator is asymptotically statistically efficient for large number of snapshots L. Hence the efficiency condition for the univariate case in [25] can be extended to the multivariate case as well. When L, K, M and N are fixed, the MSE of the multivariate ML estimator is proportional to $\mathrm{CRB}(\mathbf{B})$. Hence it is expected that the MSE-versus-SNR lines will be parallel to the CRB-versus-SNR lines. This theoretical result will be verified via numerical simulations in Section 2.4.

Furthermore, the CRB of $\mathbf{B}$ depends on $(\mathbf{S}^*\mathbf{S}^T)^{-1}$ and $(\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{A})^{-1}$. As we show in Appendix A.1, orthogonalities among the rows of $\mathbf{S}$ and among the columns of $\mathbf{Q}^{-\frac{1}{2}}\mathbf{A}$ lead to small diagonal elements for $(\mathbf{S}^*\mathbf{S}^T)^{-1}$ and $(\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{A})^{-1}$, respectively, which in turn reduce the CRB.

We also note that when $M = N$, which implies that $\mathbf{A}$ is a square matrix, the multivariate ML estimator is efficient. However, we should not think of it as a significant advantage to make N as large as possible. As we show in Appendix A.1, in the case that the columns of $\mathbf{Q}^{-\frac{1}{2}}\mathbf{A}$ are not orthogonal to each other, which often happens in practice, large N causes CRB to increase.

Now we summarize the statistical properties of the multivariate Capon estimator by the following theorem.

Theorem 2.1. *For the data model in (2–1), the multivariate ML estimate of* $\mathbf{B}$*, given by (2–14), is unbiased and asymptotically statistically efficient for large number of data samples. Its MSE matrix can be expressed as*

$$MSE(\hat{\mathbf{B}}_{ML}) \triangleq E[vec(\hat{\mathbf{B}}_{ML})vec(\hat{\mathbf{B}}_{ML})^H] = \frac{L-K}{L-K-M+N} CRB(\mathbf{B}) \tag{2–43}$$

where

$$CRB(\mathbf{B}) = (\mathbf{S}^*\mathbf{S}^T)^{-1} \otimes (\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{A})^{-1}, \tag{2–44}$$

$vec(\cdot)$, $(\cdot)^*$, $(\cdot)^T$ *and* $\otimes$ *denote the direct operator (stacking the columns of a matrix on top of each other), complex conjugate, transpose and Kronecker product of matrices, respectively.*

2.3.2 Performance Analysis of the Multivariate Capon Estimator

We now establish the theoretical properties of the multivariate Capon estimator.

2.3.2.1 Bias analysis

In Section 2.3.1, we know that the multivariate ML estimator is unbiased. We will investigate the bias of the multivariate Capon estimator by studying the relationship between the two estimators.

Comparing (2–2) and (2–13), we note that

$$\mathbf{R} = \mathbf{T} + \mathbf{X}\mathbf{S}^H(\mathbf{S}\mathbf{S}^H)^{-1}\mathbf{S}\mathbf{X}^H. \tag{2–45}$$

Applying the matrix inversion lemma gives

$$\mathbf{A}^H\mathbf{R}^{-1}\mathbf{X}\mathbf{S}^H(\mathbf{S}\mathbf{S}^H)^{-1} = \mathbf{A}^H\mathbf{T}^{-1}\mathbf{X}\mathbf{S}^H(\mathbf{S}\mathbf{X}^H\mathbf{T}^{-1}\mathbf{X}\mathbf{S}^H + \mathbf{S}\mathbf{S}^H)^{-1} \tag{2–46}$$

and

$$
\begin{aligned}
&(\mathbf{A}^H\mathbf{R}^{-1}\mathbf{A})^{-1} \\
&= [\mathbf{A}^H\mathbf{T}^{-1}\mathbf{A} - \mathbf{A}^H\mathbf{T}^{-1}\mathbf{X}\mathbf{S}^H(\mathbf{S}\mathbf{X}^H\mathbf{T}^{-1}\mathbf{X}\mathbf{S}^H + \mathbf{S}\mathbf{S}^H)^{-1}\mathbf{S}\mathbf{X}^H\mathbf{T}^{-1}\mathbf{A}]^{-1} \\
&= (\mathbf{A}^H\mathbf{T}^{-1}\mathbf{A})^{-1} - (\mathbf{A}^H\mathbf{T}^{-1}\mathbf{A})^{-1}\mathbf{A}^H\mathbf{T}^{-1}\mathbf{X}\mathbf{S}^H \\
&\quad \left[\mathbf{S}\mathbf{X}^H\mathbf{T}^{-1}\mathbf{A}(\mathbf{A}^H\mathbf{T}^{-1}\mathbf{A})^{-1}\mathbf{A}^H\mathbf{T}^{-1}\mathbf{X}\mathbf{S}^H - \mathbf{S}\mathbf{X}^H\mathbf{T}^{-1}\mathbf{X}\mathbf{S}^H - \mathbf{S}\mathbf{S}^H\right]^{-1} \\
&\quad \mathbf{S}\mathbf{X}^H\mathbf{T}^{-1}\mathbf{A}(\mathbf{A}^H\mathbf{T}^{-1}\mathbf{A})^{-1}.
\end{aligned}
\tag{2–47}
$$

Substituting (2–46) and (2–47) into (2–8), and after some straightforward manipulations, we get

$$
\hat{\mathbf{B}}_{\text{Capon}} = \hat{\mathbf{B}}_{\text{ML}}\mathbf{\Lambda}, \tag{2–48}
$$

where

$$
\mathbf{\Lambda} = [\mathbf{I} + \mathbf{V}]^{-1}, \tag{2–49}
$$

with

$$
\mathbf{V} = \mathbf{S}\mathbf{X}^H\mathbf{T}^{-1}\mathbf{X}\mathbf{S}^H(\mathbf{S}\mathbf{S}^H)^{-1} - \mathbf{S}\mathbf{X}^H\mathbf{T}^{-1}\mathbf{A}(\mathbf{A}^H\mathbf{T}^{-1}\mathbf{A})^{-1}\mathbf{A}^H\mathbf{T}^{-1}\mathbf{X}\mathbf{S}^H(\mathbf{S}\mathbf{S}^H)^{-1}. \tag{2–50}
$$

Then inserting (2–1) into (2–50) gives

$$
\mathbf{V} = \mathbf{Z}_S^H[\mathbf{T}^{-1} - \mathbf{T}^{-1}\mathbf{A}(\mathbf{A}^H\mathbf{T}^{-1}\mathbf{A})^{-1}\mathbf{A}^H\mathbf{T}^{-1}]\mathbf{Z}_S(\mathbf{S}\mathbf{S}^H)^{-1}. \tag{2–51}
$$

Note that there are two random matrices, i.e., $\mathbf{T}$ and $\mathbf{Z}_S$, in (2–26) and (2–51). Since the columns of $\mathbf{Z}$ are statistically independent zero-mean Guassian random vectors while the columns of $\mathbf{S}^H$ are orthogonal to those of $\mathbf{I} - \mathbf{S}^H(\mathbf{S}\mathbf{S}^H)\mathbf{S}$, by the property of joint Gaussian distribution we know that $\mathbf{Z}_S$ and $\mathbf{Z}_S^{\perp} \triangleq \mathbf{Z}[\mathbf{I} - \mathbf{S}^H(\mathbf{S}\mathbf{S}^H)\mathbf{S}]$ are two independent Gaussian random matrices. Hence, $\mathbf{Z}_S$ and $\mathbf{T} = \mathbf{Z}_S^{\perp}(\mathbf{Z}_S^{\perp})^H$ are also independent of each other. By Lemmas 1.9 and 1.11 in [18], which can be readily extended to the complex-valued case, we have $\mathbf{Z}_S \sim \text{CN}(\mathbf{0}; \mathbf{Q}, \mathbf{S}\mathbf{S}^H)$ and $\mathbf{T} \sim \text{CW}(L - K, M; \mathbf{Q})$.

Since $\mathbf{Z}_S$ and $\mathbf{T}$ are statistically independent of each other and by (2–26) and (2–51), we know that $(\hat{\mathbf{B}}_{\text{ML}} - \mathbf{B})\mathbf{\Lambda}$ is an odd function with respect to $\mathbf{Z}_S$. Hence replacing $\mathbf{Z}_S$ with $-\mathbf{Z}_S$ yields

$$E[(\hat{\mathbf{B}}_{\text{ML}} - \mathbf{B})\mathbf{\Lambda}]|_{\mathbf{Z}_S=-\mathbf{Z}_S} = -E[(\hat{\mathbf{B}}_{\text{ML}} - \mathbf{B})\mathbf{\Lambda}]. \tag{2–52}$$

On the other hand, since $\mathbf{Z}_S$ is a zero-mean Gaussian random matrix, $-\mathbf{Z}_S$, as a random matrix transformed from $\mathbf{Z}_S$, retains all the statistical properties of $\mathbf{Z}_S$. Hence, replacing $\mathbf{Z}_S$ by $-\mathbf{Z}_S$ will not change the expectation of $(\hat{\mathbf{B}}_{\text{ML}} - \mathbf{B})\mathbf{\Lambda}$, i.e.,

$$E[(\hat{\mathbf{B}}_{\text{ML}} - \mathbf{B})\mathbf{\Lambda}]|_{\mathbf{Z}_S=-\mathbf{Z}_S} = E[(\hat{\mathbf{B}}_{\text{ML}} - \mathbf{B})\mathbf{\Lambda}]. \tag{2–53}$$

It follows from (2–52) and (2–53) that

$$E[(\hat{\mathbf{B}}_{\text{ML}} - \mathbf{B})\mathbf{\Lambda}] = \mathbf{0}. \tag{2–54}$$

Therefore, by (2–48) and (2–54) we have

$$E(\hat{\mathbf{B}}_{\text{Capon}}) = \mathbf{B}E(\mathbf{\Lambda}). \tag{2–55}$$

Now we follow the same technique used in the previous subsection to simplify $\mathbf{\Lambda}$ and $\mathbf{V}$ via transformation of random matrices.

Following the definitions in (2–29), (2–30) and (2–31) and inserting them into (2–51), we get

$$\mathbf{V} = (\mathbf{S}\mathbf{S}^H)^{\frac{1}{2}}\tilde{\mathbf{Z}}_S^H\left[\tilde{\mathbf{T}}^{-1} - \tilde{\mathbf{T}}^{-1}\tilde{\mathbf{A}}(\tilde{\mathbf{A}}^H\tilde{\mathbf{T}}^{-1}\tilde{\mathbf{A}})^{-1}\tilde{\mathbf{A}}^H\tilde{\mathbf{T}}^{-1}\right]\tilde{\mathbf{Z}}_S(\mathbf{S}\mathbf{S})^{-\frac{1}{2}}. \tag{2–56}$$

Then we adopt the decomposition in (2–33), the definitions in (2–34) and (2–35), and the partitions in (2–36), and insert them into (2–56). By the inversion

lemma of partitioned matrices, we obtain

$$\begin{aligned}\mathbf{V} &= (\mathbf{S}\mathbf{S}^H)^{\frac{1}{2}}\boldsymbol{\Xi}^H\left[\boldsymbol{\Gamma}^{-1} - \boldsymbol{\Gamma}^{-1}\mathbf{P}(\mathbf{P}^H\boldsymbol{\Gamma}^{-1}\mathbf{P})^{-1}\mathbf{P}^H\boldsymbol{\Gamma}^{-1}\right]\boldsymbol{\Xi}(\mathbf{S}\mathbf{S}^H)^{-\frac{1}{2}} \\ &= (\mathbf{S}\mathbf{S}^H)^{\frac{1}{2}}\boldsymbol{\Xi}^H\left\{\begin{bmatrix}\boldsymbol{\Gamma}^{11} & \boldsymbol{\Gamma}^{12} \\ \boldsymbol{\Gamma}^{21} & \boldsymbol{\Gamma}^{22}\end{bmatrix} - \begin{bmatrix}\boldsymbol{\Gamma}^{11} & \boldsymbol{\Gamma}^{12} \\ \boldsymbol{\Gamma}^{21} & \boldsymbol{\Gamma}^{21}(\boldsymbol{\Gamma}^{11})^{-1}\boldsymbol{\Gamma}^{12}\end{bmatrix}\right\}\boldsymbol{\Xi}(\mathbf{S}\mathbf{S}^H)^{-\frac{1}{2}} \\ &= (\mathbf{S}\mathbf{S}^H)^{\frac{1}{2}}\boldsymbol{\Xi}_2^H\boldsymbol{\Gamma}_{22}^{-1}\boldsymbol{\Xi}_2(\mathbf{S}\mathbf{S}^H)^{-\frac{1}{2}}.\end{aligned} \tag{2–57}$$

From (2–49) and (2–57) and by the matrix inversion lemma, it follows that

$$\begin{aligned}\boldsymbol{\Lambda} &= (\mathbf{S}\mathbf{S}^H)^{\frac{1}{2}}\left[\mathbf{I} + \boldsymbol{\Xi}_2^H\boldsymbol{\Gamma}_{22}^{-1}\boldsymbol{\Xi}_2\right]^{-1}(\mathbf{S}\mathbf{S}^H)^{-\frac{1}{2}} \\ &= \mathbf{I} - (\mathbf{S}\mathbf{S}^H)^{\frac{1}{2}}\boldsymbol{\Xi}_2^H(\boldsymbol{\Gamma}_{22} + \boldsymbol{\Xi}_2\boldsymbol{\Xi}_2^H)^{-1}\boldsymbol{\Xi}_2(\mathbf{S}\mathbf{S}^H)^{-\frac{1}{2}}.\end{aligned} \tag{2–58}$$

To calculate the expectation of $\boldsymbol{\Lambda}$, we use the following lemma.

Lemma 2.4. *Let $\boldsymbol{\Upsilon}_{n\times p}$ and $\boldsymbol{\Psi}_{n\times n}$ be two independent random matrices, and*

$$\boldsymbol{\Upsilon} \sim CN(0; \mathbf{I}_n, \mathbf{I}_p), \qquad \boldsymbol{\Psi} \sim CW(l, n; \mathbf{I}_n). \tag{2–59}$$

Denote $\boldsymbol{\Pi} = \boldsymbol{\Upsilon}^H(\boldsymbol{\Psi} + \boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^H)^{-1}\boldsymbol{\Upsilon}$. Then the expectation and correlation matrices of the random matrix $\boldsymbol{\Pi}$ are

$$E(\boldsymbol{\Pi}) = \frac{n}{l+p}\mathbf{I}_p, \tag{2–60}$$

$$E(vec(\boldsymbol{\Pi})vec(\boldsymbol{\Pi})^H) = \frac{n(n+1)}{(l+p)(l+p+1)}vec(\mathbf{I}_p)\,vec(\mathbf{I}_p)^H + \eta\,\mathbf{D}_p, \tag{2–61}$$

where η is a scalar and approximately equal to $\frac{n(l+p-n)}{(l+p)^2(l+p+1)}$ for large $l+p$, and D_p is a $p^2\times p^2$ matrix with its element at the $[(c_1-1)p+r_1]$th row and the $[(c_2-1)p+r_2]$th column $(r_1, c_1, r_2, c_2 = 1, 2, \ldots p)$ being

$$d_{(c_1-1)p+r_1,(c_2-1)p+r_2} = \begin{cases} 1 & \text{when } r_1 = r_2,\ c_1 = c_2 \text{ but } r_1 \neq c_1 \\ -1 & \text{when } r_1 = c_1,\ r_2 = c_2 \text{ but } r_1 \neq r_2 \\ 0 & \text{otherwise.}\end{cases} \tag{2–62}$$

Proof: Appendix A.5.

Applying the above lemma to $\mathbf{\Xi}$ and $\mathbf{\Gamma}$, which by construction satisfy the assumptions in the lemma, we have immediately

$$E(\mathbf{\Lambda}) = \left(1 - \frac{M-N}{L}\right)\mathbf{I}. \tag{2–63}$$

Inserting (2–63) into (2–55), we get

$$E(\hat{\mathbf{B}}_{\text{Capon}}) = \left(1 - \frac{M-N}{L}\right)\mathbf{B}. \tag{2–64}$$

The above equation shows that the multivariate Capon estimator shares the same properties as the univariate Capon in [25]. In other words, it is biased downward for finite snapshot number L. However, for large L, it is asymptotically unbiased. It is also worth noting that the bias of the Capon estimator is not related to K, which means that increasing the number of rows in the temporal information matrix $\mathbf{S}$ will not cause higher bias.

Moreover, we note that when $M = N$, the multivariate Capon estimator becomes unbiased as the multivariate ML estimator. In this case, both ML and Capon reduce to the same estimator $\mathbf{A}^{-1}\mathbf{X}\mathbf{S}^H(\mathbf{S}\mathbf{S}^H)^{-1}$. Hence, for the same reason that we have stated in Section 2.3.1, this unbiasedness of the multivariate Capon estimator should not be seen as a significant advantage.

2.3.2.2 Mean-squared-error analysis

We investigate the MSE of the multivariate Capon estimator below. Using the same technique to obtain (2–54), we can prove that $\Delta\mathbf{B}\mathbf{\Lambda}$ and $\mathbf{B}(\mathbf{I}-\mathbf{\Lambda})$ are uncorrelated. Hence, from (2–26) and (2–48), we have

$$\begin{aligned}
&\text{MSE}(\hat{\mathbf{B}}_{\text{Capon}}) \\
\triangleq &E[\text{vec}(\hat{\mathbf{B}}_{\text{Capon}} - \mathbf{B})\,\text{vec}(\hat{\mathbf{B}}_{\text{Capon}} - \mathbf{B})^H] \\
= &E\{\text{vec}[\Delta\mathbf{B}\mathbf{\Lambda} - \mathbf{B}(\mathbf{I}-\mathbf{\Lambda})]\,\text{vec}[\Delta\mathbf{B}\mathbf{\Lambda} - \mathbf{B}(\mathbf{I}-\mathbf{\Lambda})]^H\} \\
= &E[\text{vec}(\Delta\mathbf{B}\mathbf{\Lambda})\,\text{vec}(\Delta\mathbf{B}\mathbf{\Lambda})^H] + E[\text{vec}(\mathbf{B}(\mathbf{I}-\mathbf{\Lambda}))\,\text{vec}(\mathbf{B}(\mathbf{I}-\mathbf{\Lambda}))^H].
\end{aligned} \tag{2–65}$$

We first calculate $E[\text{vec}(\Delta\mathbf{B}\mathbf{\Lambda})\text{vec}(\Delta\mathbf{B}\mathbf{\Lambda})^H]$.

By (2–37) and (2–58), we have

$$\Delta\mathbf{B}\mathbf{\Lambda} = (\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{A})^{-\frac{1}{2}}[\mathbf{\Xi}_1 - \mathbf{\Gamma}_{12}\mathbf{\Gamma}_{22}^{-1}\mathbf{\Xi}_2][\mathbf{I} + \mathbf{\Xi}_2^H\mathbf{\Gamma}_{22}^{-1}\mathbf{\Xi}_2]^{-1}(\mathbf{S}\mathbf{S}^H)^{-\frac{1}{2}}. \tag{2–66}$$

Using the fact that $\text{vec}(\mathbf{XYZ}) = (\mathbf{Z}^T \otimes \mathbf{X})\text{vec}(\mathbf{Y})$ as well as (2–20) yields

$$E[\text{vec}(\Delta\mathbf{B}\mathbf{\Lambda})\text{vec}(\Delta\mathbf{B}\mathbf{\Lambda})^H] = [\text{CRB}(\mathbf{B})]^{\frac{1}{2}}\,\mathbf{F}\,[\text{CRB}(\mathbf{B})]^{\frac{1}{2}}, \tag{2–67}$$

where

$$\begin{aligned}\mathbf{F} =& E\Big\{\text{vec}[(\mathbf{\Xi}_1 - \mathbf{\Gamma}_{12}\mathbf{\Gamma}_{22}^{-1}\mathbf{\Xi}_2)(\mathbf{I} + \mathbf{\Xi}_2^H\mathbf{\Gamma}_{22}^{-1}\mathbf{\Xi}_2)^{-1}] \\ & \text{vec}[(\mathbf{\Xi}_1 - \mathbf{\Gamma}_{12}\mathbf{\Gamma}_{22}^{-1}\mathbf{\Xi}_2)(\mathbf{I} + \mathbf{\Xi}_2^H\mathbf{\Gamma}_{22}^{-1}\mathbf{\Xi}_2)^{-1}]^H\Big\}.\end{aligned} \tag{2–68}$$

Then using the lemma that $\text{vec}(\mathbf{XY}) = (\mathbf{Y}^T \otimes \mathbf{I})\text{vec}(\mathbf{X})$ and the fact that $\mathbf{\Xi}_1$ and $\mathbf{\Xi}_2$ are independent standard matrix-variate Gaussian distributions, after some manipulations, we get

$$\begin{aligned}\mathbf{F} =& E\Big\{\big[\big(\mathbf{I} + (\mathbf{\Xi}_2^H\mathbf{\Gamma}_{22}^{-1}\mathbf{\Xi}_2)^*\big)^{-1} \otimes \mathbf{I}\big] \\ & \big[\mathbf{I} + E_{\mathbf{\Gamma}_{12}|\mathbf{\Gamma}_{22}}\big(\text{vec}(\mathbf{\Gamma}_{12}\mathbf{\Gamma}_{22}^{-1}\mathbf{\Xi}_2)\,\text{vec}(\mathbf{\Gamma}_{12}\mathbf{\Gamma}_{22}^{-1}\mathbf{\Xi}_2)^H\big)\big] \\ & \big[\big(\mathbf{I} + (\mathbf{\Xi}_2^H\mathbf{\Gamma}_{22}^{-1}\mathbf{\Xi}_2)^*\big)^{-1} \otimes \mathbf{I}\big]\Big\}.\end{aligned} \tag{2–69}$$

Note that

$$\begin{aligned}& E_{\mathbf{\Gamma}_{12}|\mathbf{\Gamma}_{22}}[\text{vec}(\mathbf{\Gamma}_{12}\mathbf{\Gamma}_{22}^{-1}\mathbf{\Xi}_2)\,\text{vec}(\mathbf{\Gamma}_{12}\mathbf{\Gamma}_{22}^{-1}\mathbf{\Xi}_2)^H] \\ =& [(\mathbf{\Gamma}_{22}^{-1}\mathbf{\Xi}_2)^T \otimes \mathbf{I}]E_{\mathbf{\Gamma}_{12}|\mathbf{\Gamma}_{22}}[\text{vec}(\mathbf{\Gamma}_{12})\,\text{vec}(\mathbf{\Gamma}_{12})^H][(\mathbf{\Gamma}_{22}^{-1}\mathbf{\Xi}_2)^* \otimes \mathbf{I}] \\ =& [(\mathbf{\Gamma}_{22}^{-1}\mathbf{\Xi}_2)^T \otimes \mathbf{I}][\mathbf{\Gamma}_{22}^* \otimes \mathbf{I}][(\mathbf{\Gamma}_{22}^{-1}\mathbf{\Xi}_2)^* \otimes \mathbf{I}] \\ =& (\mathbf{\Xi}_2^H\mathbf{\Gamma}_{22}^{-1}\mathbf{\Xi}_2)^* \otimes \mathbf{I}.\end{aligned} \tag{2–70}$$

To get the above equation, we have utilized Lemma 2 in [76], i.e., $\mathbf{\Gamma}_{12}|\mathbf{\Gamma}_{22} \sim \text{CN}(\mathbf{0}; \mathbf{I}, \mathbf{\Gamma}_{22})$. Hence, by the complex-valued counterpart of Lemma 1.8 in [18], we know that the covariance matrix of $\text{vec}(\mathbf{\Gamma}_{12})$ given $\mathbf{\Gamma}_{22}$ is $\mathbf{\Gamma}_{22}^* \otimes \mathbf{I}$.

Inserting (2–70) into (2–69) and recalling (2–58) and (2–63) yield

$$\begin{aligned}\mathbf{F} &= E\left\{(\mathbf{I}+(\boldsymbol{\Xi}_2^H\boldsymbol{\Gamma}_{22}^{-1}\boldsymbol{\Xi}_2)^*)^{-1}\right\}\otimes\mathbf{I}\\ &= [(\mathbf{S}\mathbf{S}^H)^{-\frac{1}{2}}E(\boldsymbol{\Lambda}^*)(\mathbf{S}\mathbf{S}^H)^{\frac{1}{2}}]\otimes\mathbf{I}\\ &= \left(1-\frac{M-N}{L}\right)\mathbf{I}.\end{aligned} \tag{2–71}$$

From (2–67) and (2–71), the equation

$$E[\mathrm{vec}(\Delta\mathbf{B}\boldsymbol{\Lambda})\,\mathrm{vec}(\Delta\mathbf{B}\boldsymbol{\Lambda})^H] = \frac{L-M+N}{L}\mathrm{CRB}(\mathbf{B}) \tag{2–72}$$

follows directly.

Now we consider the second term in (2–65). By (2–58), we know that

$$\mathbf{B}(\mathbf{I}-\boldsymbol{\Lambda}) = \mathbf{B}(\mathbf{S}\mathbf{S}^H)^{\frac{1}{2}}\boldsymbol{\Xi}_2^H(\boldsymbol{\Gamma}_{22}+\boldsymbol{\Xi}_2\boldsymbol{\Xi}_2^H)^{-1}\boldsymbol{\Xi}_2(\mathbf{S}\mathbf{S}^H)^{-\frac{1}{2}}. \tag{2–73}$$

We know that $\boldsymbol{\Xi}_2$ and $\boldsymbol{\Gamma}_{22}$ are independent of each other with $\mathrm{CN}(\mathbf{0};\mathbf{I}_{M-N},\mathbf{I}_K)$ and $\mathrm{CW}(L-K,M-N;\mathbf{I}_{M-N})$ distributions, respectively. Then using the fact that $\mathrm{vec}(\mathbf{XYZ}) = (\mathbf{Z}^T\otimes\mathbf{X})\mathrm{vec}(\mathbf{Y})$, the following equation is obtained following Lemma 2.2.

$$\begin{aligned}&E\left\{\mathrm{vec}[\mathbf{B}(\mathbf{I}-\boldsymbol{\Lambda})]\,\mathrm{vec}[\mathbf{B}(\mathbf{I}-\boldsymbol{\Lambda})]^H\right\}\\ &= \left\{(\mathbf{S}^*\mathbf{S}^T)^{-\frac{1}{2}}\otimes[\mathbf{B}(\mathbf{S}\mathbf{S}^H)^{\frac{1}{2}}]\right\}E\left\{\mathrm{vec}[\boldsymbol{\Xi}_2^H(\boldsymbol{\Gamma}_{22}+\boldsymbol{\Xi}_2\boldsymbol{\Xi}_2^H)^{-1}\boldsymbol{\Xi}_2]\right.\\ &\quad\left.\mathrm{vec}[\boldsymbol{\Xi}_2^H(\boldsymbol{\Gamma}_{22}+\boldsymbol{\Xi}_2\boldsymbol{\Xi}_2^H)^{-1}\boldsymbol{\Xi}_2]^H\right\}\left\{(\mathbf{S}^*\mathbf{S}^T)^{-\frac{1}{2}}\otimes[\mathbf{B}(\mathbf{S}\mathbf{S}^H)^{\frac{1}{2}}]\right\}^H\\ &= \frac{(M-N)(M-N+1)}{L(L+1)}\mathrm{vec}(\mathbf{B})\,\mathrm{vec}(\mathbf{B})^H\\ &\quad+\zeta\left\{(\mathbf{S}^*\mathbf{S}^T)^{-\frac{1}{2}}\otimes[\mathbf{B}(\mathbf{S}\mathbf{S}^H)^{\frac{1}{2}}]\right\}\mathbf{D}_K\left\{(\mathbf{S}^*\mathbf{S}^T)^{-\frac{1}{2}}\otimes[(\mathbf{S}\mathbf{S}^H)^{\frac{1}{2}}\mathbf{B}^H]\right\},\end{aligned} \tag{2–74}$$

where $\mathbf{D}_K$ is a $K^2\times K^2$ matrix defined as (2–62), and ζ is a scalar and approximately equal to $\frac{(M-N)(L-M+N)}{L^2(L+1)}$ for large L.

By (2–65), (2–72) and (2–74), we get the MSE of the multivariate Capon estimator

$$\begin{aligned}\mathrm{MSE}(\hat{\mathbf{B}}_{\text{Capon}}) = &\frac{L-M+N}{L}\mathrm{CRB}(\mathbf{B}) \\ &+ \frac{(M-N)(M-N+1)}{L(L+1)}\mathrm{vec}(\mathbf{B})\,\mathrm{vec}(\mathbf{B})^H \\ &+ \zeta\left\{(\mathbf{S}^*\mathbf{S}^T)^{-\frac{1}{2}} \otimes [\mathbf{B}(\mathbf{S}\mathbf{S}^H)^{\frac{1}{2}}]\right\}\mathbf{D}_K\left\{(\mathbf{S}^*\mathbf{S}^T)^{-\frac{1}{2}} \otimes [(\mathbf{S}\mathbf{S}^H)^{\frac{1}{2}}\mathbf{B}^H]\right\}.\end{aligned} \tag{2–75}$$

Equation (2–75) gives an approximate closed-form expression of the MSE of the multivariate Capon estimator. In this equation, we note that the MSE consists of three terms. The first term is proportional to CRB($\mathbf{B}$). The second term is proportional to the outer-product of vec($\mathbf{B}$) and is not related to the parameter K and the temporal information matrix $\mathbf{S}$. In the third term, although there is no explicit dependence of the parameter K, the number of non-zero elements in $\mathbf{D}_K$ is dependent of K. Hence, the third term will increase as K increases. Moreover, the third term is a function of $(\mathbf{S}^*\mathbf{S}^T)$, which depends on the the correlation among the rows of $\mathbf{S}$. As we will see in the following numerical simulations, for $\mathbf{S}$ with correlated rows, the MSE of an element of $\mathbf{B}$ increases as K increases and/or as the other elements in $\mathbf{B}$ increase. On the contrary, when $K = 1$, the third term is zero because the matrix $\mathbf{D}$ becomes a scalar 0 according to its definition. If we further set $N = 1$, then (2–75) reduces to the conclusion in the univariate case in [25].

We also note that when the number of snapshots L is large, the last two terms approach zero while the first term approaches CRB($\mathbf{B}$). Hence, the multivariate Capon estimator is also asymptotically statistically efficient for large L.

Furthermore, we note that when $M = N$, the MSE of the multivariate Capon estimator is simplified to CRB($\mathbf{B}$) like the multivariate ML estimator. This is consistent with our conclusion in the above subsection that the two multivariate methods reduce to the same estimator when $M = N$. For the same reason that we

stated in Section 2.3.1, this efficiency of the multivariate Capon estimator should not be seen as a significant advantage.

Now we summarize the statistical properties of the multivariate Capon estimator by the following theorem.

Theorem 2.2. *For the data model in (2–1), the multivariate Capon estimate of* $\mathbf{B}$ *in (2–8) is biased downward. However, for large number of data samples, it is asymptotically unbiased and statistically efficient. Its bias and MSE matrices are given by (2–64) and (2–75), respectively.*

2.4 Numerical Examples

In this section, several numerical examples are presented to verify the performance analysis results of the two multivariate estimators. We consider a uniform linear array with $M = 4$ sensors and half-wavelength spacing. We assume $N = 2$ signals arriving at the sensor array with DOAs (Direction Of Arrival) of $0°$ and $15°$ relative to the array normal. Unless specified otherwise, we assume that $L = 16$ and SNR $= 10$ dB and $\mathbf{S}$ is formed by $K = 2$ complex sinusoids with unit amplitudes and frequencies 0.10 Hz and 0.125 Hz, respectively. Except in Fig. 2–6, the elements in $\mathbf{B}$ are all set to be 1. The interference and noise term in our data model in (2–1) is temporally white but spatially colored zero-mean circularly symmetric complex Gaussian with the spatial covariance matrix $\mathbf{Q}$ given by

$$[\mathbf{Q}]_{ij} = \rho\,(0.9)^{|i-j|}, \tag{2–76}$$

where $\rho = 1/\text{SNR}$ and $[\cdot]_{ij}$ denotes the ith row and jth column element of a matrix. The figure below are all for $[\mathbf{B}]_{11}$. The figures for other elements of $\mathbf{B}$ are similar. We obtain the empirical results in Fig. 2–2 using 10000 Monte Carlo trials while the others 1000 trails.

We first investigate the bias performance. Fig. 2–1 shows the bias properties of the two multivariate estimators (denoted by "MV-ML" and "MV-Capon") from

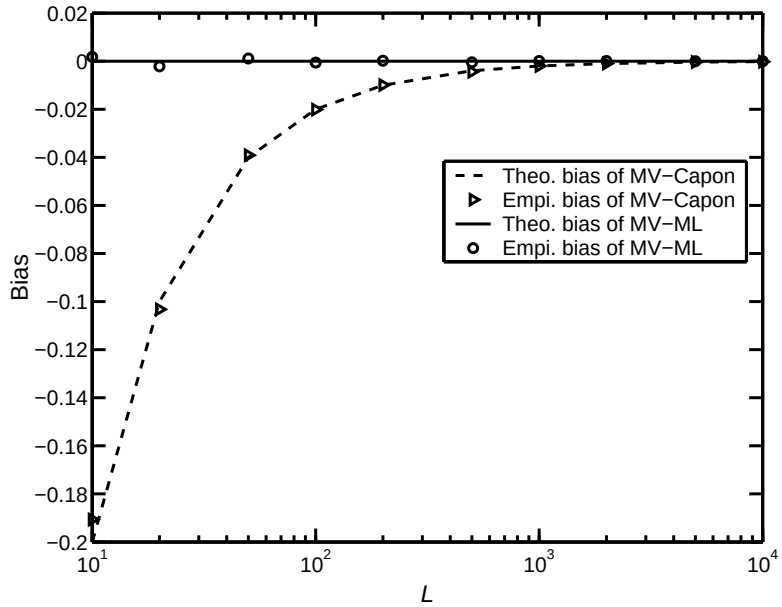


Figure 2–1: Bias versus L when SNR = 10 dB, $K = 2$, $N = 2$.

both theoretical predictions (denoted by "Theo.") and Monte Carlo trials (denoted by "Empi."). As expected, the multivariate ML is unbiased whereas Capon is biased downward for finite snapshots. However, when the number of snapshots L is large, the bias of the multivariate Capon approaches zero, as predicted by our theoretical analysis.

Fig. 2–2 illustrates the relationship between the bias and the number of rows of the temporal information matrix $\mathbf{S}$, i.e., K, when the frequency difference of the complex sinusoids in $\mathbf{S}$ is 0.04 Hz. As predicted by our theoretical analysis, the bias of the multivariate Capon estimator is independent of K.

Fig. 2–3 illustrates the MSEs of the multivariate estimators as well as the CRB as a function of L. As illustrated, the theoretical and empirical MSEs are consistent. The performance of the multivariate ML estimator is better than the multivariate Capon and very close to the corresponding CRB. As we have predicted in Section 2.3 that both multivariate estimators are asymptotically statistically

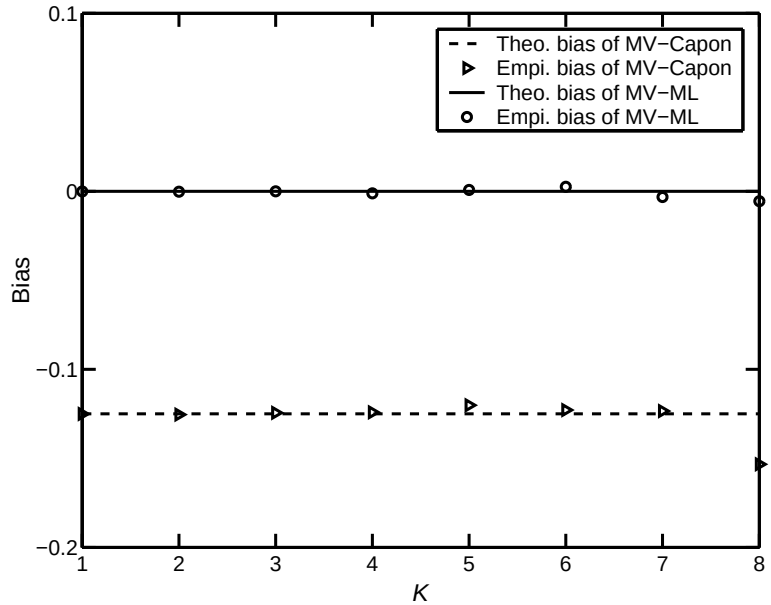


Figure 2–2: Bias versus K when SNR = 10 dB, $L = 16$, $N = 2$.

efficient for large number of snapshots, and the performance curves of the two estimators approach the CRB as L increases.

Fig. 2–4 shows the relationship between the MSE and SNR. Note that the error floor occurs at high SNR for the multivariate Capon estimator due to its bias. As shown in our theoretical analyses, for a fixed M, L, N and K, the MSE of ML is proportional to CRB($\mathbf{B}$), and hence no "threshold effect" occurs. Note also that, like in the univariate case, the Capon estimate can provide a smaller MSE than ML at low SNR. At such a low SNR, though, both ML and Capon perform poorly.

Fig. 2–5 gives the MSEs of the multivariate Capon and ML estimators as well as the corresponding CRB as a function of K when the frequency difference of the complex sinusoids in $\mathbf{S}$ is 0.04 Hz. As we can see, both the CRB and the MSEs of the two multivariate estimators increase as K increases. However, due to the contribution of the third term in (2–75), the MSE of Capon increases more quickly than the CRB and the MSE of ML.

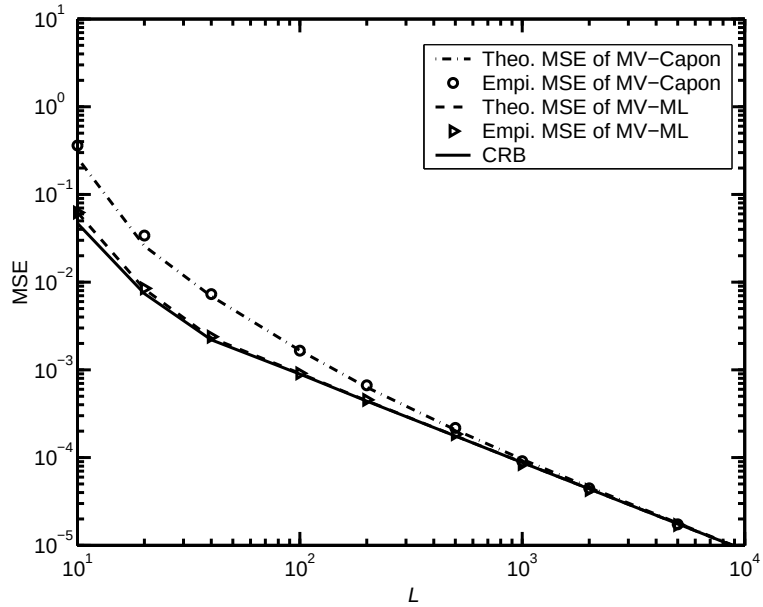


Figure 2–3: MSE versus L when SNR = 10 dB, $K = 2$, $N = 2$.

In Fig. 2–6, we consider the case where $\mathbf{B}$ has unequal elements. We set $N = K = 2$, $[\mathbf{B}]_{11} = [\mathbf{B}]_{21} = 1$ and $[\mathbf{B}]_{12} = [\mathbf{B}]_{22} = \alpha^{-\frac{1}{2}}$, where α is the power ratio between the two complex sinusoids in $\mathbf{S}$. Fig. 2–6 gives the CRB and MSEs of $[\mathbf{B}]_{11}$ as α varies. As illustrated, the MSE of the multivariate ML estimator is almost constant with respect to α, while the MSE of the multivariate Capon estimator increases rapidly when α is decreased to be lower than 0 dB due to its biased nature.

2.5 Conclusions

We have investigated the theoretical performance of two multivariate parameter estimators, namely the multivariate Capon and ML estimators. Through theoretical analysis and numerical simulations, we conclude that the multivariate ML estimator is unbiased, whereas the multivariate Capon estimator is biased downward for finite snapshots; both estimators are asymptotically statistically efficient when the number of snapshots is large.

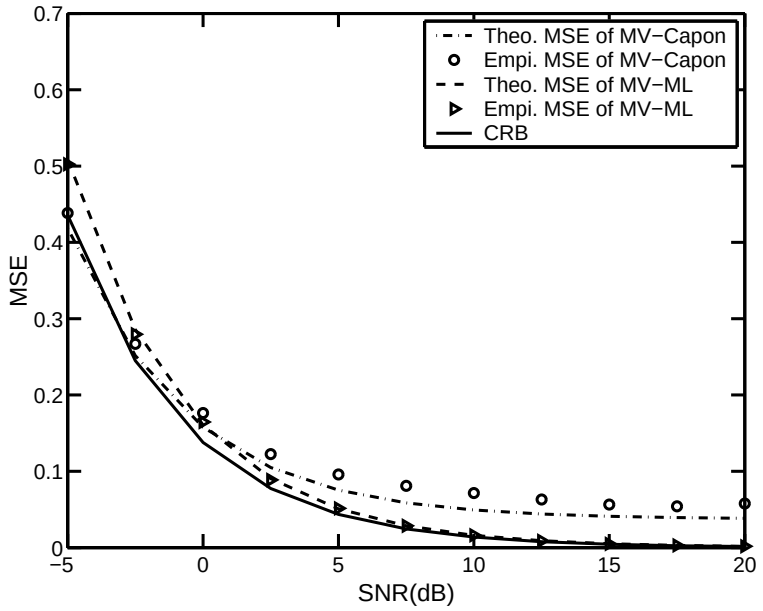


Figure 2–4: MSE versus SNR when $L = 16$, $K = 2$, $N = 2$.

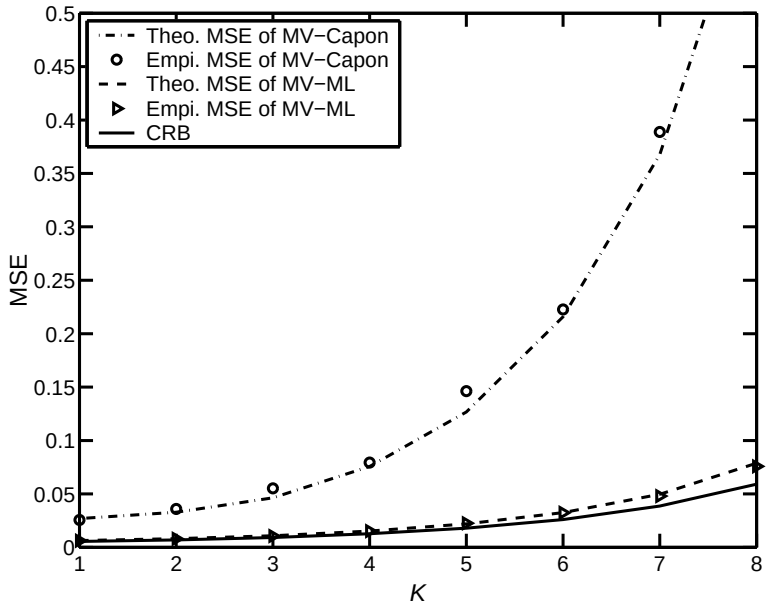


Figure 2–5: MSE versus K when SNR = 10 dB, $L = 16$, $N = 2$.

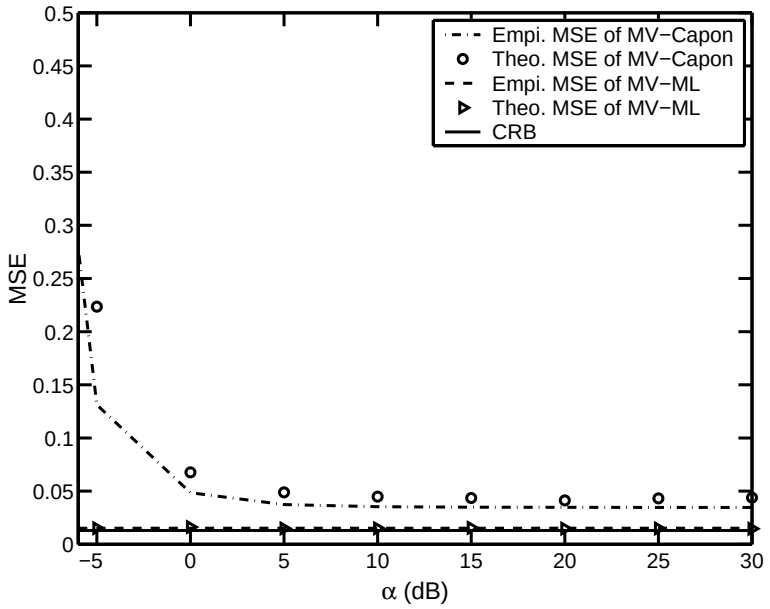


Figure 2–6: MSE versus α when SNR = 10 dB, $L = 16$, $K = 2$, $N = 2$.

CHAPTER 3
DIAGONAL GROWTH-CURVE MODEL

3.1 Introduction and Problem Formulation

In this chapter, we consider a variation of the growth-curve (GC) model, referred to as the diagonal growth-curve (DGC) model,

$$\mathbf{X} = \mathbf{ABS} + \mathbf{Z} \quad \text{with } \mathbf{B} = \text{diag}(\boldsymbol{\beta}), \tag{3–1}$$

where diag($\boldsymbol{\beta}$) denotes a diagonal matrix with its diagonal formed by the elements of the vector $\boldsymbol{\beta}$. In (3–1), $\mathbf{X} \in C^{M\times L}$ denotes the observed snapshots with L being the snapshot number and M the snapshot dimension. The columns in $\mathbf{A} \in C^{M\times N}$ are the known spatial information vectors, referred to as the steering vectors. The rows in $\mathbf{S} \in C^{N\times L}$ are the known temporal information vectors, referred to as waveforms. The elements in $\boldsymbol{\beta} \in C^{N\times 1}$ are the unknown complex amplitudes. The columns of the interference and noise matrix $\mathbf{Z} \in C^{M\times L}$ are independently and identically distributed (i.i.d.) circularly symmetric complex Gaussian random vectors with zero-mean and unknown covariance matrix $\mathbf{Q}$. Throughout this chapter, we assume that $M + r \leq L$ with r being the row rank of $\mathbf{S}$. The problem of interest is to estimate the unknown complex amplitudes. Note that in the DGC model we do not need to assume the linear independence of the steering vectors or waveforms, unlike in the GC model.

The only difference between DGC and GC is that the complex amplitude matrix $\mathbf{B}$ in DGC is constrained to be diagonal while it is arbitrary in GC. However, this seemingly minor difference makes the derivations of the multivariate ML estimator in [21] and [22] invalid. In fact, to our knowledge, no closed-form ML estimator for $\boldsymbol{\beta}$ in (3–1) exists in the literature. In this chapter, we propose

an approximate maximum likelihood (AML) estimator for $\boldsymbol{\beta}$ in this model. We also investigate its statistical properties via theoretical analyses and numerical simulations.

We remark that although in this chapter we focus on the complex amplitude estimation of signals with known steering vectors and waveforms, the proposed DGC model and AML estimator can also be used in the case where both the steering vectors and waveforms are parameterized by unknown parameters. Using the proposed AML method, we can construct a concentrated (approximate) likelihood function of the unknown parameters [27] [30] [31]. The unknown parameters in the steering vectors and waveforms can then be estimated via maximizing the concentrated likelihood function.

The remainder of the chapter is organized as follows. In Section 3.2, we introduce the AML estimator for $\boldsymbol{\beta}$ based on the DGC model in (3–1). Section 3.3 presents the performance analysis of the AML estimator and the CRB for the unknown complex amplitudes. Numerical examples are presented in Section 3.4. Finally, Section 3.5 contains our conclusions.

3.2 Approximate Maximum Likelihood Estimation

In this section, we derive an approximate maximum likelihood (AML) estimator for $\boldsymbol{\beta}$ in (3–1). This approach is based on the ML principle, but an approximation is made to get a closed-form solution.

It follows from (3–1) that the negative log-likelihood function of the observed data samples (to within an additive constant) is

$$f(\boldsymbol{\beta}, \mathbf{Q}) = L\ln|\mathbf{Q}| + \mathrm{tr}[\mathbf{Q}^{-1}(\mathbf{X} - \mathbf{ABS})(\mathbf{X} - \mathbf{ABS})^H], \tag{3–2}$$

where $\mathrm{tr}(\cdot)$, $|\cdot|$ and $(\cdot)^H$ denote the trace, the determinant and the conjugate transpose of a matrix, respectively. Minimizing the negative log-likelihood function

with respect to $\mathbf{Q}$ yields

$$\hat{\mathbf{Q}} = \frac{1}{L}(\mathbf{X} - \mathbf{ABS})(\mathbf{X} - \mathbf{ABS})^H. \tag{3–3}$$

Note that $\hat{\mathbf{Q}}$ is nonsingular, i.e., $|\hat{\mathbf{Q}}| \neq 0$, with probability one when $M \leq L$. Substituting (3–3) into (3–2), the ML estimate of $\boldsymbol{\beta}$ can be formulated as

$$\hat{\boldsymbol{\beta}}_{\text{ML}} = \arg\min_{\boldsymbol{\beta}} \ln|(\mathbf{X} - \mathbf{ABS})(\mathbf{X} - \mathbf{ABS})^H| \quad \text{with} \quad \mathbf{B} = \text{diag}(\boldsymbol{\beta}). \tag{3–4}$$

In general, the optimization problem in (3–4) does not appear to admit a closed-form solution. Here, we use the technique in [25] and [27] to solve this problem approximately. Let $\mathbf{\Pi}_S$ and $\mathbf{\Pi}_S^{\perp}$ denote the orthogonal projection matrices given by

$$\mathbf{\Pi}_S \triangleq \mathbf{S}^H(\mathbf{SS}^H)^{-}\mathbf{S}, \tag{3–5}$$

and

$$\mathbf{\Pi}_S^{\perp} \triangleq \mathbf{I} - \mathbf{\Pi}_S, \tag{3–6}$$

where $(\cdot)^{-}$ denotes the generalized inverse of a matrix [74]. Note that when the waveforms, i.e., the rows of $\mathbf{S}$, are not linearly independent of each other, the matrix $(\mathbf{SS}^H)$ is singular and hence its generalized inverse is not unique. However, $\mathbf{\Pi}_S$ and $\mathbf{\Pi}_S^{\perp}$ are unique [74].

Using (3–5) and (3–6), it follows that

$$\begin{aligned} &|(\mathbf{X} - \mathbf{ABS})(\mathbf{X} - \mathbf{ABS})^H| \\ =&|(\mathbf{X} - \mathbf{ABS})(\mathbf{\Pi}_S + \mathbf{\Pi}_S^{\perp})(\mathbf{X} - \mathbf{ABS})^H| \\ =&|(\mathbf{X\Pi}_S - \mathbf{ABS})(\mathbf{X\Pi}_S - \mathbf{ABS})^H + \mathbf{T}| \\ =&|(\mathbf{X\Pi}_S - \mathbf{ABS})(\mathbf{X\Pi}_S - \mathbf{ABS})^H\mathbf{T}^{-1} + \mathbf{I}|\,|\mathbf{T}| \end{aligned} \tag{3–7}$$

where the matrix $\mathbf{T}$ is defined by

$$\mathbf{T} \triangleq \mathbf{X\Pi}_S^{\perp}\mathbf{X}^H. \tag{3–8}$$

Note that the rank of $\mathbf{\Pi}_S^{\perp}$ is $L - r$. Hence, $\mathbf{T}$ has full rank M with probability one when $M + r \leq L$. Note also that $\mathbf{T}$ is an estimate of the covariance matrix $\mathbf{Q}$ (to within a multiplicative constant) obtained by projecting out the desired signal components from $\mathbf{X}$.

Using (3–7) and Lemma 4 in [25] (or Theorem 1 in [27]), the solution of the optimization problem in (3–4) can be approximated for a large snapshot number as follows.

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{\text{AML}} &= \arg\min_{\boldsymbol{\beta}} \operatorname{tr}[(\mathbf{X\Pi}_S - \mathbf{ABS})^H \mathbf{T}^{-1} (\mathbf{X\Pi}_S - \mathbf{ABS})] \\ &= \arg\min_{\boldsymbol{\beta}} \parallel \operatorname{vec}(\mathbf{T}^{-\frac{1}{2}} \mathbf{X\Pi}_S) - \operatorname{vec}(\mathbf{T}^{-\frac{1}{2}} \mathbf{ABS}) \parallel^2\end{aligned} \tag{3–9}$$

where $\mathbf{T}^{-\frac{1}{2}}$ is the Hermitian square root of $\mathbf{T}^{-1}$, $\parallel \cdot \parallel$ denotes the Euclidean vector norm and $\operatorname{vec}(\cdot)$ denotes the vectorization operator (stacking the columns of a matrix on top of each other).

To solve the above optimization problem, we first introduce the following lemmas.

Lemma 3.1. *Let* $\mathbf{U}$ *and* $\mathbf{V}$ *be* $m \times p$ *and* $n \times p$ *matrices, respectively, and let* $\mathbf{G}$ *be a* $p \times p$ *diagonal matrix. Then*

$$vec(\mathbf{UGV}^T) = (\mathbf{V} \boxdot \mathbf{U}) vecd(\mathbf{G}), \tag{3–10}$$

where $(\cdot)^T$ *denotes the transpose of a matrix,* $vecd(\cdot)$ *denotes a column vector formed by the diagonal elements of a matrix, and* $\boxdot$ *denotes the Khatri-Rao matrix product [77] [78].*

Proof: This result is not new but we include a simple proof in Appendix B.1 for the completeness of this chapter.

Lemma 3.2. *Let* $\mathbf{U}$, $\mathbf{G}$ *and* $\mathbf{V}$ *be matrices with dimensions* $m \times p$, $p \times n$ *and* $n \times m$, *respectively. Then*

$$vecd(\mathbf{UGV}) = (\mathbf{V} \boxdot \mathbf{U}^T)^T vec(\mathbf{G}). \tag{3–11}$$

Furthermore, if $p = n$ and $\mathbf{G}$ is a diagonal matrix, then

$$vecd(\mathbf{UGV}) = (\mathbf{U} \odot \mathbf{V}^T) vecd(\mathbf{G}), \tag{3–12}$$

where $\odot$ denotes the Hadamard product (elementwise multiplication) of two matrices.

Proof: Appendix B.2.

Lemma 3.3. *Let $\mathbf{U}$ and $\mathbf{V}$ be $m \times p$ and $n \times p$ matrices, respectively. Then*

$$(\mathbf{U} \boxdot \mathbf{V})^H (\mathbf{U} \boxdot \mathbf{V}) = (\mathbf{U}^H \mathbf{U}) \odot (\mathbf{V}^H \mathbf{V}). \tag{3–13}$$

Proof: This lemma is a straightforward extension of Lemma A2 in [78].

By Lemma 3.1, it follows from (3–9) that

$$\hat{\boldsymbol{\beta}}_{\text{AML}} = \arg\min_{\boldsymbol{\beta}} \| \operatorname{vec}(\mathbf{T}^{-\frac{1}{2}} \mathbf{X} \boldsymbol{\Pi}_S) - \boldsymbol{\Gamma} \boldsymbol{\beta} \|^2 \tag{3–14}$$

with $\boldsymbol{\Gamma} \triangleq \mathbf{S}^T \boxdot (\mathbf{T}^{-\frac{1}{2}} \mathbf{A})$. Note that (3–14) is in a quadratic function of $\boldsymbol{\beta}$. Minimizing (3–14) with respect to $\boldsymbol{\beta}$ yields

$$\hat{\boldsymbol{\beta}}_{\text{AML}} = (\boldsymbol{\Gamma}^H \boldsymbol{\Gamma})^{-1} \boldsymbol{\Gamma}^H \operatorname{vec}(\mathbf{T}^{-\frac{1}{2}} \mathbf{X} \boldsymbol{\Pi}_S). \tag{3–15}$$

In (3–15), we have assumed that the columns of $\boldsymbol{\Gamma}$ are linearly independence of each other and hence $\boldsymbol{\Gamma}^H \boldsymbol{\Gamma}$ is invertible. Note that this assumption is much weaker than that in the GC model.

On the other hand, using Lemmas 3.3 and 3.2, respectively, we have that

$$\boldsymbol{\Gamma}^H \boldsymbol{\Gamma} = (\mathbf{S}\mathbf{S}^H)^T \odot (\mathbf{A}^H \mathbf{T}^{-1} \mathbf{A}) \tag{3–16}$$

and

$$\boldsymbol{\Gamma}^H \operatorname{vec}(\mathbf{T}^{-\frac{1}{2}} \mathbf{X} \boldsymbol{\Pi}_S) = \operatorname{vecd}(\mathbf{A}^H \mathbf{T}^{-1} \mathbf{X} \mathbf{S}^H). \tag{3–17}$$

Substituting (3–16) and (3–17) into (3–14) yields the following expression for the AML estimate of $\boldsymbol{\beta}$

$$\hat{\boldsymbol{\beta}}_{\text{AML}} = [(\mathbf{A}^H\mathbf{T}^{-1}\mathbf{A}) \odot (\mathbf{S}\mathbf{S}^H)^T]^{-1}\text{vecd}(\mathbf{A}^H\mathbf{T}^{-1}\mathbf{X}\mathbf{S}^H). \tag{3–18}$$

Note that (3–18) does not require the existence of the inverse of $\mathbf{A}^H\mathbf{T}^{-1}\mathbf{A}$. Moreover, (3–18) does not require $\mathbf{S}\mathbf{S}^H$ to be invertible. Therefore, unlike the estimators in the GC model [22], the AML estimator in DGC does not require the linear independence of the steering vectors or of the waveforms.

Appendix B.3 gives another interpretation of AML from a generalized least-squares (GLS) point of view.

3.3 Performance Analysis

We now establish the theoretical properties of the AML estimator.

3.3.1 Bias Analysis

Substituting (3–1) into (3–18) gives

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{\text{AML}} =& [(\mathbf{A}^H\mathbf{T}^{-1}\mathbf{A}) \odot (\mathbf{S}\mathbf{S}^H)^T]^{-1}\text{vecd}(\mathbf{A}^H\mathbf{T}^{-1}\mathbf{A}\mathbf{B}\mathbf{S}\mathbf{S}^H) \\ &+ [(\mathbf{A}^H\mathbf{T}^{-1}\mathbf{A}) \odot (\mathbf{S}\mathbf{S}^H)^T]^{-1}\text{vecd}(\mathbf{A}^H\mathbf{T}^{-1}\mathbf{Z}\mathbf{S}^H).\end{aligned} \tag{3–19}$$

By Lemma 3.2 we get

$$\text{vecd}[(\mathbf{A}^T\mathbf{T}^{-1}\mathbf{A})\mathbf{B}(\mathbf{S}\mathbf{S}^H)] = [(\mathbf{A}^H\mathbf{T}^{-1}\mathbf{A}) \odot (\mathbf{S}\mathbf{S}^H)^T]\boldsymbol{\beta}. \tag{3–20}$$

Hence, from (3–19) we obtain the estimation error

$$\hat{\boldsymbol{\beta}}_{\text{AML}} - \boldsymbol{\beta} = [(\mathbf{A}^H\mathbf{T}^{-1}\mathbf{A}) \odot (\mathbf{S}\mathbf{S}^H)^T]^{-1}\text{vecd}(\mathbf{A}^H\mathbf{T}^{-1}\mathbf{Z}\mathbf{S}^H). \tag{3–21}$$

On the other hand, substituting (3–1) into (3–8) yields

$$\mathbf{T} = \mathbf{Z}\boldsymbol{\Pi}_S^{\perp}\mathbf{Z}^H. \tag{3–22}$$

From (3–22), we see that $\mathbf{T}$ is an even function of $\mathbf{Z}$, and hence $\hat{\boldsymbol{\beta}}_{\text{AML}} - \boldsymbol{\beta}$ is an odd function of $\mathbf{Z}$. Moreover, we have assumed that $\mathbf{Z}$ is a zero-mean Gaussian random matrix. Using the statistical properties of the zero-mean Gaussian distribution [75], we can readily show that the expectation of (3–21) is zero, i.e., $\hat{\boldsymbol{\beta}}_{\text{AML}}$ is unbiased.

3.3.2 Mean-Squared-Error Analysis

The exact MSE of $\hat{\boldsymbol{\beta}}_{\text{AML}}$ is difficult to determine, if not impossible. Herein, we provide an approximate expression for the MSE which holds for a large snapshot number.

From the theory on the linear statistical model [16], we know that $\frac{\mathbf{T}}{L-r}$ is an unbiased and consistent estimate of the noise covariance matrix $\mathbf{Q}$, which converges to $\mathbf{Q}$ when L approaches infinity. Hence, for a large snapshot number L, the estimation error in (3–21) can be approximated by

$$\hat{\boldsymbol{\beta}}_{\text{AML}} - \boldsymbol{\beta} \approx \text{CRB}(\boldsymbol{\beta})\,\text{vecd}(\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{Z}\mathbf{S}^H), \tag{3–23}$$

where

$$\text{CRB}(\boldsymbol{\beta}) = [(\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{A}) \odot (\mathbf{S}\mathbf{S}^H)^T]^{-1} \tag{3–24}$$

is the Cramér-Rao bound for $\boldsymbol{\beta}$ that is the lowest possible MSE of any unbiased estimator of $\boldsymbol{\beta}$ (see Appendix B.3 for the derivation of CRB).

From (3–23) and Lemma 3.2, it follows that

$$\begin{aligned}
\text{var}(\hat{\boldsymbol{\beta}}_{\text{AML}}) \triangleq& \text{E}[(\hat{\boldsymbol{\beta}}_{\text{AML}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_{\text{AML}} - \boldsymbol{\beta})^H] \\
\approx& \text{CRB}(\boldsymbol{\beta})\,\text{E}[\text{vecd}(\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{Z}\mathbf{S}^H)\,\text{vecd}(\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{Z}\mathbf{S}^H)^H]\,\text{CRB}(\boldsymbol{\beta}) \\
=& \text{CRB}(\boldsymbol{\beta})\,[\mathbf{S}^H \boxdot (\mathbf{A}^H\mathbf{Q}^{-1})^T]^T\,\text{E}[\text{vec}(\mathbf{Z})\text{vec}(\mathbf{Z})^H]\,[\mathbf{S}^H \boxdot (\mathbf{A}^H\mathbf{Q}^{-1})^T]^*\,\text{CRB}(\boldsymbol{\beta}) \\
=& \text{CRB}(\boldsymbol{\beta})\,[\mathbf{S}^T \boxdot (\mathbf{Q}^{-1}\mathbf{A})]^H\,(\mathbf{I} \otimes \mathbf{Q})\,[\mathbf{S}^T \boxdot (\mathbf{Q}^{-1}\mathbf{A})]\,\text{CRB}(\boldsymbol{\beta})
\end{aligned} \tag{3–25}$$

where $(\cdot)^*$ and $\otimes$ denote the complex conjugate and the Kronecker matrix product, respectively. From (3–25), similarly to Equation (B–7) in Appendix B.3, we get the

MSE of $\hat{\boldsymbol{\beta}}_{\text{AML}}$ as

$$\text{var}(\hat{\boldsymbol{\beta}}_{\text{AML}}) = [(\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{A}) \odot (\mathbf{S}\mathbf{S}^H)^T]^{-1} = \text{CRB}(\boldsymbol{\beta}). \tag{3–26}$$

We see that the AML estimate of $\boldsymbol{\beta}$ is asymptotically statistically efficient for a large snapshot number. This theoretical result is verified by the numerical examples in Section 3.4.

3.4 Numerical Examples

In this section, several numerical examples of two different applications of the DGC model are presented to demonstrate the performance of the proposed AML estimator.

3.4.1 Examples in Array Signal Processing

We consider a uniform linear array with $M = 6$ elements, and half-wavelength spacing between adjacent sensors. The sensor elements are assumed to be omni-directional. The incident signals are quadrature phase shift keyed (QPSK) sequences, and the complex amplitudes are equal to one. The columns of the interference and noise matrix $\mathbf{Z}$ are i.i.d. circularly symmetric complex Gaussian random vectors with zero-mean and an unknown covariance matrix $\mathbf{Q}$ given by

$$[\mathbf{Q}]_{m,n} = \rho\, 0.99^{|m-n|} e^{j\frac{(m-n)\pi}{2}} \tag{3–27}$$

where $\rho = 1/\text{SNR}$ and $[\cdot]_{m,n}$ denotes the $(m,\, n)$th element of a matrix.

In the following examples, we present the MSE of the AML estimator in (3–18) as well as the corresponding CRB in (3–24). For comparison, we also show the MSE of the ML estimator in [21] and [22] based on the GC model, which ignores the diagonal constraint on $\mathbf{B}$, and of the least squares (LS) estimator which assumes that the interference and noise vectors in $\mathbf{Z}$ are uncorrelated both spatially and temporally. The LS estimator can be obtained immediately via replacing $\mathbf{T}$ in

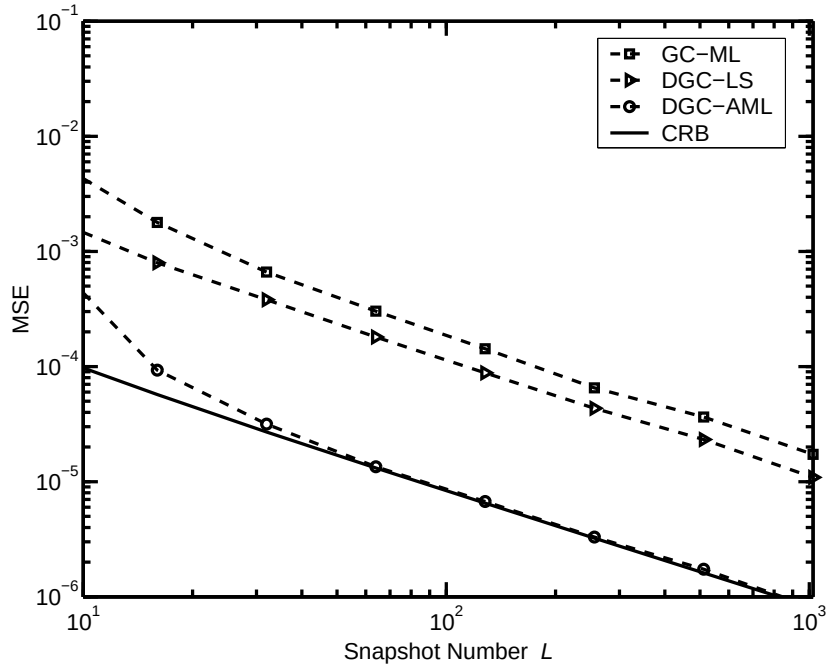


Figure 3–1: Empirical MSE's and the CRB versus L when SNR $=$ 10 dB, M $=$ 6, and N $=$ 3 with linearly independent steering vectors and linearly independent waveforms.

(3–18) by an identity matrix, i.e.,

$$\hat{\beta}_{\text{LS}} = [(\mathbf{A}^H\mathbf{A}) \odot (\mathbf{S}\mathbf{S}^H)^T]^{-1}\text{vecd}(\mathbf{A}^H\mathbf{X}\mathbf{S}^H). \tag{3–28}$$

The empirical estimation performances are all obtained by 1000 Monte-Carlo simulations.

We first consider the case where the steering vectors are linearly independent of each other and so are the waveforms. Suppose that $N = 3$ signals with known waveforms arrive at the sensor array from $\theta_1 = -10°$, $\theta_2 = 5°$ and $\theta_3 = 10°$, respectively. Assume that the DOAs of the signals have been estimated accurately and therefore they can be considered to be known. The signal waveforms are generated independently and hence are uncorrelated with each other. Due to space limitations, we only show below the MSE of the complex amplitude estimate of the signal arriving from $\theta_2 = 5°$. The performances for other signals are similar.

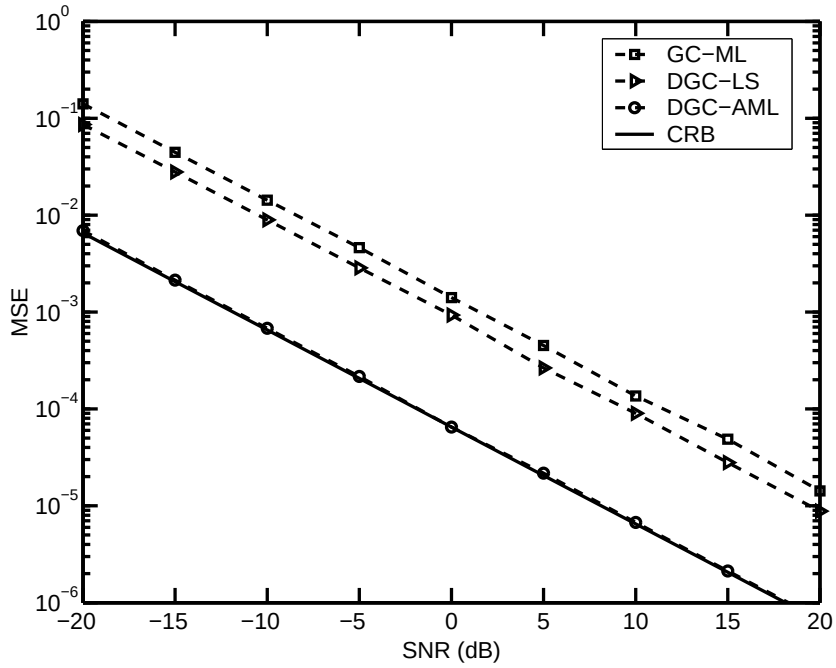


Figure 3–2: Empirical MSE's and the CRB versus SNR when $L = 128$, $M = 6$, and $N = 3$ with linearly independent steering vectors and linearly independent waveforms.

Fig. 3–1 and Fig. 3–2 show the estimation performance as a function of the snapshot number L and the SNR, respectively. The SNR is fixed at 10 dB in Fig. 3–1 while the snapshot number is fixed at 128 in Fig. 3–2. Note that the ML estimator based on the GC model gives relatively poor estimation performance. This method ignores the *a priori* information about the diagonal structure of the complex amplitude matrix $\mathbf{B}$, which increases the number of unknowns from N to N^2 and hence results in much larger estimation variances. In Fig. 3–1, we see that the AML estimator outperforms LS significantly, and approaches the CRB rapidly, as predicted theoretically. In Fig. 3–2, we see that the AML estimator outperforms the other methods and is very close to the CRB in the whole range of SNR considered.

Fig. 3–3 and Fig. 3–4 show the estimation performance when the waveforms are linearly dependent of each other. We retain the same simulation parameters as

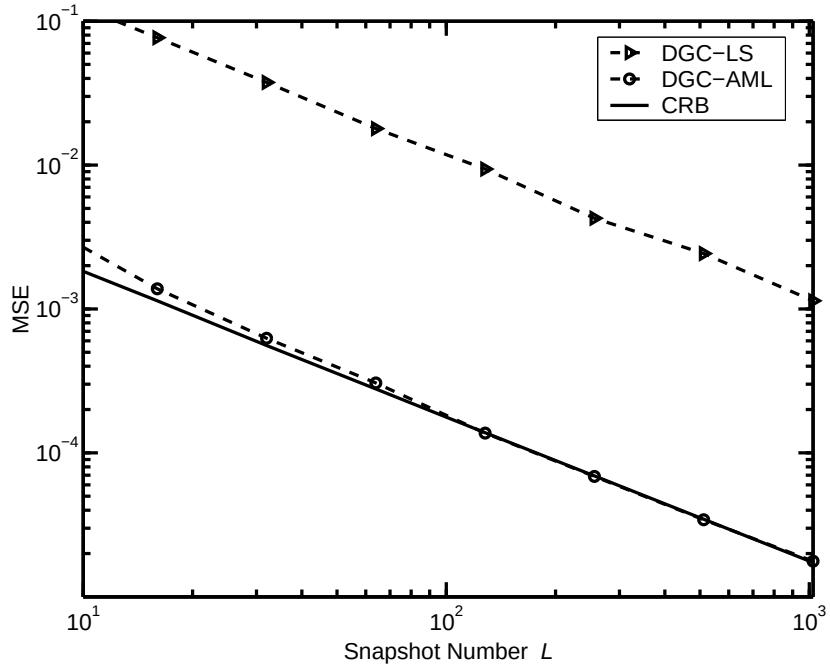


Figure 3–3: Empirical MSE's and the CRB versus L when SNR $=$ 10 dB, $M = 6$, and $N = 3$ with identical waveforms.

in Fig. 3–1 and Fig. 3–2 except that the waveforms of the three incident signals are identical in this case. Since the ML estimator based on the GC model requires that the waveforms are linearly independent of each other [22], it fails to work properly in this case. On the other hand, while the MSE's of AML are somewhat larger than those for linearly independent waveforms, the AML estimator remains asymptotically statistically efficient and provides higher estimation accuracy than LS.

Next we consider an example where $N = 13$ incident signals impinge on the sensor array from $\theta = -30°, -25°, \cdots 30°$. The other simulation parameters are the same as those for Fig. 3–1 and Fig. 3–2. Note that in this case $N > M$, which in particular means that the steering vectors are linearly dependent and hence it does not satisfy the assumptions required by the GC model; consequently again the ML estimator based on the GC model cannot be used. From Fig. 3–5 and Fig. 3–6, we

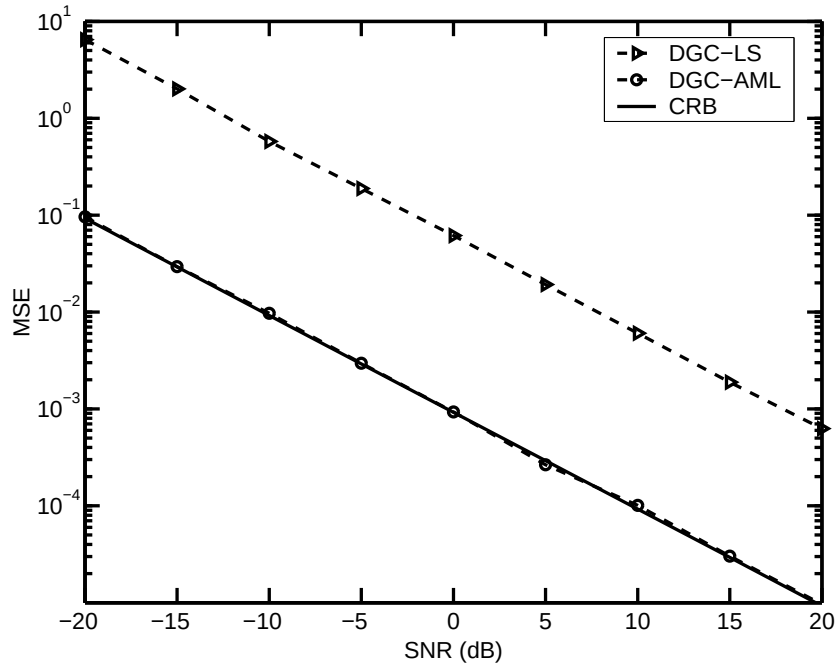


Figure 3–4: Empirical MSE's and the CRB versus SNR when $L = 128$, $M = 6$, and $N = 3$ with identical waveforms.

see that the AML estimator has a better estimation accuracy than LS and remains asymptotically statistically efficient.

3.4.2 Spectral Analysis Examples

In this subsection we consider the problem of estimating the complex amplitudes of sinusoidal signals from observations corrupted by colored noise. This problem has been studied, for example, in [24]. We apply the AML estimator to this 1-D spectral estimation problem. As we show below, the proposed AML estimator can achieve better estimation accuracy and exhibit greater robustness than the methods in [24].

Consider the noise-corrupted observations of N complex-valued sinusoids

$$x(l) = \sum_{n=1}^{N} \beta_n e^{j\omega_n l} + z(l), \quad l = 0,\ 1,\ \cdots\ L_0 - 1, \tag{3–29}$$

where β_n is the complex amplitude of the nth sinusoid with frequency ω_n; L_0 is the number of data samples; $z(l)$ is the observation noise, which is complex valued and

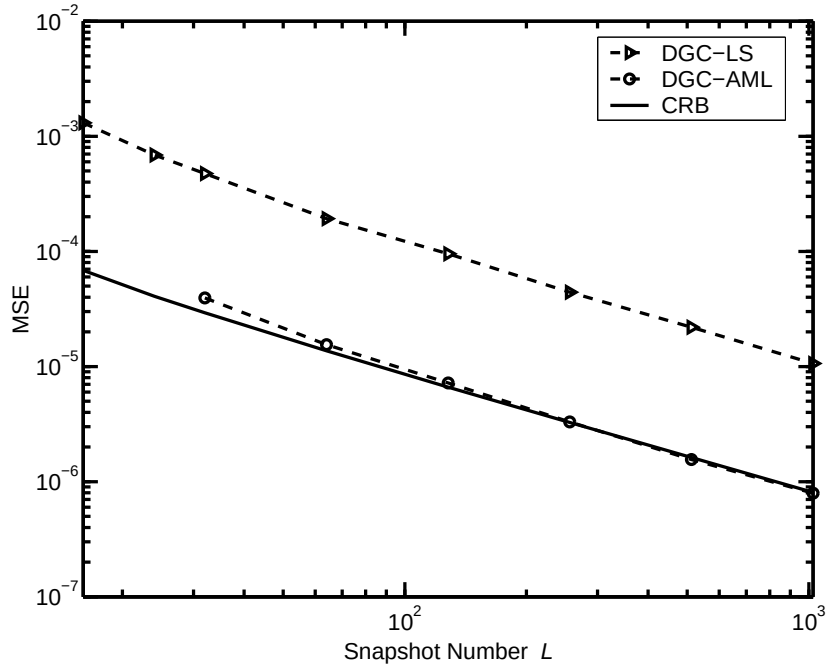


Figure 3–5: Empirical MSE's and the CRB versus L when SNR $=$ 10 dB, $M = 6$, and $N = 13$ with linearly dependent steering vectors

assumed to be stationary and possibly colored with zero-mean and unknown finite power spectral density (PSD). We assume that $\{\omega_n\}_{n=1}^{N}$ are known, with $\omega_n \neq \omega_k$ for $n \neq k$. The problem of interest is to estimate $\{\beta_n\}_{n=1}^{N}$ from the observations $\{x(l)\}_{l=0}^{L_0-1}$.

To solve this problem, we divide the data sequence into overlapping subsequences with shorter lengths [60] [23]. Then we reformulate the data equation in (3–29) into the form of a DGC model, and use the proposed AML approach to estimate the complex amplitudes of the sinusoids.

Let

$$\mathbf{X} = \begin{bmatrix} x(0) & x(1) & \cdots & x(L_0 - M) \\ x(1) & x(2) & \cdots & x(L_0 - M + 1) \\ \vdots & \vdots & \ddots & \vdots \\ x(M-1) & x(M) & \cdots & x(L_0 - 1) \end{bmatrix}, \tag{3–30}$$

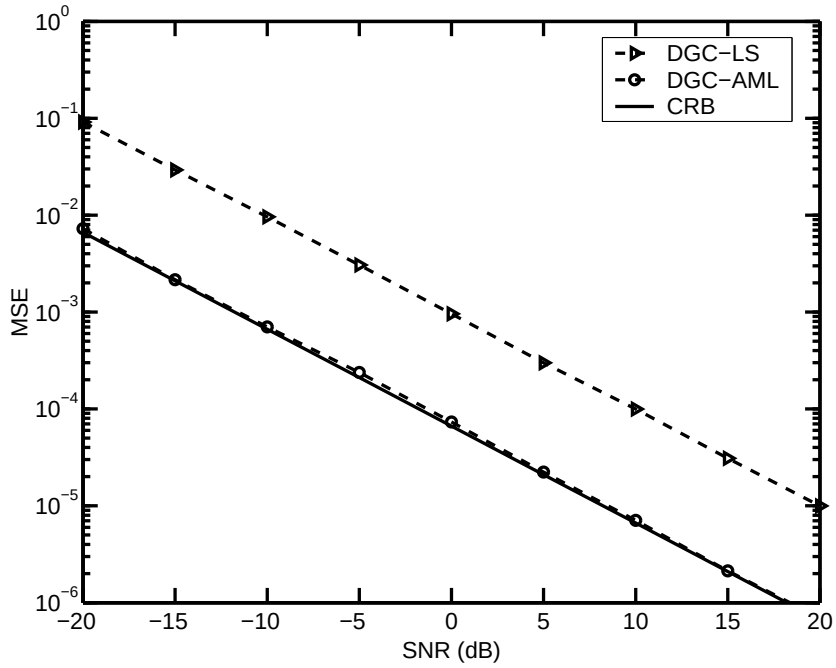


Figure 3–6: Empirical MSE's and the CRB versus SNR when $L = 128$, $M = 6$, and $N = 13$ with linearly dependent steering vectors.

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{j\omega_1} & e^{j\omega_2} & \cdots & e^{j\omega_N} \\ \vdots & \vdots & \ddots & \vdots \\ e^{j(M-1)\omega_1} & e^{j(M-1)\omega_2} & \cdots & e^{j(M-1)\omega_N} \end{bmatrix}, \tag{3–31}$$

and

$$\mathbf{S} = \begin{bmatrix} 1 & e^{j\omega_1} & \cdots & e^{j(L_0-M)\omega_1} \\ 1 & e^{j\omega_2} & \cdots & e^{j(L_0-M)\omega_2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{j\omega_N} & \cdots & e^{j(L_0-M)\omega_N} \end{bmatrix}. \tag{3–32}$$

Then, (3–29) can be readily written in the DGC form in (3–1) with $\boldsymbol{\beta} = [\beta_1,\ \beta_2,\ \cdots\ \beta_N]^T$, $L = L_0 - M + 1$ and the noise matrix $\mathbf{Z}$ defined similarly to $\mathbf{X}$ in (3–30). Hence, the AML estimator in (3–18) can be applied directly. In doing so we ignore the fact that the columns of the interference and noise matrix $\mathbf{Z}$

are correlated due to the overlapping of data samples, which is commonly done in the literature to retain the simplicity of the parameter estimation algorithm.

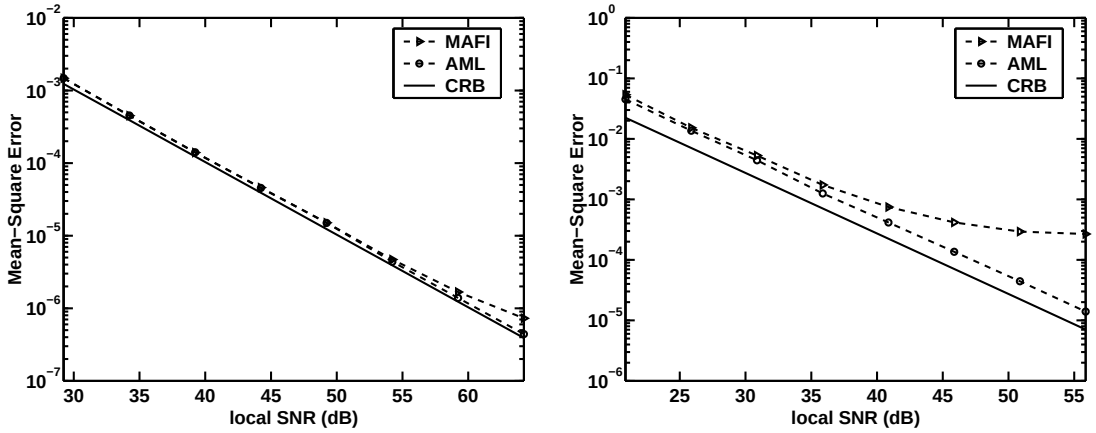
We consider a numerical example used in [24]. The observed data with $L_0 = 32$ consists of $N = 3$ complex sinusoids with frequencies $f_1 = 0.10$, $f_2 = 0.11$, $f_3 = 0.30$ ($f_k = \omega_k/2\pi$) and complex amplitudes $\beta_1 = e^{j\frac{\pi}{4}}$, $\beta_2 = e^{j\frac{\pi}{3}}$ and $\beta_3 = e^{j\frac{\pi}{4}}$, respectively. The colored noise $z(l)$ is described by the following autoregressive (AR) equation

$$z(l) = 0.99z(l-1) + e(l), \tag{3–33}$$

where $e(l)$ is a complex white Gaussian noise with zero-mean and variance σ^2. The PSD of the test data is shown in Fig. 1 of [24] when $\sigma^2 = 0.01$.

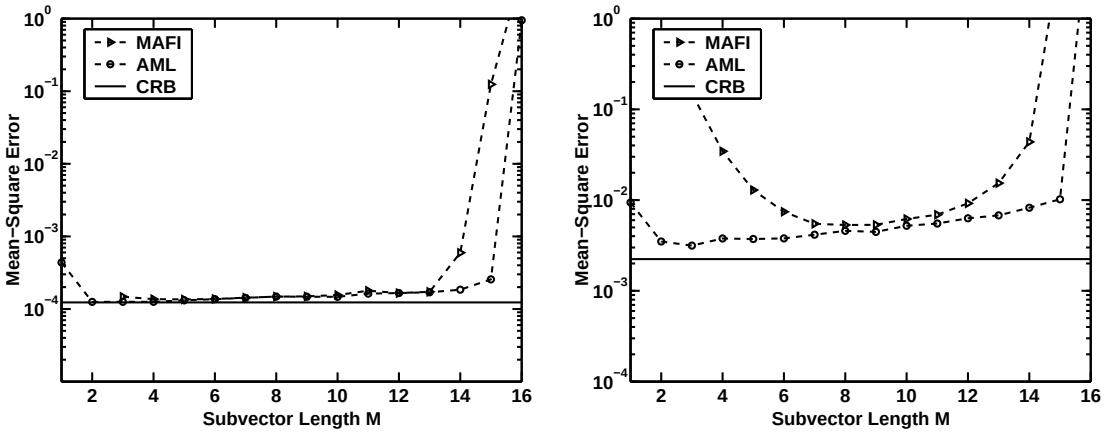
For comparison, we provide the estimation performance of the proposed AML method and of the matched-filter bank (MAFI) approach [24], which is the most competitive one among the algorithms presented in [24], as well as the corresponding CRB. Note that in this application the columns of the interference and noise matrix $\mathbf{Z}$ are not statistically independent, and hence the CRB in (3–24) is not applicable. Instead, we utilize the CRB formula presented in Equation (9) of [24]. The MSE values presented in the following examples are all obtained via 1000 Monte-Carlo simulations. Since the PSD of the AR noise varies in the frequency domain, we utilize the local SNR as a measure of the signal quality for a particular sinusoid [24] [79].

Fig. 3–7 shows the MSE's of the two amplitude estimators for β_3 and β_1 along with the corresponding CRB as the corresponding local SNR varies, when $M = L_0/4 = 8$. (The results for β_2 are omitted because they resemble those for β_1.) In Fig. 3–7(a), MAFI and AML both provide high estimation accuracy and are very close to the CRB. However, the estimation performance in Fig. 3–7(b) is significantly poorer than that in Fig. 3–7(a) for both estimators due to the interference from the second sinusoid at f_2. (Note that $f_2 - f_1 = 0.01$, which is



(a) (b)

Figure 3–7: Empirical MSE's and the CRB versus local SNR when $L_0 = 32$, $M = 8$, and the observation noise is colored. (a) For β_3 and (b) for β_1.



(a) (b)

Figure 3–8: Empirical MSE's and the CRB versus M when $L_0 = 32$, $\sigma^2 = 0.01$, and the observation noise is colored. (a) For β_3 and (b) for β_1.

smaller than the Fourier resolution limit, i.e., $1/L_0 \approx 0.03$.) As we can see, MAFI deviates away from the CRB at high SNR because of the interference at f_2, which introduces a bias into the estimate of β_1 that dominates the MSE at high SNR. The AML estimator provides better estimation accuracy than MAFI, especially at high SNR. Note that in the presented spectral analysis application, the columns of the interference and noise matrix $\mathbf{Z}$ are not i.i.d. and hence the theoretical analysis in Section 3.3 showing that AML is asymptotically statistically efficient is no longer valid. However, as we can see from Fig. 3–7(a) and Fig. 3–7(b), the MSE of AML remains close to the CRB which shows the high interference-resistant capability of AML.

Intuitively, we can expect that as M increases, the AML estimator as well as MAFI (and also other methods presented in [24]) can deal better with the case of closely spaced sinusoids, but their statistical accuracy, in general, decreases [71]. Hence, there is a tradeoff when choosing M. The following example examines the effect of M on the performance of these estimators. The scenario is similar to the example in Fig. 3–7, except that we fix $\sigma^2 = 0.01$, which corresponds to a local SNR of 30.8 dB for the first sinusoid (at $f_1 = 0.1$) and 39.2 dB for the third sinusoid (at $f_3 = 0.3$). The subvector length M is varied from 1 to 16 for AML and from 3 to 16 for MAFI (MAFI requires that $M \geq N = 3$). The MSE's of the amplitude estimates of β_3 and β_1 and the corresponding CRB's are shown in Fig. 3–8(a) and Fig. 3–8(b). As we can see, when no sinusoids are close to the one being estimated, such as the third sinusoid in this example, both AML and MAFI perform quite well for a wide range of M values. For the more difficult case shown in Fig. 3–8(b), the choice of M becomes critical. As we can see, the AML estimator outperforms MAFI over a wide range of M values and demonstrates its robustness to the choice of M.

3.5 Conclusions

We have presented an approximate maximum likelihood estimator for the diagonal growth-curve model where the steering vectors and the waveforms of the signals are known and the unknown complex amplitude matrix is constrained to be diagonal. Via a theoretical analysis, we have shown that the AML estimator is unbiased and asymptotically statistically efficient for a large snapshot number. We have applied the AML estimator to complex amplitude estimation problems in both array signal processing and spectral analysis. The presented numerical examples provide compelling evidence that the AML method can achieve better estimation accuracy and exhibit greater robustness than the best existing methods.

CHAPTER 4
BLOCK DIAGONAL GROWTH-CURVE MODEL

4.1 Introduction and Problem Formulation

In this chapter, we consider a general variation of the GC model, referred to as the block diagonal growth-curve (BDGC) model

$$\mathbf{X} = \mathbf{ABS} + \mathbf{Z} \quad \text{with} \quad \mathbf{B} = \text{Diag}(\mathbf{B}_1,\ \mathbf{B}_2,\ \cdots,\mathbf{B}_J), \tag{4–1}$$

where

$$\mathbf{A} = [\mathbf{A}_1 \quad \mathbf{A}_2 \quad \cdots \quad \mathbf{A}_J], \tag{4–2}$$

and

$$\mathbf{S} = [\mathbf{S}_1^T \quad \mathbf{S}_2^T \quad \cdots \quad \mathbf{S}_J^T]^T. \tag{4–3}$$

In (4–1), $\mathbf{X} \in C^{M\times L}$ contains the observed data samples with M being the snapshot dimension and L being the snapshot number. The columns in $\mathbf{A}_j \in C^{M\times N_j}$ and the rows in $\mathbf{S}_j \in C^{K_j\times L}$ are known and assumed to be linearly independent of each other. The matrices $\mathbf{B}_j \in C^{N_j\times K_j}$ contain the unknown regression coefficients. Throughout this chapter, we assume that $M \leq N_j$ and $M + r \leq L$ with r being the row rank of $\mathbf{S}$. The columns of the error matrix $\mathbf{Z}$ are assumed to be i.i.d. circularly symmetric complex Gaussian random vectors with zero-mean and unknown covariance matrix $\mathbf{Q}$. The problem of interest herein is to estimate the unknown block-diagonal matrix $\mathbf{B}$.

Note that in the BDGC model the linear independence among the columns of $\mathbf{A}_j$ and among the rows of $\mathbf{S}_j$ is only required within the submatrices. In other words, the columns from different submatrices of $\mathbf{A}$ and the rows from different submatrices of $\mathbf{S}$ can be linearly dependent on each other. Note also that

the BDGC model unifies the GC model [18] - [21], and the DGC model in [36]. We remark that in this dissertation we focus on the complex-valued parameter estimation problem; however the proposed AML estimator can be used directly for real-valued parameter estimation as required in some statistical applications [11].

The remainder of this chapter is organized as follows. We first give some preliminary matrix results in Section 4.2, and then derive the AML estimator in Section 4.3. The theoretical performance analysis and numerical examples are provided in Sections 4.4 and 4.5, respectively. Finally, Section 4.6 contains the conclusions.

4.2 Preliminary Results

In this section, we introduce several partitioned matrix operations and lemmas, which will be used frequently in the derivation and performance analysis of the AML estimator.

Definition 4.1. *Let* $\mathbf{G}$ *be a partitioned matrix with* $\mathbf{G}_{i,j}$ *being the* $(i,\ j)th \quad (i, j = 1,\ 2,\ \cdots J)$ *submatrix of* $\mathbf{G}$*. Then the block-diagonal vectorization operation is defined by*

$$vecb(\mathbf{G}) \triangleq [vec(\mathbf{G}_{1,1})^T\ vec(\mathbf{G}_{2,2})^T\ \cdots\ vec(\mathbf{G}_{J,J})^T]^T, \tag{4–4}$$

where $vec(\cdot)$ *denotes the matrix vectorization operator (stacking the columns of a matrix on top of each other).*

Definition 4.2. *Let* $\mathbf{\Sigma}$ *and* $\mathbf{\Omega}$ *be two partitioned matrices with conformal partitioning and with* $\mathbf{\Sigma}_{i,j}$ *and* $\mathbf{\Omega}_{i,j}$ *being the* $(i,\ j)th$ *submatrices of* $\mathbf{\Sigma}$ *and* $\mathbf{\Omega}$*, respectively. Then the generalized Khatri-Rao product is defined by*

$$[\mathbf{\Sigma} \circledast \mathbf{\Omega}]_{i,j} \triangleq \mathbf{\Sigma}_{i,j} \otimes \mathbf{\Omega}_{i,j} \tag{4–5}$$

where $[\cdot]_{i,j}$ *denotes the* $(i, j)th$ *submatrix of the given partitioned matrix, and* $\otimes$ *denotes the Kronecker matrix product (the generalized Khatri-Rao product was also used, e.g., in [80]).*

Note that the block-diagonal vectorization vecb$(\cdot)$ and the generalized Khatri-Rao product $\circledast$ are defined based on a particular matrix partitioning, i.e., different matrix partitionings will lead to different results. Note also that the standard vectorization vec$(\cdot)$ [74] and the diagonal vectorization vecd$(\cdot)$ [36] are both special cases of the block-diagonal vectorization vecb$(\cdot)$, while the Kronecker product, the Hadamard product and the Khatri-Rao product [77] [78] are all special cases of the generalized Khatri-Rao product based on different matrix partitionings. It is also worth pointing out that matrix partitioning may be " inherited " through matrix operations. For example, for the partitioned matrix $\mathbf{A}$ given by (4–2), $\mathbf{A}^H\mathbf{A}$ is a partitioned matrix with the $(i,\ j)$th $(i, j = 1,\ 2\ , \cdots\ , J)$ submatrix being $\mathbf{A}_i^H\mathbf{A}_j$; hereafter, the superscript H denotes the conjugate transpose of a matrix.

Next, we give two lemmas on partitioned matrix operations.

Lemma 4.1. *Let $\mathbf{\Sigma}$ and $\mathbf{\Omega}$ be two partitioned matrices with K block rows and J block columns, and let $\mathbf{G}$ be a block-diagonal matrix with compatible dimensions and conformal partitioning with $\mathbf{\Sigma}$ and $\mathbf{\Omega}$. Then*

$$vecb(\mathbf{\Sigma G \Omega}^T) = (\mathbf{\Omega} \circledast \mathbf{\Sigma})vecb(\mathbf{G}). \tag{4–6}$$

Proof: We note that $\mathbf{\Sigma G \Omega}^T$ is a partitioned matrix with the (k, k)th $(k = 1,\ 2,\ \cdots, K)$ submatrix being

$$[\mathbf{\Sigma G \Omega}^T]_{k,k} = \sum_{j=1}^{J}[\mathbf{\Sigma}]_{k,j}[\mathbf{G}]_{j,j}[\mathbf{\Omega}]_{k,j}^T. \tag{4–7}$$

Hence,

$$\begin{aligned}
\text{vec}([\mathbf{\Sigma G \Omega}^T]_{k,k}) &= \sum_{j=1}^{J}\text{vec}([\mathbf{\Sigma}]_{k,j}[\mathbf{G}]_{j,j}[\Omega]_{k,j}^T) \\
&= \sum_{j=1}^{J}([\mathbf{\Omega}]_{k,j} \otimes [\mathbf{\Sigma}]_{k,j})\text{vec}([\mathbf{G}]_{j,j}) \\
&= \begin{bmatrix} ([\mathbf{\Omega}]_{k,1} \otimes [\mathbf{\Sigma}]_{k,1}) & \cdots & ([\mathbf{\Omega}]_{k,J} \otimes [\mathbf{\Sigma}]_{k,J}) \end{bmatrix} \text{vecb}(\mathbf{G}),
\end{aligned} \tag{4–8}$$

where we have used the fact that $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}(\mathbf{B})$ [74]. Arranging the vectors $\text{vec}([\mathbf{\Sigma G \Omega}^T]_{k,k})$ $(k = 1,\ 2,\ \cdots\ J)$ in (4–8) into a column vector yields (4–6).

Lemma 4.2. *Let* $\mathbf{U}$ *and* $\mathbf{V}$ *be two partitioned matrices with* 1 *block row and* J *block columns, and let* $\mathbf{H}$ *and* $\mathbf{F}$ *be two partitioned matrices with* 1 *block row and* K *block columns and with compatible dimensions with* $\mathbf{U}$ *and* $\mathbf{V}$*, respectively. Then,*

$$(\mathbf{U} \circledast \mathbf{V})^H(\mathbf{H} \circledast \mathbf{F}) = (\mathbf{U}^H\mathbf{H}) \circledast (\mathbf{V}^H\mathbf{F}). \tag{4–9}$$

Proof: Note that $\mathbf{U} \circledast \mathbf{V}$ is a partitioned matrix with 1 block row and J block columns, and with the $(1, j)$th submatrix being

$$[\mathbf{U} \circledast \mathbf{V}]_{1,j} = [\mathbf{U}]_{1,j} \otimes [\mathbf{V}]_{1,j}. \tag{4–10}$$

Similarly, $\mathbf{H} \circledast \mathbf{F}$ is a partitioned matrix with 1 block row and K block columns, and with the $(1, k)$th submatrix being

$$[\mathbf{H} \circledast \mathbf{F}]_{1,k} = [\mathbf{H}]_{1,k} \otimes [\mathbf{F}]_{1,k}. \tag{4–11}$$

Hence, $(\mathbf{U} \circledast \mathbf{V})^H(\mathbf{H} \circledast \mathbf{F})$ is a partitioned matrix with J block rows and K block columns, and with the (j, k)th submatrix being

$$\begin{aligned}[(\mathbf{U} \circledast \mathbf{V})^H(\mathbf{H} \circledast \mathbf{F})]_{j,k} &= ([\mathbf{U}]_{1,j} \otimes [\mathbf{V}]_{1,j})^H([\mathbf{H}]_{1,k} \otimes [\mathbf{F}]_{1,k}) \\ &= ([\mathbf{U}]_{1,j}^H[\mathbf{H}]_{1,k}) \otimes ([\mathbf{V}]_{1,j}^H[\mathbf{F}]_{1,k}),\end{aligned} \tag{4–12}$$

where we have used the fact that $(\mathbf{A} \otimes \mathbf{B})^H(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{A}^H\mathbf{C}) \otimes (\mathbf{B}^H\mathbf{D})$ [74].

On the other hand, $\mathbf{U}^H\mathbf{H}$ and $\mathbf{V}^H\mathbf{F}$ are two partitioned matrices with J block rows and K block columns, and with the (j, k)th submatrices being

$$[\mathbf{U}^H\mathbf{H}]_{j,k} = [\mathbf{U}]_{1,j}^H[\mathbf{H}]_{1,k} \tag{4–13}$$

$$[\mathbf{V}^H\mathbf{F}]_{j,k} = [\mathbf{V}]_{1,j}^H[\mathbf{F}]_{1,k}. \tag{4–14}$$

From (4–12), (4–13) and (4–14), we obtain

$$[(\mathbf{U} \circledast \mathbf{V})(\mathbf{H} \circledast \mathbf{F})]_{j,k} = [\mathbf{U}^H\mathbf{H}]_{j,k} \otimes [\mathbf{V}^H\mathbf{F}]_{j,k}, \tag{4–15}$$

which yields (4–9) immediately.

We remark that Lemmas 3.1 and 3.2 in Chapter 3 are special cases of Lemma 4.1 in this chapter, while Lemmas A1 and A2 in [78] and Lemma 3.3 in [36] are special cases of Lemma 4.2 in this chapter.

4.3 Approximate Maximum Likelihood Estimation

In this section, we derive an approximate maximum likelihood (AML) estimator for the unknown block-diagonal regression coefficient matrix $\mathbf{B}$ in (4–1). Our approach is based on the ML principle, but an approximation is made to get a closed-form solution. In the following derivation, we utilize the partitioned matrix operations and lemmas of Section 4.2. Note that the partitioned matrix operations used in the derivations are based on the matrix partitionings of $\mathbf{A}$, $\mathbf{S}$ and $\mathbf{B}$ in (4–2), (4–3) and (4–1), respectively. The matrices $\mathbf{X}$ and $\mathbf{Z}$ are both treated as non-partitioned matrices.

For convenience, we arrange the unknowns in $\mathbf{B}$ into a column vector, i.e., $\boldsymbol{\beta} = \text{vecb}(\mathbf{B})$. It follows from (4–1) that the negative log-likelihood function of the observed data samples is (to within an additive constant)

$$f(\boldsymbol{\beta}, \mathbf{Q}) = L\ln|\mathbf{Q}| + \text{tr}[\mathbf{Q}^{-1}(\mathbf{X} - \mathbf{ABS})(\mathbf{X} - \mathbf{ABS})^H], \tag{4–16}$$

where $\text{tr}(\cdot)$ and $|\cdot|$ denote the trace and the determinant of a matrix, respectively.

Minimizing the negative log-likelihood function with respect to $\mathbf{Q}$ yields

$$\hat{\mathbf{Q}} = \frac{1}{L}(\mathbf{X} - \mathbf{ABS})(\mathbf{X} - \mathbf{ABS})^H. \tag{4–17}$$

For $M \leq L$, $\hat{\mathbf{Q}}$ is nonsingular with probability one.

Substituting (4–17) into (4–16) yields the ML estimate of $\boldsymbol{\beta}$

$$\hat{\boldsymbol{\beta}}_{\text{ML}} = \arg\min_{\boldsymbol{\beta}} \ln|(\mathbf{X} - \mathbf{ABS})(\mathbf{X} - \mathbf{ABS})^H|. \tag{4–18}$$

In general, the optimization problem in (4–18) does not appear to admit a closed-form solution. Here, we use the technique in [25] and [27] to solve this problem approximately. Let $\mathbf{\Pi}_S$ and $\mathbf{\Pi}_S^{\perp}$, respectively, denote the orthogonal projection matrices given by

$$\mathbf{\Pi}_S \triangleq \mathbf{S}^H(\mathbf{S}\mathbf{S}^H)^{-}\mathbf{S}, \tag{4–19}$$

and

$$\mathbf{\Pi}_S^{\perp} \triangleq \mathbf{I} - \mathbf{\Pi}_S, \tag{4–20}$$

where $(\cdot)^{-}$ denotes the pseudo or generalized inverse of a matrix [74]. Note that when the rows of $\mathbf{S}$ are not linearly independent of each other, the matrix $(\mathbf{S}\mathbf{S}^H)$ is singular and its generalized inverse is not unique. However, $\mathbf{\Pi}_S$ and $\mathbf{\Pi}_S^{\perp}$ are unique [74].

Using (4–19) and (4–20), we get

$$\begin{aligned} &|(\mathbf{X} - \mathbf{ABS})(\mathbf{X} - \mathbf{ABS})^H| \\ =&|(\mathbf{X} - \mathbf{ABS})(\mathbf{\Pi}_S + \mathbf{\Pi}_S^{\perp})(\mathbf{X} - \mathbf{ABS})^H| \\ =&|(\mathbf{X}\mathbf{\Pi}_S - \mathbf{ABS})(\mathbf{X}\mathbf{\Pi}_S - \mathbf{ABS})^H + \mathbf{T}| \\ =&|(\mathbf{X}\mathbf{\Pi}_S - \mathbf{ABS})(\mathbf{X}\mathbf{\Pi}_S - \mathbf{ABS})^H\mathbf{T}^{-1} + \mathbf{I}|\,|\mathbf{T}|, \end{aligned} \tag{4–21}$$

where the matrix $\mathbf{T}$ is defined by

$$\mathbf{T} \triangleq \mathbf{X}\mathbf{\Pi}_S^{\perp}\mathbf{X}^H. \tag{4–22}$$

Note that the rank of $\mathbf{\Pi}_S^{\perp}$ is $L - r$. Hence, $\mathbf{T}$ has full rank M with probability one when $M + r \leq L$.

Using (4–21) and Lemma 4 in [25] (or Theorem 1 in [27]), the solution of the optimization problem in (4–18) can be approximated, for a large snapshot number L, as follows.

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{\text{AML}} &= \arg\min_{\boldsymbol{\beta}} \operatorname{tr}[(\mathbf{X}\boldsymbol{\Pi}_S - \mathbf{ABS})^H \mathbf{T}^{-1}(\mathbf{X}\boldsymbol{\Pi}_S - \mathbf{ABS})] \\ &= \arg\min_{\boldsymbol{\beta}} \parallel \operatorname{vec}(\mathbf{T}^{-\frac{1}{2}}\mathbf{X}\boldsymbol{\Pi}_S) - \operatorname{vec}(\mathbf{T}^{-\frac{1}{2}}\mathbf{ABS}) \parallel^2,\end{aligned} \tag{4–23}$$

where $\mathbf{T}^{-\frac{1}{2}}$ is the Hermitian square root of $\mathbf{T}^{-1}$ and $\parallel \cdot \parallel$ denotes the Euclidean vector norm .

By using Lemma 4.1 (with $K = 1$), we have

$$\operatorname{vec}(\mathbf{T}^{-\frac{1}{2}}\mathbf{ABS}) = \operatorname{vecb}(\mathbf{T}^{-\frac{1}{2}}\mathbf{ABS}) = [\mathbf{S}^T \circledast (\mathbf{T}^{-\frac{1}{2}}\mathbf{A})]\boldsymbol{\beta}. \tag{4–24}$$

Substituting (4–24) into (4–23) yields a quadratic function of $\boldsymbol{\beta}$, whose minimizer is given by

$$\hat{\boldsymbol{\beta}}_{\text{AML}} = (\boldsymbol{\Gamma}^H\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}^H \operatorname{vec}(\mathbf{T}^{-\frac{1}{2}}\mathbf{X}\boldsymbol{\Pi}_S) \tag{4–25}$$

where

$$\boldsymbol{\Gamma} \triangleq \mathbf{S}^T \circledast (\mathbf{T}^{-\frac{1}{2}}\mathbf{A}). \tag{4–26}$$

To guarantee that the matrix $\boldsymbol{\Gamma}^H\boldsymbol{\Gamma}$ is invertible, we assume that the columns of $\boldsymbol{\Gamma}$ are linearly independent of each other, which requires the linear independence among the columns within each submatrix of $\mathbf{A}$ and among the rows within each submatrix of $\mathbf{S}$. However, the columns from different submatrices of $\mathbf{A}$ and the rows from different submatrices of $\mathbf{S}$ can be linearly dependent on each other.

By Lemmas 4.2 and 4.1, respectively, we have that

$$\boldsymbol{\Gamma}^H\boldsymbol{\Gamma} = (\mathbf{SS}^H)^T \circledast (\mathbf{A}^H\mathbf{T}^{-1}\mathbf{A}) \tag{4–27}$$

and

$$\boldsymbol{\Gamma}^H \operatorname{vec}(\mathbf{T}^{-\frac{1}{2}}\mathbf{X}\boldsymbol{\Pi}_S) = \boldsymbol{\Gamma}^H \operatorname{vecb}(\mathbf{T}^{-\frac{1}{2}}\mathbf{X}\boldsymbol{\Pi}_S) = \operatorname{vecb}(\mathbf{A}^H\mathbf{T}^{-1}\mathbf{XS}^H). \tag{4–28}$$

Substituting (4–27) and (4–28) into (4–25) yields the AML estimate of $\boldsymbol{\beta}$

$$\hat{\boldsymbol{\beta}}_{\text{AML}} = [(\mathbf{S}\mathbf{S}^H)^T \circledast (\mathbf{A}^H\mathbf{T}^{-1}\mathbf{A})]^{-1}\text{vecb}(\mathbf{A}^H\mathbf{T}^{-1}\mathbf{X}\mathbf{S}^H). \tag{4–29}$$

We note that in (4–29) when $J = 1$ the generalized Khatri-Rao product reduces to the Kronecker product and vecb$(\cdot)$ reduces to vec$(\cdot)$. Then, using the fact that $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}(\mathbf{B})$, it can be readily shown that the AML estimator in (4–29) reduces to

$$\hat{\mathbf{B}}_{\text{GC-ML}} = (\mathbf{A}^H\mathbf{T}^{-1}\mathbf{A})^{-1}\mathbf{A}^H\mathbf{T}^{-1}\mathbf{X}\mathbf{S}^H(\mathbf{S}\mathbf{S}^H)^{-1}, \tag{4–30}$$

which is the exact ML estimator based on the GC model [17] [18] [21] [22]. On the other hand, when $N_j = K_j = 1$ for $j = 1, 2, \cdots J$, the generalized Khatri-Rao product reduces to the Hadamard product and vecb$(\cdot)$ reduces to vecd$(\cdot)$. Hence, (4–29) reduces to the AML estimator for the DGC model in [36], namely

$$\hat{\boldsymbol{\beta}}_{\text{DGC-AML}} = [(\mathbf{A}^H\mathbf{T}^{-1}\mathbf{A}) \odot (\mathbf{S}\mathbf{S}^H)^T]^{-1}\text{vecd}(\mathbf{A}^H\mathbf{T}^{-1}\mathbf{X}\mathbf{S}^H), \tag{4–31}$$

where $\odot$ denotes the Hadamard product (elementwise multiplication) between two matrices, and vecb$(\cdot)$ denotes a column vector formed by the diagonal elements of a matrix.

4.4 Performance Analysis

In this section, we establish the theoretical properties of the AML estimator.

4.4.1 Bias Analysis

Substituting (4–1) into (4–29) and using Lemma 4.1, we have that

$$\hat{\boldsymbol{\beta}}_{\text{AML}} - \boldsymbol{\beta} = [(\mathbf{S}\mathbf{S}^H)^T \circledast (\mathbf{A}^H\mathbf{T}^{-1}\mathbf{A})]^{-1}\text{vecb}(\mathbf{A}^H\mathbf{T}^{-1}\mathbf{Z}\mathbf{S}^H). \tag{4–32}$$

On the other hand, substituting (4–1) into (4–22) yields

$$\mathbf{T} = \mathbf{Z}\boldsymbol{\Pi}_S^{\perp}\mathbf{Z}^H. \tag{4–33}$$

From (4–33), we see that $\mathbf{T}$ is an even function of $\mathbf{Z}$, and hence $\hat{\boldsymbol{\beta}}_{\text{AML}} - \boldsymbol{\beta}$ is an odd function of $\mathbf{Z}$. Moreover, we have assumed that $\mathbf{Z}$ is a zero-mean Gaussian random matrix. Using the statistical properties of the zero-mean Gaussian distribution, we can readily show that the expectation of (4–32) is zero, i.e., the AML estimator based on the BDGC model is unbiased.

4.4.2 Mean-Squared-Error (MSE) Analysis

Before calculating the MSE of the AML estimate, we first discuss the best possible performance for any unbiased estimate of β, i.e., the Cramér-Rao bound (CRB). By utilizing Lemma 4.1, (4–1) can be rewritten as

$$\text{vec}(\mathbf{X}) = (\mathbf{S}^T \circledast \mathbf{A})\boldsymbol{\beta} + \text{vec}(\mathbf{Z}). \tag{4–34}$$

Note that (4–34) is a linear statistical model with unknown noise covariance matrix $\mathbf{I} \otimes \mathbf{Q}$. It can be easily verified that the Fisher information matrix for this model is a block-diagonal matrix with respect to $\boldsymbol{\beta}$ and $\mathbf{Q}$ [73]. Hence, the unknowns in $\mathbf{Q}$ do not affect the CRB for $\boldsymbol{\beta}$. It follows that the CRB for $\boldsymbol{\beta}$ can be readily written [73] as

$$\text{CRB}(\boldsymbol{\beta}) = [(\mathbf{S}^T \circledast \mathbf{A})^H (\mathbf{I} \otimes \mathbf{Q})^{-1} (\mathbf{S}^T \circledast \mathbf{A})]^{-1}. \tag{4–35}$$

Then, by using Lemma 4.2, (4–35) can be simplified as

$$\text{CRB}(\boldsymbol{\beta}) = [(\mathbf{S}\mathbf{S}^H)^T \circledast (\mathbf{A}^H \mathbf{Q}^{-1} \mathbf{A})]^{-1}. \tag{4–36}$$

Next, we turn to the MSE analysis of $\hat{\boldsymbol{\beta}}_{\text{AML}}$. The exact MSE of $\hat{\boldsymbol{\beta}}_{\text{AML}}$ is difficult to determine, if not impossible. Herein, we provide an approximate expression for the MSE of $\hat{\boldsymbol{\beta}}_{\text{AML}}$, which holds for a large snapshot number L.

From the theory of the linear statistical model [16], we know that $\frac{\mathbf{T}}{L-r}$ is an unbiased and consistent (in L) estimate of the noise covariance matrix $\mathbf{Q}$. Hence, for a large snapshot number L, the estimation error in (4–32) can be approximated

by

$$\hat{\boldsymbol{\beta}}_{\text{AML}} - \boldsymbol{\beta} \approx \text{CRB}(\boldsymbol{\beta})\,\text{vecd}(\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{Z}\mathbf{S}^H). \tag{4–37}$$

Using Lemmas 4.1 and 4.2 (and viewing $\otimes$ as a special case of $\circledast$), it follows from (4–37) that

$$\begin{aligned}
\text{MSE}(\hat{\boldsymbol{\beta}}_{\text{AML}}) \triangleq& \text{E}[(\hat{\boldsymbol{\beta}}_{\text{AML}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_{\text{AML}} - \boldsymbol{\beta})^H] \\
\approx& \text{CRB}(\boldsymbol{\beta})\,\text{E}[\text{vecb}(\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{Z}\mathbf{S}^H)\,\text{vecb}(\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{Z}\mathbf{S}^H)^H]\,\text{CRB}(\boldsymbol{\beta}) \\
=& \text{CRB}(\boldsymbol{\beta})\,[\mathbf{S}^* \circledast (\mathbf{A}^H\mathbf{Q}^{-1})]\,\text{E}[\text{vec}(\mathbf{Z})\text{vec}(\mathbf{Z})^H]\,[\mathbf{S}^T \circledast (\mathbf{A}^H\mathbf{Q}^{-1})^H]\,\text{CRB}(\boldsymbol{\beta}) \\
=& \text{CRB}(\boldsymbol{\beta})\,[\mathbf{S}^* \circledast (\mathbf{A}^H\mathbf{Q}^{-1})]\,(\mathbf{I} \otimes \mathbf{Q})\,[\mathbf{S}^T \circledast (\mathbf{Q}^{-1}\mathbf{A})]\,\text{CRB}(\boldsymbol{\beta}) \\
=& \text{CRB}(\boldsymbol{\beta}),
\end{aligned} \tag{4–38}$$

where $(\cdot)^*$ denotes the complex conjugate.

We see from (4–38) that the AML estimate of $\boldsymbol{\beta}$ is asymptotically statistically efficient for a large snapshot number L. This theoretical result is illustrated numerically in Section 4.5. It can be easily verified that the CRB for the GC model in Equation (13) of [22] and the CRB for the DGC model in Equation (24) of [36]) are special cases of (4–36).

4.5 Numerical Results

In this section, numerical examples illustrating the application of the BDGC model in wireless communications are presented to demonstrate the performance of the AML estimator and to verify the theoretical results in Section 4.4.

We consider a DS-CDMA receiver with a uniform linear array consisting of $M = 6$ antennas with half-wavelength spacing between adjacent antennas. The array elements are assumed to be omni-directional. Consider $J = 2$ transmitters, whose signals are modulated by two different pseudo-noise (PN) sequences [81], respectively. The PN sequences are known *a priori* by the receiver. Suppose that the signal from the 1st transmitter arrives at the array through $N_1 = 2$ paths

with directions of arrival (DOAs) of $\theta_1 = -10°$ and $\theta_2 = 5°$, while the signal from the 2nd transmitter arrives at the array through $N_2 = 1$ path with DOA of $\theta_3 = 10°$. We assume that the DOAs have been estimated accurately using, e.g., the method of direction estimation (MODE) algorithm [82] [83], and therefore they can be considered to be known. The problem of interest is to estimate the unknown complex amplitudes β_1 and β_2 for the two paths of the 1st transmitter, and β_3 for the 2nd transmitter, which contain the transmitted information. The estimates of $\{\beta_j\}_{j=1}^{3}$ can be used for symbol detection. The error matrix $\mathbf{Z}$ contains the noise and interference, whose columns are assumed to be i.i.d. circularly symmetric complex Gaussian random vectors with zero-mean and an unknown covariance matrix $\mathbf{Q}$ given by

$$Q_{m,n} = \rho\, 0.9^{|m-n|} e^{j\frac{(m-n)\pi}{2}} \tag{4–39}$$

where $\rho = 1/\text{SNR}$ with SNR being the signal-to-noise ratio and $Q_{m,n}$ denotes the $(m,\, n)$th element of $\mathbf{Q}$.

Let

$$\mathbf{A}_1 = [\mathbf{a}(\theta_1) \quad \mathbf{a}(\theta_2)] \qquad \text{and} \qquad \mathbf{A}_2 = [\mathbf{a}(\theta_3)], \tag{4–40}$$

where $\mathbf{a}(\theta) = [1\ e^{-j\pi sin(\theta)}\ e^{-j2\pi sin(\theta)}\ \cdots\ e^{-j(M-1)\pi sin(\theta)}]^T$ is the steering vector of the signal with DOA of θ. Let

$$\mathbf{B}_1 = [\beta_1 \quad \beta_2]^T \qquad \text{and} \qquad \mathbf{B}_2 = [\beta_3]. \tag{4–41}$$

Let $\mathbf{S}_1 \in C^{1\times L}$ and $\mathbf{S}_2 \in C^{1\times L}$ be the known PN sequences for the two transmitters, respectively, with L being the length of each PN sequence. Under the previous assumptions we can describe the received signal using a BDGC model with $M = 6$, $J = 2$, $N_1 = 2$, $N_2 = 1$ and $K_1 = K_2 = 1$. Hence, the AML estimator in (4–29) can be applied directly.

We present the MSE of the AML estimator in (4–29) as well as the corresponding CRB in (4–36). For comparison purposes, we also show the performances of the GC

method, the least squares (LS) and the exact ML estimators. In the GC method, we estimate the full matrix of $\mathbf{B}$ using (4–30) [17] [18] [21] [22], and then pick up the corresponding block-diagonal submatrices of $\hat{\mathbf{B}}_{\text{GC-ML}}$ as the estimate of $\mathbf{B}_j$. The LS estimator assumes that the interference and noise vectors in $\mathbf{Z}$ are uncorrelated both spatially and temporally, and can be obtained immediately from (4–29) by replacing $\mathbf{T}$ there by an identity matrix, i.e.,

$$\hat{\beta}_{\text{LS}} = [(\mathbf{S}\mathbf{S}^H)^T \circledast (\mathbf{A}^H\mathbf{A})]^{-1}\text{vecb}(\mathbf{A}^H\mathbf{X}\mathbf{S}^H). \tag{4–42}$$

The exact ML estimates are obtained by applying the cyclic maximization (CM) technique [84] to the cost function in (4–18) with respect to various $\mathbf{B}_j$. In each step of this iterative algorithm, we assume that all regression coefficient submatrices, except for $\mathbf{B}_k$, are known, which means that $\mathbf{B}_k$ can be readily estimated by using (4–30). We use the AML estimates as the initial values of the regression coefficient matrices in the exact ML estimator. The empirical estimation performances are all obtained from 500 Monte-Carlo simulations.

Figs. 4–1 and 4–1 show the estimation performance as functions of the length of the PN sequences L and the SNR, respectively. The SNR is fixed at 10 dB in Fig. 4–1 while L is fixed to be 128 in Fig. 4–2. Note that the ML estimator based on the GC model has a relatively poor estimation performance. This method ignores the *a priori* information about the block-diagonal structure of the complex amplitude matrix $\mathbf{B}$, which doubles the number of unknowns and hence results in much larger estimation variances. Note also that due to the (quasi-)orthogonal property of the PN sequences we have $\mathbf{R}_{SS} = L\,\mathbf{I}$ (approximately), and hence the CRB in (4–36) reduces to $\frac{1}{L}[\mathbf{I} \circledast (\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{A})]^{-1}$ in this example. As shown in Figs. 4–1(a)-4–1(c), the CRB decreases linearly as the snapshot number L increases, when both CRB and L are presented in log-scales. From Figs. 4–1(a) - 4–1(c), we can also see that the AML estimator achieves a very similar performance as the

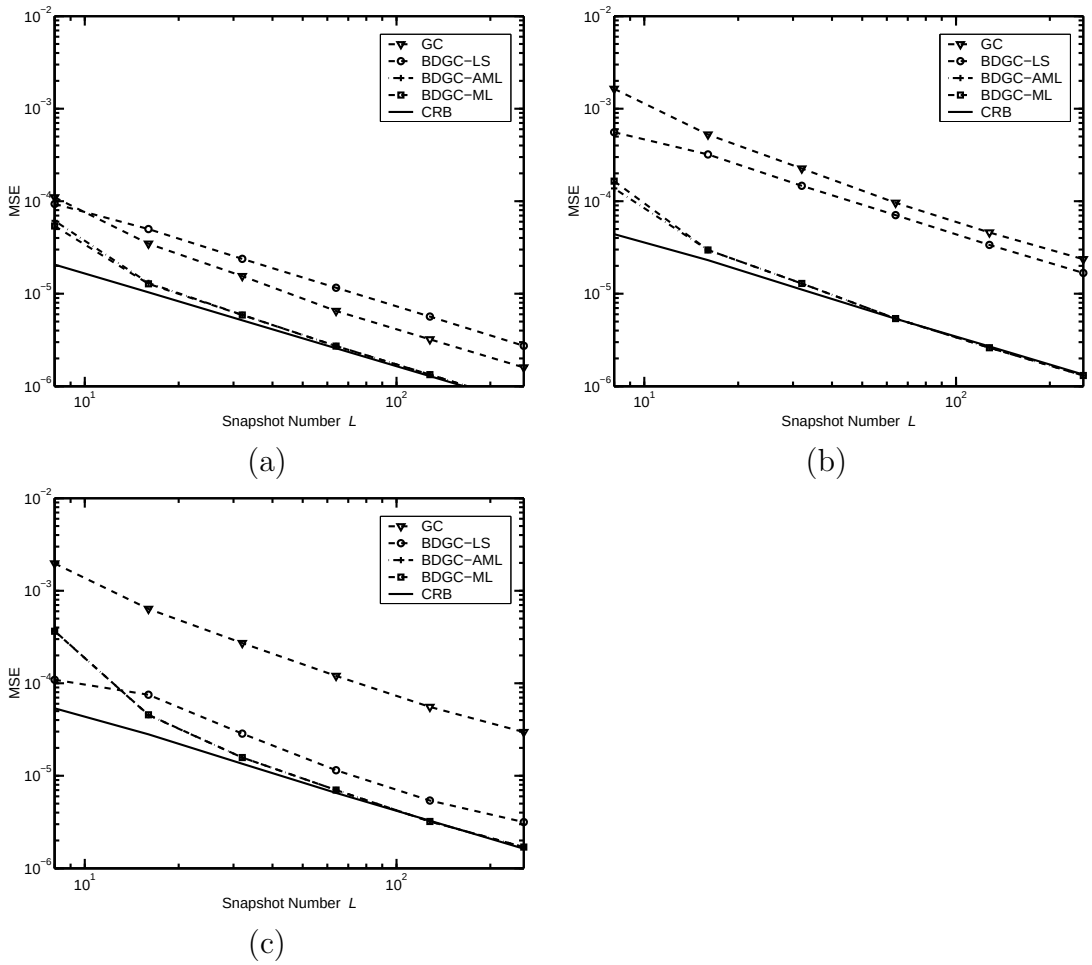


Figure 4–1: Empirical MSE's and the CRB versus L when SNR $=$ 10 dB. (a) For β_1, (b) for β_2, and (c) for β_3.

exact ML; and it outperforms LS and GC significantly. As predicted theoretically, both the exact ML and AML are asymptotically statistically efficient for a large snapshot number, and they approach the corresponding CRB rapidly. From Fig. 4–2, we note that AML and the exact ML provide almost identical performances, and their estimates are very close to the CRB for the entire range of SNR considered. Again, AML outperforms the LS and GC estimators significantly.

4.6 Conclusions

We have presented an approximate maximum likelihood estimator for the block-diagonal growth-curve model where the unknown regression coefficient

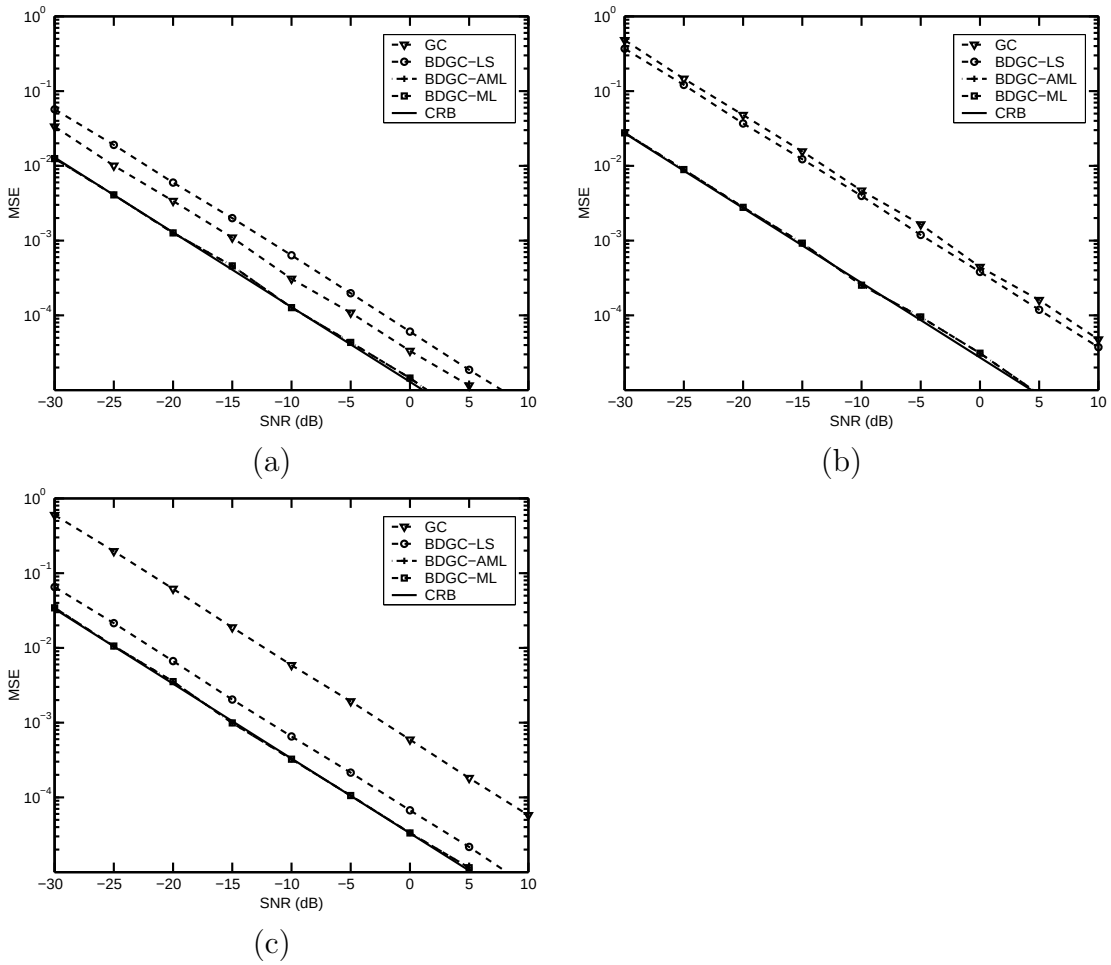


Figure 4–2: Empirical MSE's and the CRB versus SNR (dB) when $L = 128$. (a) For β_1, (b) for β_2, and (c) for β_3.

matrix is constrained to be block-diagonal. Via a theoretical analysis, we have shown that the AML estimator is unbiased and asymptotically statistically efficient for a large snapshot number. We have applied the AML estimator to a complex amplitude estimation problem in wireless communications. The numerical examples provide compelling evidence that the AML method can achieve excellent estimation accuracy.

CHAPTER 5
ITERATIVE GENERALIZED LIKELIHOOD RATIO TEST FOR MIMO RADAR

5.1 Introduction and Signal Model

In Chapter 1, we have discussed several multiple-input multiple-output (MIMO) radar systems. According to their antenna configurations, the MIMO radars discussed can be grouped into two classes. One is the conventional radar array, in which both transmitting and receiving antennas are closely spaced for coherent transmission and detection [44] - [57]. The other is the diverse antenna configuration, where the antennas are separated far away from each other to achieve spatial diversity gain [40] - [43]. To reap the benefits of both schemes, in this chapter, we consider a general antenna configuration, i.e., both the transmitting and receiving antenna arrays consist of several well-separated subarrays with each subarray containing closely-spaced antennas. We establish the growth-curve models in Chapters 2 - 4 and devise several estimators for the proposed MIMO radar system.

Consider a narrow-band MIMO radar system with $\tilde{N}$ and $\tilde{M}$ subarrays for transmitting and receiving, respectively. The nth transmit and mth receive subarrays have, respectively, N_n and M_m closely-spaced antennas, $n = 1,\ 2,\ \cdots,\ \tilde{N}$, $m = 1,\ 2,\ \cdots,\ \tilde{M}$. We assume that the subarrays are sufficiently separated, and hence for each target its radar cross-sections (RCS) for different transmit and receive subarray pairs are statistically independent of each other. Let $\mathbf{v}_n(\theta)$ and $\mathbf{a}_m(\theta)$ be the steering vectors of the nth transmitting subarray and the mth receiving subarray, respectively, where θ denotes the target location parameter, for example its angular location. Let the rows of $\mathbf{\Phi}_n$ be the waveforms transmitted from the antennas of the nth transmit subarray. We assume that the arrival time is

known. Then, the signal received by the mth subarray due to the reflection of the target at θ can be written as

$$\mathbf{X}_m = \sum_{n=1}^{\tilde{N}} \mathbf{a}_m(\theta)\beta_{mn,\theta}\, \mathbf{v}_n^T(\theta)\mathbf{\Phi}_n + \mathbf{Z}_m, \quad m = 1, \cdots, \tilde{M}, \tag{5–1}$$

where $\beta_{mn,\theta}$ is the complex amplitude proportional to the RCS for the (m, n)th receive and transmit subarray pair and for the target at the location θ, and the matrix $\mathbf{Z}_m$ denotes the residual term containing the unmodelled noise, interferences from targets other than θ and at other range bins, and intentional or unintentional jamming. For notational simplicity, we will not show explicitly the dependence of $\mathbf{Z}_m$ on θ.

Let

$$\mathbf{X} = [\mathbf{X}_1^T \ \cdots \ \mathbf{X}_{\tilde{M}}^T]^T \in C^{M\times L}, \tag{5–2}$$

$$\mathbf{A}(\theta) = \mathrm{Diag}[\mathbf{a}_1(\theta), \cdots, \mathbf{a}_{\tilde{M}}(\theta)] \in C^{M\times\tilde{M}}, \tag{5–3}$$

$$\mathbf{V}(\theta) = \mathrm{Diag}[\mathbf{v}_1(\theta), \cdots, \mathbf{v}_{\tilde{N}}(\theta)] \in C^{N\times\tilde{N}}, \tag{5–4}$$

and

$$\mathbf{\Phi} = [\mathbf{\Phi}_1^T \ \cdots \ \mathbf{\Phi}_{\tilde{N}}^T]^T \in C^{N\times L}, \tag{5–5}$$

where $M = M_1 + \cdots + M_{\tilde{M}}$ and $N = N_1 + \cdots + N_{\tilde{N}}$ are the total numbers of receive and transmit antennas, respectively, L is the number of data samples of the transmitted waveforms, $(\cdot)^T$ denotes the transpose operator, and $\mathrm{Diag}(\mathbf{a}_1, \cdots, \mathbf{a}_{\tilde{M}})$ is a block-diagonal matrix with $\mathbf{a}_1, \cdots, \mathbf{a}_{\tilde{M}}$ being its diagonal submatrices. Then, (5–1) can be readily rewritten in the growth-curve (GC) model in Chapter 2, i.e.,

$$\mathbf{X} = \mathbf{A}(\theta)\mathbf{B}_\theta\mathbf{S}(\theta) + \mathbf{Z} \tag{5–6}$$

where the (m, n)th element of the $\tilde{M} \times \tilde{N}$ matrix $\mathbf{B}_\theta$ is $\beta_{mn,\theta}$, $\mathbf{Z}$ is defined similarly to $\mathbf{X}$ in (5–2), and the rows of $\mathbf{S}(\theta)$ are the reflected waveforms by the target at

location θ, i.e.,

$$\mathbf{S}(\theta) = \mathbf{V}^T(\theta)\mathbf{\Phi}. \tag{5–7}$$

Note that when $\tilde{N} = \tilde{M} = 1$, the signal model in (5–6) reduces to the MIMO radar model in [55] - [57], whereas when $N = \tilde{N}$ and $M = \tilde{M}$ it reduces to the diversity data model in [40] - [42]. Based on this data model, We below propose two classes of nonparametric methods, i.e., spatial spectral estimation and generalized likelihood ratio test (GLRT), for target detection and localization.

The remainder of this Chapter is organized as follows. In Section 5.2, we introduce several adaptive spatial spectral estimators including Capon [60] and APES [23]. In Section 5.3, we describe a generalized likelihood ratio test (GLRT) and a conditional generalized likelihood ratio test (cGLRT), and we then propose an iterative GLRT (iGLRT) procedure for target detection and parameter estimation. Numerical examples are provided in Section 5.4. We first compare the Cramér-Rao bounds (CRBs) for MIMO radars with different antenna configurations, and then present the detection and localization performance of the proposed methods. Finally, Section 5.5 contains the conclusions.

5.2 Several Spatial Spectral Estimators

We introduce several spatial spectral estimators for the proposed MIMO radar system. We use these methods to estimate the complex amplitudes in $\mathbf{B}_\theta$ for each θ of interest from the observed data matrix $\mathbf{X}$. The Frobenius norm of the so-obtained $\hat{\mathbf{B}}_\theta$ forms a spatial spectrum in the 1D case or a radar image in the 2D case. We can then estimate the number of targets and their locations by searching for the peaks in the so-obtained spectrum (or image).

A simple way to estimate $\mathbf{B}_\theta$ in (5–6) is via the Least-Squares (LS) method, i.e.,

$$\hat{\mathbf{B}}_{\text{LS},\theta} = [\mathbf{A}^H(\theta)\mathbf{A}(\theta)]^{-1}\mathbf{A}(\theta)\mathbf{X}\mathbf{S}^H(\theta)[\mathbf{S}(\theta)\mathbf{S}^H(\theta)]^{-1}, \tag{5–8}$$

where $(\cdot)^H$ denotes the conjugate transpose. However, as any other data-independent beamforming-type method, the LS method suffers from high-sidelobes and low resolution. In the presence of strong interference and jamming, the method completely fails to work. We introduce below two adaptive spatial spectral estimation approaches that offer much higher resolution and interference suppression capabilities.

5.2.1 Capon

The Capon estimator for $\mathbf{B}_\theta$ in (5–6) consists of two main steps [60] [85] [22]. The first is a generalized Capon beamforming step. The second is a LS estimation step, which involves basically a matched filtering to the known waveform $\mathbf{S}(\theta)$.

The generalized Capon beamformer can be formulated as

$$\min_{\mathbf{W}} \operatorname{tr}(\mathbf{W}^H\mathbf{R}\mathbf{W}) \quad \text{subject to} \quad \mathbf{W}^H\mathbf{A}(\theta) = \mathbf{I}, \tag{5–9}$$

where $\mathbf{W} \in C^{M\times\tilde{M}}$ is the weighting matrix used to achieve noise, interference and jamming suppression while keeping the desired signal undistorted, $\operatorname{tr}(\cdot)$ denotes the trace of a matrix, and

$$\hat{\mathbf{R}} = \frac{1}{L}\mathbf{X}\mathbf{X}^H \tag{5–10}$$

is the sample covariance matrix with L being the number of data samples.

Solving the optimization problem in (5–9), we can readily have

$$\hat{\mathbf{W}}_{\text{Capon}} = \hat{\mathbf{R}}^{-1}\mathbf{A}(\theta)[\mathbf{A}^H(\theta)\hat{\mathbf{R}}^{-1}\mathbf{A}(\theta)]^{-1}. \tag{5–11}$$

By using (5–11) and (5–6), the output of the Capon beamformer can be written as

$$[\mathbf{A}^H(\theta)\hat{\mathbf{R}}^{-1}\mathbf{A}(\theta)]^{-1}\mathbf{A}^H(\theta)\hat{\mathbf{R}}^{-1}\mathbf{X} = \mathbf{B}_\theta\mathbf{S}(\theta)+[\mathbf{A}^H(\theta)\hat{\mathbf{R}}^{-1}\mathbf{A}(\theta)]^{-1}\mathbf{A}^H(\theta)\hat{\mathbf{R}}^{-1}\mathbf{Z}. \tag{5–12}$$

By applying the least-squares (LS) method to (5–12), the Capon estimate of $\mathbf{B}_\theta$ follows readily, i.e.,

$$\hat{\mathbf{B}}_{\text{Capon},\theta} = [\mathbf{A}^H(\theta)\hat{\mathbf{R}}^{-1}\mathbf{A}(\theta)]^{-1}\mathbf{A}^H(\theta)\hat{\mathbf{R}}^{-1}\mathbf{X}\mathbf{S}^H(\theta)[\mathbf{S}(\theta)\mathbf{S}^H(\theta)]^{-1}. \tag{5–13}$$

5.2.2 APES

The generalized APES method is a straightforward extension of the APES method [23] [24], which can be formulated as

$$\min_{\mathbf{W},\mathbf{B}} \parallel \mathbf{W}^H\mathbf{X} - \mathbf{B}_\theta\mathbf{S}(\theta) \parallel^2 \quad \text{subject to} \quad \mathbf{W}^H\mathbf{A}(\theta) = \mathbf{I}, \tag{5–14}$$

where $\parallel \cdot \parallel$ denotes the Frobenius norm, and $\mathbf{W}$ is the weighting matrix. Minimizing the cost function in (5–14) with respect to $\mathbf{B}_\theta$ yields

$$\hat{\mathbf{B}}_{\text{APES},\theta} = \mathbf{W}^H\mathbf{X}\mathbf{S}^H(\theta)[\mathbf{S}(\theta)\mathbf{S}^H(\theta)]^{-1}. \tag{5–15}$$

Then the optimization problem reduces to

$$\min \operatorname{tr}(\mathbf{W}^H\hat{\mathbf{Q}}\mathbf{W}) \quad \text{subject to} \quad \mathbf{W}^H\mathbf{A}(\theta) = \mathbf{I}, \tag{5–16}$$

with

$$\hat{\mathbf{Q}} = \hat{\mathbf{R}} - \frac{1}{L}\mathbf{X}\mathbf{S}^H(\theta)[\mathbf{S}(\theta)\mathbf{S}^H(\theta)]^{-1}\mathbf{S}(\theta)\mathbf{X}^H. \tag{5–17}$$

For notional simplicity, we have omitted the dependence of $\hat{\mathbf{Q}}$ on θ.

Solving the optimization problem of (5–16) gives the generalized APES beamformer weighting matrix

$$\hat{\mathbf{W}}_{\text{APES},\theta} = [\mathbf{A}^H(\theta)\hat{\mathbf{Q}}\mathbf{A}(\theta)]^{-1}\hat{\mathbf{Q}}^{-1}\mathbf{A}(\theta). \tag{5–18}$$

Inserting (5–18) in (5–15), we readily get the APES estimate of $\mathbf{B}_\theta$ as

$$\hat{\mathbf{B}}_{\text{APES},\theta} = [\mathbf{A}^H(\theta)\hat{\mathbf{Q}}^{-1}\mathbf{A}(\theta)]^{-1}\mathbf{A}^H(\theta)\hat{\mathbf{Q}}^{-1}\mathbf{X}\mathbf{S}^H(\theta)[\mathbf{S}(\theta)\mathbf{S}^H(\theta)]^{-1}. \tag{5–19}$$

Interestingly, we note that (5–19) has the same form as the ML estimate in Chapter 2. However, the APES estimate is derived based on the beamforming method, and, unlike the ML in Chapter 2, it does not need probability density function (pdf) of $\mathbf{Z}$.

5.3 Generalized Likelihood Ratio Test

Generalized likelihood ratio test (GLRT) has been used widely for target detection and localization. We derive below a GLRT and a conditional generalized likelihood ratio test (cGLRT) for the proposed MIMO radar, and then propose an iterative GLRT (iGLRT) procedure for improved performance.

5.3.1 Generalized Likelihood Ratio Test (GLRT)

Throughout this section, we assume that the columns of the interference and noise term $\mathbf{Z}$ in (5–6) are independently and identically distributed (i.i.d.) circularly symmetric complex Gaussian random vectors with mean zero and an unknown covariance matrix $\mathbf{Q}$.

Consider the following hypothesis test problem

$$\begin{cases} \mathrm{H}_0: & \mathbf{X} = \mathbf{Z} \\ \mathrm{H}_1: & \mathbf{X} = \mathbf{A}(\theta)\mathbf{B}_\theta\mathbf{S}(\theta) + \mathbf{Z}, \end{cases} \tag{5–20}$$

i.e., we want to test if there exists a target at location θ or not. Similarly to [65] and [86], we define a generalized likelihood ratio (GLR)

$$\rho(\theta) = \left\{ 1 - \left[\frac{\max_{\mathbf{Q}} f(\mathbf{X}|\mathrm{H}_0)}{\max_{\mathbf{B}_\theta,\mathbf{Q}} f(\mathbf{X}|\mathrm{H}_1)} \right]^{\frac{1}{L}} \right\}^L, \tag{5–21}$$

where $f(\mathbf{X}|\mathrm{H}_i)$ $(i = 0, 1)$ is the pdf of $\mathbf{X}$ under the hypothesis H_i. From (5–21), we note that the value of the GLR, $\rho(\theta)$, lies between 0 and 1. If there is a target at a location θ of interest, we have $\max_{\mathbf{B},\mathbf{Q}} f(\mathbf{X}|H_1) \gg \max_{\mathbf{Q}} f(\mathbf{X}|H_0)$, i.e., $\rho \approx 1$; otherwise $\rho \approx 0$.

Under Hypothesis H_0, we have

$$f(\mathbf{X}|\mathrm{H}_0) = \frac{1}{\pi^{LM}|\mathbf{Q}|^L} \exp\{\mathrm{tr}(\mathbf{Q}^{-1}\mathbf{X}\mathbf{X}^H)\}, \tag{5–22}$$

where $|\cdot|$ denotes the determinant of a matrix. Maximizing (5–22) with respect to $\mathbf{Q}$ yields

$$\max_{\mathbf{Q}} f(\mathbf{X}|\mathrm{H}_0) = (\pi e)^{-LM}|\hat{\mathbf{R}}|^{-L}, \tag{5–23}$$

where $\hat{\mathbf{R}}$ is defined in (5–10).

Similarly, under Hypothesis H_1, we have

$$f(\mathbf{X}|\mathrm{H}_1) = \frac{1}{\pi^{LM}|\mathbf{Q}|^L} \exp \mathrm{tr}\{\mathbf{Q}^{-1}[\mathbf{X} - \mathbf{A}(\theta)\mathbf{B}_\theta\mathbf{S}(\theta)][\mathbf{X} - \mathbf{A}(\theta)\mathbf{B}_\theta\mathbf{S}(\theta)]^H\}. \tag{5–24}$$

Maximizing (5–24) with respect to $\mathbf{Q}$ yields

$$\max_{\mathbf{Q}} f(\mathbf{X}|\mathrm{H}_1) = (\pi e)^{-LM} \left|\frac{1}{L}[\mathbf{X} - \mathbf{A}(\theta)\mathbf{B}_\theta\mathbf{S}(\theta)][\mathbf{X} - \mathbf{A}(\theta)\mathbf{B}_\theta\mathbf{S}(\theta)]^H\right|^{-L}. \tag{5–25}$$

Hence, the optimization problem in the denominator of (5–21) reduces to

$$\min_{\mathbf{B}_\theta} \left|\frac{1}{L}[\mathbf{X} - \mathbf{A}(\theta)\mathbf{B}_\theta\mathbf{S}(\theta)][\mathbf{X} - \mathbf{A}(\theta)\mathbf{B}_\theta\mathbf{S}(\theta)]^H\right|. \tag{5–26}$$

Following [21] and [22] and dropping the dependence of $\mathbf{A}$, $\mathbf{S}$ and $\mathbf{B}$ on θ for notional convenience, we have

$$
\begin{aligned}
& \left|\frac{1}{L}[\mathbf{X}-\mathbf{ABS}][\mathbf{X}-\mathbf{ABS}]^H\right| \\
= & \left|\frac{1}{L}\left[\mathbf{AB}_\theta-\mathbf{XS}^H(\mathbf{SS}^H)^{-1}\right](\mathbf{SS}^H)\left[\mathbf{AB}_\theta-\mathbf{XS}^H(\mathbf{SS}^H)^{-1}\right]^H+\hat{\mathbf{Q}}\right| \\
= & |\hat{\mathbf{Q}}|\left|\mathbf{I}+\frac{1}{L}\hat{\mathbf{Q}}^{-\frac{1}{2}}\left[\mathbf{AB}_\theta-\mathbf{XS}^H(\mathbf{SS}^H)^{-1}\right](\mathbf{SS}^H)\left[\mathbf{AB}_\theta-\mathbf{XS}^H(\mathbf{SS}^H)^{-1}\right]^H\hat{\mathbf{Q}}^{-\frac{1}{2}}\right| \\
= & |\hat{\mathbf{Q}}|\left|\mathbf{I}+\frac{1}{L}(\mathbf{SS}^H)^{\frac{1}{2}}\left[\mathbf{AB}_\theta-\mathbf{XS}^H(\mathbf{SS}^H)^{-1}\right]^H\right. \\
& \left.\hat{\mathbf{Q}}^{-1}\left[\mathbf{AB}_\theta-\mathbf{XS}^H(\mathbf{SS}^H)^{-1}\right](\mathbf{SS}^H)^{\frac{1}{2}}\right| \qquad (5\text{–}27) \\
= & |\hat{\mathbf{Q}}|\left|\mathbf{I}+\frac{1}{L}(\mathbf{SS}^H)^{-\frac{1}{2}}\mathbf{SX}^H\left[\hat{\mathbf{Q}}^{-1}-\hat{\mathbf{Q}}^{-1}\mathbf{A}(\mathbf{A}^H\hat{\mathbf{Q}}^{-1}\mathbf{A})^{-1}\mathbf{A}^H\hat{\mathbf{Q}}^{-1}\right]\mathbf{XS}^H(\mathbf{SS}^H)^{-\frac{1}{2}}\right. \\
& +\frac{1}{L}(\mathbf{SS}^H)^{\frac{1}{2}}\left[\mathbf{B}_\theta-(\mathbf{A}^H\hat{\mathbf{Q}}^{-1}\mathbf{A})^{-1}\mathbf{A}^H\hat{\mathbf{Q}}^{-1}\mathbf{XS}^H(\mathbf{SS}^H)^{-1}\right]^H(\mathbf{A}^H\hat{\mathbf{Q}}^{-1}\mathbf{A}) \\
& \left.\left[\mathbf{B}_\theta-(\mathbf{A}^H\hat{\mathbf{Q}}^{-1}\mathbf{A})^{-1}\mathbf{A}^H\hat{\mathbf{Q}}^{-1}\mathbf{XS}^H(\mathbf{SS}^H)^{-1}\right](\mathbf{SS}^H)^{\frac{1}{2}}\right| \\
\geq & |\hat{\mathbf{Q}}|\left|\mathbf{I}+\frac{1}{L}(\mathbf{SS}^H)^{-\frac{1}{2}}\mathbf{SX}^H[\hat{\mathbf{Q}}^{-1}-\hat{\mathbf{Q}}^{-1}\mathbf{A}(\mathbf{A}^H\hat{\mathbf{Q}}^{-1}\mathbf{A})^{-1}\mathbf{A}^H\hat{\mathbf{Q}}^{-1}]\right. \\
& \left.\mathbf{XS}^H(\mathbf{SS}^H)^{-\frac{1}{2}}\right| \qquad (5\text{–}28)
\end{aligned}
$$

where $\hat{\mathbf{Q}}$ is defined in (5–17). To get (5–27), we have used the fact that $|\mathbf{I}+\mathbf{XY}| = |\mathbf{I}+\mathbf{YX}|$ [74], and the equality in (5–28) holds when $\mathbf{B}$ equates to the APES estimate in (5–19). Note that

$$
\begin{aligned}
& |\hat{\mathbf{Q}}|\left|\mathbf{I}+\frac{1}{L}(\mathbf{SS}^H)^{-\frac{1}{2}}\mathbf{SX}^H\ [\hat{\mathbf{Q}}^{-1}-\hat{\mathbf{Q}}^{-1}\mathbf{A}(\mathbf{A}^H\hat{\mathbf{Q}}^{-1}\mathbf{A})^{-1}\mathbf{A}^H\hat{\mathbf{Q}}^{-1}]\mathbf{XS}^H(\mathbf{SS}^H)^{-\frac{1}{2}}\right| \\
= & |\hat{\mathbf{Q}}|\,|\mathbf{I}+[\hat{\mathbf{Q}}^{-1}-\hat{\mathbf{Q}}^{-1}\mathbf{A}(\mathbf{A}^H\hat{\mathbf{Q}}^{-1}\mathbf{A})^{-1}\mathbf{A}^H\hat{\mathbf{Q}}^{-1}](\hat{\mathbf{R}}-\hat{\mathbf{Q}})| \\
= & |\hat{\mathbf{R}}-\mathbf{A}(\mathbf{A}^H\hat{\mathbf{Q}}^{-1}\mathbf{A})^{-1}\mathbf{A}^H\hat{\mathbf{Q}}^{-1}(\hat{\mathbf{R}}-\hat{\mathbf{Q}})| \\
= & |\hat{\mathbf{R}}||\mathbf{I}-\mathbf{A}^H(\mathbf{A}^H\hat{\mathbf{Q}}^{-1}\mathbf{A})^{-1}\mathbf{A}^H(\hat{\mathbf{Q}}^{-1}-\hat{\mathbf{R}}^{-1})| \\
= & |\hat{\mathbf{R}}||(\mathbf{A}^H\hat{\mathbf{Q}}^{-1}\mathbf{A})^{-1}(\mathbf{A}^H\hat{\mathbf{R}}^{-1}\mathbf{A})|, \qquad (5\text{–}29)
\end{aligned}
$$

From (5–25), (5–28) and (5–29), it follows that

$$
\max_{\mathbf{B}_\theta,\mathbf{Q}} f(\mathbf{X}|\mathrm{H}_1) = (\pi e)^{-LM}|\hat{\mathbf{R}}|^{-L}|\mathbf{A}^H(\theta)\hat{\mathbf{Q}}^{-1}\mathbf{A}(\theta)|^{L}|\mathbf{A}^H(\theta)\hat{\mathbf{R}}^{-1}\mathbf{A}(\theta)|^{-L}. \qquad (5\text{–}30)
$$

Substituting (5–23) and (5–30) into (5–21) yields

$$\rho(\theta) = \left\{ 1 - \frac{|\mathbf{A}^H(\theta)\hat{\mathbf{R}}^{-1}\mathbf{A}(\theta)|}{|\mathbf{A}^H(\theta)\hat{\mathbf{Q}}^{-1}\mathbf{A}(\theta)|} \right\}^L . \tag{5–31}$$

We remark that when there are multiple targets, and the number of targets (say K) are known *a priori*, the GLRT in (5–31) can be extended to a multivariate counterpart by considering the following hypothesis testing problem

$$\begin{cases} \mathrm{H}_0 : & \mathbf{X} = \mathbf{Z} \\ \mathrm{H}_K : & \mathbf{X} = \sum_{k=1}^{K} \mathbf{A}(\theta_k)\mathbf{B}_{\theta_k}\mathbf{S}(\theta_k) + \mathbf{Z}. \end{cases} \tag{5–32}$$

As a parametric method, this multivariate GLRT can provide better target detection and parameter estimation performance than its univariate counterpart. However, the multivariate GLRT is computationally intensive due to the fact that it needs to search in the $K-$dimensional parameter space $\{\theta_k\}_{k=1}^{K}$. Moreover, the number of targets is hardly known *a priori* in practice.

We propose below an iterative GLRT (iGLRT), which require only one-dimensional search (like the univariate GLRT), but provides a target detection and parameter estimation performance close to the multivariate GLRT.

5.3.2 Conditional Generalized Likelihood Ratio Test (cGLRT)

Before we describe the iGLRT procedure, we first consider the following hypothesis testing problem, referred to as the conditional generalized likelihood ratio test (cGLRT). Suppose that we have known (or detected) p targets at the estimated locations $\{\hat{\theta}_k\}_{k=1}^{p}$, and we want to determine if there are any additional targets. This problem can be formulated in the following hypothesis testing problem

$$\begin{cases} \mathrm{H}_p : & \mathbf{X} = \sum_{k=1}^{p} \mathbf{A}(\hat{\theta}_k)\mathbf{B}_{\hat{\theta}_k}\mathbf{S}(\hat{\theta}_k) + \mathbf{Z} \\ \mathrm{H}_{p+1} : & \mathbf{X} = \mathbf{A}(\theta)\mathbf{B}_{\theta}\mathbf{S}(\theta) + \sum_{k=1}^{p} \mathbf{A}(\hat{\theta}_k)\mathbf{B}_{\hat{\theta}_k}\mathbf{S}(\hat{\theta}_k) + \mathbf{Z}. \end{cases} \tag{5–33}$$

Note that both the equations in (5–33) are in the form of the block diagonal growth curve (BDGC) model studied in Chapter 4. For convenience, we rewrite (5–33) as

$$\begin{cases} \mathrm{H}_p: & \mathbf{X} = \tilde{\mathbf{A}}_p \tilde{\mathbf{B}}_p \tilde{\mathbf{S}}_p + \mathbf{Z} \\ \mathrm{H}_{p+1}: & \mathbf{X} = \tilde{\mathbf{A}}_{p+1} \tilde{\mathbf{B}}_{p+1} \tilde{\mathbf{S}}_{p+1} + \mathbf{Z}, \end{cases} \tag{5–34}$$

where

$$\tilde{\mathbf{B}}_p = \mathrm{Diag}(\mathbf{B}_{\hat{\theta}_1}, \cdots, \mathbf{B}_{\hat{\theta}_p}) \tag{5–35}$$

$$\tilde{\mathbf{A}}_p = [\mathbf{A}(\hat{\theta}_1) \cdots \mathbf{A}(\hat{\theta}_p)], \tag{5–36}$$

$$\tilde{\mathbf{S}}_p = [\mathbf{S}^T(\hat{\theta}_1) \cdots \mathbf{S}^T(\hat{\theta}_p)]^T, \tag{5–37}$$

$$\tilde{\mathbf{B}}_{p+1} = \mathrm{Diag}(\mathbf{B}_\theta, \mathbf{B}_{\hat{\theta}_1}, \cdots, \mathbf{B}_{\hat{\theta}_p}) \tag{5–38}$$

$$\tilde{\mathbf{A}}_{p+1} = [\mathbf{A}(\theta)\ \mathbf{A}(\hat{\theta}_1) \cdots \mathbf{A}(\hat{\theta}_p)], \tag{5–39}$$

$$\tilde{\mathbf{S}}_{p+1} = [\mathbf{S}^T(\theta)\ \mathbf{S}^T(\hat{\theta}_1) \cdots \mathbf{S}^T(\hat{\theta}_p)]^T, \tag{5–40}$$

and $\mathrm{Diag}(\Omega_1, \cdots, \Omega_K)$ denotes a block diagonal matrix formed from $\Omega_1, \cdots, \Omega_K$.

Similarly, we define a conditional generalized likelihood ratio (cGLR)

$$\rho(\theta|\{\hat{\theta}_k\}_{k=1}^p) = \left\{ 1 - \left[\frac{\max_{\tilde{\mathbf{B}}_p, \mathbf{Q}} f(\mathbf{X}|\mathrm{H}_p)}{\max_{\tilde{\mathbf{B}}_{p+1}, \mathbf{Q}} f(\mathbf{X}|\mathrm{H}_{p+1})} \right]^{\frac{1}{L}} \right\}^L, \tag{5–41}$$

where $f(\mathbf{X}|\mathrm{H}_i)$ is the pdf of $\mathbf{X}$ under the H_i hypothesis, and $\mathbf{Q}$ is the covariance matrix of the columns of $\mathbf{Z}$.

We first consider the optimization problem of the numerator in (5–41). Maximizing $f(\mathbf{X}|\mathrm{H}_p)$ with respect to $\mathbf{Q}$ yields

$$\max_{\mathbf{Q}} f(\mathbf{X}|\mathrm{H}_p) = (\pi e)^{-ML} \left| \frac{1}{L} (\mathbf{X} - \tilde{\mathbf{A}}_p \tilde{\mathbf{B}}_p \tilde{\mathbf{S}}_p)(\mathbf{X} - \tilde{\mathbf{A}}_p \tilde{\mathbf{B}}_p \tilde{\mathbf{S}}_p)^H \right|^{-L}. \tag{5–42}$$

Hence, the optimization problem reduces to

$$\min_{\tilde{\mathbf{B}}_p} \left| \frac{1}{L} (\mathbf{X} - \tilde{\mathbf{A}}_p \tilde{\mathbf{B}}_p \tilde{\mathbf{S}}_p)(\mathbf{X} - \tilde{\mathbf{A}}_p \tilde{\mathbf{B}}_p \tilde{\mathbf{S}}_p)^H \right| \quad \text{with} \quad \tilde{\mathbf{B}}_p = \mathrm{Diag}(\mathbf{B}_{\hat{\theta}_1}, \cdots, \mathbf{B}_{\hat{\theta}_p}). \tag{5–43}$$

The optimization problem of (5–43) does not appear to admit a closed-form solution, due to the constraint that $\tilde{\mathbf{B}}_p$ is a block-diagonal matrix. Herein, we adopt a technique used in Chapters 3 and 4 to get approximate closed-form solution.

Note that

$$\begin{aligned}
&\left|\frac{1}{L}(\mathbf{X}-\tilde{\mathbf{A}}_p\tilde{\mathbf{B}}_p\tilde{\mathbf{S}}_p)(\mathbf{X}-\tilde{\mathbf{A}}_p\tilde{\mathbf{B}}_p\tilde{\mathbf{S}}_p)^H\right| \\
=&\left|\frac{1}{L}(\mathbf{X}-\tilde{\mathbf{A}}_p\tilde{\mathbf{B}}_p\tilde{\mathbf{S}}_p)(\mathbf{\Pi}_{\tilde{\mathbf{S}}_p}+\mathbf{\Pi}^{\perp}_{\tilde{\mathbf{S}}_p})(\mathbf{X}-\tilde{\mathbf{A}}_p\tilde{\mathbf{B}}_p\tilde{\mathbf{S}}_p)^H\right| \\
=&\left|\frac{1}{L}(\mathbf{X}\mathbf{\Pi}_{\tilde{\mathbf{S}}_p}-\tilde{\mathbf{A}}_p\tilde{\mathbf{B}}_p\tilde{\mathbf{S}}_p)(\mathbf{X}\mathbf{\Pi}_{\tilde{\mathbf{S}}_p}-\tilde{\mathbf{A}}_p\tilde{\mathbf{B}}_p\tilde{\mathbf{S}}_p)^H+\tilde{\mathbf{Q}}_p\right| \\
=&\left|\frac{1}{L}(\mathbf{X}\mathbf{\Pi}_{\tilde{\mathbf{S}}_p}-\tilde{\mathbf{A}}_p\tilde{\mathbf{B}}_p\tilde{\mathbf{S}}_p)(\mathbf{X}\mathbf{\Pi}_{\tilde{\mathbf{S}}_p}-\tilde{\mathbf{A}}_p\tilde{\mathbf{B}}_p\tilde{\mathbf{S}}_p)^H\tilde{\mathbf{Q}}_p^{-1}+\mathbf{I}\right|\,|\tilde{\mathbf{Q}}_p|,
\end{aligned} \tag{5–44}$$

where

$$\mathbf{\Pi}_{\tilde{\mathbf{S}}_p}=\tilde{\mathbf{S}}_p^H(\tilde{\mathbf{S}}_p\tilde{\mathbf{S}}_p^H)^-\tilde{\mathbf{S}}_p,\quad \mathbf{\Pi}^{\perp}_{\tilde{\mathbf{S}}_p}=\mathbf{I}-\mathbf{\Pi}_{\tilde{\mathbf{S}}_p}, \tag{5–45}$$

and

$$\tilde{\mathbf{Q}}_p=\frac{1}{L}\mathbf{X}\mathbf{\Pi}^{\perp}_{\tilde{\mathbf{S}}_p}\mathbf{X}^H \tag{5–46}$$

with $(\cdot)^-$ denoting the generalized matrix inverse.

Consider the idempotent matrices $\mathbf{\Pi}_{\tilde{\mathbf{S}}_p}$ and $\mathbf{\Pi}^{\perp}_{\tilde{\mathbf{S}}_p}$. Assume that the number of data samples is large enough, i.e., $L \gg \tilde{N}p$. Note that $\tilde{\mathbf{S}}_p$ is an $\tilde{N}p \times L$ matrix. Hence, we have

$$\mathrm{rank}(\mathbf{\Pi}_{\tilde{\mathbf{S}}_p}) \le \tilde{N}p \quad \text{and} \quad \mathrm{rank}(\mathbf{\Pi}^{\perp}_{\tilde{\mathbf{S}}_p}) \ge L-\tilde{N}p, \tag{5–47}$$

with $\mathrm{rank}(\cdot)$ denoting the rank of a matrix. Then, we have

$$\tilde{\mathbf{Q}}_p = O(1), \tag{5–48}$$

$$\begin{aligned} & \frac{1}{L}(\mathbf{X}\mathbf{\Pi}_{\tilde{\mathbf{S}}_p} - \tilde{\mathbf{A}}_p\tilde{\mathbf{B}}_p\tilde{\mathbf{S}}_p)(\mathbf{X}\mathbf{\Pi}_{\tilde{\mathbf{S}}_p} - \tilde{\mathbf{A}}_p\tilde{\mathbf{B}}_p\tilde{\mathbf{S}}_p)^T \\ = \; & \frac{1}{L}(\mathbf{X} - \tilde{\mathbf{A}}_p\tilde{\mathbf{B}}_p\tilde{\mathbf{S}}_p)\mathbf{\Pi}_{\tilde{\mathbf{S}}_p}(\mathbf{X} - \tilde{\mathbf{A}}_p\tilde{\mathbf{B}}_p\tilde{\mathbf{S}}_p)^T = O\left(\frac{1}{L}\right). \end{aligned} \tag{5–49}$$

Therefore, we get

$$\frac{1}{L}(\mathbf{X}\mathbf{\Pi}_{\tilde{\mathbf{S}}_p} - \tilde{\mathbf{A}}_p\tilde{\mathbf{B}}_p\tilde{\mathbf{S}}_p)(\mathbf{X}\mathbf{\Pi}_{\tilde{\mathbf{S}}_p} - \tilde{\mathbf{A}}_p\tilde{\mathbf{B}}_p\tilde{\mathbf{S}}_p)^T\tilde{\mathbf{Q}}_p^{-1} = O\left(\frac{1}{L}\right) \ll 1. \tag{5–50}$$

Let $\{\lambda_i\}_{i=1}^M$ be the eigenvalues of the matrix in (5–50), which obviously satisfy that $0 \leq \lambda_i \ll 1$. By using some matrix manipulations, we obtain

$$\begin{aligned} & \left|\frac{1}{L}(\mathbf{X}\mathbf{\Pi}_{\tilde{\mathbf{S}}_p} - \tilde{\mathbf{A}}_p\tilde{\mathbf{B}}_p\tilde{\mathbf{S}}_p)(\mathbf{X}\mathbf{\Pi}_{\tilde{\mathbf{S}}_p} - \tilde{\mathbf{A}}_p\tilde{\mathbf{B}}_p\tilde{\mathbf{S}}_p)^T\tilde{\mathbf{Q}}_p^{-1} + \mathbf{I}\right| \\ = \; & \prod_{i=1}^{M}(1+\lambda_i) \approx 1 + \sum_{i=1}^{M}\lambda_i \\ = \; & 1 + \frac{1}{L}\operatorname{tr}\left[(\mathbf{X}\mathbf{\Pi}_{\tilde{\mathbf{S}}_p} - \frac{1}{L}\tilde{\mathbf{A}}_p\tilde{\mathbf{B}}_p\tilde{\mathbf{S}}_p)(\mathbf{X}\mathbf{\Pi}_{\tilde{\mathbf{S}}_p} - \tilde{\mathbf{A}}_p\tilde{\mathbf{B}}_p\tilde{\mathbf{S}}_p)^T\tilde{\mathbf{Q}}_p^{-1}\right] \\ = \; & 1 + \frac{1}{L}\parallel \operatorname{vec}(\tilde{\mathbf{Q}}_p^{-\frac{1}{2}}\mathbf{X}\mathbf{\Pi}_{\tilde{\mathbf{S}}_p}) - \operatorname{vec}(\tilde{\mathbf{Q}}_p^{-\frac{1}{2}}\tilde{\mathbf{A}}_p\tilde{\mathbf{B}}_p\tilde{\mathbf{S}}_p)\parallel^2, \end{aligned} \tag{5–51}$$

where $\parallel \cdot \parallel$ and $\operatorname{vec}(\cdot)$ denote the Euclidean norm and vectorization operator (stacking the columns of a matrix on top of each other), respectively, and $\tilde{\mathbf{Q}}_p^{-\frac{1}{2}}$ is the Hermitian square root of $\tilde{\mathbf{Q}}_p^{-1}$. In (5–51), we have omitted the high-order terms of $\{\lambda_i\}$ for the approximation.

Hence, for a large number of data samples, the optimization problem in (5–43) can be approximated as

$$\begin{aligned} \min_{\tilde{\mathbf{B}}_p} \quad & \parallel \operatorname{vec}(\tilde{\mathbf{Q}}_p^{-\frac{1}{2}}\mathbf{X}\mathbf{\Pi}_{\tilde{\mathbf{S}}_p}) - \operatorname{vec}(\tilde{\mathbf{Q}}_p^{-\frac{1}{2}}\tilde{\mathbf{A}}_p\tilde{\mathbf{B}}_p\tilde{\mathbf{S}}_p)\parallel^2 \\ & \text{with} \quad \tilde{\mathbf{B}}_p = \operatorname{Diag}(\mathbf{B}_{\hat{\theta}_1}, \cdots, \mathbf{B}_{\hat{\theta}_p}). \end{aligned} \tag{5–52}$$

To solve the above optimization problem, we need the block-matrix operations and lemmas 4.1 and 4.2 in Chapter 4. Throughout this chapter, the partitioned matrix operation are all based on the partitionings in (5–35) - (5–40).

Now, let

$$\tilde{\boldsymbol{\beta}}_p = \text{vecb}(\tilde{\mathbf{B}}_p), \tag{5–53}$$

where vecb$(\cdot)$ denotes the block-diagonal vectorization operator (Definition 4.1), i.e.,

$$\tilde{\boldsymbol{\beta}}_p = \left[\text{vec}^T(\mathbf{B}_{\hat{\theta}_1}) \ \cdots \text{vec}^T(\mathbf{B}_{\hat{\theta}_p})\right]^T. \tag{5–54}$$

By using Lemma 4.1 (with $K = 1$ and $J = p$) of Chapter 4, we obtain

$$\text{vec}(\tilde{\mathbf{Q}}_p^{-\frac{1}{2}}\tilde{\mathbf{A}}_p\tilde{\mathbf{B}}_p\tilde{\mathbf{S}}_p) = [\tilde{\mathbf{S}}_p^T \circledast (\tilde{\mathbf{Q}}_p^{-\frac{1}{2}}\tilde{\mathbf{A}})]\tilde{\boldsymbol{\beta}}_p \triangleq \tilde{\boldsymbol{\Gamma}}_p\tilde{\boldsymbol{\beta}}_p, \tag{5–55}$$

where $\circledast$ denotes the generalized Khatri-Rao product (Definition 4.2). Hence,

$$\begin{aligned}
& \| \text{vec}(\tilde{\mathbf{Q}}_p^{-\frac{1}{2}}\mathbf{X}\boldsymbol{\Pi}_{\tilde{\mathbf{S}}_p}) - \text{vec}(\tilde{\mathbf{Q}}_p^{-\frac{1}{2}}\tilde{\mathbf{A}}_p\tilde{\mathbf{B}}_p\tilde{\mathbf{S}}_p) \|^2 \\
= \ & \| \text{vec}(\tilde{\mathbf{Q}}_p^{-\frac{1}{2}}\mathbf{X}\boldsymbol{\Pi}_{\tilde{\mathbf{S}}_p}) - \boldsymbol{\Gamma}_p\tilde{\boldsymbol{\beta}}_p \|^2 \\
\geq \ & \| \text{vec}(\tilde{\mathbf{Q}}_p^{-\frac{1}{2}}\mathbf{X}\boldsymbol{\Pi}_{\tilde{\mathbf{S}}_p}) \|^2 - \text{vec}^H(\tilde{\mathbf{Q}}_p^{-\frac{1}{2}}\mathbf{X}\boldsymbol{\Pi}_{\tilde{\mathbf{S}}_p})\tilde{\boldsymbol{\Gamma}}_p(\tilde{\boldsymbol{\Gamma}}_p^H\tilde{\boldsymbol{\Gamma}}_p)^{-1}\tilde{\boldsymbol{\Gamma}}_p^H\text{vec}(\tilde{\mathbf{Q}}_p^{-\frac{1}{2}}\mathbf{X}\boldsymbol{\Pi}_{\tilde{\mathbf{S}}_p}) \\
= \ & L\,\text{tr}(\tilde{\mathbf{Q}}_p^{-1}\hat{\mathbf{R}}) - LM - \text{vecb}^H(\tilde{\mathbf{A}}_p^H\tilde{\mathbf{Q}}_p^{-1}\mathbf{X}\tilde{\mathbf{S}}_p^H) \\
& \left[(\tilde{\mathbf{S}}_p\tilde{\mathbf{S}}_p^H)^T \circledast (\tilde{\mathbf{A}}_p^H\tilde{\mathbf{Q}}_p^{-1}\tilde{\mathbf{A}}_p)\right]^{-1}\text{vecb}(\tilde{\mathbf{A}}_p^H\tilde{\mathbf{Q}}_p^{-1}\mathbf{X}\tilde{\mathbf{S}}_p^H),
\end{aligned} \tag{5–56}$$

where we have used *Lemmas* 4.1 and 4.2, and the equality holds when

$$\tilde{\boldsymbol{\beta}} = (\tilde{\boldsymbol{\Gamma}}_p^H\tilde{\boldsymbol{\Gamma}}_p)^{-1}\tilde{\boldsymbol{\Gamma}}_p^H\text{vec}(\tilde{\mathbf{Q}}_p^{-\frac{1}{2}}\mathbf{X}\boldsymbol{\Pi}_{\tilde{\mathbf{S}}_p}). \tag{5–57}$$

By using (5–42), (5–44), (5–51) and (5–56), it follows that

$$\left[\max_{\mathbf{Q},\tilde{\mathbf{B}}_p} f(\mathbf{X}|\mathrm{H}_p)\right]^{\frac{1}{L}} \approx \frac{1}{(\pi e)^M\, g(\hat{\theta}_1,\ \cdots,\ \hat{\theta}_p)\,|\tilde{\mathbf{Q}}_p|}, \tag{5–58}$$

where

$$\begin{aligned}
g(\hat{\theta}_1,\ \cdots,\ \hat{\theta}_p) \ = \ & 1 - M + \text{tr}(\tilde{\mathbf{Q}}_p^{-1}\hat{\mathbf{R}}) - \text{vecb}^H(\tilde{\mathbf{A}}_p^H\hat{\mathbf{Q}}_P^{-1}\mathbf{X}\tilde{\mathbf{S}}_p^H) \\
& \left[(\tilde{\mathbf{S}}_p\tilde{\mathbf{S}}_p^H)^T \circledast (\tilde{\mathbf{A}}_p^H\tilde{\mathbf{Q}}_p^{-1}\tilde{\mathbf{A}}_p)\right]^{-1}\text{vecb}(\tilde{\mathbf{A}}_p^H\hat{\mathbf{Q}}_P^{-1}\mathbf{X}\tilde{\mathbf{S}}_p^H).
\end{aligned} \tag{5–59}$$

Similarly, we have

$$\left[\max_{\mathbf{Q},\tilde{\mathbf{B}}_{p+1}} f(\mathbf{X}|\mathrm{H}_{p+1})\right]^{\frac{1}{L}} \approx \frac{1}{(\pi e)^M\, g(\theta,\hat{\theta}_1,\cdots,\hat{\theta}_p)\,|\tilde{\mathbf{Q}}_{p+1}|}, \tag{5–60}$$

where $\tilde{\mathbf{Q}}_{p+1}$ and $g(\theta,\hat{\theta}_1,\cdots,\hat{\theta}_p)$ are defined similarly to $\tilde{\mathbf{Q}}_p$ in (5–46) and $g(\hat{\theta}_1,\cdots,\hat{\theta}_p)$ in (5–59), respectively.

Substituting (5–58) and (5–60) into (5–41) yields the conditional GLR

$$\rho(\theta|\{\hat{\theta}_j\}) = \left\{1 - \frac{g(\theta,\hat{\theta}_1,\cdots,\hat{\theta}_p)\,|\tilde{\mathbf{Q}}_{p+1}|}{g(\hat{\theta}_1,\cdots,\hat{\theta}_p)\,|\tilde{\mathbf{Q}}_p|}\right\}^L. \tag{5–61}$$

5.3.3 Iterative Generalized Likelihood Ratio Test (iGLRT)

The basic idea of the iterative generalized likelihood ratio test (iGLRT) is to detect and localize targets sequentially. In each step of the iteration, the results from the previous iterations and steps are exploited for the detection and localization of new targets by calculating cGLR. Specifically, we first perform GLRT to get the location of the dominant target, and the following targets are detected and localized by using cGLRT conditioned on the most recently available estimates. The detailed steps of iGLRT are described as follows.

Step I:

- *Calculate $\rho(\theta)$ in (5–31) for each θ.*
- *Compare $\rho(\theta)$ to a threshold (say ρ_0). If $\rho(\theta) < \rho_0$ for all θ, then Stop; otherwise, $\hat{\theta}_1 = \arg\max_\theta \rho(\theta)$, go to Step II.*

Step II: *For $k = 1,\ 2,\ \cdots$, do the following substeps:*

- *Calculate $\rho\big(\theta|\{\hat{\theta}_i\}_{i=1}^k\big)$ in (5–61) for each θ.*
- *If $\rho\big(\theta|\{\hat{\theta}_i\}_{i=1}^k\big) < \rho_0$ for all θ, then go to Step III; otherwise, $\hat{\theta}_{k+1} = \arg\max_\theta \rho\big(\theta|\{\hat{\theta}_i\}_{i=1}^k\big)$.*

Step III: *Repeat the following substeps until convergence*

for $k = 1,\ 2,\ \cdots,\ \hat{K}$ *(suppose that* $\hat{K}$ *targets are detected in Steps I and II)*

- *Calculate* $\rho\big(\theta|\{\hat{\theta}_i\}_{i\neq k}\big)$ *for each* θ.
- *Update* $\hat{\theta}_k$ *by* $\arg\max_\theta \rho\big(\theta|\{\hat{\theta}_i\}_{i\neq k}\big)$.

Once the locations of the targets are determined, the amplitudes of the reflected signals can be estimated by using the AML estimator in [37], i.e.,

$$\hat{\boldsymbol{\beta}}_{\hat{K}} = \big[(\tilde{\mathbf{A}}_{\hat{K}}^H \tilde{\mathbf{Q}}_{\hat{K}}^{-1} \tilde{\mathbf{A}}_{\hat{K}}) \circledast (\tilde{\mathbf{S}}_{\hat{K}}^H \tilde{\mathbf{S}}_{\hat{K}})^T\big]^{-1} \mathrm{vecb}(\tilde{\mathbf{A}}_{\hat{K}}^H \tilde{\mathbf{Q}}^{-1} \mathbf{X} \tilde{\mathbf{S}}_{\hat{K}}^H), \tag{5–62}$$

where $\tilde{\mathbf{A}}_{\hat{K}}$, $\tilde{\mathbf{S}}_{\hat{K}}$ and $\tilde{\mathbf{Q}}_{\hat{K}}$ are defined similarly to $\tilde{\mathbf{A}}_p$, $\tilde{\mathbf{A}}_p$ and $\tilde{\mathbf{Q}}_p$ in (5–37), (5–38) and (5–46), respectively.

We note that Step III of the above iGLRT algorithm actually minimize the function $g(\theta_1, \cdots, \theta_{\hat{K}})$ with respect to $\{\theta_k\}_{k=1}^{\hat{K}}$ by using the cyclic minimization (CM) technique [84]. Under a mild condition, i.e., $L \gg \tilde{N}\hat{K}$, we have $g(\theta_1, \cdots, \theta_{\hat{K}}) \geq 0$. Furthermore, we know that the CM algorithm monotonically decreases the cost function. Hence the iGLRT algorithm is convergent. When $\hat{K}$ is the true number of targets, iGLRT reduces to an approximate (parametric) maximum likelihood estimator. As we will show via numerical examples, the mean-squared-error (MSE) of the estimate of iGLRT approaches the corresponding Cramér-Rao bound (CRB) for a large number of data samples. On the other hand, we note that iGLRT needs only one-dimensional search and hence is computationally efficient.

5.4 Numerical Examples

5.4.1 Cramér-Rao Bound

We first study the Cramér-Rao bound under various antenna configurations. Consider a MIMO radar system with $M = N = 8$ antennas for transmitting and receiving. We assume that the receiving and transmitting antennas are

grouped into multiple subarrays (each being a uniform linear array with half-wavelength spacing between adjacent elements). We consider the following antenna configuration schemes.

MIMO Radar A: 1 subarray with 8 antennas for transmitting and receiving;

MIMO Radar B: 2 subarrays each with 4 antennas for transmitting, and 1 subarray with 8 antennas for receiving;

MIMO Radar C: 8 subarrays each with 1 antenna for transmitting, and 1 subarray with 8 antennas for receiving;

MIMO Radar D: 2 subarrays each with 4 antennas for transmitting and receiving.

We assume that the transmitted waveforms are linearly orthogonal to each other and the total transmitted power is fixed to be 1, i.e., $\mathbf{R}_{\Phi\Phi} = \frac{1}{N}\mathbf{I}$.

We consider a scenario in which $K = 3$ targets are located at $\theta_1 = -40°$, $\theta_2 = -4°$ and $\theta_3 = 0°$, and the elements of $\{\mathbf{B}_{\theta_k}\}_{k=1}^{3}$ are independently and identically distributed (i.i.d.) circularly symmetric complex Gaussian random variables with zero mean and unit variance. There is a strong jammer at 10° with amplitude 100, i.e., 40 dB above the reflected signals. The received signal has $L = 128$ snapshots and is corrupted by a zero-mean spatially colored Gaussian noise with an unknown covariance matrix. The (p, q)th element of the unknown noise covariance matrix is $\frac{1}{\text{SNR}} 0.90^{|p-q|} e^{j\frac{(p-q)\pi}{2}}$. Figs. 5–1(a) and 5–1 (b) show the cumulative density functions (CDFs) of the CRBs for MIMO radar with various antenna configurations when SNR=20 dB. (The CRB of θ_2 is similar to that of θ_3 and hence is not shown.) The CDFs are obtained by 2000 Monte-Carlo trials. In each trial, we generate the elements of $\{\mathbf{B}_{\theta_k}\}_{k=1}^{3}$ randomly, and then calculate the corresponding CRBs using (C–18) given by in the Appendix C. For comparison purposes, we also provide the CDF of the phased-array (single-input multiple-output) counterpart, i.e., the special case of the above MIMO radar when $N = 1$,

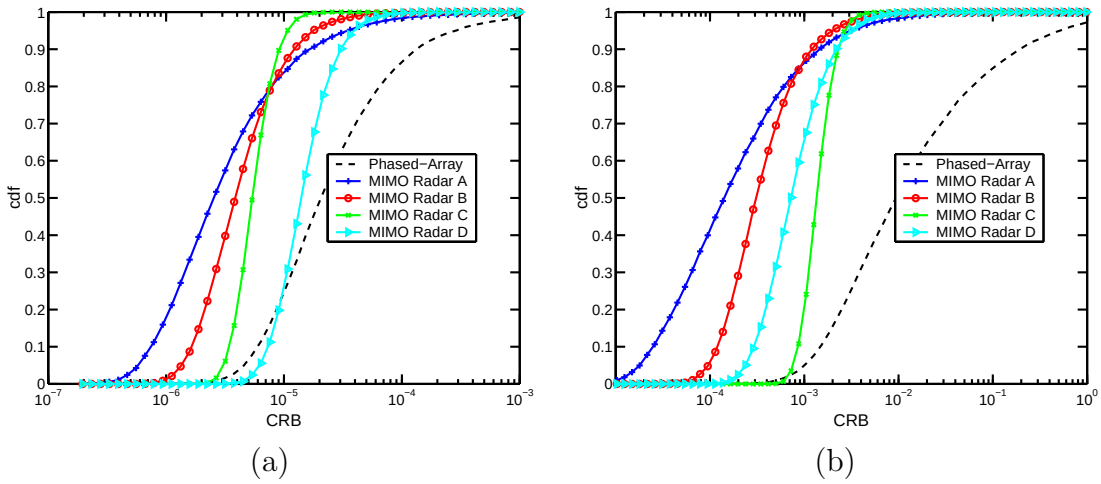


Figure 5–1: Cumulative density functions of the Cramér-Rao bounds for (a) θ_1 and (b) θ_3.

with the same total transmission power. As expected, the MIMO radar provides much better performance than the phased-array counterpart. Due to the fading effect of the elements of $\{\mathbf{B}_{\theta_k}\}_{k=1}^{3}$, the CRB of MIMO Radar A varies within a large range. Within a 95% confidence interval (i.e., when CDF varies from 2.5% to 97.5%), its CRB for θ_1 varies approximately from 5×10^{-7} to 5×10^{-5}. The CRBs for MIMO radar C varies within a small range.

To evaluate the CRB performance, we define an outage CRB [43] for a given probability p, denoted by CRB_p, as

$$P(\text{CRB} \geq \text{CRB}_p) = p. \tag{5–63}$$

Figs. 5–2(a) - 5–2(d) show the outage $\text{CRB}_{0.01}$ and $\text{CRB}_{0.1}$ of θ_1 and θ_3, as functions of SNR. As expected, the SNR gains depend on the probability p. As we can see, when $p = 0.01$, MIMO radar C outperforms the other radar configurations, and provides around 20 dB and 12 dB improvements in terms of SNR compared to the phased-array and MIMO radar A, respectively. On the other hand, Fig. 5–2(d) shows that MIMO radars A and B outperform others when $p = 0.1$.

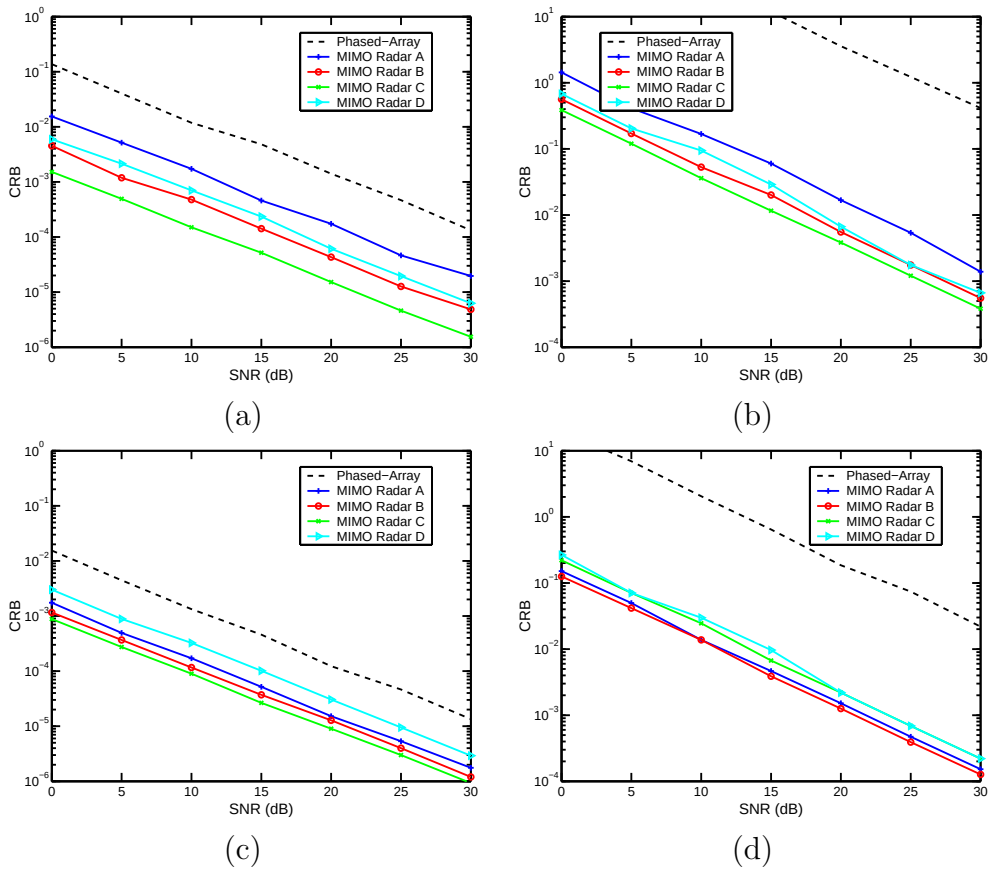


Figure 5–2: Outage CRB versus SNR. (a) $\mathrm{CRB}_{0.01}$ for θ_1, (b) $\mathrm{CRB}_{0.01}$ for θ_2, (c) $\mathrm{CRB}_{0.1}$ for θ_1, and (d) $\mathrm{CRB}_{0.1}$ for θ_2.

5.4.2 Target Detection and Localization

We focus below on MIMO radar B, i.e., a MIMO radar system with 2 subarrays (each with 4 antennas) for transmitting and 1 subarray (with 8 antennas) for receiving.

We first consider a scenario in which 3 targets are located at $\theta_1 = -40°$, $\theta_2 = -20°$ and $\theta_3 = 0°$ with the corresponding elements in $\mathbf{B}_{\theta_1}$, $\mathbf{B}_{\theta_2}$ and $\mathbf{B}_{\theta_3}$ being fixed to 2, 2 and 1, respectively. The other simulation parameters are the same as for Fig. 5–1. The Frobenius norm of the spatial spectral estimates of $\mathbf{B}_\theta$ versus θ, obtained by using LS, Capon and APES are given by Figs. 5–3(a) - 5–3(c). For comparison purposes, we show the true spatial spectrum via dashed lines in these figures. As seen from Fig. 5–3, the LS method suffers from high-sidelobes and poor resolution problems. Due to the presence of the strong jamming signal, the LS estimator fails to work properly. Capon and APES possess excellent interference and jamming suppression capabilities. The Capon method gives very narrow peaks around the target locations. However, the Capon estimates of $\mathbf{B}_{\theta_1}$, $\mathbf{B}_{\theta_2}$ and $\mathbf{B}_{\theta_3}$ are biased downward. The APES method gives more accurate estimates around the target locations but its resolution is worse than that of Capon. Note that in Figs. 5–3(a), 5–3(b) and 5–3(c), a false peak occurs at $\theta = 10°$ due to the presence of the strong jammer. Despite the fact that the jammer waveform is statistically independent of the waveforms transmitted by the MIMO radar, a false peak still exists since the jammer is 40 dB stronger than the weakest target and the number of data samples is finite. Figs. 5–3(d) and 5–3(e) give the GLRT, and the iGLRT results, as functions of the target location parameter θ. For convenience, in Fig. 5–3(e), we have included all cGLR functions obtained by iGLRT, each indicating one target. As expected, we get high GLRs (cGLRs) at the target locations and low GLRs (cGLRs) at other locations including the jammer location. By comparing the GLR with a threshold, the false peak due to the strong jammer can be readily

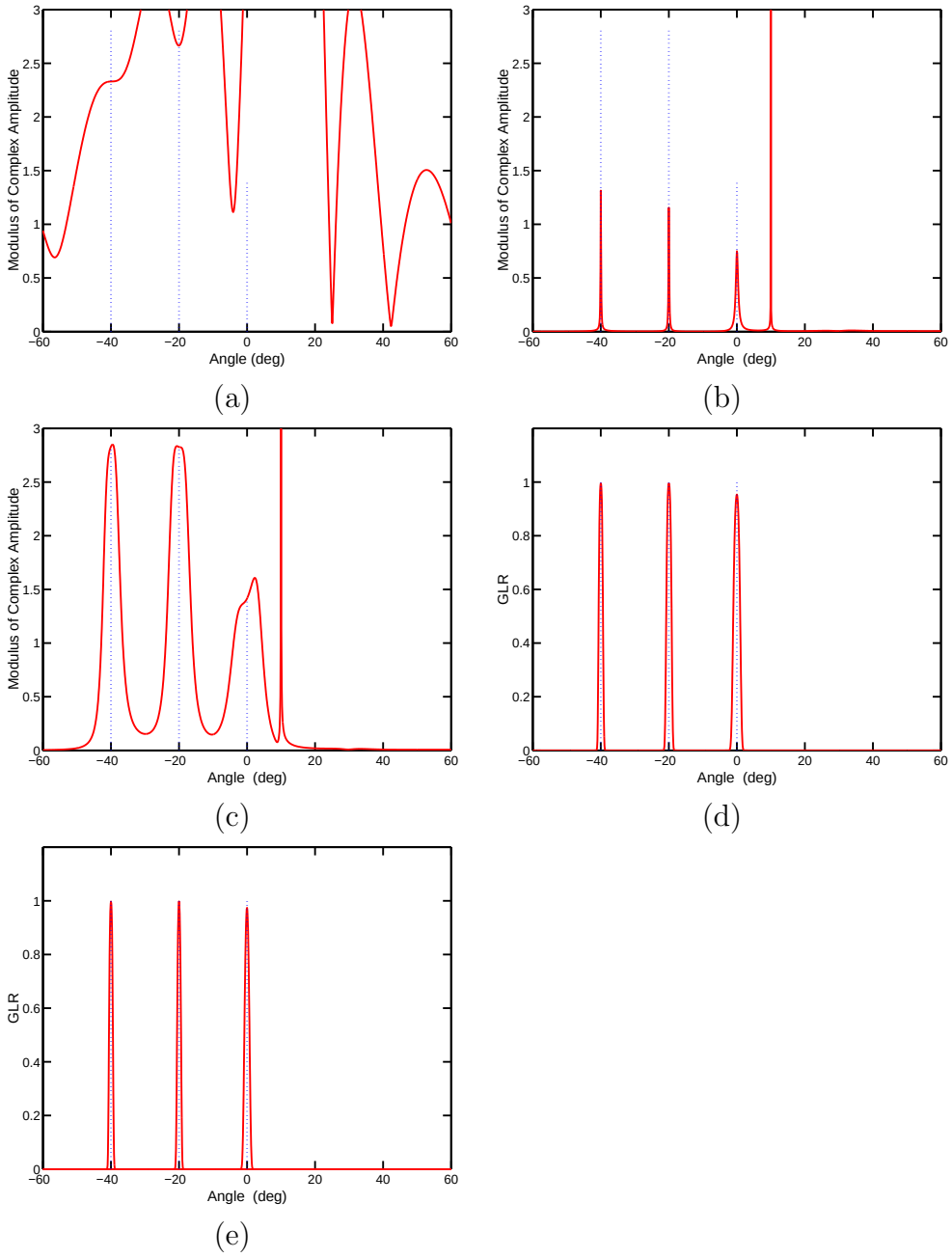


Figure 5–3: Spatial spectra, and GLR and cGLR Pseudo-Spectra ,when $\theta_1 = -40°$, $\theta_2 = -20°$, and $\theta_3 = 0°$. (a) LS, (b) Capon, (c) APES, (d) GLRT, and (e) iGLRT.

detected and rejected, and a correct estimate of the number of the targets can be obtained by both methods.

Next we consider a more challenging example where θ_2 is $-4°$ while all the other simulation parameters are the same as before. As shown in Fig. 5–4(c), in this example, the APES, Capon and GLRT methods fail to resolve the two closely spaced targets at $\theta_2 = -4°$ and $\theta_3 = 0°$. On the other hand, iGLRT gives well-resolved peaks around the true target locations. To illustrate the procedure of the iGLRT algorithm, we give the GLR, and cGLRs obtained in Steps I and II of iGLRT in Figs. 5–5(a) - 5–5(d). Figs. 5–5(a) and 5–5(b) show the GLR $\rho(\theta)$ and the cGLR $\rho(\theta|\hat{\theta}_1)$, respectively, where $\hat{\theta}_1$ is the estimated location of target 1 from $\rho(\theta)$. As we can see, there is no peak at around $\theta_3 = 0°$ in both figures. Yet a clear peak is shown in $\rho(\theta|\hat{\theta}_1, \hat{\theta}_2)$ in Fig. 5–5(c), which indicates the existence and location of target 3. The cGLR $\rho(\theta|\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)$ in Fig. 5–5(d) shows that no additional target exists other than the targets at $\hat{\theta}_1$, $\hat{\theta}_2$ and $\hat{\theta}_3$. In other words, the iGLRT method correctly estimates the number of targets to be 3.

Now we consider the elements in $\mathbf{B}_{\theta_1}$, $\mathbf{B}_{\theta_2}$ and $\mathbf{B}_{\theta_3}$ as i.i.d complex Gaussian random variables with mean zero and unit variance. The other parameters are the same as those in Fig. 5–6. The Figs. 5–6(a) and 5–6(b) present the CDFs of the MSEs of θ_1 and θ_3 as well as the CRBs, when SNR = 20 dB and $L = 128$. As we can see, the MSEs of the iGLRT are very close to the corresponding CRBs. Figs. 5–7(a) and 5–7(b) show the outage $\text{MSE}_{0.1}$ and $\text{CRB}_{0.1}$ when $p = 0.1$ as functions of SNR when $L = 128$. Again, the MSEs are very close to the corresponding the CRBs, and decreases almost linearly as SNR increases. Fig. 5–8 givse the outage $\text{MSE}_{0.1}$ and $\text{CRB}_{0.1}$ as functions of L when SNR=20 dB. As expected, the outage $\text{MSE}_{0.1}$ approaches the corresponding $\text{CRB}_{0.1}$ as L increases.

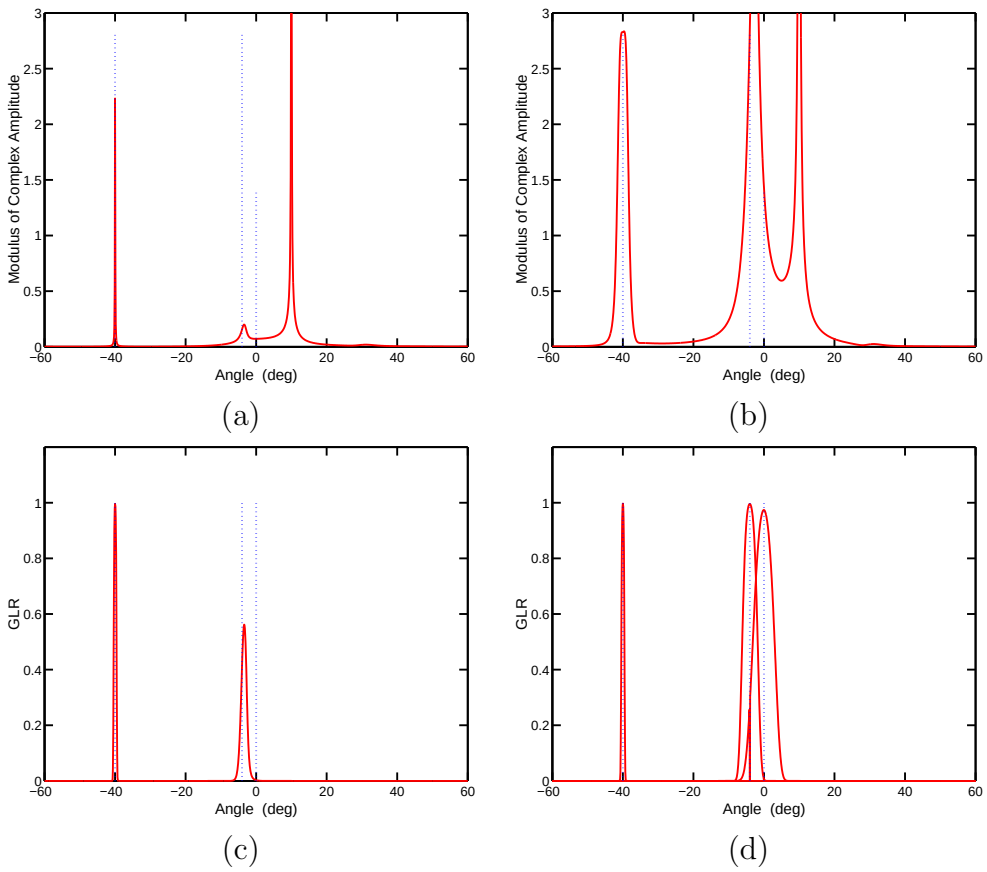


Figure 5–4: Spatial spectra, and GLR and cGLR Pseudo-Spectra, when $\theta_1 = -40°$, $\theta_2 = -4°$, and $\theta_3 = 0°$. (a) Capon, (b) APES, (c) GLRT, and (d) iGLRT.

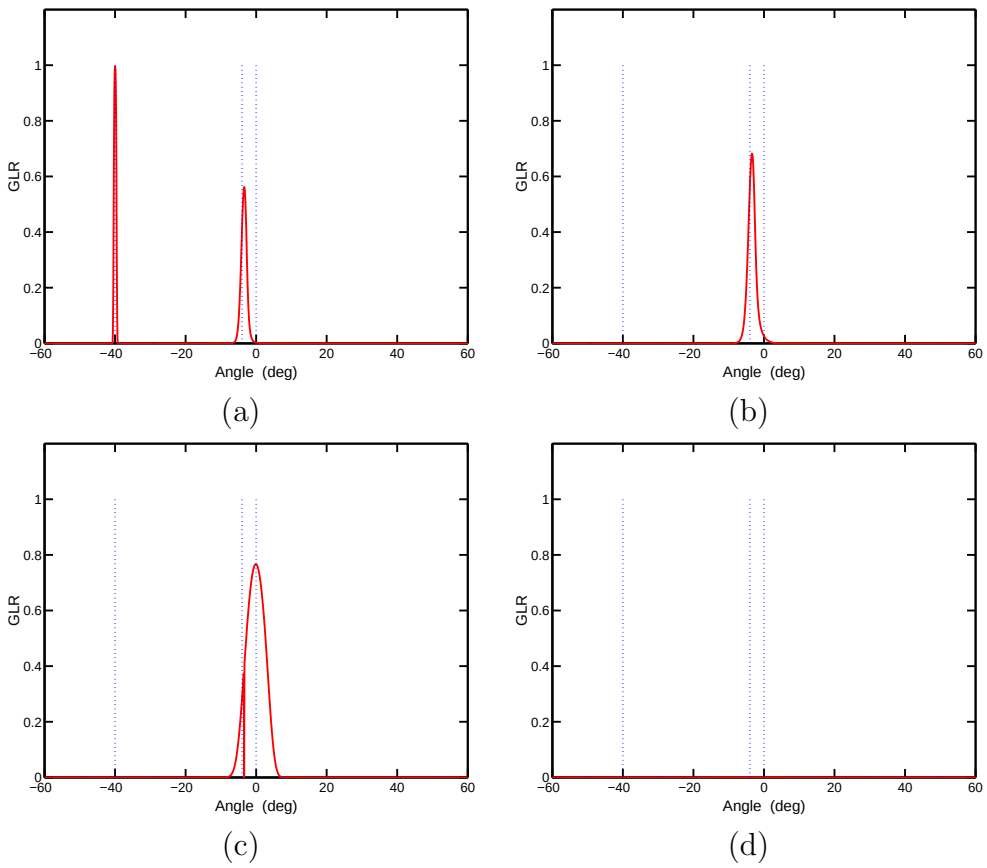


Figure 5–5: GLR and cGLR Pseudo-Spectra obtained in Steps I and II of iGLRT, when $\theta_1 = -40°$, $\theta_2 = -4°$, and $\theta_3 = 0°$. (a) $\rho(\theta)$, (b) $\rho(\theta|\hat{\theta}_1)$, (c) $\rho(\theta|\hat{\theta}_1, \hat{\theta}_2)$, and (d) $\rho(\theta|\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)$.

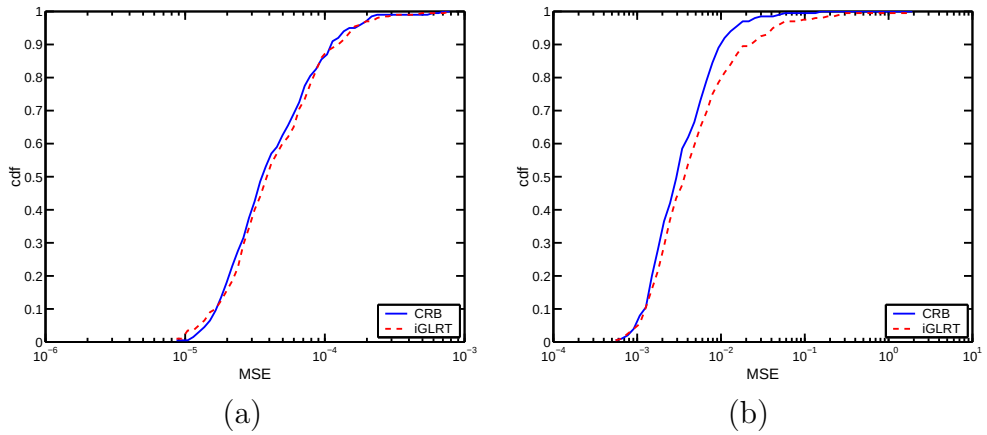
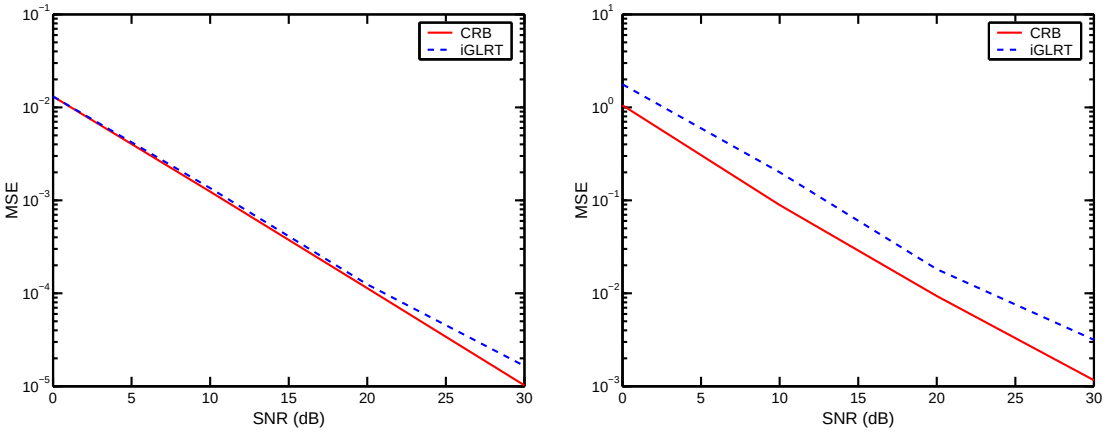
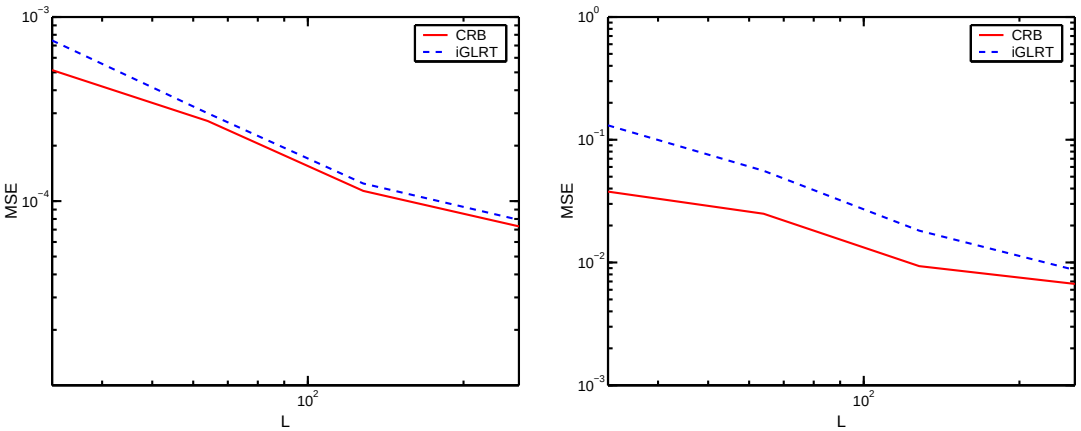


Figure 5–6: Cumulative density functions of the CRBs and MSEs for (a) θ_1 and (b) θ_3.



(a) (b)

Figure 5–7: Outage $\mathrm{CRB}_{0.1}$ and $\mathrm{MSE}_{0.1}$ versus SNR for (a) θ_1 and (b) θ_3.



(a) (b)

Figure 5–8: Outage $\mathrm{CRB}_{0.1}$ and $\mathrm{MSE}_{0.1}$ versus SNR for (a) θ_1 and (b) θ_3.

5.5 Conclusions

We have considered a multiple-input multiple-output (MIMO) radar system with a general antenna configuration that can be used to achieve both the coherent processing gain and the spatial diversity gain. We have first introduced several spatial spectral estimators, including Capon and APES, for target detection and parameter estimation. By using our results on the growth-curve models, we have provided a generalized likelihood ratio test (GLRT) and a conditional generalized likelihood ratio test (cGLRT), and then proposed an iterative GLRT (iGLRT) procedure for the MIMO radar system. Via several numerical examples, we have shown that the iGLRT method can provide excellent target detection and parameter estimation performance at a low computational cost.

CHAPTER 6
CONCLUSIONS AND FUTURE WORK

In this dissertation, we have studied several variations of the growth-curve model, including growth-curve (GC), diagonal growth-curve (DGC) and block diagonal growth-curve (BDGC). These models are in the same form but with different constraints on the unknown regression coefficient matrix. Specifically, the regression coefficient matrix in GC is unconstrained, whereas it is constrained to be diagonal in DGC and block diagonal in BDGC. We have provided a maximum likelihood (ML) estimator for the regression coefficient matrix for the GC model, and proposed two approximate maximum likelihood (AML) estimators for DGC and BDGC. We have also analyzed the statistical properties of the ML and AML estimators theoretically, and have shown that all three estimators are unbiased and asymptotically statistically efficient for a large data sample number. Several applications of the growth curve models to signal processing, including spectral analysis, array signal processing, wireless communications and multiple-input multiple-output (MIMO) radar, have also been investigated.

Motivated by the success of the MIMO wireless communications, we have investigated a MIMO radar topic in detail, which is attracting increasing attentions in both academic and industry. We have considered a MIMO radar system with a general antenna configuration, i.e., both the transmitter and receiver have multiple well-separated subarrays with each subarray containing closely-spaced antennas. Hence, both the coherent processing gain and the spatial diversity gain can be achieved by the system simultaneously. By using our results on the growth-curve model, we have provided a generalized likelihood ratio test (GLRT) and a conditional generalized likelihood ratio test (cGLRT) for the system, and

then proposed an iterative GLRT (iGLRT) procedure for target detection and parameter estimation. Via several numerical examples, we have shown that iGLRT can provide excellent detection and estimation performance at a low computational cost.

Since MIMO radar is an emerging technology, many topics and issues of MIMO radar are yet to be addressed . The following list provides of the interesting topics on MIMO radar.

Fundamental Tradeoffs in MIMO Radar

Two classes of MIMO radar have been studied for different purposes in the literature. In [42], multiple antennas are used to achieve diversity. It has been shown [42] that for a MIMO radar with M transmit and N receive antennas, the maximal diversity gain is MN. In [58], we use the MIMO radar technology to improve the parameter identifiability. We have shown [58] that the maximum number of targets that can be uniquely identified (with probability 1) by the MIMO radar is up to $\lfloor \frac{2MN-1}{3} \rfloor$, i.e., approximately M times that of its phased-array counterpart. Obviously, by using the general antenna configuration in Chapter 5, we need to make a trade-off between the spatial diversity gain and the parameter identifiability. One interesting problem is weather there is a fundamental tradeoff theory for MIMO radar, which is analogue to the one for wireless communications [87].

Target Detection and Parameter Estimation

In this dissertation, we have discussed the target detection and parameter estimation problem in a narrow-band MIMO radar system. We have considered the case that the targets are point-sources and are located in the same range bin with the arrival time known *a priori*. However, in many practical applications, these assumptions are hardly satisfied. The detection and estimation problem for MIMO

radar in more challenging environments and for various applications needs to be studied. For examples,we need to devise algorithms for

- Wide-band MIMO radar,
- MIMO radar for synthetic aperture radar (SAR) imaging,
- Knowledge-aided [88] - [90] MIMO radar,
- MIMO radar for biomedical imaging.

Probing Signal Design for MIMO Radar

Several probing signal design methods for MIMO radar have been proposed in [44] [45]. These methods mainly focus on the optimization of the spatial features of the transmitted waveforms, and hence they actually are spatial beampattern designs. On the other hand, in the literature [91] - [94] many temporal designs have been proposed for the conventional radar. In [94], we have proposed a Signal Waveform's Optimal-under-Restriction Design (SWORD) for active sensing, which can significantly increase the SINR while keeping the desired features, for example, constant modulus as well as reasonable range resolution and peak sidelobe level. In the light of these two different design philosophies, it is interesting to develop a space-time design of probing signals, which can achieve optimality both spatially and temporally.

APPENDIX A
PROOFS FOR THE GROWTH-CURVE MODEL

A.1 Cramér-Rao Bound for the GC Model

In the data model in (2–1), both $\mathbf{Q}$ and $\mathbf{B}$ are unknown. Let $\boldsymbol{\theta}$ denote the vector containing the real-valued unknowns in $\mathbf{Q}$ and $\mathbf{B}$. Then, the (i,k)-th element of the corresponding Fisher information matrix (FIM) [71] [72] is

$$\mathrm{FIM}(\theta_i,\theta_k) = L\,\mathrm{tr}\big(\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\theta_i}\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\theta_k}\big) + 2\mathrm{Re}\Big\{\mathrm{tr}\Big[\Big(\frac{\partial(\mathbf{ABS})}{\partial\theta_i}\Big)^H\mathbf{Q}^{-1}\Big(\frac{\partial(\mathbf{ABS})}{\partial\theta_k}\Big)\Big]\Big\}, \tag{A–1}$$

where $\mathrm{tr}(\cdot)$ denotes the trace of a matrix, $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ denote the real and imaginary parts of a complex number (or matrix), respectively, and θ_i denotes the ith element of $\boldsymbol{\theta}$. Because $\mathbf{Q}$ and $\mathbf{B}$ depend on the different variables in $\boldsymbol{\theta}$, FIM will be a block diagonal matrix with respect to the unknowns in $\mathbf{Q}$ and $\mathbf{B}$. Hence, we can calculate the CRBs of $\mathbf{B}$ and $\mathbf{Q}$ separately. In this Chapter, we are only interested in the CRB of $\mathbf{B}$.

Let $b_{ik,R}$ and $b_{ik,I}$ denote the real and imaginary parts of the (i,k)-th element in $\mathbf{B}$, respectively. The corresponding elements in FIM with respect to any two real-valued unknowns in $\mathbf{B}$ are

$$\mathrm{FIM}(b_{ik,R},b_{mn,R}) = 2\mathrm{Re}\big[(\mathbf{s}_k^*\mathbf{s}_n^T)(\mathbf{a}_i^H\mathbf{Q}^{-1}\mathbf{a}_m)\big], \tag{A–2}$$

$$\mathrm{FIM}(b_{ik,R},b_{mn,I}) = -2\mathrm{Im}\big[(\mathbf{s}_k^*\mathbf{s}_n^T)(\mathbf{a}_i^H\mathbf{Q}^{-1}\mathbf{a}_m)\big], \tag{A–3}$$

$$\mathrm{FIM}(b_{ik,I},b_{mn,R}) = 2\mathrm{Im}\big[(\mathbf{s}_k^*\mathbf{s}_n^T)(\mathbf{a}_i^H\mathbf{Q}^{-1}\mathbf{a}_m)\big], \tag{A–4}$$

$$\mathrm{FIM}(b_{ik,I},b_{mn,I}) = 2\mathrm{Re}\big[(\mathbf{s}_k^*\mathbf{s}_n^T)(\mathbf{a}_i^H\mathbf{Q}^{-1}\mathbf{a}_m)\big], \tag{A–5}$$

where $\mathbf{s}_k$ $(k=1,2,\ldots K)$ and $\mathbf{a}_i$ $(i=1,2,\ldots N)$ are the kth row vector in $\mathbf{S}$ and ith column vector in $\mathbf{A}$, respectively.

Arranging the complex-valued matrix $\mathbf{B}$ to form a real-valued vector, i.e.,

$$\mathbf{b} = [\mathrm{Re}(\mathrm{vec}(\mathbf{B}))^T \quad \mathrm{Im}(\mathrm{vec}(\mathbf{B}))^T]^T, \tag{A–6}$$

and arranging (A–2) to (A–5) to a matrix according to $\mathbf{b}$, we get the corresponding FIM

$$\mathrm{FIM}(\mathbf{b}) = 2 \begin{bmatrix} \mathrm{Re}(\mathbf{\Phi}) & -\mathrm{Im}(\mathbf{\Phi}) \\ \mathrm{Im}(\mathbf{\Phi}) & \mathrm{Re}(\mathbf{\Phi}) \end{bmatrix}, \tag{A–7}$$

where

$$\mathbf{\Phi} = (\mathbf{S}^*\mathbf{S}^T) \otimes (\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{A}). \tag{A–8}$$

Using the matrix inversion lemma and the inversion lemma of partitioned matrices [74], we get the CRB in real-valued form

$$\mathrm{CRB}(\mathbf{b}) = [\ \mathrm{FIM}(\mathbf{b})]^{-1} = \frac{1}{2} \begin{bmatrix} \mathrm{Re}(\mathbf{\Phi}^{-1}) & -\mathrm{Im}(\mathbf{\Phi}^{-1}) \\ \mathrm{Im}(\mathbf{\Phi}^{-1}) & \mathrm{Re}(\mathbf{\Phi}^{-1}) \end{bmatrix}. \tag{A–9}$$

Transforming the above real-valued CRB into the complex-valued form yields

$$\mathrm{CRB}(\mathbf{B}) \triangleq \mathrm{CRB}\,(\mathrm{vec}(\mathbf{B})) = (\mathbf{S}^*\mathbf{S}^T)^{-1} \otimes (\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{A})^{-1}. \tag{A–10}$$

Clearly, the diagonal elements of $\mathrm{CRB}(\mathbf{B})$ are determined by the diagonal elements of $(\mathbf{S}^*\mathbf{S}^T)^{-1}$ and $(\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{A})^{-1}$. To study the influences of the temporal information matrix $\mathbf{S}$ and the spatial information matrix $\mathbf{A}$ on $\mathrm{CRB}(\mathbf{B})$, we denote $\mathbf{R}_S = (\mathbf{S}^*\mathbf{S}^T)$ and $\mathbf{R}_A = (\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{A})$. Without loss of generality, we consider the CRB of b_{11}, i.e., the element on the first column and first row of $\mathbf{B}$. Partition $\mathbf{R}_S$, $\mathbf{R}_S^{-1}$, $\mathbf{R}_A$ and $\mathbf{R}_A^{-1}$ as follows.

$$\mathbf{R}_S = \begin{bmatrix} (r_S)_{11} & (\mathbf{r}_S)_{12} \\ (\mathbf{r}_S)_{21} & (\mathbf{R}_S)_{22} \end{bmatrix}, \qquad \mathbf{R}_S^{-1} = \begin{bmatrix} (r_S)^{11} & (\mathbf{r}_S)^{12} \\ (\mathbf{r}_S)^{21} & (\mathbf{R}_S)^{22} \end{bmatrix}, \tag{A–11}$$

$$\mathbf{R}_A = \begin{bmatrix} (r_A)_{11} & (\mathbf{r}_A)_{12} \\ (\mathbf{r}_A)_{21} & (\mathbf{R}_A)_{22} \end{bmatrix}, \qquad \mathbf{R}_A^{-1} = \begin{bmatrix} (r_A)^{11} & (\mathbf{r}_A)^{12} \\ (\mathbf{r}_A)^{21} & (\mathbf{R}_A)^{22} \end{bmatrix}. \tag{A–12}$$

Obviously, $\mathrm{CRB}(b_{11}) = (r_S)^{11}(r_A)^{11}$. By the inversion lemma of partitioned matrices [74], we have

$$(r_S)^{11} = \frac{1}{(r_S)_{11} - (\mathbf{r}_S)_{12}\,(\mathbf{R}_S)_{22}^{-1}\,(\mathbf{r}_S)_{21}}, \tag{A–13}$$

$$(r_A)^{11} = \frac{1}{(r_A)_{11} - (\mathbf{r}_A)_{12}\,(\mathbf{R}_A)_{22}^{-1}\,(\mathbf{r}_A)_{21}}. \tag{A–14}$$

In (A–13), $(r_S)_{11} = \| \mathbf{s}_1 \|^2$ is the Euclidean norm square of the first row of $\mathbf{S}$. It is easily verified that $(\mathbf{R}_S)_{22}$ is a positive definite matrix while $(\mathbf{r}_S)_{12} = (\mathbf{r}_S)_{21}^H$. Therefore, $(r_S)^{11}$ is minimized when $(\mathbf{r}_S)_{12} = \mathbf{0}$, i.e., $\mathbf{s}_1^*\mathbf{s}_k^T = 0$ $(k = 2, 3, \ldots, K)$. Hence, to minimize $\mathrm{CRB}(\mathbf{B})$, the row vectors in $\mathbf{S}$ should be orthogonal to each other.

Similarly, $(r_A)^{11}$ is minimized when $(\mathbf{r}_A)_{12} = \mathbf{0}$, i.e., $(\mathbf{Q}^{-\frac{1}{2}}\mathbf{a}_1)^H(\mathbf{Q}^{-\frac{1}{2}}\mathbf{a}_i) = 0$ $(i = 2, 3, \ldots N)$. Note that when this condition is not satisfied, large N causes $(r_A)^{11}$ to increase. Furthermore, We note that $(r_A)_{11} = \mathbf{a}_1^H\mathbf{Q}\mathbf{a}_1$. Therefore, when $\mathbf{a}_1$ is proportional to the eigenvector of $\mathbf{Q}$ corresponding to its smallest eigenvalue, $(r_A)^{11}$ is minimized. Therefore, to minimize $\mathrm{CRB}(\mathbf{B})$, the columns of $\mathbf{Q}^{-\frac{1}{2}}\mathbf{A}$ should be orthogonal to each other and the columns of $\mathbf{A}$ should correspond to the subspace spanned by the eigenvectors of $\mathbf{Q}$ corresponding to its smallest N eigenvalues. Since $\mathbf{Q}$ is unknown and usually $\mathbf{A}$ is given and cannot be changed in practice, we can only hope for these conditions of $\mathbf{A}$.

A.2 Proof of Lemma 2.1

According to [75] and [95], the pdf of the complex Wishart distribution is

$$p(\boldsymbol{\Upsilon}) = \frac{|\boldsymbol{\Upsilon}|^{l-m}}{I(\boldsymbol{\Sigma})}\exp\{-\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Upsilon})\}, \tag{A–15}$$

where $I(\mathbf{\Sigma})$ is a constant scalar dependent on $\mathbf{\Sigma}$, i.e.,

$$I(\mathbf{\Sigma}) = \pi^{\frac{1}{2}m(m-1)}\Gamma(l)\Gamma(l-1)\ldots\Gamma(l-m+1)|\mathbf{\Sigma}|^m, \tag{A–16}$$

with $\Gamma(\cdot)$ being the Gamma function.

Partition $\mathbf{\Sigma}^{-1}$ as $\mathbf{\Sigma}$ in (2–21)

$$\mathbf{\Sigma}^{-1} = \begin{pmatrix} \mathbf{\Sigma}^{11} & \mathbf{\Sigma}^{12} \\ \mathbf{\Sigma}^{21} & \mathbf{\Sigma}^{22} \end{pmatrix}. \tag{A–17}$$

Utilizing the formula that $|\mathbf{\Upsilon}| = |\mathbf{\Upsilon}_{22}||\mathbf{\Upsilon}_{11.2}|$ [74] and the fact that $\mathbf{\Upsilon}$ and $\mathbf{\Sigma}$ are both Hermitian symmetric, (A–15) can be written

$$f(\mathbf{\Upsilon}) = \frac{|\mathbf{\Upsilon}_{22}|^{l-m}}{I(\mathbf{\Sigma})}|\mathbf{\Upsilon}_{11.2}|^{l-m}\exp\big\{-\mathrm{tr}(\mathbf{\Sigma}^{11}\mathbf{\Upsilon}_{11} + \mathbf{\Sigma}^{12}\mathbf{\Upsilon}_{12}^H + (\mathbf{\Sigma}^{12})^H\mathbf{\Upsilon}_{12} + \mathbf{\Sigma}^{22}\mathbf{\Upsilon}_{22})\big\}. \tag{A–18}$$

Making a transformation from $(\mathbf{\Upsilon}_{11},\ \mathbf{\Upsilon}_{12},\ \mathbf{\Upsilon}_{22})$ to $(\mathbf{\Upsilon}_{11.2},\ \mathbf{\Upsilon}_{12},\ \mathbf{\Upsilon}_{22})$ whose Jocobian determinant is easily calculated to be 1, and after some straightforward manipulations using the inversion lemma of partitioned matrices, we get

$$f(\mathbf{\Upsilon}_{11.2}, \mathbf{\Upsilon}_{12}, \mathbf{\Upsilon}_{22}) = f_1(\mathbf{\Upsilon}_{11.2})\, f_2(\mathbf{\Upsilon}_{12})\, f_3(\mathbf{\Upsilon}_{22}), \tag{A–19}$$

where

$$f_1(\mathbf{\Upsilon}_{11.2}) = K_1|\mathbf{\Upsilon}_{11.2}|^{l-m}\exp\big\{-\mathrm{tr}[\mathbf{\Sigma}_{11.2}^{-1}\mathbf{\Upsilon}_{11.2}]\big\}, \tag{A–20}$$

$$f_2(\mathbf{\Upsilon}_{12}) = K_2|\mathbf{\Upsilon}_{22}|^{-m_1}\exp\big\{-\mathrm{tr}[\mathbf{\Sigma}_{11.2}^{-1}(\mathbf{\Upsilon}_{12} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Upsilon}_{22})\mathbf{\Upsilon}_{22}^{-1}(\mathbf{\Upsilon}_{12} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Upsilon}_{22})^H\big\}, \tag{A–21}$$

$$f_3(\mathbf{\Upsilon}_{22}) = K_3|\mathbf{\Upsilon}_{22}|^{l-m_2}\exp\big\{-\mathrm{tr}[\mathbf{\Sigma}_{22}^{-1}\mathbf{\Upsilon}_{22}]\big\}, \tag{A–22}$$

with K_1, K_2, K_3 being constant scalars.

Based on the pdfs of the complex Wishart and matrix-variate complex Gaussian distributions in [75] and [95], the conclusions in Lemma 2.1 is immediately reached from (A–19) to (A–22).

A.3 Proof of Lemma 2.2

Let $\tilde{\boldsymbol{\Upsilon}} = \boldsymbol{\Upsilon} - \mathbf{H}$. Then obviously $\tilde{\boldsymbol{\Upsilon}} \sim CN(\mathbf{0}; \boldsymbol{\Sigma}, \boldsymbol{\Omega})$ [75]. Hence,

$$E(\boldsymbol{\Upsilon}\mathbf{C}\boldsymbol{\Upsilon}^H) = E(\tilde{\boldsymbol{\Upsilon}}\mathbf{C}\tilde{\boldsymbol{\Upsilon}}^H) + \mathbf{H}\mathbf{C}\mathbf{H}^H. \tag{A–23}$$

Since $\boldsymbol{\Omega}$ is Hermitian symmetric [75], $|\boldsymbol{\Omega}|$ is a real-valued scalar, i.e., $|\boldsymbol{\Omega}|^* = |\boldsymbol{\Omega}|$. Using the formulas that $\mathrm{tr}(\mathbf{X}^T\mathbf{Y}\mathbf{Z}\mathbf{W}^T) = \mathrm{vec}(\mathbf{X})^T(\mathbf{W} \otimes \mathbf{Y})\mathrm{vec}(\mathbf{Z})$ and $|\mathbf{X}_{l\times l}|^m|\mathbf{Y}_{m\times m}|^l = |\mathbf{X} \otimes \mathbf{Y}|$ [74], we can write the pdf of $\tilde{\boldsymbol{\Upsilon}}$ as

$$f(\tilde{\boldsymbol{\Upsilon}}) = (2\pi)^{-np}|\boldsymbol{\Omega}^* \otimes \boldsymbol{\Sigma}|^{-1}\exp\{-\mathrm{vec}(\tilde{\boldsymbol{\Upsilon}})^H(\boldsymbol{\Omega}^* \otimes \boldsymbol{\Sigma})^{-1}\mathrm{vec}(\tilde{\boldsymbol{\Upsilon}})\}. \tag{A–24}$$

Equation (A–24) is a standard vector complex Gaussian pdf with zero-mean and covariance matrix $\boldsymbol{\Omega}^* \otimes \boldsymbol{\Sigma}$. Hence,

$$E(\mathrm{vec}(\tilde{\boldsymbol{\Upsilon}})\mathrm{vec}(\tilde{\boldsymbol{\Upsilon}})^H) = \boldsymbol{\Omega}^* \otimes \boldsymbol{\Sigma}, \tag{A–25}$$

i.e.,

$$E(\tilde{\boldsymbol{\upsilon}}_i\tilde{\boldsymbol{\upsilon}}_k^H) = \Omega_{ik}^*\boldsymbol{\Sigma} = \Omega_{ki}\boldsymbol{\Sigma}, \tag{A–26}$$

where $\tilde{\boldsymbol{\upsilon}}_i$ denotes the ith column vector of $\tilde{\boldsymbol{\Upsilon}}$ and Ω_{ik} is the (i,k)-th element in $\boldsymbol{\Omega}$.

Let c_{ik} be the (i,k)-th element in $\mathbf{C}$. Then we have

$$\begin{aligned} E(\tilde{\boldsymbol{\Upsilon}}\mathbf{C}\tilde{\boldsymbol{\Upsilon}}^H) &= E(\sum_{i=1}^{p}\sum_{k=1}^{p} c_{ik}\tilde{\boldsymbol{\upsilon}}_i\tilde{\boldsymbol{\upsilon}}_k^H) \\ &= \boldsymbol{\Sigma}\sum_{i=1}^{p}\sum_{k=1}^{p} c_{ik}\Omega_{ki} \\ &= \boldsymbol{\Sigma}\mathrm{tr}(\mathbf{C}\boldsymbol{\Omega}). \end{aligned} \tag{A–27}$$

By (A–23) and (A–27), Lemma 2.2 is immediately proved.

A.4 Proof of Lemma 2.3

Let $\mathbf{g}$ be any non-zero $m \times 1$ vector. To calculate the expectation of matrix $\boldsymbol{\Upsilon}^{-1}$, we first consider the expectation of the scalar random variable $\mathbf{g}^H\boldsymbol{\Upsilon}^{-1}\mathbf{g}$.

Let $\tilde{\mathbf{g}} = \mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{g}$ and $\tilde{\mathbf{\Upsilon}} = \mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{\Upsilon}\mathbf{\Sigma}^{-\frac{1}{2}}$. Obviously, we have [75]

$$\tilde{\mathbf{\Upsilon}} \sim \mathrm{CW}(l, m; \mathbf{I}), \tag{A–28}$$

and

$$\mathbf{g}^H\mathbf{\Upsilon}^{-1}\mathbf{g} = \tilde{\mathbf{g}}^H\tilde{\mathbf{\Upsilon}}^{-1}\tilde{\mathbf{g}}. \tag{A–29}$$

Decompose the vector $\tilde{\mathbf{g}}$ as

$$\tilde{\mathbf{g}} = \mathbf{U}\,\mathbf{p}, \tag{A–30}$$

where $\mathbf{U}$ is a unitary matrix with its first column being $\frac{\tilde{\mathbf{g}}}{\|\tilde{\mathbf{g}}\|}$ and $\mathbf{p}_{m\times 1} = \| \tilde{\mathbf{g}} \| [1 \quad 0 \quad 0 \quad \dots \quad 0]^T$. Let $\breve{\mathbf{\Upsilon}} = \mathbf{U}^H\tilde{\mathbf{\Upsilon}}\mathbf{U}$, which has the $\mathrm{CW}(l, m; \mathbf{I})$ distribution. From (A–29) we get

$$\mathbf{g}^H\mathbf{\Upsilon}^{-1}\mathbf{g} = \mathbf{p}^H\breve{\mathbf{\Upsilon}}^{-1}\mathbf{p}. \tag{A–31}$$

Then partition $\breve{\mathbf{\Upsilon}}$ and $\breve{\mathbf{\Upsilon}}^{-1}$ as follows.

$$\breve{\mathbf{\Upsilon}} = \begin{bmatrix} \breve{\upsilon}_{11} & \breve{\boldsymbol{\upsilon}}_{12} \\ \breve{\boldsymbol{\upsilon}}_{21} & \breve{\mathbf{\Upsilon}}_{22} \end{bmatrix}, \qquad \breve{\mathbf{\Upsilon}}^{-1} = \begin{bmatrix} \breve{\upsilon}^{11} & \breve{\boldsymbol{\upsilon}}^{12} \\ \breve{\boldsymbol{\upsilon}}^{21} & \breve{\mathbf{\Upsilon}}^{22} \end{bmatrix}, \tag{A–32}$$

where $\breve{\upsilon}_{11}$ and $\breve{\upsilon}^{11}$ are both scalars.

Let $\breve{\upsilon}_{11.2} = \breve{\upsilon}_{11} - \breve{\boldsymbol{\upsilon}}_{12}\breve{\mathbf{\Upsilon}}_{22}^{-1}\breve{\boldsymbol{\upsilon}}_{21}$. Then,

$$\begin{aligned} (\mathbf{g}^H\mathbf{\Upsilon}^{-1}\mathbf{g})^{-1} &= (\mathbf{p}^H\breve{\mathbf{\Upsilon}}^{-1}\mathbf{p})^{-1} \\ &= (\mathbf{g}^H\mathbf{\Sigma}^{-1}\mathbf{g})^{-1}(\breve{\upsilon}^{11})^{-1} \\ &= (\mathbf{g}^H\mathbf{\Sigma}^{-1}\mathbf{g})^{-1}\breve{\upsilon}_{11.2}. \end{aligned} \tag{A–33}$$

By Lemma 2.1, we know $\breve{\upsilon}_{11.2} \sim \mathrm{CW}(l - m + 1, 1; 1)$. According to the definition of the complex Wishart distribution, $\breve{\upsilon}_{11.2}$ can be expressed as the sum of the norm squares of $l - m + 1$ independent standard complex Gaussian random variables. Hence, $2\breve{\upsilon}_{11.2}$ can be expressed as the sum of the squares of $2(l - m + 1)$ independent standard real-valued Gaussian random variables, i.e., $2\breve{\upsilon}_{11.2}$ has the central χ^2 distribution with $2(l - m + 1)$ degrees of freedom [96]. According to the

proprieties of the χ^2 distribution, we know that

$$E\left[(2\breve{v}_{11.2})^{-1}\right] = \frac{1}{2(l-m)}. \tag{A–34}$$

Thus, by (A–33) and (A–34)

$$\begin{aligned} E(\mathbf{g}^H \boldsymbol{\Upsilon}^{-1}\mathbf{g}) &= (2\mathbf{g}^H \boldsymbol{\Sigma}^{-1}\mathbf{g}) E[(2\breve{v}_{11.2})^{-1}] \\ &= \mathbf{g}^H \left(\frac{\boldsymbol{\Sigma}^{-1}}{l-m}\right)\mathbf{g}. \end{aligned} \tag{A–35}$$

Hence, $\mathbf{g}^H E(\boldsymbol{\Upsilon}^{-1})\mathbf{g} = \mathbf{g}^H(\frac{\boldsymbol{\Sigma}^{-1}}{l-m})\mathbf{g}$. The equation should be satisfied for all non-zero $\mathbf{g}$, which results in $E(\boldsymbol{\Upsilon}^{-1}) = \frac{\boldsymbol{\Sigma}^{-1}}{l-m}$.

A.5 Proof of Lemma 2.4

Let $\mathbf{g}$ be any non-zero $p\times 1$ vector. We first consider the expectation of $\mathbf{g}^H \boldsymbol{\Pi}\, \mathbf{g}$.

Decompose $\mathbf{g}$ as

$$\mathbf{g} = \mathbf{U}\,\tilde{\mathbf{g}}, \tag{A–36}$$

where $\mathbf{U}$ is a $p \times p$ unitary matrix with its first column being $\frac{\mathbf{g}}{\|\mathbf{g}\|}$ and $\tilde{\mathbf{g}}_{p\times 1} = \| \mathbf{g} \|$ $[1 \quad 0 \quad 0 \quad \ldots \quad 0]^T$.

Let $\tilde{\boldsymbol{\Upsilon}} = \boldsymbol{\Upsilon}\mathbf{U}$; obviously $\tilde{\boldsymbol{\Upsilon}} \sim \mathrm{CN}(0; \mathbf{I}_n, \mathbf{I}_p)$. Let $\tilde{\boldsymbol{v}}_i$ $(i = 1, 2, \ldots p)$ be the ith column vector of $\tilde{\boldsymbol{\Upsilon}}$. Let $\tilde{\boldsymbol{\Psi}} = \boldsymbol{\Psi} + \sum_{i=2}^{p} \tilde{\boldsymbol{v}}_i \tilde{\boldsymbol{v}}_i^H$; obviously $\tilde{\boldsymbol{\Psi}} \sim \mathrm{CW}(l+p-1, n; \mathbf{I})$ [75]. Then

$$\begin{aligned} \mathbf{g}^H \boldsymbol{\Pi}\, \mathbf{g} &= \tilde{\mathbf{g}}^H \tilde{\boldsymbol{\Upsilon}}^H (\boldsymbol{\Psi} + \tilde{\boldsymbol{\Upsilon}}\tilde{\boldsymbol{\Upsilon}}^H)^{-1} \tilde{\boldsymbol{\Upsilon}}\tilde{\mathbf{g}} \\ &= \| \mathbf{g} \|^2 \tilde{\boldsymbol{v}}_1^H (\tilde{\boldsymbol{\Psi}} + \tilde{\boldsymbol{v}}_1 \tilde{\boldsymbol{v}}_1^H)^{-1} \tilde{\boldsymbol{v}}_1 \\ &= \| \mathbf{g} \|^2 \big[1 - \frac{1}{1 + \tilde{\boldsymbol{v}}_1^H \tilde{\boldsymbol{\Psi}}^{-1} \tilde{\boldsymbol{v}}_1}\big]. \end{aligned} \tag{A–37}$$

It has been shown in the appendix of [64] that

$$\frac{1}{1 + \tilde{\boldsymbol{v}}_1^H \tilde{\boldsymbol{\Psi}}^{-1} \tilde{\boldsymbol{v}}_1} \sim \mathrm{beta}(l-n+p, n). \tag{A–38}$$

Hence [96],

$$E\left(\frac{1}{1+\tilde{\boldsymbol{\upsilon}}_1^H\tilde{\boldsymbol{\Psi}}^{-1}\tilde{\boldsymbol{\upsilon}}_1}\right)=\frac{l-n+p}{l+p}. \tag{A–39}$$

From (A–37) and (A–39), we have

$$\mathbf{g}^H E(\boldsymbol{\Pi})\,\mathbf{g}=\mathbf{g}^H\left(\frac{n}{l+p}\mathbf{I}_p\right)\mathbf{g}. \tag{A–40}$$

The above equation should be satisfied for any non-zero vector $\mathbf{g}$, which means $E(\boldsymbol{\Pi})=\frac{n}{l+p}\mathbf{I}$.

To calculate $E(\mathrm{vec}(\boldsymbol{\Pi})\mathrm{vec}(\boldsymbol{\Pi})^H)$, we need to calculate

$$E(\pi_{r_1c_1}\pi^*_{r_2c_2})=E\big[\boldsymbol{\upsilon}_{r_1}^H(\boldsymbol{\Psi}+\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^H)^{-1}\boldsymbol{\upsilon}_{c_1}\boldsymbol{\upsilon}_{c_2}^H(\boldsymbol{\Psi}+\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^H)^{-1}\boldsymbol{\upsilon}_{r_2}\big], \tag{A–41}$$

where $\pi_{r_1c_1}$ denotes the (r_1,c_1)-th element in $\boldsymbol{\Pi}$, $\boldsymbol{\upsilon}_k$ denotes the kth column vector in $\boldsymbol{\Upsilon}$ and $r_1,c_1,r_2,c_2=1,2,3,\ldots,p$.

Note that (A–41) is a function of $\boldsymbol{\upsilon}_k$ $(k=1,2,\ldots,p)$. Adopting the same technique used to obtain (2–54), we can easily show that the expectation is zero when (A–41) contains odd numbers of $\boldsymbol{\upsilon}_k$, e.g., $E(\pi_{12}\pi^*_{34})=E(\pi_{11}\pi^*_{23})=E(\pi_{12}\pi^*_{23})=E(\pi_{12}\pi^*_{32})=0$.

When $r_1=c_2$, $c_1=r_2$ but $r_1\neq c_1$,

$$E(\pi_{r_1c_1}\pi^*_{c_1r_1})=E\big\{\big[\boldsymbol{\upsilon}_{r_1}^H(\boldsymbol{\Psi}+\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^H)^{-1}\boldsymbol{\upsilon}_{c_1}\big]^2\big\}. \tag{A–42}$$

Replacing $\boldsymbol{\upsilon}_{r_1}$ by $j\cdot\boldsymbol{\upsilon}_{r_1}$, where j is the unit of the imaginary number, yields

$$E(\pi_{r_1c_1}\pi^*_{c_1r_1})|_{\boldsymbol{\upsilon}_{r_1}=j\cdot\boldsymbol{\upsilon}_{r_1}}=-E(\pi_{r_1c_1}\pi^*_{c_1r_1}). \tag{A–43}$$

On the other hand, since $\boldsymbol{\upsilon}_{r_1}$ is a zero-mean circularly symmetric complex Gaussian random vector, $j\cdot\boldsymbol{\upsilon}_{r_1}$, as a random vector transformed from $\boldsymbol{\upsilon}_{r_1}$, has the same statistical property as $\boldsymbol{\upsilon}_{r_1}$. Hence

$$E(\pi_{r_1c_1}\pi^*_{c_1r_1})|_{\boldsymbol{\upsilon}_{r_1}=j\cdot\boldsymbol{\upsilon}_{r_1}}=E(\pi_{r_1c_1}\pi^*_{c_1r_1}). \tag{A–44}$$

Hence, by (A–43) and (A–44) we have $E(\pi_{r_1c_1}\pi^*_{c_1r_1}) = 0$.

Besides the above cases, there are three cases left in which $E(\mathbf{\Pi}_{r_1c_1}\mathbf{\Pi}^*_{r_2c_2})$ is non-zero, i.e.,

(i). $r_1 = c_1 = r_2 = c_2$;

(ii). $r_1 = c_1$, $r_2 = c_2$ but $r_1 \neq r_2$;

(iii). $r_1 = r_2$, $c_1 = c_2$ but $r_1 \neq c_1$.

We know that the expectation in each case is equal. For convenience, we denote the expectations in the three cases as $E(|\pi_{11}|^2)$, $E(\pi_{11}\pi^*_{22})$, and $E(|\pi_{12}|^2)$, respectively.

First, we calculate $E(|\pi_{11}|^2)$. Let $\breve{\mathbf{\Psi}} = \mathbf{\Psi} + \sum_{k=2}^{p} \boldsymbol{\upsilon}_k\boldsymbol{\upsilon}_k^H$, which obviously has the $\mathrm{CW}(l+p-1, n; \mathbf{I})$ distribution. Then,

$$\begin{aligned} E(|\pi_{11}|^2) &= E\left(\left|\boldsymbol{\upsilon}_1^H(\breve{\mathbf{\Psi}}^{-1} + \boldsymbol{\upsilon}_1\boldsymbol{\upsilon}_1^H)^{-1}\boldsymbol{\upsilon}_1\right|^2\right) \\ &= E\left(\left|1 - \frac{1}{1+\boldsymbol{\upsilon}_1^H\breve{\mathbf{\Psi}}^{-1}\boldsymbol{\upsilon}_1}\right|^2\right) \\ &= 1 - 2E\left(\frac{1}{1+\boldsymbol{\upsilon}_1^H\breve{\mathbf{\Psi}}^{-1}\boldsymbol{\upsilon}_1}\right) + E\left[\left(\frac{1}{1+\boldsymbol{\upsilon}_1^H\breve{\mathbf{\Psi}}^{-1}\boldsymbol{\upsilon}_1}\right)^2\right]. \end{aligned} \tag{A–45}$$

Again, using the conclusion in the appendix of [64], we know

$$\frac{1}{1+\boldsymbol{\upsilon}_1^H\breve{\mathbf{\Psi}}^{-1}\boldsymbol{\upsilon}_1} \sim \mathrm{beta}(l-n+p, n). \tag{A–46}$$

Calculating the expectation and the 2nd-order moment of the above beta distribution and substituting them into (A–45) yields

$$E(|\pi_{11}|^2) = \frac{n(n+1)}{(l+p)(l+p+1)}. \tag{A–47}$$

Second, we consider $E(\pi_{11}\pi^*_{22})$. It is difficult to calculate this expectation directly. However, note that $\mathbf{\Psi} + \mathbf{\Upsilon}\mathbf{\Upsilon}^H$ can be expressed as the sum of the outer-products of $l+p$ complex Gaussian random vectors with zero mean and covariance $\mathbf{I}_n$. By the Law of Large Numbers [96], for large $l+p$, it approaches $(l+p)\mathbf{I}$ in

probability. Hence,

$$E(\pi_{11}\pi_{22}^*) \approx \frac{1}{(l+p)^2} E\left[(\boldsymbol{v}_1^H \boldsymbol{v}_1)(\boldsymbol{v}_2^H \boldsymbol{v}_2)\right] = \frac{n^2}{(l+p)^2}. \tag{A–48}$$

Third, in order to calculate $E(|\pi_{12}|^2)$, we use the fact that

$$E(|\pi_{11}|^2) = E(|\pi_{12}|^2) + E(\pi_{11}\pi_{22}^*). \tag{A–49}$$

This equation can be proved as follows. Let us make a transformation from $(\boldsymbol{v}_1, \boldsymbol{v}_2)$ to

$\left(\frac{\sqrt{2}}{2}(\boldsymbol{v}_1 + \boldsymbol{v}_2), \frac{\sqrt{2}}{2}(\boldsymbol{v}_1 - \boldsymbol{v}_2)\right)$. Obviously, $\frac{\sqrt{2}}{2}(\boldsymbol{v}_1 + \boldsymbol{v}_2)$ and $\frac{\sqrt{2}}{2}(\boldsymbol{v}_1 - \boldsymbol{v}_2)$ are two independent standard complex Gaussian random vectors and retain the same statistical properties of $(\boldsymbol{v}_1, \boldsymbol{v}_2)$. Hence, replacing $(\boldsymbol{v}_1, \boldsymbol{v}_2)$ by the new random vectors in $E(|\pi_{11}|^2)$, the expectation will not change, i.e.,

$$\begin{aligned} E(|\pi_{11}|^2) &= \frac{1}{4} E\left\{|(\boldsymbol{v}_1^H + \boldsymbol{v}_2^H)(\boldsymbol{\Psi} + \boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^H)^{-1}(\boldsymbol{v}_1 + \boldsymbol{v}_2)|^2\right\} \\ &= \frac{1}{2} E(|\pi_{11}|^2) + \frac{1}{2} E(|\pi_{12}|^2) + \frac{1}{2} E(\pi_{11}\pi_{22}^*). \end{aligned} \tag{A–50}$$

From (A–50), Equation (A–49) is proved.

Using the facts of (A–47), (A–48) and (A–49) and arranging $E(\pi_{r_1c_1}\pi_{r_2c_2}^*)$ into $E(\mathrm{vec}(\boldsymbol{\Pi})\mathrm{vec}(\boldsymbol{\Pi})^H)$, (2–61) follows immediately.

APPENDIX B
PROOFS FOR THE DIAGONAL GROWTH-CURVE MODEL

B.1 Proof of Lemma 3.1

Let $\mathbf{u}_i$ and $\mathbf{v}_i$ be the ith columns of $\mathbf{U}$ and $\mathbf{V}$, respectively. Let g_i be the ith diagonal element in $\mathbf{G}$. Then, we have

$$\begin{aligned}\mathrm{vec}(\mathbf{U}\mathbf{G}\mathbf{V}^T) &= \mathrm{vec}\Big[\sum_{i=1}^{p}(g_i\mathbf{u}_i\mathbf{v}_i^T)\Big] = \sum_{i=1}^{p}\big[\mathrm{vec}(\mathbf{u}_i\, g_i\, \mathbf{v}_i^T)\big] \\ &= \sum_{i=1}^{p}\big[(\mathbf{v}_i \otimes \mathbf{u}_i)\, g_i\big] = (\mathbf{V} \boxdot \mathbf{U})\mathbf{g},\end{aligned} \tag{B–1}$$

where we have used the fact that $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\mathrm{vec}(\mathbf{B})$ [74].

B.2 Proof of Lemma 3.2

Let $\mathbf{u}_i^T$ and $\mathbf{v}_i$ be the ith row and the ith column of $\mathbf{U}$ and $\mathbf{V}$, respectively. Then, the ith diagonal element of $\mathbf{UGV}$ is

$$[\mathbf{UGV}]_{i,i} = \mathbf{u}_i^T\mathbf{G}\mathbf{v}_i = (\mathbf{v}_i^T \otimes \mathbf{u}_i^T)\mathrm{vec}(\mathbf{G}), \tag{B–2}$$

where we have used the fact that $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\mathrm{vec}(\mathbf{B})$ [74]. Arranging (B–2) into a column vector yields Equation (3–11).

To prove (3–12), let $u_{i,j}$ and $v_{i,j}$ be the $(i,\, j)$th elements of $\mathbf{U}$ and $\mathbf{V}$, respectively. Let g_j be the jth diagonal element in $\mathbf{G}$. Then

$$[\mathbf{UGV}]_{i,i} = \mathbf{u}_i^T\mathbf{G}\mathbf{v}_i = \sum_{j=1}^{p}(u_{i,j}v_{j,i})g_j = (\mathbf{u}_i^T \odot \mathbf{v}_i^T)\mathrm{vecd}(\mathbf{G}). \tag{B–3}$$

Arranging (B–3) into a column vector yields Equation (3–12).

B.3 Generalized Least-Squares (GLS) Interpretation of AML in (3–18)

In this appendix, we give an interpretation of the AML estimator from a generalized least-squares (GLS) point of view [16].

By Lemma 1, we know that

$$\mathrm{vec}(\mathbf{ABS}) = (\mathbf{S}^T \boxdot \mathbf{A})\boldsymbol{\beta}. \tag{B–4}$$

Then, the DGC model in (3–1) can be rewritten as

$$\mathrm{vec}(\mathbf{X}) = (\mathbf{S}^T \boxdot \mathbf{A})\boldsymbol{\beta} + \mathrm{vec}(\mathbf{Z}), \tag{B–5}$$

where $\mathrm{vec}(\mathbf{Z})$ is a Gaussian random vector with zero-mean and covariance matrix $(\mathbf{I} \otimes \mathbf{Q})$. From (B–5), the GLS estimator can be readily obtained as

$$\hat{\boldsymbol{\beta}}_{\text{GLS}} = [(\mathbf{S}^T \boxdot \mathbf{A})^H(\mathbf{I} \otimes \mathbf{Q})^{-1}(\mathbf{S}^T \boxdot \mathbf{A})]^{-1}(\mathbf{S}^T \boxdot \mathbf{A})^H(\mathbf{I} \otimes \mathbf{Q})^{-1}\mathrm{vec}(\mathbf{X}). \tag{B–6}$$

Furthermore, we have

$$\begin{aligned}
&(\mathbf{S}^T \boxdot \mathbf{A})^H(\mathbf{I} \otimes \mathbf{Q})^{-1}(\mathbf{S}^T \boxdot \mathbf{A}) \\
=&\left[(\mathbf{S}^T \boxdot \mathbf{A})^H(\mathbf{I} \otimes \mathbf{Q}^{-\frac{1}{2}})\right]\left[(\mathbf{I} \otimes \mathbf{Q}^{-\frac{1}{2}})(\mathbf{S}^T \boxdot \mathbf{A})\right] \\
=&[\mathbf{S}^T \boxdot (\mathbf{Q}^{-\frac{1}{2}}\mathbf{A})]^H[\mathbf{S}^T \boxdot (\mathbf{Q}^{-\frac{1}{2}}\mathbf{A})] \\
=&(\mathbf{SS}^H)^T \odot (\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{A}),
\end{aligned} \tag{B–7}$$

where we have used the fact that $(\mathbf{U} \otimes \mathbf{V})(\mathbf{G} \boxdot \mathbf{H}) = (\mathbf{UG}) \boxdot (\mathbf{VH})$ (see, Lemma A1 in [78]) and Lemma 3 in Section II. Also, by Lemma A1 in [78] and Lemma 2 we have

$$\begin{aligned}
&(\mathbf{S}^T \boxdot \mathbf{A})^H(\mathbf{I} \otimes \mathbf{Q})^{-1}\mathrm{vec}(\mathbf{X}) \\
=&[\mathbf{S}^T \boxdot (\mathbf{Q}^{-1}\mathbf{A})]^H\mathrm{vec}(\mathbf{X}) \\
=&\mathrm{vecd}(\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{XS}^H).
\end{aligned} \tag{B–8}$$

Substituting (B–7) and (B–8) into (B–6) yields

$$\hat{\boldsymbol{\beta}}_{\text{GLS}} = [(\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{A}) \odot (\mathbf{SS}^H)^T]^{-1}\mathrm{vecd}(\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{XS}^H). \tag{B–9}$$

Note that in (B–9) the covariance matrix $\mathbf{Q}$ is unknown. However, we have known that $\mathbf{T}$ is a consistent estimate of $\mathbf{Q}$ (to within a multiplicative constant). Replacing $\mathbf{Q}$ by $\mathbf{T}$ in (B–9) yields the AML estimator in (3–18).

B.4 Cramér-Rao Bound for the DGC model

Note that (B–5) is a linear statistical model with unknown noise covariance matrix $\mathbf{I} \otimes \mathbf{Q}$. It can be easily verified that the Fisher information matrix for this model is a block diagonal matrix with respect to $\boldsymbol{\beta}$ and $\mathbf{Q}$. Hence, the unknowns in $\mathbf{Q}$ do not affect the CRB of $\boldsymbol{\beta}$. Therefore, the CRB of $\boldsymbol{\beta}$ can be readily obtained as [73]

$$\mathrm{CRB}(\boldsymbol{\beta}) = (\mathbf{S}^T \boxdot \mathbf{A})^H (\mathbf{I} \otimes \mathbf{Q}^{-1})(\mathbf{S}^T \boxdot \mathbf{A}). \tag{B–10}$$

Then by (B–7) the CRB in (3–24) follows immediately.

APPENDIX C
CRAMÉR-RAO BOUND FOR THE MIMO RADAR

Consider a MIMO radar system with K targets. Then the received signal can be written as

$$\mathbf{X} = \sum_{k=1}^{K} \mathbf{A}(\theta_k)\mathbf{B}_{\theta_k}\mathbf{V}^T(\theta_k)\mathbf{\Phi} + \mathbf{Z}. \tag{C–1}$$

Let

$$\boldsymbol{\theta} = [\theta_1 \ \cdots \ \theta_K]^T, \tag{C–2}$$

$$\boldsymbol{\beta} = [\mathrm{vec}^T(\mathbf{B}_{\theta_1}) \ \cdots \ \mathrm{vec}^T(\mathbf{B}_{\theta_K})]^T, \tag{C–3}$$

and

$$\boldsymbol{\beta}_R = \mathrm{Re}(\boldsymbol{\beta}) \quad \boldsymbol{\beta}_I = \mathrm{Im}(\boldsymbol{\beta}), \tag{C–4}$$

where $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ denote the real and imaginary parts, respectively. Assume that the columns of $\mathbf{Z}$ are i.i.d. circularly symmetric complex Gaussian random vectors with zero-mean and an unknown covariance matrix $\mathbf{Q}$.

Using the same argument as in Appendix A , we know that the unknowns in $\mathbf{Q}$ will not affect the CRBs of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. Hence, we need only to calculate the following Fisher information matrix with respect to $\boldsymbol{\theta}$, $\boldsymbol{\beta}_R$ and $\boldsymbol{\beta}_I$, i.e.,

$$\mathrm{FIM} = \begin{bmatrix} \mathbf{F}(\boldsymbol{\theta},\boldsymbol{\theta}) & \mathbf{F}(\boldsymbol{\theta},\boldsymbol{\beta}_R) & \mathbf{F}(\boldsymbol{\theta},\boldsymbol{\beta}_I) \\ \mathbf{F}(\boldsymbol{\beta}_R,\boldsymbol{\theta}) & \mathbf{F}(\boldsymbol{\beta}_R,\boldsymbol{\beta}_R) & \mathbf{F}(\boldsymbol{\beta}_R,\boldsymbol{\beta}_I) \\ \mathbf{F}(\boldsymbol{\beta}_I,\boldsymbol{\theta}) & \mathbf{F}(\boldsymbol{\beta}_I,\boldsymbol{\beta}_R) & \mathbf{F}(\boldsymbol{\beta}_I,\boldsymbol{\beta}_I) \end{bmatrix}, \tag{C–5}$$

where $\mathbf{F}(\boldsymbol{\alpha},\boldsymbol{\beta})$ denotes the Fisher information matrix with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

Note that

$$F(\theta_i,\theta_j) = 2\,\mathrm{Re}\,\mathrm{tr}\left\{\frac{\partial\left[\sum_{k=1}^{K}\mathbf{A}(\theta_k)\mathbf{B}_{\theta_k}\mathbf{V}^T(\theta_k)\mathbf{\Phi}\right]^H}{\partial\theta_i}\right.$$
$$\left.\mathbf{Q}^{-1}\frac{\partial\left[\sum_{k=1}^{K}\mathbf{A}(\theta_k)\mathbf{B}_{\theta_k}\mathbf{V}^T(\theta_k)\mathbf{\Phi}\right]}{\partial\theta_j}\right\}, \tag{C–6}$$

and

$$\frac{\partial\left[\sum_{k=1}^{K}\mathbf{A}(\theta_k)\mathbf{B}_{\theta_k}\mathbf{V}^T(\theta_k)\mathbf{\Phi}\right]}{\partial\theta_i} = \dot{\mathbf{A}}(\theta_i)\mathbf{B}_{\theta_k}\mathbf{V}^T(\theta_i)\mathbf{\Phi} + \mathbf{A}(\theta_i)\mathbf{B}_{\theta_k}\dot{\mathbf{V}}^T(\theta_i)\mathbf{\Phi}, \tag{C–7}$$

where

$$\dot{\mathbf{A}}(\theta_k) = \frac{\partial\mathbf{A}(\theta_k)}{\partial\theta_k} \quad \text{and} \quad \dot{\mathbf{V}}(\theta_k) = \frac{\partial\mathbf{V}(\theta_k)}{\partial\theta_k}. \tag{C–8}$$

Inserting (C–7) into (C–6) and after some matrix manipulations, we obtain

$$\begin{aligned}
F(\theta_i,\theta_j) &= 2L\,\mathrm{Re}\,\mathrm{tr}\left\{[\dot{\mathbf{A}}^H(\theta_i)\mathbf{Q}^{-1}\dot{\mathbf{A}}(\theta_j)]\,[\mathbf{B}_{\theta_j}\mathbf{V}^T(\theta_j)\mathbf{R}_{\Phi\Phi}\mathbf{V}^*(\theta_i)\mathbf{B}_{\theta_i}^H]\right\}\\
&+ 2L\,\mathrm{Re}\,\mathrm{tr}\left\{[\dot{\mathbf{A}}^H(\theta_i)\mathbf{Q}^{-1}\mathbf{A}(\theta_j)]\,[\mathbf{B}_{\theta_j}\dot{\mathbf{V}}^T(\theta_j)\mathbf{R}_{\Phi\Phi}\mathbf{V}^*(\theta_i)\mathbf{B}_{\theta_i}^H]\right\}\\
&+ 2L\,\mathrm{Re}\,\mathrm{tr}\left\{[\mathbf{A}^H(\theta_i)\mathbf{Q}^{-1}\dot{\mathbf{A}}(\theta_j)]\,[\mathbf{B}_{\theta_j}\mathbf{V}^T(\theta_j)\mathbf{R}_{\Phi\Phi}\dot{\mathbf{V}}^*(\theta_i)\mathbf{B}_{\theta_i}^H]\right\}\\
&+ 2L\,\mathrm{Re}\,\mathrm{tr}\left\{[\mathbf{A}^H(\theta_i)\mathbf{Q}^{-1}\mathbf{A}(\theta_j)]\,[\mathbf{B}_{\theta_j}\dot{\mathbf{V}}^T(\theta_j)\mathbf{R}_{\Phi\Phi}\dot{\mathbf{V}}^*(\theta_i)\mathbf{B}_{\theta_i}^H]\right\},
\end{aligned} \tag{C–9}$$

where $\mathbf{R}_{\Phi\Phi} = \frac{1}{L}\mathbf{\Phi}\mathbf{\Phi}^H$ is the covariance matrix of the transmitted waveforms. Hence,

$$\mathbf{F}(\boldsymbol{\theta},\boldsymbol{\theta}) = 2\,\mathrm{Re}(\mathbf{F}_{\theta\theta}), \tag{C–10}$$

where $\mathbf{F}_{\theta\theta}$ is a $K\times K$ matrix with its (i,j) element being

$$\begin{aligned}
[\mathbf{F}_{\theta\theta}]_{ij} &= L\,\mathrm{tr}\left\{[\dot{\mathbf{A}}(\theta_i)^H\mathbf{Q}^{-1}\dot{\mathbf{A}}(\theta_j)]\,[\mathbf{B}_{\theta_j}\mathbf{V}^T(\theta_j)\mathbf{R}_{\Phi\Phi}\mathbf{V}^*(\theta_i)\mathbf{B}_{\theta_i}^H]\right\}\\
&+ L\,\mathrm{tr}\left\{[\dot{\mathbf{A}}^H(\theta_i)\mathbf{Q}^{-1}\mathbf{A}(\theta_j)]\,[\mathbf{B}_{\theta_j}\dot{\mathbf{V}}^T(\theta_j)\mathbf{R}_{\Phi\Phi}\mathbf{V}^T(\theta_i)\mathbf{B}_{\theta_i}^H]\right\}\\
&+ L\,\mathrm{tr}\left\{[\mathbf{A}^H(\theta_i)\mathbf{Q}^{-1}\dot{\mathbf{A}}(\theta_j)]\,[\mathbf{B}_{\theta_j}\mathbf{V}^T(\theta_j)\mathbf{R}_{\Phi\Phi}\dot{\mathbf{V}}^*(\theta_i)\mathbf{B}_{\theta_i}^H]\right\}\\
&+ L\,\mathrm{Re}\left\{[\mathbf{A}^H(\theta_i)\mathbf{Q}^{-1}\mathbf{A}(\theta_j)]\,[\mathbf{B}_{\theta_j}\dot{\mathbf{V}}^T(\theta_j)\mathbf{R}_{\Phi\Phi}\dot{\mathbf{V}}^*(\theta_i)\mathbf{B}_{\theta_i}^H]\right\}.
\end{aligned} \tag{C–11}$$

Similarly, we have

$$\mathbf{F}^T(\boldsymbol{\theta}, \boldsymbol{\beta}_R) = \mathbf{F}(\boldsymbol{\beta}_R, \boldsymbol{\theta}) = 2\,\mathrm{Re}(\mathbf{F}_{\beta\theta}), \tag{C–12}$$

$$-\,\mathbf{F}^T(\boldsymbol{\theta}, \boldsymbol{\beta}_I) = \mathbf{F}(\boldsymbol{\beta}_I, \boldsymbol{\theta}) = 2\,\mathrm{Im}(\mathbf{F}_{\beta\theta}), \tag{C–13}$$

$$\mathbf{F}(\boldsymbol{\beta}_R, \boldsymbol{\beta}_R) = \mathbf{F}(\boldsymbol{\beta}_I, \boldsymbol{\beta}_I) = 2\,\mathrm{Re}(\mathbf{F}_{\beta\beta}), \tag{C–14}$$

and

$$-\,\mathbf{F}(\boldsymbol{\beta}_R, \boldsymbol{\beta}_I) = \mathbf{F}(\boldsymbol{\beta}_I, \boldsymbol{\beta}_R) = 2\,\mathrm{Im}(\mathbf{F}_{\beta\beta}), \tag{C–15}$$

where $\mathbf{F}_{\theta\beta}$ and $\mathbf{F}_{\beta\beta}$ are both partitioned matrices formed by the following submatrices, respectively,

$$\begin{aligned}[\mathbf{F}_{\beta\theta}]_{ij} =& L\,\mathrm{vec}\left\{[\mathbf{A}^H(\theta_i)\mathbf{Q}^{-1}\dot{\mathbf{A}}(\theta_j)]\;[\mathbf{B}_{\theta_j}\mathbf{V}^T(\theta_j)\mathbf{R}_{\Phi\Phi}\mathbf{V}^*(\theta_i)]\right\} \\ &+ L\,\mathrm{vec}\left\{[\mathbf{A}^H(\theta_i)\mathbf{Q}^{-1}\mathbf{A}(\theta_j)]\;[\mathbf{B}_{\theta_j}\dot{\mathbf{V}}^T(\theta_j)\mathbf{R}_{\Phi\Phi}\mathbf{V}^*(\theta_i)]\right\}\end{aligned} \tag{C–16}$$

and

$$[\mathbf{F}_{\beta\beta}]_{ij} = L[\mathbf{V}^T(\theta_i)\mathbf{R}_{\Phi\Phi}\mathbf{V}^*(\theta_j)] \otimes [\mathbf{A}^H(\theta_i)\mathbf{Q}^{-1}\mathbf{A}(\theta_j)], \tag{C–17}$$

with $i, j = 1,\ 2,\ \cdots,\ K$ and $\otimes$ denoting the Kronecker product.

Substituting Equations (C–9) - (C–15) into (C–6), and after some matrix manipulations, we get

$$\mathrm{CRB}(\boldsymbol{\theta}) = \frac{1}{2}\Big\{\,\mathrm{Re}\left[\mathbf{F}_{\theta\theta} - \mathbf{F}_{\theta\beta}\mathbf{F}_{\beta\beta}^{-1}\mathbf{F}_{\theta\beta}^H\right]\Big\}^{-1}, \tag{C–18}$$

and

$$\mathrm{CRB}(\boldsymbol{\beta}) = \mathbf{F}_{\beta\beta}^{-1} + \mathbf{F}_{\beta\beta}^{-1}\mathbf{F}_{\theta\beta}^H\,\mathrm{CRB}(\boldsymbol{\theta})\,\mathbf{F}_{\theta\beta}\mathbf{F}_{\beta\beta}^{-1}. \tag{C–19}$$

REFERENCES

[1] R. F. Potthoff and S. N. Roy, "A generalized multivariate analysis of variance model useful especially for growth curve problems," *Biometrika*, vol. 51, pp. 313–326, 1964.

[2] C. R. Rao, "The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves," *Biometrika*, vol. 52, pp. 447–458, 1965.

[3] C. R. Rao, "Covariance adjustment and related problems in multivariate analysis," *Multivariate Analysis,* P.R. Krishnaiah, Ed., Academic Press, New York, pp. 87–103, 1966.

[4] C. R. Rao, "Least square theory using an estimated dispersion matrix and its applications to measurement of signals," *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, 1*, pp. 355–372, 1967.

[5] C. G. Khatri, "A note on a MANOVA model applied to problems in growth curve," *Annals of Institute of Statistical Mathematics*, vol. 18, pp. 75–86, 1966.

[6] L. J. Gleser and I. Olkin, "Linear models in multivariate analysis," In *Essays in Probability and Statistics.* R. C. Bose, I. M. Chakravarti, P. C. Mahalanobis, C. R. Rao and K. J. C. Smith (Eds), Chapel Hill, North Carolina: University of North Carolina Press, pp. 267–292, 1970.

[7] S. Geisser, "Bayesian analysis of growth curves," *Sankhya*, vol. A 32, pp. 53–64, 1970.

[8] D. von Rosen, "Maximum likelihood estimates in multivariate linear normal model," *Journal of Multivariate Analysis*, vol. 31, pp. 187–200, 1989.

[9] D. von Rosen, "Moments for a multivariate linear normal model with applications to growth curve model," *Journal of Multivariate Analysis*, vol. 35, pp. 243–259, 1990.

[10] D. von Rosen, "The growth curve model: a review," *Communications in Statistics: Theory and Methods*, vol. 20, no. 9, pp. 2791–2822, 1991.

[11] A. P. Verbyla and W. N. Venables, "An extension of the growth curve model," *Biometrika*, vol. 75, no. 1, pp. 129–138, 1988.

[12] M. S. Srivastava, "Nested growth curve models," *Sankhya*, vol. A 64, pp. 379–408, 2002.

[13] R. H. Jones, *Longitudinal Data with Serial Correlation: A State-Space Approach*. London: Chapman and Hall, 1993.

[14] N. M. Laird, N. Lange, and D. Stram, "Maximum likelihood computations with repeated measures: applications of the EM algorithm," *Journal of the American Statistical Association*, vol. 82, pp. 97–105, 1987.

[15] M. J. Crowder and D. J. Hand, *Analysis of Repeated Measures*. London: Chapman and Hall, 1990.

[16] C. Rao, *Linear Statistical Inference and Its Applications*. Wiley, New York, 1973.

[17] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis, second edition*. John Wiley and Sons, Inc., 1984.

[18] J.-X. Pan and K.-T. Fang, *Growth Curve Models and Statistical Diagnostics*. New York: Springer, Inc., 2002.

[19] K. A. Burgess and B. D. Van Veen, "Subspace-based adaptive generalized likelihood ratio detection," *IEEE Transactions on Signal Processing*, vol. 44, pp. 912–927, April 1996.

[20] A. Dogandžić and A. Nehorai, "Estimating range, velocity, and direction with a radar array," *in Proc. Int. Conf. Acoust., Speech., Signal processing, Phoenix, AZ*, pp. 2772–2776, Mar. 1999.

[21] A. Dogandžić and A. Nehorai, "Generalized multivariate analysis of variance," *IEEE Signal Processing Magazine*, pp. 39–54, September 2003.

[22] L. Xu and J. Li, "Performance analysis of multivariate complex amplitude estimators," *IEEE Transactions on Signal Processing*, vol. 53, pp. 3162–3171, August 2005.

[23] J. Li and P. Stoica, "An adaptive filtering approach to spectral estimation and SAR imaging," *IEEE Transactions on Signal Processing*, vol. 44, pp. 1469–1484, June 1996.

[24] P. Stoica, H. Li, and J. Li, "Amplitude estimation for sinusoidal signals with application to system identification," *IEEE Transactions on Signal Processing*, vol. 48, pp. 338–352, February 2000.

[25] Y. Jiang, P. Stoica, and J. Li, "Array signal processing in the known waveform and steering vector case," *IEEE Transactions on Signal Processing*, vol. 52, pp. 23–35, January 2004.

[26] J. Li and R. T. Compton, Jr., "Maximum likelihood angle estimation for signals with known waveforms," *IEEE Transactions on Signal Processing*, vol. ASSP-41, pp. 2850–2862, September 1993.

[27] J. Li, B. Halder, P. Stoica, and M. Viberg, "Computationally efficient angle estimation for signals with known waveforms," *IEEE Transactions on Signal Processing*, vol. 43, pp. 2154–2163, September 1995.

[28] A. Zeira and B. Friedlander, "Direction of arrival estimation using parametric signal models," *IEEE Transactions on Signal Processing*, vol. 44, pp. 339–350, February 1996.

[29] M. Cedervall and R. L. Moses, "Efficient maximum likelihood DOA estimation for signals with known waveforms in the presence of multipath," *IEEE Transactions on Signal Processing*, vol. 45, pp. 808–811, March 1997.

[30] M. Wax and A. Leshem, "Joint estimation of time delays and directions of arrival of multiple reflections of a known signal," *IEEE Transactions on Signal Processing*, pp. 2477–2484, October 1997.

[31] A. L. Swindlehurst, "Time delay and spatial signature estimation using known asynchronous signals," *IEEE Transactions on Signal Processing*, vol. 46, pp. 449–462, February 1998.

[32] A. L. Swindlehurst and P. Stoica, "Maximum likelihood methods in radar array signal processing," *Proceedings of the IEEE*, vol. 86, pp. 421–441, February 1998.

[33] A. Jakobsson, A. Swindlehurst, and P. Stoica, "Subspace-based estimation of time delays and doppler shifts," *IEEE Transactions on Signal Processing*, vol. 46, no. 9, pp. 2472–2483, September 1998.

[34] H. Li, G. Liu, and J. Li, "A novel angle estimator for signals with known waveforms," *Electronics Letters*, vol. 35, pp. 1992–1994, November 1999.

[35] U. Spagnolini, "An iterative quadratic method for high resolution delay estimation with known waveform," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, pp. 1134–1137, March 2000.

[36] L. Xu, P. Stoica, and J. Li, "A diagonal growth curve model and some signal processing applications," to appear in *IEEE Transactions on Signal Processing*, 2006.

[37] L. Xu, P. Stoica, and J. Li, "A block diagonal growth curve model," to appear in *Digital Signal Processing*, 2006.

[38] J. M. Colin, "Phased array radars in france: Present and future," *Proc. of the IEEE Int. Symp. on Phased Array Systems and Technology*, pp. 458–562, October 1996.

[39] D. Rabideau, "Ubiquitous MIMO digital array radar," *Proc. of 37th Asilomar Conf. on Signals, Systems, and Computers*, pp. 1057–1064, November 2003.

[40] E. Fishler, A. Haimovich, R. Blum, D. Chizhik, L. Cimini, and R. Valenzuela, "MIMO radar: an idea whose time has come," *Proceedings of the IEEE Radar Conference*, pp. 71–78, April 2004.

[41] E. Fishler, A. Haimovich, R. Blum, L. Cimini, D. Chizhik, and R. Valenzuela, "Performance of MIMO radar systems: advantages of angular diversity," *Proceedings of the 38th Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 305–309, Nov. 2004.

[42] E. Fishler, A. Haimovich, R. Blum, L. Cimini, D. Chizhik, and R. Valenzuela, "Spatial diversity in radars - models and detection performance," *IEEE Transactions on Signal Processing*, vol. 54, pp. 823–838, March 2006.

[43] N. Lehmann, E. Fishler, A. M. Haimovich, R. S. Blum, D. Chizhik, L. Cimini, and R. Valenzuela, "Evaluation of transmit diversity in MIMO-radar direction finding," Submitted to *IEEE Transactions on Signal Processing*, 2005.

[44] D. R. Fuhrmann and G. S. Antonio, "Transmit beamforming for MIMO radar systems using partial signal correlation," *Proceedings of the 38th Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 295–299, Nov. 2004.

[45] J. Li, P. Stoica, and Y. Xie, "On probing signal design for MIMO radar," *Asilomar (*invited), October 2006.

[46] A. S. Fletcher and F. C. Robey, "Performance bounds for adaptive coherence of sparse array radar," *Proc. of the 11th Conf. on Adaptive Sensors Array Processing*, March 2003.

[47] D. W. Bliss and K. W. Forsythe, "Multiple-input multiple-output (MIMO) radar and imaging: degrees of freedom and resolution," *Proceedings of the 37th Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 54–59, Nov. 2003.

[48] F. C. Robey, S. Coutts, D. D. Weikle, J. C. McHarg, and K. Cuomo, "MIMO radar theory and exprimental results," *Proceedings of the 38th Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 300–304, Nov. 2004.

[49] K. W. Forsythe, D. W. Bliss, and G. S. Fawcett, "Multiple-input multiple-output (MIMO) radar: performance issues," *Proceedings of the 38th Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 310–314, Nov. 2004.

[50] L. B. White and P. S. Ray, "Signal design for MIMO diversity systems," *Proceedings of the 38th Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 973–977, November 2004.

[51] I. Bekkerman and J. Tabrikian, "Spatially coded signal model for active arrays," *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004.*, vol. 2, pp. 209–212, March 2004.

[52] J. Tabrikian and I. Bekkerman, "Transmission diversity smoothing for multi-target localization," *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 4, pp. 1041–1044, March 2005.

[53] J. Tabrikian, "Barankin bounds for target localization by mimo radars," *Fourth IEEE Workshop on Sensor Array and Multi-channel Processing, SAM 2006*, July 2006.

[54] I. Bekkerman and J. Tabrikian, "Space-time coding for active arrays," to appear in *IEEE Transactions on Signal Processing*, 2006.

[55] L. Xu, J. Li, and P. Stoica, "Radar imaging via adaptive MIMO techniques," *Proceedings of the 14th European Signal Processing Conference* (invited), *Florence, Italy*, 2006.

[56] L. Xu, J. Li, and P. Stoica, "Adaptive techniques for MIMO radar," *Proceedings of the Fourth IEEE Sensor Array and Multichannel Signal Processing Workshop* (invited), Waltham, Massachusetts, USA, 2006.

[57] L. Xu, J. Li, and P. Stoica, "Adaptive MIMO radar," Submitted to *IEEE Transactions on Aerospace and Electronic Systems*, 2006.

[58] J. Li, P. Stoica, L. Xu, and W. Roberts, "On parameter identifiability of MIMO radar," to appear in *IEEE Signal Processing Letters*, 2006.

[59] M. I. Skolnik, *Introduction to Radar System, 3rd ed.* New York, NY: McGraw-Hill, 2002.

[60] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, pp. 1408–1418, August 1969.

[61] L. Xu and J. Li, "Iterative generalized likelihood ratio test (iGLRT) for MIMO radar," submitted to *IEEE Transactions on Signal Processing*, 2006.

[62] M. Siotani, T. Hayakawa, and Y. Fujikoshi, *Modern Multivariate Statistical Analysis: A Graduate Course and Handbook.* Columbus, Ohio: American Sciences Press, Inc., 1985.

[63] R. J. Muirhead, *Aspects of Multivariate Statistical Theory.* New York, NY: John Wiley & Sons, Inc., 1982.

[64] E. J. Kelly, "An adaptive detection algorithm," *IEEE Transactions on Aerospace and Electronics Systems*, vol. 22, pp. 115–127, March 1986.

[65] H. Wang and L. Cai, "On adaptive multiband signal detection with GLR algorithm," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 27, pp. 225–233, Mar. 1991.

[66] P. Stoica and T. Sundin, "Nonparametric NMR spectroscopy," *J. Magn. Reson.*, vol. 152, pp. 57–69, Sept. 2001.

[67] A. Dogandžić and A. Nehorai, "Space-time fading channel estimation and symbol detection in unknown spatially correlated noise," *IEEE Transactions on Signal Processing*, pp. 457–474, Mar 2002.

[68] E. G. Larsson, J. Liu, and J. Li, "Demodulation of space-time codes in the presence of interference," *IEE Electronics Letters*, vol. 37, pp. 697–698, May 2001.

[69] E. G. Larsson, P. Stoica, and J. Li, "On maximum-likelihood detection and decoding for space-time coding systems," *IEEE Transactions on Signal Processing*, vol. 50, pp. 937–944, April 2002.

[70] E. G. Larsson, P. Stoica, and J. Li, "Orthogonal space-time block codes: Maximum-likelihood detection for unknown channels and unstructured interference," *IEEE Transactions on Signal Processing*, vol. 51, pp. 362–372, Feb. 2003.

[71] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis.* Englewood Cliffs, NJ: Prentice-Hall, 1997.

[72] H. L. Van Trees, *Optimum Array Processing. Part IV of Detection, Estimation, and Modulation Theory.* New York, NY: John Wiley and Sons, Inc., 2002.

[73] T. Söderström and P. Stoica, *System Identification.* London, U.K.: Prentice-Hall International, 1989.

[74] D. A. Harville, *Matrix Algebra from a Statistician's Perspective.* New York, NY: Springer, Inc., 1997.

[75] H. H. Andersen, M. Højbjerre, D. Sørensen, and P. S. Eriksen, *Linear and Graphical Models for Multivariate Complex Normal Distribution.* New York, NY: Springer-Verlag, Inc., 1995.

[76] C. G. Khatri and C. R. Rao, "Effects of estimated noise covariance matrix optimal signal detection," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, pp. 671–679, May 1987.

[77] C. G. Khatri and C. R. Rao, "Solutions to some functional equations and their applications to characterization of probability distributions," *Sankhya*, vol. 30, pp. 167–180, 1968.

[78] C. R. Rao, "Estimation of heteroscedastic variances in linear models," *Journal of the American Statistical Association*, vol. 65, pp. 161–172, Mar. 1970.

[79] S. M. Kay, *Modern Spectral Estimation: Theory and Application.* Englewood Cliffs, N.J.: Prentice-Hall, 1988.

[80] S. Liu, "Matrix results on the Khatri-Rao and Tracy-Singh products," *Linear Algebra and its Applications*, vol. 289, pp. 267–277, 1999.

[81] J. G. Proakis, *Digital Communications.* McGraw-Hill Inc., fourth edition, 2001.

[82] P. Stoica and K. C. Sharman, "Novel eigenanalysis method for direction estimation," *IEE Proceedings, Pt. F*, vol. 137, pp. 19–26, February 1990.

[83] P. Stoica and K. C. Sharman, "Maximum likelihood methods for direction-of-arrival estimation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-38, pp. 1132–1143, July 1990.

[84] P. Stoica and Y. Selén, "Cyclic minimizers, majorization techniques, and expectation-maximization algorithm: A refresher," *IEEE Signal Processing Magazine*, pp. 112–114, January 2004.

[85] P. Stoica and R. L. Moses, *Spectral Analysis of Signals.* Englewood Cliffs, NJ: Prentice-Hall, 2005.

[86] E. J. Kelly, "Finite-sum expressions for signal detection probabilities," Technical Report 566, Lincoln Laboratory, M.I.T, May 1981.

[87] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Transactions on Information Theory*, vol. 49, pp. 1073–1096, May 2003.

[88] J. R. Guerci and E. J. Baranoski, "Knowledge-aided adaptive radar at darpa," *IEEE Signal Processing Magazine*, pp. 41–50, January 2006.

[89] G. T. Capraro, A. Farina, H. Griffiths, and M. C. Wicks, "Knowledge-based radar signal and data processing," *IEEE Signal Processing Magazine*, pp. 18–29, January 2006.

[90] S. Miranda, C. Baker, K. Woodbridge, and H. Griffiths, "Knowledge-based resource management for multifunction radar," *IEEE Signal Processing Magazine*, pp. 65–76, January 2006.

[91] C. A. Stutt and L. J. Spafford, "A 'best' mismatched filter response for radar clutter discrimination," *IEEE Transactions on Information Theory*, vol. 14, pp. 280–287, March 1968.

[92] M. R. Bell, "Information theory and radar waveform design," *IEEE Transactions on Information Theory*, vol. 39, pp. 1578–1597, September 1993.

[93] D. A. Garren, M. K. Osborn, A. C. Odom, J. S. Goldstein, S. U. Pillai, and J. Guerci, "Enhanced target detection and identification via optimised radar transmission pulse shape," *IEE Proceedings - Radar, Sonar, and Navigation*, vol. 148, pp. 130–138, June 2001.

[94] J. Li, J. R. Guerci, and L. Xu, "Signal waveform's optimal-under-restriction design for active sensing," to appear in *IEEE Signal Processing Letters*, 2006.

[95] N. R. Goodman, "Statistical analysis based on a certain multi-variate complex Gaussian distribution," *Ann. Math. Stat.*, vol. 34, pp. 152–177, March, 1963.

[96] G. Casella and R. L. Berger, *Statistical Inference (2nd Edition).* Thomson Learning, Inc., 2002.

BIOGRAPHICAL SKETCH

Luzhou Xu received the B.Eng. and M.S. degrees in electrical engineering from Zhejiang University, Hangzhou, China, in 1996 and 1999, respectively, and he is expected to receive the Ph.D degree in electrical engineering from University of Florida, Gainesville, in 2006. From 1999 to 2001, he was with Zhongxing Research and Development Institute, Shanghai, where he was involved in the system and algorithm design of mobile communications. From 2001 to 2003, he was with Philips Research. His research interests include statistical and array signal processing and their applications.