

# Array Signal Processing in the Known Waveform and Steering Vector Case

Yi Jiang, *Student Member, IEEE*, Petre Stoica, *Fellow, IEEE*, and Jian Li, *Senior Member, IEEE*

**Abstract**—The amplitude estimation of a signal that is known only up to an unknown scaling factor, with interference and noise present, is of interest in several applications, including using the emerging quadrupole resonance (QR) technology for explosive detection. In such applications, a sensor array is often deployed for interference suppression. This paper considers the complex amplitude estimation of a known waveform signal whose array response is also known *a priori*. Two approaches, viz., the Capon and the maximum likelihood (ML) methods, are considered for the signal amplitude estimation in the presence of temporally white but spatially colored interference and noise. We derive closed-form expressions for the expected values and mean-squared errors (MSEs) of the two estimators. A comparative study shows that the ML estimate is unbiased, whereas the Capon estimate is biased downwards for finite data sample lengths. We show that both methods are asymptotically statistically efficient when the number of data samples is large but not when the signal-to-noise ratio (SNR) is high. Furthermore, we consider a more general scenario where the interference and noise are both spatially and temporally correlated. We model the interference and noise vector as a multichannel autoregressive (AR) random process. An alternating least squares (ALS) method for parameter estimation is presented. We show that in most cases, the ALS method is superior to the model-mismatched ML ( $M^3L$ ) method, which ignores the temporal correlation of the interference and noise.

**Index Terms**—Asymptotic analysis, capon, interference suppression, maximum likelihood, multichannel autoregressive random process, parameter estimation, quadrupole resonance.

## I. INTRODUCTION

**E**STIMATING the signal parameters in the presence of interference and noise via array processing is often encountered in practical applications (see, e.g., [1], [2], and the references therein). It is well known that temporal information about the signal can be utilized to effectively suppress the interference and noise and, hence, to significantly improve the estimation accuracy. For example, [3] and [4] have studied the estimation of directions of arrival (DOAs) of signals with known waveforms and showed that significant improvements in accuracy, interference suppression capability, and spatial resolution can be obtained. Through Cramér–Rao bound (CRB) analysis, [5] has studied the DOA estimation of a parametric signal and

Manuscript received August 30, 2002; revised March 19, 2003. This work was supported in part by the National Science Foundation under Grant CCR-0104887 and the Swedish Science Council. The associate editor coordinating the review of this paper and approving it for publication was Dr. Alex B. Gershman.

Y. Jiang and J. Li are with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611 USA (e-mail: yjiang@dsp.ufl.edu; li@dsp.ufl.edu).

P. Stoica is with the Department of Systems and Control, Uppsala University, Uppsala, Sweden (e-mail: ps@syscon.uu.se).

Digital Object Identifier 10.1109/TSP.2003.820074

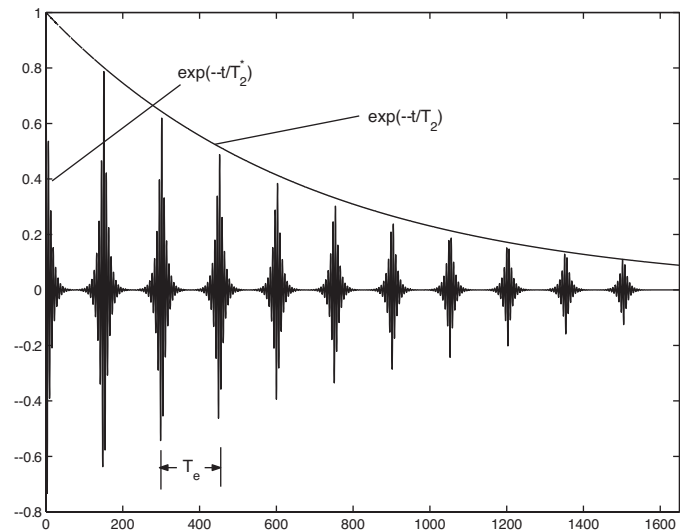


Fig. 1. QR response of  $^{14}\text{N}$  in the TNT.

shown that exploiting temporal information about the signal can improve the DOA estimation. Additional work dealing with the estimation of the parameters of known waveform signals can be found in [6]–[10].

All of the aforementioned papers concentrate on the estimation of signal parameters by exploiting the temporal information only. Exploiting both the temporal and spatial information on the signal for interference suppression and signal parameter estimation has yet to be fully investigated in the previous literature, yet it is of practical importance in some applications, such as using the emerging quadrupole resonance (QR) technology for explosive detection [11], [12]. With reference to the QR application, the  $^{14}\text{N}$  in the TNT, when stimulated by a sequence of pulses, gives a characteristic response specific to the TNT consisting of a sequence of echoes. We refer to this response as the QR signal. Each echo of the QR signal is a back-to-back exponentially damped sinusoid separated by an interval  $T_e$  from the adjacent echoes and with a damping rate determined by a parameter  $T_2^*$  (see Fig. 1). The echo amplitude also damps exponentially as  $e^{-t/T_2}$ . The QR signal frequency  $T_2$  and  $T_2^*$  are all known quite accurately *a priori*, and the precise echo timing  $T_e$  is also available in practice. Hence, the QR signal is *a priori* known to within a multiplicative constant. The major challenge of using the QR technology for landmine detection is that the QR signal frequency falls within the frequency band for the AM/FM radio signals, and the QR signal frequency cannot be changed. To suppress the radio interferences, an antenna array can be deployed with one of the sensors receiving the QR signal as well as the interference and noise (we refer to it as the main

antenna), whereas the remaining sensors receive the interference and noise only (we refer to them as the reference antennas). Hence, one of the elements of the array steering vector for the QR signal is equal to one, and the remaining elements are zero. Thus, in this application, both temporal and spatial information are available *a priori*.

Motivated by the QR application, we study herein the problem of amplitude estimation of a signal with known waveform and steering vector since it is a mandatory step for detection. We consider two approaches [the Capon and the maximum likelihood (ML) methods] that utilize both the temporal and spatial information on the signal for amplitude estimation in the presence of temporally white but spatially colored interference and noise. Based on the results in [13] and [14], we derive closed-form expressions for the expected values and mean-squared errors (MSEs) of the two estimators. A comparative study shows that the ML estimate is unbiased, whereas the Capon estimate is biased downwards for finite data sample lengths. We also show that both methods approach the corresponding Cramér–Rao bound (CRB), i.e., they are asymptotically statistically efficient, when the number of data samples is large. However, they are not asymptotically statistically efficient when the signal-to-noise ratio (SNR) is high. Furthermore, we consider a more general scenario where the interference and noise vectors are both spatially and temporally correlated. We model the interference and noise vector as a multichannel autoregressive (AR) random process. An alternating least squares (ALS) method is proposed to tackle the amplitude estimation problem in this situation. We show that in most cases, the ALS method is superior to the model-mismatched ML (M<sup>3</sup>L) method that ignores the temporal correlation of the interference and noise. Finally, numerical examples are presented to illustrate the theoretical properties and demonstrate the practical performance of the estimators.

The remainder of the paper is organized as follows. Section II formulates the problem of interest. The ML and Capon estimators are also given in that section. Section III gives the closed-form expressions of the expected values and MSEs of the ML and Capon estimators and compares their statistical properties. In Section IV, we propose an alternating least squares (ALS) method to deal with the more general scenario of both spatially and temporally correlated interference and noise. Our theoretical findings are verified via numerical examples in Section V. Finally, Section VI gives our conclusions.

## II. PRELIMINARIES

We consider the problem of estimating the complex amplitude of a known waveform signal in the presence of interference and noise:

$$\mathbf{x}_l = \mathbf{a}\beta s_l + \mathbf{e}_l \quad (1)$$

where  $\mathbf{x}_l \in \mathcal{C}^{M \times 1}$ ,  $l = 1, 2, \dots, L$  denotes the  $l$ th array output vector (where  $M$  is the number of sensors and  $L$  is the number of snapshots), the array steering vector  $\mathbf{a} \in \mathcal{C}^{M \times 1}$  of the signal of interest is known, and  $\beta$  is the unknown complex amplitude of the signal whose temporal waveform  $\{s_l\}_{l=1}^L$  is known.

First, we model the interference and noise term  $\mathbf{e}_l \in \mathcal{C}^{M \times 1}$  as a zero-mean temporally white but spatially colored circularly symmetric complex Gaussian random process with an unknown and arbitrary, but fixed, spatial covariance matrix  $\mathbf{Q}$ . We define the SNR by lumping the interference and noise together in a “generalized noise” term:

$$\text{SNR} = \frac{M|\beta|^2 P_s}{\text{tr}(\mathbf{Q})} \quad (2)$$

where  $\text{tr}(\cdot)$  denotes the trace of a matrix, and

$$P_s = \frac{1}{L} \sum_{l=1}^L |s_l|^2 \quad (3)$$

is the average power of the known waveform. We will discuss the extension to the case where the interference and noise term is both temporally and spatially correlated in Section IV.

The Capon and ML estimators are two widely used methods in array processing [2]. The Capon method [also known as the minimum variance distortionless response (MVDR) beamformer] estimates the signal amplitude via [15]

$$\hat{\beta}_{\text{Capon}} = \hat{\mathbf{w}}_{\text{Capon}}^H \frac{\bar{\mathbf{x}}}{P_s} \quad (4)$$

where

$$\hat{\mathbf{w}}_{\text{Capon}} = \arg \min_{\mathbf{w}} \mathbf{w}^H \hat{\mathbf{R}} \mathbf{w} \quad \text{subject to} \quad \mathbf{w}^H \mathbf{a} = 1 \quad (5)$$

with  $(\cdot)^H$  denoting the conjugate transpose

$$\hat{\mathbf{R}} = \frac{1}{L} \sum_{l=1}^L \mathbf{x}_l \mathbf{x}_l^H \quad (6)$$

and

$$\bar{\mathbf{x}} = \frac{1}{L} \sum_{l=1}^L \mathbf{x}_l s_l^*. \quad (7)$$

It is easy to solve (5) and show that [2]

$$\hat{\beta}_{\text{Capon}} = \frac{\mathbf{a}^H \hat{\mathbf{R}}^{-1} \bar{\mathbf{x}}}{P_s \mathbf{a}^H \hat{\mathbf{R}}^{-1} \mathbf{a}}. \quad (8)$$

The ML method estimates the signal amplitude by maximizing the likelihood function of the random vectors  $\{\mathbf{x}_l\}_{l=1}^L$ . We show in Appendix A that

$$\hat{\beta}_{\text{ML}} = \frac{\mathbf{a}^H \mathbf{T}^{-1} \bar{\mathbf{x}}}{P_s \mathbf{a}^H \mathbf{T}^{-1} \mathbf{a}} \quad (9)$$

where

$$\mathbf{T} = \hat{\mathbf{R}} - \frac{\bar{\mathbf{x}} \bar{\mathbf{x}}^H}{P_s}. \quad (10)$$

Note that the only difference between the Capon and ML estimators is that the matrix  $\hat{\mathbf{R}}$  in (8) is replaced by  $\mathbf{T}$  in (9). We will

show in the following sections that this seemingly minor difference in fact leads to significant and interesting performance differences between the two estimators.

### III. PERFORMANCE ANALYSIS OF ML AND CAPON

#### A. Performance Analysis of the ML Estimator

We present below a statistical performance analysis of the ML estimator. We prove that the ML estimator is unbiased. By deriving the MSE of the ML estimate and comparing it with the corresponding CRB, we show that the ML estimator is asymptotically statistically efficient when the number of snapshots is large but, in addition, that this is not the case for high SNR.

1) *Bias Analysis*: The matrix  $\mathbf{T}$  defined in (10) and the vector  $\bar{\mathbf{x}}$  in (7) are both functions of the vectors  $\{\mathbf{x}_l\}_{l=1}^L$ , which might suggest that they are correlated with each other. However, the lemma below somewhat surprisingly shows that they are in fact statistically independent of each other.

*Lemma 1*: Under the assumption made on the data model in (1), the vector  $\bar{\mathbf{x}}$  and the matrix  $\mathbf{T}$  are statistically independent of each other.

*Proof*: See Appendix B.  $\blacksquare$

Utilizing this lemma and the conditional expectation rule, we can easily show that the ML estimator is unbiased, i.e.,

$$\mathcal{E}[\hat{\beta}_{ML}] = \mathcal{E}_{\mathbf{T}}[\mathcal{E}_{\bar{\mathbf{x}}|\mathbf{T}}[\hat{\beta}_{ML}]] = \mathcal{E}_{\mathbf{T}}[\beta] = \beta \quad (11)$$

where  $\mathcal{E}_{\bar{\mathbf{x}}|\mathbf{T}}[\cdot]$  denotes calculating the expected value with respect to  $\bar{\mathbf{x}}$  for a fixed  $\mathbf{T}$ .

Lemma 1 and its proof will also be helpful in the performance analyzes that follow.

2) *Mean-Squared Error Analysis*: Before calculating the MSE of the ML estimate, we first introduce the best possible performance bound for any unbiased estimator of  $\beta$ , i.e., the CRB. Appendix C shows that the CRB has the following compact form:

$$\text{CRB}(\beta) = \frac{1}{LP_s \mathbf{a}^H \mathbf{Q}^{-1} \mathbf{a}}. \quad (12)$$

Note that

$$\begin{aligned} \hat{\beta}_{ML} &= \frac{\mathbf{a}^H \mathbf{T}^{-1} \bar{\mathbf{x}}}{P_s \mathbf{a}^H \mathbf{T}^{-1} \mathbf{a}} \\ &= \frac{\mathbf{a}^H \mathbf{T}^{-1} \frac{1}{L} \sum_{l=1}^L (\mathbf{a} \beta s_l s_l^* + \mathbf{e}_l s_l^*)}{P_s \mathbf{a}^H \mathbf{T}^{-1} \mathbf{a}} \\ &= \beta + \frac{\mathbf{a}^H \mathbf{T}^{-1} \frac{1}{L} \sum_{l=1}^L \mathbf{e}_l s_l^*}{P_s \mathbf{a}^H \mathbf{T}^{-1} \mathbf{a}} \end{aligned} \quad (13)$$

i.e., the error of the ML estimate of  $\beta$  is

$$\hat{\beta}_{ML} - \beta = \frac{\mathbf{a}^H \mathbf{T}^{-1} \frac{1}{L} \sum_{l=1}^L \mathbf{e}_l s_l^*}{P_s \mathbf{a}^H \mathbf{T}^{-1} \mathbf{a}}. \quad (14)$$

Hence, the MSE of the ML estimate is

$$\text{MSE}(\hat{\beta}_{ML}) = \mathcal{E} \left[ \frac{\mathbf{a}^H \mathbf{T}^{-1} \left( \frac{1}{L} \sum_{l=1}^L \mathbf{e}_l s_l^* \right) \left( \frac{1}{L} \sum_{l=1}^L \mathbf{e}_l s_l^* \right)^H \mathbf{T}^{-1} \mathbf{a}}{(\mathbf{a}^H \mathbf{T}^{-1} \mathbf{a})^2 P_s^2} \right] \quad (15)$$

$$= \mathcal{E} \left[ \frac{\mathbf{a}^H \mathbf{T}^{-1} \mathbf{Q} \mathbf{T}^{-1} \mathbf{a}}{(\mathbf{a}^H \mathbf{T}^{-1} \mathbf{a})^2 LP_s} \right]. \quad (16)$$

To obtain (16) from (15), we have utilized the facts that  $\mathbf{T}$  and  $\sum_{l=1}^L \mathbf{e}_l s_l^*$  are statistically independent of each other (see Lemma 1) and that  $\{\mathbf{e}_l\}_{l=1}^L$  are independently and identically distributed.

Let

$$\zeta = \frac{\mathbf{a}^H \mathbf{T}^{-1} \mathbf{Q} \mathbf{T}^{-1} \mathbf{a} \mathbf{a}^H \mathbf{Q}^{-1} \mathbf{a}}{(\mathbf{a}^H \mathbf{T}^{-1} \mathbf{a})^2}. \quad (17)$$

According to (16) and (12)

$$\mathcal{E}[\zeta] = \frac{\text{MSE}(\hat{\beta}_{ML})}{\text{CRB}(\beta)}. \quad (18)$$

As shown in the proof of Lemma 1

$$L\mathbf{T} = \sum_{l=2}^L \mathbf{z}_l \mathbf{z}_l^H \quad (19)$$

where  $\mathbf{z}_l \sim N(\mathbf{0}, \mathbf{Q})$   $l = 2, \dots, L$  are independently and identically distributed. Hence,  $\mathbf{T}$  has a complex Wishart distribution [13], [14], [16], i.e.,  $\mathbf{T} \sim CW(L-1, M; \mathbf{Q}/L)$ . In [13], the probability density function (PDF) of  $\rho \triangleq 1/\zeta$  has been provided:

$$f(\rho) = \frac{(L-1)!}{(M-2)!(L-M)!} (1-\rho)^{M-2} \rho^{L-M} \text{ for } 0 \leq \rho \leq 1. \quad (20)$$

Thus

$$\begin{aligned} \mathcal{E}[\zeta] &= \mathcal{E} \left[ \frac{1}{\rho} \right] \\ &= \frac{(L-1)!}{(M-2)!(L-M)!} \int_0^1 \rho^{L-M-1} (1-\rho)^{M-2} d\rho \quad (21) \\ &= \frac{L-1}{L-M}. \end{aligned} \quad (22)$$

Therefore, it follows from (18) and (22) that

$$\text{MSE}(\hat{\beta}_{ML}) = \frac{L-1}{L-M} \text{CRB}(\beta). \quad (23)$$

The above equation shows that 1) the ML estimate is asymptotically statistically efficient for large snapshot lengths  $L$ , which was expected, and 2) the ML estimate is not asymptotically statistically efficient for high SNR values when  $L$  is fixed. In fact, we can infer from (23) that the MSE (in decibels)-versus-SNR (in decibels) line is parallel to the

CRB-versus-SNR line, and thus, there is no ‘‘threshold effect’’ at low SNR. This theoretical result is verified via a numerical example in Section V.

### B. Performance Analysis of the Capon Estimator

We now establish the theoretical properties of the Capon estimator via an analysis that parallels the one in Section III-A. Our analysis relies on the results in [13], [14], and [17].

1) *Bias Analysis:* We have proved that the ML estimate is unbiased. We investigate the bias of the Capon estimate by studying the relationship between the two estimators.

*Lemma 2:* Under the assumptions made on the data model in (1), we have

$$\hat{\beta}_{\text{Capon}} = \lambda \hat{\beta}_{ML}, \quad \lambda = \frac{1}{1+u} \leq 1 \quad (24)$$

where

$$u = \frac{1}{P_s} \left( \bar{\mathbf{x}}^H \mathbf{T}^{-1} \bar{\mathbf{x}} - \frac{\mathbf{a}^H \mathbf{T}^{-1} \bar{\mathbf{x}} \bar{\mathbf{x}}^H \mathbf{T}^{-1} \mathbf{a}}{\mathbf{a}^H \mathbf{T}^{-1} \mathbf{a}} \right) \quad (25)$$

and the equality holds iff  $\bar{\mathbf{x}} = \alpha \mathbf{a}$  for some constant  $\alpha$ .

*Proof:* See Appendix D. ■

*Lemma 3:* Under the assumption made on the data model in (1), any polynomial function of the  $\lambda$  defined in (24) is uncorrelated with the ML estimate, i.e.,

$$\mathcal{E} \left[ f(\lambda) (\hat{\beta}_{ML} - \beta) \right] = 0$$

where  $f(\lambda)$  is an arbitrary polynomial.

*Proof:* See Appendix E. ■

Based on Lemmas 2 and 3, we have

$$\begin{aligned} \mathcal{E}[\hat{\beta}_{\text{Capon}}] &= \mathcal{E}[\lambda \hat{\beta}_{ML}] \\ &= \mathcal{E}[\lambda] \beta. \end{aligned} \quad (26)$$

We derive below the PDF of  $\lambda$ . Recall that  $\bar{\mathbf{e}} \sim N(\mathbf{0}, \mathbf{Q}/L)$  [cf. (126)] and  $\mathbf{T} \sim CW(L-1, M; \mathbf{Q}/L)$ . Following the techniques used in [13] and [14], we define

$$\boldsymbol{\xi} = L^{1/2} \mathbf{Q}^{-1/2} \bar{\mathbf{e}} \quad (27)$$

and

$$\mathbf{C} = L \mathbf{Q}^{-1/2} \mathbf{T} \mathbf{Q}^{-1/2}. \quad (28)$$

Then,  $\boldsymbol{\xi} \sim N(\mathbf{0}, \mathbf{I})$  and  $\mathbf{C} \sim CW(L-1, M; \mathbf{I})$  [16].

Inserting (27) and (28) into (127) yields

$$u = \boldsymbol{\xi}^H \left( \mathbf{C}^{-1} - \frac{\mathbf{C}^{-1} \mathbf{d} \mathbf{d}^H \mathbf{C}^{-1}}{\mathbf{d}^H \mathbf{C}^{-1} \mathbf{d}} \right) \boldsymbol{\xi} \quad (29)$$

where

$$\mathbf{d} = L^{1/2} \mathbf{Q}^{-1/2} \mathbf{a}. \quad (30)$$

We next introduce an  $M \times M$  unitary matrix  $\tilde{\mathbf{U}}$  such that

$$\mathbf{t} = \tilde{\mathbf{U}} \mathbf{d} = [||\mathbf{d}||, \mathbf{0}_{1 \times M-1}]^T. \quad (31)$$

Note that  $\tilde{\mathbf{U}} \boldsymbol{\xi}$  and  $\tilde{\mathbf{U}} \mathbf{C} \tilde{\mathbf{U}}^H$  have the same PDFs as  $\boldsymbol{\xi}$  and  $\mathbf{C}$ , respectively. Then, (29) can be further simplified as

$$u = \boldsymbol{\xi}^H \mathbf{C}^{-1} \boldsymbol{\xi} - \boldsymbol{\xi}^H \frac{\mathbf{C}^{-1} \mathbf{t} \mathbf{t}^H \mathbf{C}^{-1}}{\mathbf{t}^H \mathbf{C}^{-1} \mathbf{t}} \boldsymbol{\xi}. \quad (32)$$

We partition  $\mathbf{C}$ ,  $\mathbf{C}^{-1}$ , and  $\boldsymbol{\xi}$  as

$$\mathbf{C} = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \quad (33)$$

$$\mathbf{C}^{-1} = \begin{pmatrix} C^{11} & C^{12} \\ C^{21} & C^{22} \end{pmatrix} \quad (34)$$

$$\boldsymbol{\xi} = \begin{pmatrix} \xi_1 \\ \boldsymbol{\xi}_2 \end{pmatrix} \quad (35)$$

where  $C_{11}$ ,  $C^{11}$ , and  $\xi_1$  are scalars,  $C_{21} = C_{12}^H$ ,  $C^{21} = [C^{12}]^H$  and  $\boldsymbol{\xi}_2$  are  $(M-1) \times 1$  vectors, and both  $C_{22}$  and  $C^{22}$  are  $(M-1) \times (M-1)$  matrices. Some useful relationships are as follows [13], [14]:

$$C^{11} = (C_{11} - C_{12} C_{22}^{-1} C_{21})^{-1} \quad (36)$$

$$C^{21} = -C_{22}^{-1} C_{21} C^{11} \quad (37)$$

$$C^{22} = (C_{22} - C_{21} C_{11}^{-1} C_{12})^{-1} \quad (38)$$

$$= C_{22}^{-1} + \frac{C_{22}^{-1} C_{21} C_{12} C_{22}^{-1}}{C_{11} - C_{12} C_{22}^{-1} C_{21}}. \quad (39)$$

To obtain (39) from (38), we have used the matrix inversion lemma.

Based on the partitions defined in (33)–(35), (32) can be written as

$$u = \frac{\xi_1^* C^{11} \xi_1 + \xi_1^* C^{12} \boldsymbol{\xi}_2 + \boldsymbol{\xi}_2^H C^{21} \xi_1 + \boldsymbol{\xi}_2^H C^{22} \boldsymbol{\xi}_2}{|\xi_1^* C^{11} + \boldsymbol{\xi}_2^H C^{21}|^2} C^{11}$$

$$= \frac{\xi_1^* C^{11} \xi_1 + \xi_1^* C^{12} \boldsymbol{\xi}_2 + \boldsymbol{\xi}_2^H C^{21} \xi_1 + \boldsymbol{\xi}_2^H C^{22} \boldsymbol{\xi}_2}{|\xi_1^* - \boldsymbol{\xi}_2^H C_{22}^{-1} C_{21}|^2} C^{11}$$

$$= \frac{\boldsymbol{\xi}_2^H C^{22} \boldsymbol{\xi}_2 - \boldsymbol{\xi}_2^H C_{22}^{-1} C_{21} C^{11} C_{12} C_{22}^{-1} \boldsymbol{\xi}_2}{|\xi_1^* - \boldsymbol{\xi}_2^H C_{22}^{-1} C_{21}|^2} C^{11} \quad (40)$$

$$= \frac{\boldsymbol{\xi}_2^H C^{22} \boldsymbol{\xi}_2 - \frac{\boldsymbol{\xi}_2^H C_{22}^{-1} C_{21} C_{12} C_{22}^{-1} \boldsymbol{\xi}_2}{C_{11} - C_{12} C_{22}^{-1} C_{21}}}{|\xi_1^* - \boldsymbol{\xi}_2^H C_{22}^{-1} C_{21}|^2} C^{11} \quad (41)$$

$$= \boldsymbol{\xi}_2^H C_{22}^{-1} \boldsymbol{\xi}_2. \quad (42)$$

Hence, we have

$$\lambda = \frac{1}{1 + \boldsymbol{\xi}_2^H C_{22}^{-1} \boldsymbol{\xi}_2} \quad (43)$$

where  $\boldsymbol{\xi}_2 \sim N(\mathbf{0}, \mathbf{I})$  and  $C_{22} \sim CW(L-1, M-1; \mathbf{I})$ , and they are statistically independent of each other. It can be shown (see the [14, App.]) that the PDF of  $\lambda$  is

$$f(\lambda) = \frac{(L-1)!}{(L-M)!(M-2)!} (1-\lambda)^{M-2} \lambda^{L-M} \quad \text{for } 0 \leq \lambda \leq 1 \quad (44)$$

and the first and second moments of this distribution are

$$\mathcal{E}[\lambda] = \frac{L-M+1}{L} \quad (45)$$

$$\mathcal{E}[\lambda^2] = \frac{(L-M+1)(L-M+2)}{L(L+1)}. \quad (46)$$

Thus, it follows from (45) and (26) that

$$\mathcal{E}[\hat{\beta}_{\text{Capon}}] = \frac{L-M+1}{L} \beta. \quad (47)$$

The above equation shows that the Capon estimate of  $\beta$  is biased downwards for a finite data sample number  $L$ , whereas it

is asymptotically unbiased for large  $L$ . We can also see that the biasedness of the Capon estimate is not related to SNR. We note that the result in (47) is not completely new. In [18] and [19], results similar to (47) when studying the convergence rate of the Capon beamformer, which was used to estimate the signal power and waveform, are given. However, our assumptions are different from those in [18] and [19], where the signals were assumed to be independently and identically distributed complex Gaussian random vectors.

2) *Mean-Squared Error Analysis*: Based on Lemma 2, we can obtain the  $\text{MSE}(\hat{\beta}_{\text{Capon}})$  as

$$\begin{aligned} & \text{MSE}(\hat{\beta}_{\text{Capon}}) \\ &= \mathcal{E} \left[ |\beta - \lambda \hat{\beta}_{ML}|^2 \right] \end{aligned} \quad (48)$$

$$= \mathcal{E} \left[ |(1 - \lambda)\beta - \lambda(\hat{\beta}_{ML} - \beta)|^2 \right] \quad (49)$$

$$= \mathcal{E} \left[ \lambda^2 |\hat{\beta}_{ML} - \beta|^2 \right] + \mathcal{E} \left[ (1 - \lambda)^2 |\beta|^2 \right] \quad (50)$$

where to obtain (50) from (49), we have used Lemma 3 as well as (11).

Combining (14) and (126) gives

$$\hat{\beta}_{ML} - \beta = \frac{\mathbf{a}^H \mathbf{T}^{-1} \sqrt{P_s} \mathbf{e}}{P_s \mathbf{a}^H \mathbf{T}^{-1} \mathbf{a}}. \quad (51)$$

Using the notations defined in (27) – (39), we can rewrite (51) as

$$\hat{\beta}_{ML} - \beta = \frac{\mathbf{t}^H \mathbf{C}^{-1} \boldsymbol{\xi}}{\sqrt{P_s} \mathbf{t}^H \mathbf{C}^{-1} \mathbf{t}} \quad (52)$$

$$= \frac{C^{11} \xi_1 + C^{12} \xi_2}{\sqrt{LP_s \mathbf{a}^H \mathbf{Q}^{-1} \mathbf{a} C^{11}}} \quad (53)$$

$$= \frac{\xi_1 - C_{12} C_{22}^{-1} \xi_2}{\sqrt{LP_s \mathbf{a} \mathbf{Q}^{-1} \mathbf{a}}}. \quad (54)$$

Thus

$$\begin{aligned} & \mathcal{E} \left[ \lambda^2 |\hat{\beta}_{ML} - \beta|^2 \right] \\ &= \mathcal{E} \left[ \left| \frac{\xi_1 - C_{12} C_{22}^{-1} \xi_2}{1 + \xi_2^H C_{22}^{-1} \xi_2} \right|^2 \right] \text{CRB}(\beta) \end{aligned} \quad (55)$$

$$= \mathcal{E} \left[ \frac{|\xi_1|^2 + \xi_2^H C_{22}^{-1} C_{21} C_{12} C_{22}^{-1} \xi_2}{(1 + \xi_2^H C_{22}^{-1} \xi_2)^2} \right] \text{CRB}(\beta) \quad (56)$$

$$= \mathcal{E} \left[ \frac{1 + \xi_2^H C_{22}^{-1} \xi_2}{(1 + \xi_2^H C_{22}^{-1} \xi_2)^2} \right] \text{CRB}(\beta) \quad (57)$$

$$= \mathcal{E} [\lambda] \text{CRB}(\beta) \quad (58)$$

$$= \frac{L - M + 1}{L} \text{CRB}(\beta). \quad (59)$$

We obtained (57) from (56) using the fact that

$$\mathcal{E}_{C_{21}|C_{22}} [C_{21} C_{12}] = C_{22}. \quad (60)$$

It follows from (45) and (46) that the second term of (50) is

$$\begin{aligned} & \mathcal{E} \left[ (1 - \lambda)^2 |\beta|^2 \right] \\ &= \left[ 1 - \frac{2(L - M + 1)}{L} + \frac{(L - M + 1)(L - M + 2)}{L(L + 1)} \right] |\beta|^2 \end{aligned} \quad (61)$$

$$= \frac{M^2 - M}{L^2 + L} |\beta|^2. \quad (62)$$

TABLE I  
M SENSORS, L SNAPSHOTS, AND SIGNAL AMPLITUDE  $\beta$

M sensors, L snapshots, and signal amplitude $\beta$ .		
	ML	Capon
Expected value	$\beta$	$\frac{L - M + 1}{L} \beta$
Mean-Squared Error	$\frac{L - 1}{L - M} \text{CRB}$	$\frac{L - M + 1}{L} \text{CRB} + \frac{M^2 - M}{L^2 + L}  \beta ^2$

Thus

$$\text{MSE}(\hat{\beta}_{\text{Capon}}) = \frac{L - M + 1}{L} \text{CRB}(\beta) + \frac{M^2 - M}{L^2 + L} |\beta|^2. \quad (63)$$

We can conclude from (63) that the Capon estimate is asymptotically statistically efficient for a large data sample number  $L$ . In the finite  $L$  case, the MSE of the Capon estimate may be smaller than the CRB for any unbiased estimator if the first term in (63) dominates the second one. However, the MSE of the Capon estimate has an error floor equal to  $(M^2 - M)|\beta|^2/(L^2 + L)$  at high SNR.

### C. Summary of the Capon and ML Statistical Properties

- The ML estimate is unbiased.
- The ML estimate is asymptotically statistically efficient for large number of snapshots.
- The ML estimate is *not* asymptotically statistically efficient for high SNR.
- The Capon estimate is biased downwards.
- The Capon estimate is asymptotically unbiased and statistically efficient for large number of snapshots.
- The Capon estimate is *neither* asymptotically unbiased *nor* asymptotically statistically efficient for high SNR.

The above properties are summarized in Table I.

## IV. EXTENSION TO MULTICHANNEL AR INTERFERENCE AND NOISE

The previous study assumed that the interference and noise term in (1) is spatially colored but temporally white, despite the fact that the interference and noise can be temporally correlated [1], [20]. In this section, we model the interference and noise vector as a multichannel autoregressive (AR) random process and propose an alternating least squares (ALS) method based on the cyclic optimization approach [21]. For discussion purposes, we refer to the ML method in Section II that ignores the temporal correlation of the interference and noise as the model-mismatched ML ( $M^3L$ ) method.

### A. Data Model

Consider the data model:

$$\mathbf{x}_l = \mathbf{a} \beta s_l + \mathbf{v}_l, \quad l = 1, 2, \dots, L \quad (64)$$

which is the same as the one in (1), except that the interference and noise term now satisfies the following AR equation [22]

$$\mathbf{A}(z^{-1}) \mathbf{v}_l = \mathbf{e}_l \quad (65)$$

where  $z^{-1}$  is the unit delay operator

$$\mathbf{A}(z^{-1}) = \mathbf{I} + \mathbf{A}_1 z^{-1} + \mathbf{A}_2 z^{-2} + \dots + \mathbf{A}_p z^{-p} \quad (66)$$

and

$$\mathcal{E}[\mathbf{e}_l \mathbf{e}_n^H] = \mathbf{Q} \delta_{ln} \quad (67)$$

where  $\delta_{ln}$  denotes the Kronecker delta function:

$$\delta_{ln} = \begin{cases} 1, & l = n \\ 0, & l \neq n. \end{cases} \quad (68)$$

Note that if the interference component in  $\mathbf{v}_l$  is a multichannel AR process while the noise component in  $\mathbf{v}_l$  is white temporally, then the interference and noise term will be a multichannel autoregressive and moving average (ARMA) random process, which can still be approximated by a multichannel AR process. The SNR for the data model in (64) is defined as

$$\text{SNR} = \frac{MP_s |\beta|^2}{\text{tr}(\mathbf{R}_v)} \quad (69)$$

where  $\mathbf{R}_v$  is the covariance matrix of  $\{\mathbf{v}_l\}$ .

### B. ALS Algorithm

Conditioned on the first  $p$  data vectors  $\{\mathbf{x}_l\}_{l=1}^p$ , the log-likelihood function is proportional to

$$C_1 = -(L-p) \ln |\mathbf{Q}| - \text{tr} \left[ \mathbf{Q}^{-1} \sum_{l=p+1}^L [\mathbf{A}(z^{-1})(\mathbf{x}_l - \mathbf{a}\beta_{s_l})] [\mathbf{A}(z^{-1})(\mathbf{x}_l - \mathbf{a}\beta_{s_l})]^H \right]. \quad (70)$$

Maximizing the above function with respect to  $\mathbf{Q}$  gives (see, e.g., [23])

$$\hat{\mathbf{Q}} = \frac{1}{L-p} \times \sum_{l=p+1}^L [\mathbf{A}(z^{-1})(\mathbf{x}_l - \mathbf{a}\beta_{s_l})] [\mathbf{A}(z^{-1})(\mathbf{x}_l - \mathbf{a}\beta_{s_l})]^H. \quad (71)$$

After substituting (71) into (70), we need to minimize

$$C_2 = \left| \sum_{l=p+1}^L [\mathbf{A}(z^{-1})(\mathbf{x}_l - \mathbf{a}\beta_{s_l})] [\mathbf{A}(z^{-1})(\mathbf{x}_l - \mathbf{a}\beta_{s_l})]^H \right| \quad (72)$$

with respect to both  $\beta$  and  $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_p]$ . Hence, the optimization problem becomes more complicated than the one in Section II. Here, we propose an alternating least squared (ALS) approach to solve this problem.

To begin with, we obtain an initial estimate  $\hat{\beta}^{(0)}$  of  $\beta$  by using the M<sup>3</sup>L method [cf. (9)]. For a given estimate  $\hat{\beta}$ , let  $\mathbf{z}_l = \mathbf{x}_l - \mathbf{a}\hat{\beta}_{s_l}$ . From (72), we get

$$\hat{\mathbf{A}} = \arg \min_{\mathbf{A}} \left| \sum_{l=p+1}^L [\mathbf{A}(z^{-1})\mathbf{z}_l] [\mathbf{A}(z^{-1})\mathbf{z}_l]^H \right| \quad (73)$$

where

$$\mathbf{A}(z^{-1})\mathbf{z}_l = \mathbf{z}_l + [\mathbf{A}_1 \dots \mathbf{A}_p] \begin{bmatrix} \mathbf{z}_{l-1} \\ \vdots \\ \mathbf{z}_{l-p} \end{bmatrix} \triangleq \mathbf{z}_l + \mathbf{A}\phi_l.$$

Note that

$$\begin{aligned} & \sum_{l=p+1}^L [\mathbf{A}(z^{-1})\mathbf{z}_l] [\mathbf{A}(z^{-1})\mathbf{z}_l]^H \\ &= \sum_{l=p+1}^L [\mathbf{z}_l + \mathbf{A}\phi_l] [\mathbf{z}_l + \mathbf{A}\phi_l]^H \\ &= \sum_{l=p+1}^L \mathbf{z}_l \mathbf{z}_l^H + \sum_{l=p+1}^L \mathbf{z}_l \phi_l^H \mathbf{A}^H + \mathbf{A} \sum_{l=p+1}^L \phi_l \mathbf{z}_l^H \\ & \quad + \mathbf{A} \left( \sum_{l=p+1}^L \phi_l \phi_l^H \right) \mathbf{A}^H \\ & \triangleq \hat{\mathbf{R}}_z + \hat{\mathbf{R}}_{z\phi} \mathbf{A}^H + \mathbf{A} \hat{\mathbf{R}}_{z\phi}^H + \mathbf{A} \hat{\mathbf{R}}_{\phi} \mathbf{A}^H \\ &= \left( \mathbf{A} + \hat{\mathbf{R}}_{z\phi} \hat{\mathbf{R}}_{\phi}^{-1} \right) \hat{\mathbf{R}}_{\phi} \left( \mathbf{A} + \hat{\mathbf{R}}_{z\phi} \hat{\mathbf{R}}_{\phi}^{-1} \right)^H \\ & \quad + \hat{\mathbf{R}}_z - \hat{\mathbf{R}}_{z\phi} \hat{\mathbf{R}}_{\phi}^{-1} \hat{\mathbf{R}}_{z\phi}^H. \end{aligned} \quad (74)$$

Hence, the solution to (73) is

$$\hat{\mathbf{A}} = -\hat{\mathbf{R}}_{z\phi} \hat{\mathbf{R}}_{\phi}^{-1} \quad (75)$$

which is recognized as the multichannel Prony estimate of  $\mathbf{A}$  [22]. We assume that the order  $p$  of the multichannel random process AR( $p$ ) is known. If  $p$  is unknown, it can be estimated, for instance, by using the generalized Akaike information criterion (GAIC) [23].

For a given  $\hat{\mathbf{A}}$ , we obtain an improved estimate of  $\beta$  as follows:

$$\hat{\beta} = \arg \min_{\beta} \left| \sum_{l=p+1}^L \left[ \hat{\mathbf{A}}(z^{-1})(\mathbf{x}_l - \mathbf{a}\beta_{s_l}) \right] \left[ \hat{\mathbf{A}}(z^{-1})(\mathbf{x}_l - \mathbf{a}\beta_{s_l}) \right]^H \right|. \quad (76)$$

First, we consider the case of a known damped (or undamped) sinusoidal signal, i.e.,  $s_l = e^{(-\alpha_s + j\omega_s)l}$  with known frequency  $\omega_s$  and damping factor  $\alpha_s$ .

Let

$$\begin{aligned} \mathbf{y}_l &= \hat{\mathbf{A}}(z^{-1})\mathbf{x}_l, \quad l = p+1, \dots, L, \\ \mathbf{b} &= \hat{\mathbf{A}}(z^{-1})|_{z=e^{-\alpha_s + j\omega_s}} \mathbf{a}. \end{aligned}$$

Note that the length of the new data sequence  $\mathbf{y}_l$ ,  $l = p+1, \dots, L$  is  $L-p$  instead of  $L$ . The solution to the above problem is given by the ML estimator proposed in Section II:

$$\hat{\beta} = \frac{\mathbf{b}^H \mathbf{T}_y^{-1} \bar{\mathbf{y}}}{P_s \mathbf{b}^H \mathbf{T}_y^{-1} \mathbf{b}} \quad (77)$$

where

$$\mathbf{T}_y = \frac{1}{L-p} \sum_{l=p+1}^L \mathbf{y}_l \mathbf{y}_l^H - \frac{\bar{\mathbf{y}} \bar{\mathbf{y}}^H}{P_s}, \quad \bar{\mathbf{y}} = \frac{1}{L-p} \sum_{l=p+1}^L \mathbf{y}_l s_l^* \quad (78)$$

and  $P_s = \frac{1}{L-p} \sum_{l=p+1}^L |s_l|^2$  is the average power of the known waveform  $\{s_l\}_{l=p+1}^L$ . The ALS approach maximizes the likelihood function cyclically. We set  $\hat{\mathbf{A}}^{(0)} = \mathbf{0}$  and obtain  $\hat{\beta}^{(0)} = \hat{\beta}_{M^3L}$ . We then iterate the following two steps until the solution converges, i.e., until the two consecutive estimates  $\hat{\beta}^{(i)}$  and  $\hat{\beta}^{(i+1)}$  are sufficiently close:

$$\hat{\mathbf{A}}^{(i+1)} = \arg \max_{\mathbf{A}} f(\mathbf{x}|\mathbf{A}; \hat{\beta}^{(i)}, \{\mathbf{x}_l\}_{l=1}^p) \quad (79)$$

which is given by (75) with  $\hat{\beta}$  replaced by  $\hat{\beta}^{(i)}$ , and

$$\hat{\beta}^{(i+1)} = \arg \max_{\beta} f(\mathbf{x}|\beta; \hat{\mathbf{A}}^{(i+1)}, \{\mathbf{x}_l\}_{l=1}^p) \quad (80)$$

which is given by (77) with  $\hat{\mathbf{A}}$  replaced by  $\hat{\mathbf{A}}^{(i+1)}$ .

Obviously, the likelihood function never decreases in any iteration. In the simulations reported in the next section, we found that ALS converges in two or three iterations, although there is no proof of convergence in general. Hence, the ALS estimator is computationally quite efficient.

Next, we consider the case of an arbitrary known waveform signal. Let

$$\mathbf{y}_l = \hat{\mathbf{A}}(z^{-1})\mathbf{x}_l, \quad \mathbf{a}_l = \hat{\mathbf{A}}(z^{-1})\mathbf{a}_s l \\ l = p+1, \dots, L$$

$$\mathbf{Y}_{M \times (L-p)} = [\mathbf{y}_{p+1} \mathbf{y}_{p+2} \dots \mathbf{y}_L] \\ \text{and } \mathbf{G}_{M \times (L-p)} = [\mathbf{a}_{p+1} \mathbf{a}_{p+2} \dots \mathbf{a}_L].$$

In addition, let  $\mathbf{P}$  be an orthogonal projection matrix defined as

$$\mathbf{P} = \mathbf{G}^H(\mathbf{G}^H)^\dagger \quad (81)$$

where  $(\mathbf{G}^H)^\dagger$  is the Moore–Penrose pseudo-inverse of  $\mathbf{G}^H$  [24], and let  $\mathbf{P}^\perp = \mathbf{I} - \mathbf{P}$ .

Then, (76) can be written concisely in a matrix form:

$$\hat{\beta} = \arg \min_{\beta} |(\mathbf{Y} - \beta\mathbf{G})(\mathbf{Y} - \beta\mathbf{G})^H| \\ = \arg \min_{\beta} |(\mathbf{Y} - \beta\mathbf{G})(\mathbf{P} + \mathbf{P}^\perp)(\mathbf{Y} - \beta\mathbf{G})^H| \quad (82)$$

$$= \arg \min_{\beta} |(\mathbf{Y}\mathbf{P} - \beta\mathbf{G})(\mathbf{Y}\mathbf{P} - \beta\mathbf{G})^H + \mathbf{Y}\mathbf{P}^\perp\mathbf{Y}^H| \quad (83)$$

$$\triangleq \arg \min_{\beta} |(\mathbf{Y}\mathbf{P} - \beta\mathbf{G})(\mathbf{Y}\mathbf{P} - \beta\mathbf{G})^H + \mathbf{T}_y| \quad (84)$$

$$= \arg \min_{\beta} |(\mathbf{Y}\mathbf{P} - \beta\mathbf{G})(\mathbf{Y}\mathbf{P} - \beta\mathbf{G})^H \mathbf{T}_y^{-1} + \mathbf{I}| \|\mathbf{T}_y\|. \quad (85)$$

Note that minimizing the cost function in (85) requires a two-dimensional (2-D) search (since  $\beta$  is complex-valued). To avoid the search, we use Lemma 4 to obtain an approximate estimate of  $\beta$ .

*Lemma 4:* For a large data sample number  $L$ , minimizing  $F_1 = |(\mathbf{Y}\mathbf{P} - \beta\mathbf{G})(\mathbf{Y}\mathbf{P} - \beta\mathbf{G})^H \mathbf{T}_y^{-1} + \mathbf{I}|$  is asymptotically equivalent to minimizing

$$F_2 = \text{tr} [(\mathbf{Y}\mathbf{P} - \beta\mathbf{G})^H \mathbf{T}_y^{-1} (\mathbf{Y}\mathbf{P} - \beta\mathbf{G})]. \quad (86)$$

*Proof:* See Appendix F.

It follows from (86) that

$$F_2 = |\beta|^2 \text{tr}(\mathbf{G}^H \mathbf{T}_y^{-1} \mathbf{G}) - \text{tr}(\mathbf{G}^H \mathbf{T}_y^{-1} \mathbf{Y}) \beta^* \\ - \beta \text{tr}(\mathbf{Y}^H \mathbf{T}_y^{-1} \mathbf{G}) + \text{tr}(\mathbf{P} \mathbf{Y}^H \mathbf{T}_y^{-1} \mathbf{Y}). \quad (87)$$

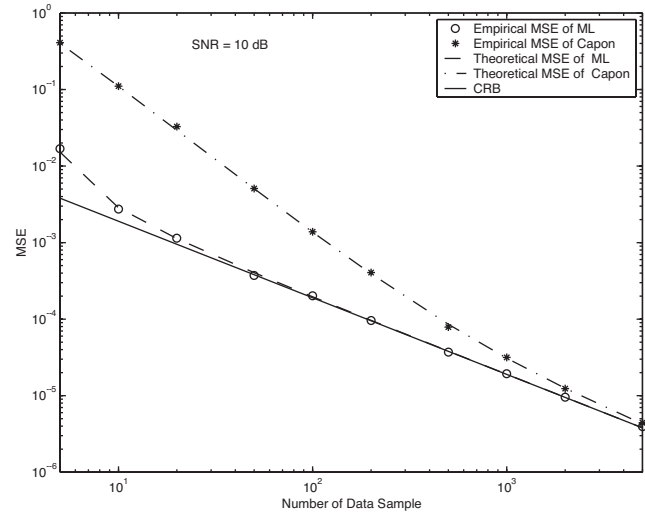


Fig. 2. MSEs of  $\hat{\beta}_{ML}$  and  $\hat{\beta}_{\text{Capon}}$  and the corresponding CRB versus  $L$  when  $\text{SNR} = 10$  dB.

Minimizing  $F_2$  with respect to  $\beta$  yields

$$\hat{\beta} \approx \frac{\text{tr}(\mathbf{G}^H \mathbf{T}_y^{-1} \mathbf{Y})}{\text{tr}(\mathbf{G}^H \mathbf{T}_y^{-1} \mathbf{G})} \quad (88)$$

where we remind the reader that  $\mathbf{T}_y = \mathbf{Y}\mathbf{P}^\perp\mathbf{Y}^H$ . Because (88) is only an approximate solution to (85) in this more general case, ALS is not theoretically guaranteed to yield a more accurate solution than the  $M^3L$  method. However, in our numerical examples, ALS outperforms  $M^3L$  in most cases, even for modest data sample lengths. To avoid any “divergence problem” in this case in which ALS is no longer an iterative maximizer, we simply preimpose the number of iterations to be 3.

## V. NUMERICAL AND EXPERIMENTAL EXAMPLES

We provide both simulated-data and real-life data examples to demonstrate the performance the ML and Capon estimates. In all the simulated-data examples, we consider the case where the steering vector is given by  $\mathbf{a} = [1 \ 0 \ 0 \ 0]^T$  and  $\beta = 1$ . This corresponds to the case where the first of the  $M = 4$  sensors receives the signal as well as the interference and noise, whereas the other three sensors receive the interference and noise only. In all but the last example, we assume that  $s_l = 1$ ,  $l = 1, 2, \dots, L$ , for simplicity. We obtain the empirical MSEs of the estimates by using 500 Monte Carlo trials.

### A. Spatially Colored but Temporally White Interference and Noise

We first consider simulated-data examples. We assume that the interference and noise term is a spatially colored but temporally white Gaussian random vector with the spatial covariance matrix  $\mathbf{Q}$  given by

$$[\mathbf{Q}]_{ij} = \rho \cdot (0.9)^{|i-j|} \quad (89)$$

where  $\rho = 1/\text{SNR}$ .

Fig. 2 shows the MSEs of the Capon and ML estimates obtained from both theoretical predictions [based on (23) and (63)] and Monte Carlo trials as well as the corresponding CRB as a function of  $L$  when the SNR is 10 dB. As expected, both the

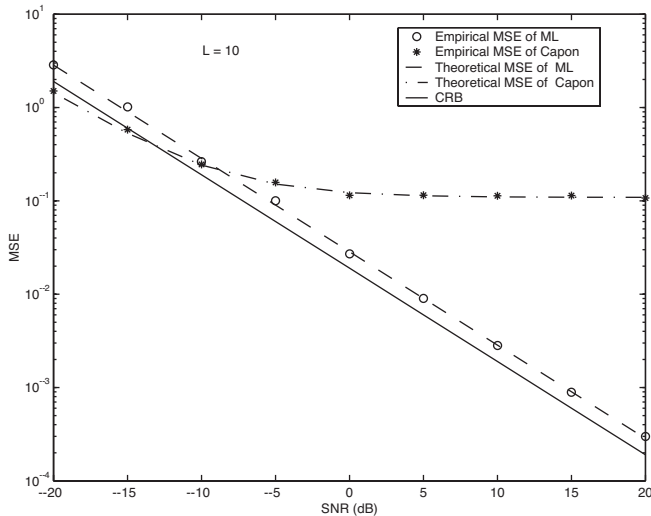


Fig. 3. MSEs of  $\hat{\beta}_{ML}$  and  $\hat{\beta}_{Capon}$  and the corresponding CRB versus SNR when  $L = 10$ .

ML and Capon estimates approach the corresponding CRB as  $L$  increases since both methods are asymptotically statistically efficient for large  $L$ . Fig. 3 gives the MSEs of the Capon and ML estimates as well as the corresponding CRB as a function of SNR when  $L = 10$ . Note the error floor of the Capon estimate at high SNR due to its bias. Note also that the biased Capon estimate can have lower MSE than the unbiased ML estimate at low SNR (yet this happens at MSE values that are too large to be of practical value). As predicted by our theoretical analyses, for a fixed data length  $L$ ,  $MSE(\hat{\beta}_{ML})$  is parallel to the CRB, and no “threshold effect” occurs. Furthermore, the ML estimate is not asymptotically statistically efficient for high SNR.

Next, we present a real-life data example based on experimentally measured QR data. The main antenna of a QR landmine detector receives a QR signal that consists of 40 echoes as well as AM/FM interferences. We apply a fast Fourier transform (FFT) to each echo and only pick the value corresponding to the echo frequency  $\omega_e$ . In this way, we compress the QR signal into a signal with known waveform  $s_l = e^{-lT_e/T_2}$ ,  $l = 1, \dots, 40$ . Next, the data received at the three reference antennas is segmented into 40 blocks that occupy the same period of time as the 40 echoes. We apply FFT to each block and pick three values corresponding to  $\omega_e$  and the two adjacent frequency bins. Doing so, we get a virtual array with one main antenna and nine reference antennas, i.e.,  $\mathbf{a} = [1 \ \mathbf{0}_{1 \times 9}]^T$ . Although the aforementioned preprocessing method might seem somewhat *ad hoc*, it worked well in our experiments. Fig. 4 shows the ML and Capon estimates of the QR signal amplitude in 30 experimental trials. Since we do not know the true value of the signal amplitude, we cannot compare the MSEs of the two estimators. Nevertheless, we let

$$\epsilon(\hat{\beta}) = \frac{\sigma[\hat{\beta}]}{\mathcal{E}[\hat{\beta}]} \quad (90)$$

where  $\sigma[\hat{\beta}]$  and  $\mathcal{E}[\hat{\beta}]$  are the empirical standard deviation and the mean of  $\hat{\beta}$  (over the 30 trials). A small  $\epsilon(\hat{\beta})$  is desirable

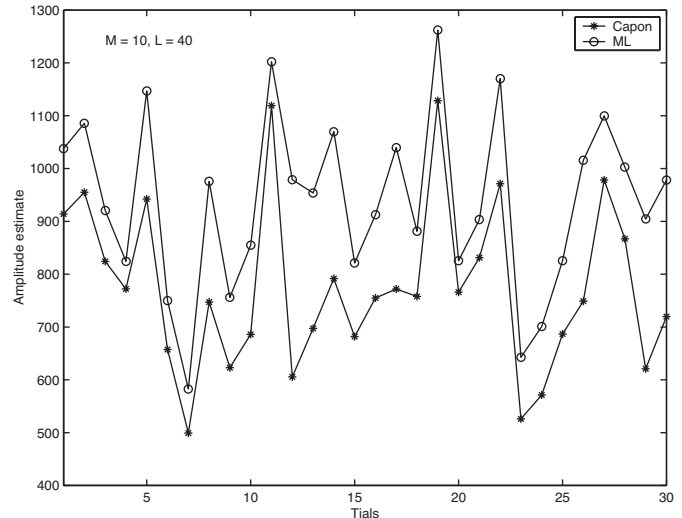


Fig. 4. Amplitude estimates of a QR signal obtained via ML and Capon using experimentally measured data.

for signal detection. Based on the 30 trials, we get  $\epsilon(\hat{\beta}_{ML}) = 0.1746$ , which is smaller than  $\epsilon(\hat{\beta}_{Capon}) = 0.2039$ .

### B. Both Spatially and Temporally Colored Interference and Noise

We now consider the case of spatially and temporally correlated interference and noise. We generate a multichannel AR(2) random process with the method in [20]. The autocorrelation matrices are given by

$$[\mathbf{R}_v(l)]_{mn} = \mathcal{E}[\mathbf{v}_t \mathbf{v}_{t-l}^H] = \rho \rho_s^{|m-n|} \exp\{-\rho_t l^2 + j(m-n+l)\omega\} \quad (91)$$

and

$$\mathbf{R}_v(l) = \mathbf{R}_v^H(-l), \quad l = 0, 1, \dots, p \quad (92)$$

where  $\rho = 1/\text{SNR}$ ,  $\rho_s$  controls the spatial correlation,  $\rho_t$  partly decides the temporal correlation, and  $\omega$  defines the spectral peak location of the colored interference and noise in each channel. The data sample number is  $L = 50$ . When we use the true autoregressive matrix  $\mathbf{A}$  in the ALS instead of the estimated one, we refer to the method as the known-AR ML (KML) approach. We include KML for comparison purposes only. Note that unlike the temporally white interference and noise case discussed previously, the performance of the  $M^3L$  and ALS estimators depends on the temporal frequency characteristics of the known signal. The following simulations are performed for both a constant signal and a known BPSK signal.

First, we consider the relationship between the cost function  $C_2$  defined in (72) and  $\beta$ . Because  $\mathbf{A}(z^{-1})$  can be concentrated out by using its estimate given in (75),  $C_2$  is a function of  $\beta$  only. Consider the constant signal case. Fig. 5 shows the mesh plot of  $-C_2$  versus the real and imaginary part of  $\hat{\beta}$ . We can see that there is only one local maximum around the true value of  $\beta = 1$ .



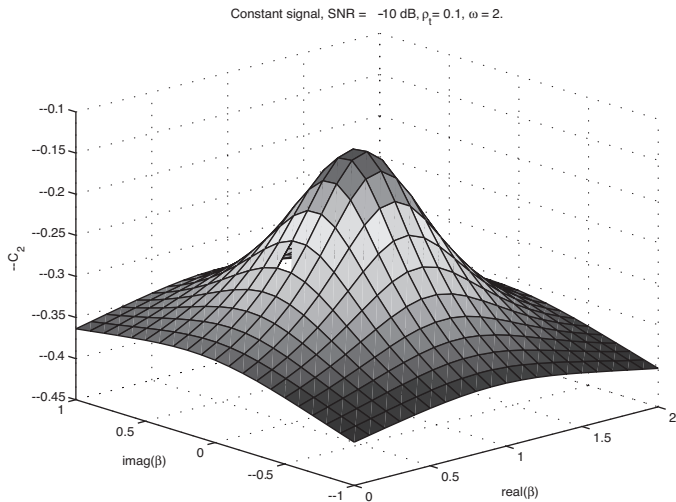


Fig. 5. Mesh plot of  $-C_2$  (with  $C_2$  defined in (72)) for a constant signal versus the real and imaginary part of  $\beta$  when SNR =  $-10$  dB,  $\rho_t = 0.1$ ,  $\rho_s = 0.6$ ,  $\omega = 2$ , and the true value of  $\beta$  is 1.

For the constant signal case, our simulations show that the spatial correlation coefficient  $\rho_s$  is not closely related to the gap between the performance of  $M^3L$  and ALS. However, the temporal correlation coefficient  $\rho_t$  and the position of the spectral peak  $\omega$  have an impact on the relative performance of the two methods (see Figs. 6 and 7). We summarize our observations as follows:

- A: Both ALS and  $M^3L$  work better for large  $\omega$  and/or small  $\rho_t$ .
- B: ALS is slightly worse than  $M^3L$  for small  $\omega$  and/or large  $\rho_t$ .
- C: ALS is significantly better than  $M^3L$  for large  $\omega$  and/or small  $\rho_t$ .

To explain these observations, we examine the signal as well as the interference and noise term in the temporal frequency domain. The signal is a constant and, hence, has power at zero frequency only. The power of the interference and noise is concentrated around  $\omega$ , especially for small  $\rho_t$ . For large  $\omega$ , the signal is separated from the interference and noise in the temporal frequency domain, which benefits both methods. Similarly, smaller  $\rho_t$  means higher correlation in the temporal domain or more peaky spectra in the temporal frequency domain. Hence, both estimators perform better for this case when  $\omega$  is away from zero. This explains Observation A. Next, we note that a large  $\rho_t$  means low correlation in the temporal domain, and hence, the interference and noise vector is approximately temporally white. For small  $\omega$ , the signal and the interference and noise terms are not well separated in the temporal frequency domain. KML behaves approximately as  $M^3L$  in either case. Since ALS is inferior to KML, ALS is also slightly worse than  $M^3L$ . This explains Observation B. Observation C was expected since ALS estimates the temporal correlation of the interference and noise and can suppress the interference and noise more efficiently in this case in which the temporal correlation is significant.

Finally, we consider the known BPSK signal case. Because a BPSK signal is wideband in the temporal frequency domain, the impact of the interference spectral peak location  $\omega$  on the performance of the two methods is not as significant as in the constant

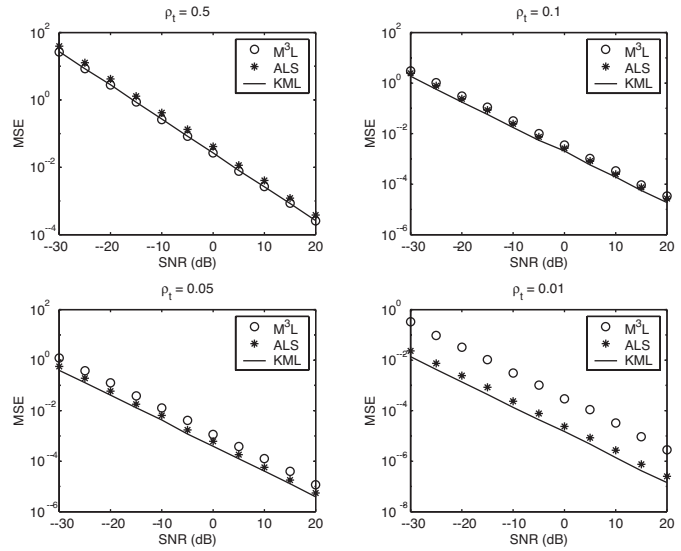


Fig. 6. MSEs of ALS  $M^3L$  and KML estimates for a constant signal versus SNR when  $\omega = 1$  and  $\rho_s = 0.6$ .

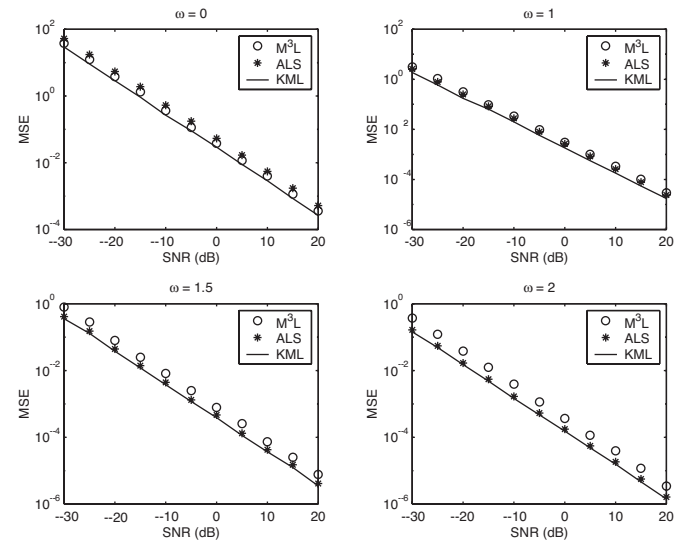


Fig. 7. MSEs of ALS  $M^3L$  and KML estimates for a constant signal versus SNR when  $\rho_t = 0.1$  and  $\rho_s = 0.6$ .

signal case, which was verified in our simulations. However, the temporal correlation parameter  $\rho_t$  still controls the relative performance of the two methods as shown in Fig. 8. We also see from Fig. 8 that the ALS method significantly outperforms  $M^3L$  (by over 10 dB in SNR) even for modestly temporally correlated interference and noise ( $\rho_t = 0.1$ ), although it performs similarly to  $M^3L$  when the temporal correlation of the interference and noise is weak ( $\rho_t = 2$ ). Our simulations also suggest that a known wideband signal makes suppressing temporally correlated interference and noise easier than a narrowband one in the sense that better estimates of  $\beta$  can be obtained in the wideband case.

## VI. CONCLUSIONS

We have investigated the problem of amplitude estimation for a signal with known waveform and steering vector in the pres-

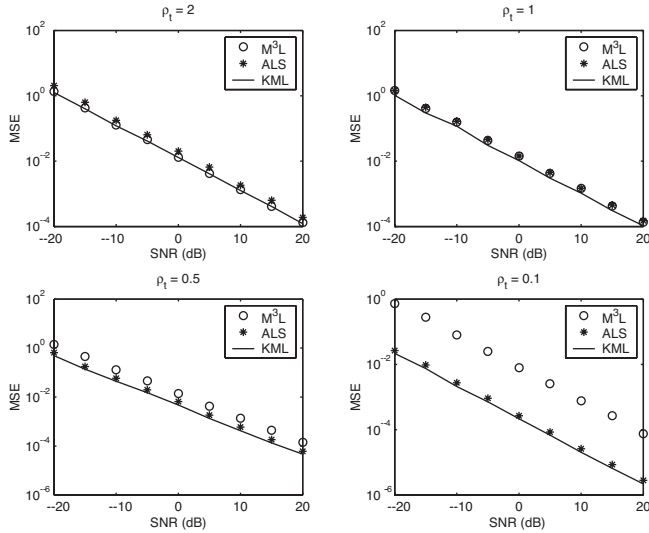


Fig. 8. MSEs of ALS, M<sup>3</sup>L and KML estimates for a known BPSK signal versus SNR when  $\omega = 0$  and  $\rho_s = 0.6$ .

ence of interference and noise. We first assumed that the interference and noise vector was spatially colored but temporally white. The ML and Capon methods as well as the closed-form expressions of the expected values and MSEs of the two estimators have been derived. We have shown that the ML estimate is unbiased as well as asymptotically statistically efficient for large data sample sets but, in addition, that it is not asymptotically statistically efficient for high SNR. We have also shown that the Capon method is biased downwards, but it is asymptotically unbiased and efficient for large data sample lengths. The bias of the Capon estimate dominates its variance for high SNR, which results in an error floor that does not decrease with SNR. At low SNR, however, Capon can outperform ML as well as the CRB for any unbiased estimator. We then considered a more general scenario where the interference and noise vector was both spatially and temporally colored. We have proposed an ALS method based on the idea of cyclic optimization. We have shown that in most cases, ALS outperforms the M<sup>3</sup>L estimator, which ignores the temporal correlation of the interference and noise.

#### APPENDIX A DERIVATION OF THE ML ESTIMATOR

The normalized log-likelihood function of  $\{\mathbf{x}_l\}_{l=1}^L$  is proportional to

$$C = -\ln |\mathbf{Q}| - \text{tr} \left[ \mathbf{Q}^{-1} \frac{1}{L} \sum_{l=1}^L (\mathbf{x}_l - \mathbf{a}\beta s_l)(\mathbf{x}_l - \mathbf{a}\beta s_l)^H \right] \quad (93)$$

where  $|\cdot|$  denotes the determinant of a matrix. Maximizing the above cost function with respect to  $\mathbf{Q}$  gives (see, e.g., [1])

$$\hat{\mathbf{Q}}_{ML} = \frac{1}{L} \sum_{l=1}^L (\mathbf{x}_l - \mathbf{a}\beta s_l)(\mathbf{x}_l - \mathbf{a}\beta s_l)^H. \quad (94)$$

Hence, the ML estimate of  $\beta$  is obtained as follows:

$$\begin{aligned} \hat{\beta}_{ML} &= \arg \min_{\beta} \left| \frac{1}{L} \sum_{l=1}^L (\mathbf{x}_l - \mathbf{a}\beta s_l)(\mathbf{x}_l - \mathbf{a}\beta s_l)^H \right| \quad (95) \\ &= \arg \min_{\beta} \left| \frac{1}{L} \sum_{l=1}^L \mathbf{x}_l \mathbf{x}_l^H - \left( \frac{1}{L} \sum_{l=1}^L \mathbf{x}_l s_l^* \right) (\mathbf{a}\beta)^H \right. \\ &\quad \left. - (\mathbf{a}\beta) \left( \frac{1}{L} \sum_{l=1}^L \mathbf{x}_l s_l^* \right)^H + \mathbf{a}\beta \beta^H \mathbf{a}^H \frac{1}{L} \sum_{l=1}^L |s_l|^2 \right| \\ &\triangleq \arg \min_{\beta} \left| \hat{\mathbf{R}} - \bar{\mathbf{x}}(\mathbf{a}\beta)^H - (\mathbf{a}\beta)\bar{\mathbf{x}}^H + (\mathbf{a}\beta)P_s(\mathbf{a}\beta)^H \right| \quad (96) \\ &= \arg \min_{\beta} \left| \hat{\mathbf{R}} - \frac{\bar{\mathbf{x}}\bar{\mathbf{x}}^H}{P_s} + \left( \mathbf{a}\beta\sqrt{P_s} - \frac{\bar{\mathbf{x}}}{\sqrt{P_s}} \right) \left( \mathbf{a}\beta\sqrt{P_s} - \frac{\bar{\mathbf{x}}}{\sqrt{P_s}} \right)^H \right| \end{aligned}$$

$$\triangleq \arg \min_{\beta} \left| \mathbf{T} \left| \mathbf{I} + \mathbf{T}^{-1} \left( \mathbf{a}\beta\sqrt{P_s} - \frac{\bar{\mathbf{x}}}{\sqrt{P_s}} \right) \left( \mathbf{a}\beta\sqrt{P_s} - \frac{\bar{\mathbf{x}}}{\sqrt{P_s}} \right)^H \right| \right| \quad (97)$$

$$= \arg \min_{\beta} \left( \mathbf{a}\beta\sqrt{P_s} - \frac{\bar{\mathbf{x}}}{\sqrt{P_s}} \right)^H \mathbf{T}^{-1} \left( \mathbf{a}\beta\sqrt{P_s} - \frac{\bar{\mathbf{x}}}{\sqrt{P_s}} \right) \quad (98)$$

$$= \frac{\mathbf{a}^H \mathbf{T}^{-1} \bar{\mathbf{x}}}{P_s \mathbf{a}^H \mathbf{T}^{-1} \mathbf{a}} \quad (99)$$

where from (97) to (98), we have used the fact that  $|\mathbf{I} + \mathbf{A}\mathbf{B}| = |\mathbf{I} + \mathbf{B}\mathbf{A}|$  (see, e.g., [2]).

#### APPENDIX B PROOF OF LEMMA 1

In the following, a subscript is used to indicate the dimension of a matrix for the sake of clarity and is dropped whenever convenient. Consider an  $M \times L$  matrix

$$\mathbf{X}_{M \times L} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_L] = \mathbf{a}\beta \mathbf{s}^T + \mathbf{E}_{M \times L} \quad (100)$$

where  $\mathbf{s} = [s_1, \dots, s_L]^T$ , with  $(\cdot)^T$  denoting the transpose and

$$\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_L]. \quad (101)$$

We can construct a unitary matrix

$$\mathbf{U}_{L \times L} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_L]$$

whose first column  $\mathbf{u}_1$  is chosen as  $\mathbf{u}_1 = 1/\sqrt{LP_s} [s_1, s_2, \dots, s_L]^H$ .

Let

$$\mathbf{Y}_{M \times L} = \mathbf{X}_{M \times L} \mathbf{U}_{L \times L} \quad (102)$$

and

$$\mathbf{Z}_{M \times L} = \mathbf{E}_{M \times L} \mathbf{U}_{L \times L}. \quad (103)$$

Inserting (100) into (102) gives

$$\mathbf{Y} = \left[ \sqrt{LP_s} \beta \mathbf{a}, \mathbf{0}_{M \times (L-1)} \right] + \mathbf{Z}. \quad (104)$$

From (7), the first column of  $\mathbf{Y}$

$$\mathbf{y}_1 = \sqrt{LP_s} \beta \mathbf{a} + \mathbf{z}_1 = \sqrt{\frac{L}{P_s}} \bar{\mathbf{x}} \quad (105)$$

and

$$\mathbf{y}_l = \mathbf{z}_l, \quad l = 2, \dots, L \quad (106)$$

where  $\mathbf{y}_l$  (or  $\mathbf{z}_l$ ) denotes the  $l$ th column of  $\mathbf{Y}$  (or  $\mathbf{Z}$ ). Since  $\mathbf{U}\mathbf{U}^H = \mathbf{I}$ ,

$$\mathbf{Y}\mathbf{Y}^H = \mathbf{X}\mathbf{U}\mathbf{U}^H\mathbf{X}^H = \mathbf{X}\mathbf{X}^H. \quad (107)$$

It follows from (6), (10), and (107) that

$$\begin{aligned} L\mathbf{T} &= \mathbf{X}\mathbf{X}^H - \frac{L}{P_s} \bar{\mathbf{x}}\bar{\mathbf{x}}^H \\ &= \mathbf{Y}\mathbf{Y}^H - \mathbf{y}_1\mathbf{y}_1^H \\ &= \sum_{l=2}^L \mathbf{y}_l\mathbf{y}_l^H \\ &= \sum_{l=2}^L \mathbf{z}_l\mathbf{z}_l^H. \end{aligned} \quad (108)$$

Hence,  $\mathbf{T}$  is a function of  $\mathbf{z}_l, l = 2, 3, \dots, L$ , whereas  $\bar{\mathbf{x}} = \sqrt{P_s/L} \mathbf{z}_1$ .

We now show that  $\{\mathbf{z}_l\}_{l=1}^L$  are statistically independent of each other. The cross correlation matrix of any two columns of  $\mathbf{Z}$  has the form

$$\mathbf{B} = \mathcal{E}[\mathbf{z}_i\mathbf{z}_j^H] = \mathcal{E}[\mathbf{E}\mathbf{u}_i\mathbf{u}_j^H\mathbf{E}^H] \quad (109)$$

which is an  $M \times M$  matrix with the  $(k, l)$ th element

$$\begin{aligned} [\mathbf{B}]_{kl} &= \mathcal{E}[\tilde{\mathbf{e}}_k^H \mathbf{u}_i \mathbf{u}_j^H \tilde{\mathbf{e}}_l] \\ &= \text{tr}\{[\mathbf{u}_i \mathbf{u}_j^H] \mathcal{E}[\tilde{\mathbf{e}}_l \tilde{\mathbf{e}}_k^H]\} \end{aligned} \quad (110)$$

where  $\text{tr}(\cdot)$  denotes the trace, and  $\tilde{\mathbf{e}}_k^H$  denotes the  $k$ th row of  $\mathbf{E}$ . To obtain (110), we have used  $\text{tr}(\mathbf{A}\mathbf{B}) = \text{tr}(\mathbf{B}\mathbf{A})$ . Next, note that

$$\begin{aligned} [\mathcal{E}[\tilde{\mathbf{e}}_l \tilde{\mathbf{e}}_k^H]]_{ij} &= \mathcal{E}[e_{li}^* e_{kj}] \\ &= q_{kl} \delta_{ij}, \end{aligned} \quad (111)$$

where  $q_{kl}$  is the  $(k, l)$ th element of the covariance matrix  $\mathbf{Q}$ , and  $\delta_{ij}$  is the Kronecker delta function defined in (68). Substituting (111) into (110) yields

$$[\mathbf{B}]_{kl} = \text{tr}\{[\mathbf{u}_i \mathbf{u}_j^H] q_{kl} \mathbf{I}_L\} \quad (112)$$

$$= q_{kl} \delta_{ij}, \quad \text{for } k, l = 1, 2, \dots, M. \quad (113)$$

Therefore

$$\mathcal{E}[\mathbf{z}_i \mathbf{z}_j^H] = \mathbf{Q} \delta_{ij}. \quad (114)$$

Furthermore, based on the circularly symmetric property of  $\{\mathbf{e}_l\}_{l=1}^L$ , it is easy to show that  $\mathcal{E}[\mathbf{z}_i \mathbf{z}_j^T] = \mathbf{0}$  for  $1 \leq i, j \leq L$ . Since  $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_L$  are Gaussian random vectors, they are statistically independent of each other. Since  $\mathbf{T}$  is a function of  $\mathbf{z}_l, l = 2, \dots, L$ , and  $\bar{\mathbf{x}}$  is a function of  $\mathbf{z}_1$ , we conclude that  $\mathbf{T}$  and  $\bar{\mathbf{x}}$  are statistically independent of each other.

#### APPENDIX C CRAMÉR–RAO BOUND

Let  $\boldsymbol{\eta}$  be a vector containing all of the real-valued unknowns in the data model in (1). Let  $\boldsymbol{\mu}_l = \mathbf{a} \beta s_l$ . Then, the Fisher information matrix (FIM) for  $\boldsymbol{\eta}$  is [2]

$$\begin{aligned} \text{FIM}_{ij}(\boldsymbol{\eta}) &= L \text{tr} \left( \mathbf{Q}^{-1} \frac{\partial \mathbf{Q}}{\partial \eta_i} \mathbf{Q}^{-1} \frac{\partial \mathbf{Q}}{\partial \eta_j} \right) \\ &\quad + 2 \text{Re} \sum_{l=1}^L \left[ \left( \frac{\partial \boldsymbol{\mu}_l}{\partial \eta_i} \right)^H \mathbf{Q}^{-1} \left( \frac{\partial \boldsymbol{\mu}_l}{\partial \eta_j} \right) \right] \end{aligned} \quad (115)$$

where  $\eta_i$  denotes the  $i$ th element of  $\boldsymbol{\eta}$ . Because  $\boldsymbol{\mu}_l$  and  $\mathbf{Q}$  depend on different elements of  $\boldsymbol{\eta}$ ,  $\text{FIM}(\boldsymbol{\eta})$  will be block diagonal with respect to  $[\text{Re}\{\beta\} \quad \text{Im}\{\beta\}]^T$  and the elements of  $\mathbf{Q}$ , where  $\text{Re}$  and  $\text{Im}$  denote the real and imaginary parts of a complex variable, respectively. Hence, the first term of (115) does not affect the CRB of  $\beta$ , and

$$\begin{aligned} &\text{FIM}([\text{Re}\{\beta\} \quad \text{Im}\{\beta\}]^T) \\ &= 2 \text{Re} \begin{bmatrix} \mathbf{a}^H \mathbf{Q}^{-1} \mathbf{a} \sum_{l=1}^L |s_l|^2 & j \mathbf{a}^H \mathbf{Q}^{-1} \mathbf{a} \sum_{l=1}^L |s_l|^2 \\ -j \mathbf{a}^H \mathbf{Q}^{-1} \mathbf{a} \sum_{l=1}^L |s_l|^2 & \mathbf{a}^H \mathbf{Q}^{-1} \mathbf{a} \sum_{l=1}^L |s_l|^2 \end{bmatrix} \end{aligned} \quad (116)$$

$$= 2 \begin{bmatrix} \mathbf{a}^H \mathbf{Q}^{-1} \mathbf{a} \sum_{l=1}^L |s_l|^2 & 0 \\ 0 & \mathbf{a}^H \mathbf{Q}^{-1} \mathbf{a} \sum_{l=1}^L |s_l|^2 \end{bmatrix}. \quad (117)$$

Hence

$$\text{CRB}_\beta = \text{CRB}_{\text{Re}\{\beta\}} + \text{CRB}_{\text{Im}\{\beta\}} \quad (118)$$

$$= \frac{1}{\mathbf{a}^H \mathbf{Q}^{-1} \mathbf{a} \sum_{l=1}^L |s_l|^2} \quad (119)$$

$$= \frac{1}{\mathbf{a}^H \mathbf{Q}^{-1} \mathbf{a} L P_s}. \quad (120)$$

#### APPENDIX D PROOF OF LEMMA 2

We know from (10) that

$$\hat{\mathbf{R}} = \mathbf{T} + \frac{\bar{\mathbf{x}}\bar{\mathbf{x}}^H}{P_s}. \quad (121)$$

Using the matrix inversion lemma on (121) gives

$$\hat{\mathbf{R}}^{-1} = \mathbf{T}^{-1} - \frac{\mathbf{T}^{-1}\bar{\mathbf{x}}\bar{\mathbf{x}}^H\mathbf{T}^{-1}}{P_s + \bar{\mathbf{x}}^H\mathbf{T}^{-1}\bar{\mathbf{x}}}. \quad (122)$$

Substituting (122) into (8), we have

$$\begin{aligned} \hat{\beta}_{\text{Capon}} &= \frac{\mathbf{a}^H\mathbf{T}^{-1}\bar{\mathbf{x}} - \frac{\mathbf{a}^H\mathbf{T}^{-1}\bar{\mathbf{x}}\bar{\mathbf{x}}^H\mathbf{T}^{-1}\bar{\mathbf{x}}}{P_s + \bar{\mathbf{x}}^H\mathbf{T}^{-1}\bar{\mathbf{x}}}}{P_s(\mathbf{a}^H\mathbf{T}^{-1}\mathbf{a} - \frac{\mathbf{a}^H\mathbf{T}^{-1}\bar{\mathbf{x}}\bar{\mathbf{x}}^H\mathbf{T}^{-1}\mathbf{a}}{P_s + \bar{\mathbf{x}}^H\mathbf{T}^{-1}\bar{\mathbf{x}}})} \\ &= \hat{\beta}_{ML} \frac{1 - \frac{\bar{\mathbf{x}}^H\mathbf{T}^{-1}\bar{\mathbf{x}}}{P_s + \bar{\mathbf{x}}^H\mathbf{T}^{-1}\bar{\mathbf{x}}}}{1 - \frac{\mathbf{a}^H\mathbf{T}^{-1}\bar{\mathbf{x}}\bar{\mathbf{x}}^H\mathbf{T}^{-1}\mathbf{a}}{\mathbf{a}^H\mathbf{T}^{-1}\mathbf{a}(P_s + \bar{\mathbf{x}}^H\mathbf{T}^{-1}\bar{\mathbf{x}})}} \\ &= \hat{\beta}_{ML} \frac{1}{1 + \frac{(\bar{\mathbf{x}}^H\mathbf{T}^{-1}\bar{\mathbf{x}} - \frac{\mathbf{a}^H\mathbf{T}^{-1}\bar{\mathbf{x}}\bar{\mathbf{x}}^H\mathbf{T}^{-1}\mathbf{a}}{P_s})}{\mathbf{a}^H\mathbf{T}^{-1}\mathbf{a}}} \end{aligned} \quad (123)$$

$$= \hat{\beta}_{ML} \frac{1}{1+u} \quad (124)$$

where  $u$  is defined in (25) and by the Cauchy-Schwartz inequality  $u \geq 0$ . Hence, the lemma is proved.

#### APPENDIX E PROOF OF LEMMA 3

We rewrite the  $\bar{\mathbf{x}}$  in (7) as

$$\bar{\mathbf{x}} = P_s\mathbf{a}\beta + \sqrt{P_s}\bar{\mathbf{e}} \quad (125)$$

where

$$\bar{\mathbf{e}} = \frac{1}{L\sqrt{P_s}} \sum_{l=1}^L \mathbf{e}_l s_l^* \quad (126)$$

and  $\bar{\mathbf{e}} \sim N(0, \mathbf{Q}/L)$ . Then, (25) can be reduced to

$$u = \bar{\mathbf{e}}^H \left( \mathbf{T}^{-1} - \frac{\mathbf{T}^{-1}\mathbf{a}\mathbf{a}^H\mathbf{T}^{-1}}{\mathbf{a}^H\mathbf{T}^{-1}\mathbf{a}} \right) \bar{\mathbf{e}}. \quad (127)$$

From (14) and (126), we get

$$\hat{\beta}_{ML} - \beta = \frac{\mathbf{a}^H\mathbf{T}^{-1}\bar{\mathbf{e}}}{\sqrt{P_s}\mathbf{a}^H\mathbf{T}^{-1}\mathbf{a}}. \quad (128)$$

According to the conditional expectation rule

$$\mathcal{E} \left[ \lambda^n \left( \hat{\beta}_{ML} - \beta \right) \right] = \mathcal{E}_{\mathbf{T}} \left\{ \mathcal{E}_{\bar{\mathbf{e}}|\mathbf{T}} \left[ \lambda^n \left( \hat{\beta}_{ML} - \beta \right) \right] \right\}. \quad (129)$$

Since  $\bar{\mathbf{e}}$  and  $\mathbf{T}$  are statistically independent of each other, the conditional probability density function  $f_{\bar{\mathbf{e}}|\mathbf{T}}(\bar{\mathbf{e}}|\mathbf{T}) = f_{\bar{\mathbf{e}}}(\bar{\mathbf{e}})$  is an even function of  $\bar{\mathbf{e}}$ . It follows from (127), (128), and (24) that

$$\mathcal{E}_{\bar{\mathbf{e}}|\mathbf{T}} \left[ \lambda^n \left( \hat{\beta}_{ML} - \beta \right) \right] = 0 \quad (130)$$

since  $\lambda^n \left( \hat{\beta}_{ML} - \beta \right)$  is an odd function of  $\bar{\mathbf{e}}$ . From (129), we obtain

$$\mathcal{E} \left[ \lambda^n \left( \hat{\beta}_{ML} - \beta \right) \right] = 0, \quad \text{for } n = 0, 1, 2, \dots \quad (131)$$

Hence, the lemma is proved.

#### APPENDIX F PROOF OF LEMMA 4

Let  $\tilde{\mathbf{E}} = \mathbf{Y} - \beta\mathbf{G}$ . Then, we have  $(\mathbf{Y}\mathbf{P} - \beta\mathbf{G})(\mathbf{Y}\mathbf{P} - \beta\mathbf{G})^H = \tilde{\mathbf{E}}\tilde{\mathbf{P}}\tilde{\mathbf{E}}^H$ . Let  $\mathbf{P} = \mathbf{V}\mathbf{\Gamma}\mathbf{V}^H$  denote the singular value decomposition of  $\mathbf{P}$ . Since  $\mathbf{P}$  is an orthogonal projection matrix, the first  $r$  ( $r = \text{rank}(\mathbf{P}) \leq M$ ) diagonal elements of  $\mathbf{\Gamma}$  are ones, and the rest are zeros. Hence, we get  $\mathbf{P} = \tilde{\mathbf{V}}_{L \times r} \tilde{\mathbf{V}}^H$ , where  $\tilde{\mathbf{V}} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r]$  consists of the first  $r$  columns of the unitary matrix  $\mathbf{V}$ . Let  $\tilde{\mathbf{e}}_i^H$  denote the  $i$ th row of  $\tilde{\mathbf{E}}$ . Then

$$\left[ \tilde{\mathbf{E}}\tilde{\mathbf{P}}\tilde{\mathbf{E}}^H \right]_{ii} = \sum_{j=1}^r |\tilde{\mathbf{e}}_i^H \mathbf{v}_j|^2. \quad (132)$$

Because  $\mathcal{E}\{\tilde{\mathbf{e}}_i\tilde{\mathbf{e}}_i^H\} \rightarrow q_{ii}\mathbf{I}_L$  as  $L \rightarrow \infty$ , where  $q_{ii}$  denotes the  $i$ th diagonal element of the  $\mathbf{Q}$  defined in (67), we have  $\mathcal{E}\{|\tilde{\mathbf{e}}_i^H \mathbf{v}_j|^2\} \rightarrow q_{ii}\|\mathbf{v}_j\|^2 = q_{ii}$  as  $L \rightarrow \infty$ . It follows from (132) that the diagonal elements of  $\tilde{\mathbf{E}}\tilde{\mathbf{P}}\tilde{\mathbf{E}}^H$  are bounded. Since  $\tilde{\mathbf{E}}\tilde{\mathbf{P}}\tilde{\mathbf{E}}^H$  is a positive semi-definite matrix, we have  $\tilde{\mathbf{E}}\tilde{\mathbf{P}}\tilde{\mathbf{E}}^H = O(1)$ . However,  $\mathbf{T}_y = \tilde{\mathbf{E}}\mathbf{P}^\perp\tilde{\mathbf{E}}^H = \tilde{\mathbf{E}}\tilde{\mathbf{E}}^H - \tilde{\mathbf{E}}\tilde{\mathbf{P}}\tilde{\mathbf{E}}^H = O(L)$ . Hence,  $\mathbf{C} = (\mathbf{Y}\mathbf{P} - \beta\mathbf{G})(\mathbf{Y}\mathbf{P} - \beta\mathbf{G})^H\mathbf{T}_y^{-1} = O(1/L)$ . Let  $\{\lambda_m\}_{m=1}^M$  denote the eigenvalues of  $\mathbf{C}$ . Then

$$F_1 = |(\mathbf{Y}\mathbf{P} - \beta\mathbf{G})(\mathbf{Y}\mathbf{P} - \beta\mathbf{G})^H\mathbf{T}_y^{-1} + \mathbf{I}| \quad (133)$$

$$= \prod_{m=1}^M (1 + \lambda_m). \quad (134)$$

Since  $\lambda_m \ll 1$  for large  $L$ , (134) can be approximated as

$$F_1 \approx 1 + \sum_{m=1}^M \lambda_m. \quad (135)$$

[Note that if  $\lambda_m = 0$  for  $m = 2, 3, \dots, M$ , which is true for the case of a known damped or undamped sinusoidal signal, (134) and (135) are exactly equal.] Thus, minimizing  $F_1$  is asymptotically (for large  $L$ ) equivalent to minimizing

$$F_2 = \text{tr} \left[ (\mathbf{Y}\mathbf{P} - \beta\mathbf{G})^H\mathbf{T}_y^{-1}(\mathbf{Y}\mathbf{P} - \beta\mathbf{G}) \right]. \quad (136)$$

The lemma is proved.

#### ACKNOWLEDGMENT

The authors are grateful to two anonymous reviewers who provided us with a large number of detailed suggestions for improving the submitted manuscript.

#### REFERENCES

- [1] A. L. Swindlehurst and P. Stoica, "Maximum likelihood methods in radar array signal processing," *Proc. IEEE*, vol. 86, pp. 421–441, Feb. 1998.
- [2] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1997.
- [3] J. Li and R. T. Compton Jr., "Maximum likelihood angle estimation for signals with known waveforms," *IEEE Trans. Signal Processing*, vol. 41, pp. 2850–2862, Sept. 1993.
- [4] J. Li, B. Halder, P. Stoica, and M. Viberg, "Computationally efficient angle estimation for signals with known waveforms," *IEEE Trans. Signal Processing*, vol. 43, pp. 2154–2163, Sept. 1995.
- [5] A. Zeira and B. Friedlander, "Direction of arrival estimation using parametric signal models," *IEEE Trans. Signal Processing*, vol. 44, pp. 339–350, Feb. 1996.

- [6] M. Cedervall and R. L. Moses, "Efficient maximum likelihood DOA estimation for signals with known waveforms in the presence of multipath," *IEEE Trans. Signal Processing*, vol. 45, pp. 808–811, Mar. 1997.
- [7] M. Wax and A. Leshem, "Joint estimation of time delays and directions of arrival of multiple reflections of a known signal," *IEEE Trans. Signal Processing*, pp. 2477–2484, Oct. 1997.
- [8] A. L. Swindlehurst, "Time delay and spatial signature estimation using known asynchronous signals," *IEEE Trans. Signal Processing*, vol. 46, pp. 449–462, Feb. 1998.
- [9] A. Jakobsson, A. Swindlehurst, and P. Stoica, "Subspace-based estimation of time delays and doppler shifts," *IEEE Trans. Signal Processing*, vol. 46, pp. 2472–2483, Sept. 1998.
- [10] U. Spagnolini, "An iterative quadratic method for high resolution delay estimation with known waveform," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, pp. 1134–1137, Mar. 2000.
- [11] M. Rowe and J. Smith, "Mine detection by nuclear quadrupole resonance," in *EUREL International Conference*, vol. 431, 1996, pp. 62–66.
- [12] S. Tantum, L. Collines, and L. Carin, "Signal processing for NQR discrimination of buried landmines," *Proc. SPIE*, vol. 3710, 1999.
- [13] I. S. Reed, J. D. Mallett, and L. E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-10, pp. 853–863, Nov. 1974.
- [14] E. J. Kelly, "An adaptive detection algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 22, pp. 115–127, Mar. 1986.
- [15] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, pp. 1408–1418, Aug. 1969.
- [16] N. R. Goodman, "Statistical analysis based on a certain multi-variate complex Gaussian distribution," *Ann. Math. Stat.*, vol. 34, pp. 152–177, Mar. 1963.
- [17] J. Capon and N. R. Goodman, "Probability distribution for estimators of the frequency-wavenumber spectrum," *Proc. IEEE*, vol. 58, pp. 1785–1786, Oct. 1970.
- [18] B. Van Veen, "Adaptive convergence of linearly constrained beamformers based on the sample covariance matrix," *IEEE Trans. Signal Processing*, vol. 39, pp. 1470–1473, June 1991.
- [19] J. L. Krolik and D. N. Swingler, "On the mean-square error performance of adaptive minimum variance beamformers based on the sample covariance matrix," *IEEE Trans. Signal Processing*, vol. 42, pp. 445–448, Feb. 1994.
- [20] J. H. Michels, P. Varshney, and D. D. Weiner, "Synthesis of correlated multichannel random processes," *IEEE Trans. Signal Processing*, vol. 42, pp. 367–375, Feb. 1994.
- [21] W. I. Zangwill, *Nonlinear Programming: A Unified Approach*. Englewood Cliffs, NJ: Prentice-Hall, 1969.
- [22] S. L. Marple Jr., *Digital Spectral Analysis With Applications*. Upper Saddle River, NJ: Prentice-Hall, 1987.
- [23] T. Söderström and P. Stoica, *System Identification*. London, U.K.: Prentice-Hall Int., 1989.
- [24] S. R. Searle, *Matrix Algebra Useful for Statistics*. New York: Wiley, 1982.



**Petre Stoica** (F'94) received the D.Sc. degree in automatic control from the Polytechnic Institute of Bucharest (BPI), Bucharest, Romania, in 1979 and an honorary doctorate degree in science from Uppsala University (UU), Uppsala, Sweden, in 1993.

He is Professor of system modeling with the Department of Systems and Control at UU. Previously, he was a Professor of system identification and signal processing with the Faculty of Automatic Control and Computers at BPI. He held longer visiting positions with Eindhoven University of Technology, Eindhoven, The Netherlands; Chalmers University of Technology, Gothenburg, Sweden (where he held a Jubilee Visiting Professorship); UU; The University of Florida, Gainesville; and Stanford University, Stanford, CA. His main scientific interests are in the areas of system identification, time series analysis and prediction, statistical signal and array processing, spectral analysis, wireless communications, and radar signal processing. He has published seven books, ten book chapters, and some 500 papers in archival journals and conference records on these topics. The most recent book he co-authored, with R. Moses, is entitled *Introduction to Spectral Analysis* (Englewood Cliffs, NJ: Prentice-Hall, 1997). Recently, he edited two books on signal processing advances in wireless communications and mobile communications (Englewood Cliffs, NJ: Prentice-Hall, 2001). He is on the editorial boards of five journals in the field: *Journal of Forecasting; Signal Processing; Circuits, Signals, and Signal Processing; Digital Signal Processing—A Review Journal; and Multidimensional Systems and Signal Processing*. He was a Co-Guest Editor for several special issues on system identification, signal processing, spectral analysis, and radar for some of the aforementioned journals, as well as for PROCEEDINGS OF THE IEEE.

Dr. Stoica was co-recipient of the IEEE ASSP Senior Award for a paper on statistical aspects of array signal processing. He was also recipient of the Technical Achievement Award of the IEEE Signal Processing Society for fundamental contributions to statistical signal processing with applications in time-series analysis, system identification, and array signal processing. In 1998, he was the recipient of a Senior Individual Grant Award of the Swedish Foundation for Strategic Research. He was also co-recipient of the 1998 EURASIP Best Paper Award for Signal Processing for a work on parameter estimation of exponential signals with time-varying amplitude, a 1999 IEEE Signal Processing Society Best Paper Award for a paper on parameter and rank estimation of reduced-rank regression, a 2000 IEEE Third Millennium Medal, and the 2000 W. R. G. Baker Prize Paper Award for a paper on maximum likelihood methods for radar. He was a member of the international program committees of many topical conferences. From 1981 to 1986, he was the Director of the International Time-Series Analysis and Forecasting Society, and he has been a member of the IFAC Technical Committee on Modeling, Identification, and Signal Processing since 1994. He is also a member of the Romanian Academy and a Fellow of the Royal Statistical Society.



**Jian Li** (S'87–M'91–SM'97) received the M.Sc. and Ph.D. degrees in electrical engineering from The Ohio State University (OSU), Columbus, in 1987 and 1991, respectively.

From April 1991 to June 1991, she was an Adjunct Assistant Professor with the Department of Electrical Engineering, OSU. From July 1991 to June 1993, she was an Assistant Professor with the Department of Electrical Engineering, University of Kentucky, Lexington. Since August 1993, she has been with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, where she is currently a Professor. Her current research interests include spectral estimation, array signal processing, and their applications.

Dr. Li is a member of Sigma Xi and Phi Kappa Phi. She received the 1994 National Science Foundation Young Investigator Award and the 1996 Office of Naval Research Young Investigator Award. She was an Executive Committee Member of the 2002 International Conference on Acoustics, Speech, and Signal Processing, Orlando, FL, May 2002. She has been an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING since 1999. She is presently a member of the Signal Processing Theory and Methods (SPTM) Technical Committee of the IEEE Signal Processing Society.



**Yi Jiang** (S'02) was born in Jiangsu Province, China, in 1978. He received the B.Sc. degree from the University of Science and Technology of China (USTC), Hefei, in 2001 and the M.Sc. degree from the University of Florida (UF), Gainesville, in 2003, both in electrical engineering. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, UF.

His research interests include spectral estimation, array signal processing, and information theory.