Maximum Likelihood Methods in Radar Array Signal Processing

A. LEE SWINDLEHURST, MEMBER, IEEE, AND PETRE STOICA, FELLOW, IEEE

We consider robust and computationally efficient maximum likelihood algorithms for estimating the parameters of a radar target whose signal is observed by an array of sensors in interference with unknown second-order statistics. Two data models are described: one that uses the target direction of arrival and signal amplitude as parameters and one that is a simpler, unstructured model that uses a generic target "spatial signature." An extended invariance principle is invoked to show how the less accurate maximum likelihood estimates obtained from the simple model may be refined to achieve asymptotically the performance available using the structured model. The resulting algorithm requires two one-dimensional (1-D) searches rather than a two-dimensional search, as with previous approaches for the structured case. If a uniform linear array is used, only a single 1-D search is needed. A generalized likelihood ratio test for target detection is also derived under the unstructured model. The principal advantage of this approach is that it is computationally simple and robust to errors in the model (calibration) of the array response.

Keywords*—*Antenna arrays, calibration errors, direction of arrival, Doppler frequency, extended invariance principle, likelihood ratio detection test, maximum likelihood estimation, radar cross section, radar signal processing, robust estimation and detection, structured/unstructured modeling.

I. INTRODUCTION

In active radar systems, the primary goal is to detect the presence and estimate the parameters of targets in the presence of noise, ground clutter, and electronic countermeasures (e.g., jamming). For airborne radar applications, the parameters of interest include the target signal amplitude (related to its radar cross section), direction of arrival (DOA), range, and Doppler frequency. Even when only a single target is present, this detection and estimation problem is quite daunting due to the large volume of the parameter space that must be searched. For this reason, computational ease and efficiency is of the utmost importance.

Manuscript received November 25, 1996; revised May 29, 1997. This work was supported by the Office of Naval Research under Contract N00014-96-1-0934 and by the Swedish Individual Grant program of the Swedish Foundation for Strategic Research.

A. L. Swindlehurst is with the Department of Electrical and Computer Engineering, Brigham Young University, Provo, UT 84602 USA.

P. Stoica is with the Systems and Control Group, Uppsala University, Uppsala SE-751 03 Sweden.

Publisher Item Identifier S 0018-9219(98)01295-X.

Standard solutions to this problem involve the use of classical space-time filters, either data independent (e.g., as with a delay-and-sum beam former) or "adaptive" (e.g., using linear constraints, maximum signal-to-interference-plus-noise ratio criteria, etc.) [1]. A discussion of the application of such techniques to radar signal processing can be found in [2]–[5] and references therein. Maximum likelihood (ML) approaches have not been extensively considered for this problem, mainly because they are perceived to be too computationally complex and because there is some reluctance in accepting the required modeling assumptions. Recent work has shown that for a single target source in Gaussian interference with unknown spatial covariance, the ML solution can be obtained via a two-dimensional (2-D) search over target DOA and Doppler, followed by a generalized likelihood ratio test (GLRT) [6], [7].

One of the goals of this paper is to demonstrate that the 2-D search required for the ML solution can be replaced with two simpler one-dimensional (1-D) searches without affecting the asymptotic accuracy of the estimates. If the array is uniform and linear, then only a single 1-D search is required. The resulting computational savings have the potential for making the ML approach more feasible for radar applications. The key idea behind the simplification of the algorithm is the use of the so-called *extended invariance principle* (EXIP) in estimation first introduced in [8]. This technique involves reparameterizing the ML criterion in a way that admits a simple solution and then refining that solution by means of a weighted least squares (WLS) fit. In particular, we initially use an unstructured model for the array response instead of one parameterized by the DOA of the target signal. Under this model, the ML solution reduces to a search over only the target Doppler frequency. Refined estimates of the target parameters, including DOA, are then obtained via WLS, where the optimal weighting is simply the Fisher information matrix (FIM) corresponding to the unstructured ML criterion. As will be seen in Section IV, this second step involves a search only for the DOA parameter.

An additional advantage of using an unstructured model for the array is that it provides robustness to errors in

the model for the array response. The array response is usually determined either by empirical measurements, a process referred to as array *calibration*, or by making certain assumptions about the sensors in the array and their geometry (e.g., identical sensors in known locations). Unfortunately, an array cannot be perfectly calibrated, and idealized assumptions made about the array geometry are never satisfied in practice. Due to changes in antenna location, temperature, and the surrounding environment, the response of the array may be significantly different than when it was last calibrated. Furthermore, the calibration measurements themselves are subject to gain and phase errors, and they can only be obtained for discrete angles, thus necessitating interpolation techniques for uncalibrated directions. For the case of analytically calibrated arrays of nominally identical, identically oriented elements, errors result since the elements are not really identical and their locations are not precisely known.

Depending on the degree to which the actual array response differs from its model, detection and estimation performance may be significantly degraded. For this reason, we develop a GLRT for target detection under the unstructured array model, which by its very nature does not depend on the availability of array calibration data. We then compare the unstructured GLRT with the standard test via simulation for various levels of array perturbation. Even in the absence of such perturbations, the unstructured GLRT may be useful in providing a more rapid "initial scan" of the environment prior to application of the structured model. For example, one could apply the unstructured GLRT with a relatively high false-alarm rate and then use the complete model to process data sets that showed an initial detection. We also show that taking array calibration errors into account in the estimation of the target parameters can improve performance as well. In particular, we demonstrate that information about the second-order statistics of the array perturbation can be used to "regularize" the EXIP solution and make it robust to such errors.

Before describing the proposed detection and estimation algorithms, we introduce the data model assumed in this work and the EXIP in Sections II and III, respectively. The two-step estimation procedure employing the EXIP is then discussed in Section IV. The GLRT for both the structured and unstructured models is outlined in Section V and the results of some simulations are included in Section VI.

II. Radar Data Model

Suppose an m-element antenna array transmits a series of N pulses using a certain set of transmit weights. The transmit weights are usually chosen to steer a beam in some nominal direction, say, θ_t. The output of each receiving antenna element is sampled at a number of time instants following each pulse, the time difference between each sampling instant and the pulse transmission corresponding to the distance, or range, between the array and a potential scattering source. The data collected for a given set of transmit weights is referred to as a *coherent processing interval* (CPI). During a given CPI, a typical radar system collects the N samples from each range "bin" and interrogates them for the presence/absence of a target. If present, the DOA, Doppler frequency, and radar cross section of the detected target are estimated. This procedure is continually repeated for every range bin from numerous CPI's in an attempt to provide constant surveillance over a wide area.

A. Structured Array Model

A target present in a particular range bin during some CPI may be modeled as producing the following base-band vector signal (after pulse compression and demodulation) [5]–[7]:

$$\text{target signal} = b_0\boldsymbol{a}(\theta_0)e^{j\omega_0 t}, \qquad t = 1, \cdots, N \tag{1}$$

where b_0 is the complex amplitude of the signal, ω_0 is the Doppler shift due to the relative motion between the array platform and the target, and $\boldsymbol{a}(\theta_0)$ is the response of the array to a plane wave arriving from azimuthal direction[1] θ_0. The vector $\boldsymbol{a}(\theta)$ is assumed to be scaled such that its first element is unity for all θ.

Assuming that at most one target is present in any given range bin, the sampled vector output of the array at time t, or "snapshot," may be written as

$$\boldsymbol{x}(t) = b_0\boldsymbol{a}(\theta_0)e^{j\omega_0 t} + \boldsymbol{e}(t) \in \mathbb{C}^m, \qquad t = 1, \cdots, N \tag{2}$$

where b_0 is taken to be zero if no target is present and $\boldsymbol{e}(t)$ represents interference due to clutter, receiver noise, and jamming. Since $\boldsymbol{e}(t)$ depends on so many physical variables, deriving a deterministic model for it is problematic and usually impractical. To begin with, we choose to model $\boldsymbol{e}(t)$ as a stationary, temporally white Gaussian random process over the N samples of the CPI, with zero-mean and unknown spatial covariance $\boldsymbol{Q}$. As explained below, extensions to the case where $\boldsymbol{e}(t)$ is not temporally white will also be considered. The Gaussian hypothesis is made for the convenience of the ML approach presented later, but it is not critical in that the ML method derived under this assumption provides accurate estimates for many other data/noise distributions as well.

The above model for $\boldsymbol{e}(t)$ is rather simplistic, since often the clutter component cannot be considered to be temporally white. When the interference term is dominated by clutter, a common approach [5] is to assume that an estimate of both the temporal and spatial statistics of $\boldsymbol{e}(t)$ is available from data in adjacent target-free range bins and CPI's. However, forming an estimate of the full space-time covariance of $\boldsymbol{e}(t)$ typically involves hundreds of correlation measurements, a large amount of secondary data, and significant computation. In this paper, we propose a computationally efficient alternative. In particular, we suggest a two-stage operation: 1) estimating a vector autoregressive (VAR) model for the temporal variation of the

[1] In principle, θ_0 could be a vector representing both azimuth and elevation DOA's. In many applications where this model is used, however, the distance between the (usually airborne) target and array is large enough that both may be assumed to be in the same horizontal plane.

clutter using secondary data from adjacent range and CPI cells and (2) using the estimated VAR model to prewhiten the data prior to processing. The computational advantage of this approach stems from the fact that the model for the temporal statistics of the clutter is embedded in a relatively small number of AR parameters rather than the comparatively large number of temporal covariance lags that may be necessary for the methods of [5]. Additionally, the proposed algorithms estimate the spatial covariance $\boldsymbol{Q}$ along with the target parameters, and thus no prior estimate of $\boldsymbol{Q}$ is required. In the discussion that follows, we will use the simple assumption of temporal whiteness to derive the algorithms and other main results of this paper. The generalization of these results using the VAR approach is described in Section IV-C.

With the model of (2) in hand, the problem considered in this paper may be succinctly stated as follows.

Given a collection of samples $\boldsymbol{x}(1), \cdots, \boldsymbol{x}(N)$ as in (2), determine whether or not a target is present, and if present, estimate the parameters

$$\boldsymbol{\eta}_0 = \{\mathrm{Re}\{b_0\}, \mathrm{Im}\{b_0\}, \omega_0, \theta_0, \boldsymbol{Q}\} \in D_\eta.$$

The set D_η represents the domain of the generic parameter vector $\boldsymbol{\eta} = \{\mathrm{Re}\{b\}, \mathrm{Im}\{b\}, \omega, \theta, \boldsymbol{Q}\}$. While the above problem statement implies that detection precedes estimation, most solutions to the problem operate in reverse: the parameters are estimated as though a target were present and then a statistical test is performed to compare the likelihood of the noise-only and signal-plus-noise models.

B. An Unstructured Array Model

In the discussion above, we have written the array response explicitly in terms of θ_0. Of course, to estimate θ_0, the vector valued function $\boldsymbol{a}(\theta)$ must be known for all θ of interest. The function $\boldsymbol{a}(\theta)$ can be obtained by means of an experimental or analytical calibration of the array, but such calibration procedures are never entirely accurate due to errors in the gain and phase response of the antenna elements, mutual coupling, quantization and interpolation errors, and drifts in the receiver electronic characteristics due to changes in weather, environment, etc. In some cases, these errors can cause the actual and calibrated array responses to be quite different, which can lead to a significant loss in detection and estimation performance.

Another difficulty with the parameterization of (2) is the nonlinear dependence of the data on the parameters θ_0 and ω_0. As a result, most algorithms for estimating θ_0 and ω_0 require some type of nonlinear 2-D search procedure that may be too time consuming to implement. These problems can be alleviated through the use of a less structured but still identifiable model. The price paid is, of course, a loss of performance in situations where the model of (2) holds, but that loss can be compensated for by the approach to be described in Section IV. In particular, consider the following model for the data, where the signal amplitude and array response have been combined into a single unstructured vector term:

$$\boldsymbol{x}(t) = \boldsymbol{\alpha}_0 e^{j\omega_0 t} + \boldsymbol{e}(t), \qquad t = 1, \cdots, N. \tag{3}$$

The vector $\boldsymbol{\alpha}_0$, referred to as the "spatial signature" of the target, takes the place of $b_0\boldsymbol{a}(\theta_0)$ and in this model is simply treated as a set of m unknown complex constants. The parameter vector for this model will be denoted by $\tilde{\boldsymbol{\eta}} = \{\mathrm{Re}\{\boldsymbol{\alpha}\}, \mathrm{Im}\{\boldsymbol{\alpha}\}, \omega, \boldsymbol{Q}\}$ and is defined on the domain $D_{\tilde{\eta}}$. In the following sections, we discuss the use of (3) together with (2) in developing estimation and detection algorithms that are computationally and statistically efficient and robust to array imperfections.

III. Extended Invariance Principle

Suppose that to estimate the radar parameters described above, we have a loss function (for example, the ML criterion) that can be parameterized in terms of either $\boldsymbol{\eta}$ or $\tilde{\boldsymbol{\eta}}$. Denote the resulting criteria as $V_N(\boldsymbol{\eta})$ and $\tilde{V}_N(\tilde{\boldsymbol{\eta}})$, respectively (the subscript N explicitly indicates that the criteria depend on N data samples). If an (invertible) mapping existed such that

$$\boldsymbol{\eta} = g(\tilde{\boldsymbol{\eta}}) \in D_\eta, \qquad \forall \tilde{\boldsymbol{\eta}} \in D_{\tilde{\eta}} \tag{4}$$

the classical *invariance principle* of estimation theory (e.g., see [9]) could be invoked to prove the equivalence of minimizing $V_N(\boldsymbol{\eta})$ and $\tilde{V}_N(\tilde{\boldsymbol{\eta}})$. Of course, no such function g exists in our problem, since $\boldsymbol{\alpha}$ may be factored as $b\boldsymbol{a}(\theta)$ only on a set of measure zero in $D_{\tilde{\eta}}$. One could choose to constrain the minimization of $\tilde{V}_N(\tilde{\boldsymbol{\eta}})$ over this measure-zero set but there would be no advantage, computational or otherwise, in doing so. Instead, we choose to consider an unconstrained minimization of $\tilde{V}_N(\tilde{\boldsymbol{\eta}})$ and apply the so-called *extended* invariance principle [8], [10] to estimate $\boldsymbol{\eta}$ from the estimated $\tilde{\boldsymbol{\eta}}$.

We present the EXIP by means of the following theorem.

Theorem 1: Define

$$\hat{\boldsymbol{\eta}} = \arg\min_{\eta} V_N(\boldsymbol{\eta}) \tag{5}$$

$$\hat{\tilde{\boldsymbol{\eta}}} = \arg\min_{\tilde{\eta}} \tilde{V}_N(\tilde{\boldsymbol{\eta}}) \tag{6}$$

and assume that a function f exists satisfying

$$\tilde{\boldsymbol{\eta}} = f(\boldsymbol{\eta}) \in D_{\tilde{\eta}} \qquad \forall \boldsymbol{\eta} \in D_\eta. \tag{7}$$

Assume that f is one to one and well defined over the entire domain of its input argument. If [2]

$$\lim_{N\to\infty} \hat{\tilde{\boldsymbol{\eta}}} = \lim_{N\to\infty} f(\hat{\boldsymbol{\eta}}) \tag{8}$$

then

$$\hat{\boldsymbol{\eta}} = \arg\min_{\eta} [\hat{\tilde{\boldsymbol{\eta}}} - f(\boldsymbol{\eta})]^T \boldsymbol{W} [\hat{\tilde{\boldsymbol{\eta}}} - f(\boldsymbol{\eta})] \tag{9}$$

is asymptotically (for large N) equivalent to the estimate $\hat{\boldsymbol{\eta}}$, where

$$\boldsymbol{W} = \mathcal{E}\left\{ \frac{\partial \tilde{V}_N(\tilde{\boldsymbol{\eta}})}{\partial \tilde{\boldsymbol{\eta}} \partial \tilde{\boldsymbol{\eta}}^T} \right\}\Bigg|_{\tilde{\eta}=\hat{\tilde{\eta}}} \tag{10}$$

[2] The most common way in which (8) is satisfied is if both $\hat{\tilde{\boldsymbol{\eta}}}$ and $f(\hat{\boldsymbol{\eta}})$ are consistent estimators of $\bar{\boldsymbol{\eta}}_0$.

and $\mathcal{E}\{\cdot\}$ denotes expectation. The term "asymptotically equivalent" is to be understood to mean that

$$||\hat{\tilde{\boldsymbol{\eta}}} - \hat{\boldsymbol{\eta}}|| = o_p(||\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0||)$$

where $||\cdot||$ is any consistent vector norm and $o_p(\xi)$ denotes a random variable that converges to zero in probability at a rate faster than $\xi \to 0$.

Proof: See [8] and [10]. ■

The idea behind the EXIP is that if the parameterization corresponding to $\tilde{\boldsymbol{\eta}}$ is suitably chosen, it may be much simpler to solve both (6) and (9) rather than (5) directly. The EXIP approach thus involves the following two-step procedure.

1) Minimize the criterion $\tilde{V}_N(\tilde{\boldsymbol{\eta}})$ for the unstructured model to obtain the estimate $\hat{\tilde{\boldsymbol{\eta}}}$.
2) Use the unstructured estimate $\hat{\tilde{\boldsymbol{\eta}}}$ in (9) to find an estimate of the structured parameter vector $\boldsymbol{\eta}$. This refined estimate will asymptotically have the same statistical properties as the structured estimate $\hat{\boldsymbol{\eta}}$ from (5).

In the next section, we show how the EXIP technique applies for the case of ML loss functions and how it leads to a computationally efficient (and asymptotically optimal) estimator for $\boldsymbol{\eta}_0$.

IV. Maximum Likelihood Estimation

A. Structured Model

The negative log-likelihood function for N samples of data from the fully parameterized model is easily shown to be, within a multiplicative and additive constant

$$V_N(\boldsymbol{\eta}) = \log |\boldsymbol{Q}| + \mathrm{Tr}\{\boldsymbol{Q}^{-1}\boldsymbol{C}(b,\theta,\omega)\} \tag{11}$$

where

$$\boldsymbol{C}(b,\theta,\omega) = \frac{1}{N}\sum_{t=1}^{N}(\boldsymbol{x}(t) - b\boldsymbol{a}(\theta)e^{j\omega t})(\boldsymbol{x}(t) - b\boldsymbol{a}(\theta)e^{j\omega t})^*. \tag{12}$$

$|\cdot|$ denotes the determinant operation and $*$ a conjugate transpose. The minimization of (11) with respect to $\boldsymbol{Q}$ and b may be performed explicitly. Using standard matrix calculus results (see, e.g., [11]), the gradient of the criterion with respect to $\boldsymbol{Q}$ is easily shown to be

$$\frac{\partial V_N(\boldsymbol{\eta})}{\partial \boldsymbol{Q}} = \boldsymbol{Q}^{-1} - \boldsymbol{Q}^{-1}\boldsymbol{C}(b,\theta,\omega)\boldsymbol{Q}^{-1}$$

from which it is clear that the ML estimate of $\boldsymbol{Q}$ is given by[3]

$$\hat{\boldsymbol{Q}}_s = \boldsymbol{C}(b,\theta,\omega) \tag{13}$$

where the subscript s denotes a "structured" estimate. When (13) is substituted into (11), the concentrated criterion

$$V_N(b,\theta,\omega) = \log |\boldsymbol{C}(b,\theta,\omega)| + m \tag{14}$$

[3]We assume that $N \geq m$ so that the matrix $\boldsymbol{C}(b,\theta,\omega)$ is invertible (with probability one) at any point in D_η.

results. As shown in [6], [12], and [13], $V_N(b,\theta,\omega)$ may be further concentrated with respect to b, yielding

$$\hat{b}_s = \frac{\boldsymbol{a}^*(\hat{\theta}_s)\tilde{\boldsymbol{R}}_\omega^{-1}\boldsymbol{y}(\hat{\omega}_s)}{\boldsymbol{a}^*(\hat{\theta}_s)\tilde{\boldsymbol{R}}_\omega^{-1}\boldsymbol{a}(\hat{\theta}_s)} \tag{15}$$

$$\hat{\theta}_s, \hat{\omega}_s = \arg\max_{\theta,\omega} \mathcal{V}_N(\theta,\omega) = \arg\max_{\theta,\omega} \frac{\left|\boldsymbol{a}^*(\theta)\hat{\boldsymbol{R}}^{-1}\boldsymbol{y}(\omega)\right|^2}{\left(1 - \boldsymbol{y}^*(\omega)\hat{\boldsymbol{R}}^{-1}\boldsymbol{y}(\omega)\right)\boldsymbol{a}^*(\theta)\hat{\boldsymbol{R}}^{-1}\boldsymbol{a}(\theta)} \tag{16}$$

where

$$\boldsymbol{y}(\omega) = \frac{1}{N}\sum_{t=1}^{N}\boldsymbol{x}(t)e^{-j\omega t} \tag{17}$$

$$\tilde{\boldsymbol{R}}_\omega = \hat{\boldsymbol{R}} - \boldsymbol{y}(\omega)\boldsymbol{y}^*(\omega) \tag{18}$$

$$\hat{\boldsymbol{R}} = \frac{1}{N}\sum_{t=1}^{N}\boldsymbol{x}(t)\boldsymbol{x}^*(t) \tag{19}$$

and $\mathcal{V}_N(\theta,\omega)$ denotes the concentrated criterion. Estimates of θ and ω are obtained by maximizing $\mathcal{V}_N(\theta,\omega)$ using some type of 2-D search algorithm.

A simpler estimation strategy than the 2-D search of (16) is suggested in [7], where it is proposed that θ be set equal to the pointing angle θ_t of the transmit beam and $\mathcal{V}_N(\theta_t,\omega)$ maximized with respect to ω only. This is a reasonable idea since it is likely that a measurable return will be possible only if the target is illuminated by main beam energy. However, such an approach can in general only yield accurate estimates if the transmit beam is sufficiently narrow.

The statistical properties of the above ML estimator are outlined in the following theorem.

Theorem 2: The structured ML estimator defined by (14)–(16) is consistent and statistically efficient. In particular, the asymptotic variance of the ML (signal) parameter estimates

$$\boldsymbol{\eta}_s = \begin{bmatrix} \mathrm{Re}\{b\} \\ \mathrm{Im}\{b\} \\ \theta \\ \omega \end{bmatrix}$$

achieves the Cram ér–Rao bound (CRB) [see (20) at the bottom of the next page] where

$$\beta_N = \frac{(N+1)}{2} \tag{21}$$

$$\gamma_N = \frac{(N+1)(N+\frac{1}{2})}{3} \tag{22}$$

and for brevity we have written $b = b_0, \boldsymbol{a} = \boldsymbol{a}(\theta_0)$, and $\boldsymbol{d} = \partial\boldsymbol{a}(\theta)/\partial\theta|_{\theta=\theta_0}$.

The asymptotic variances of the ML DOA and Doppler estimates are given by

$$\mathcal{E}\left\{(\hat{\theta}_s-\theta_0)^2\right\} = \frac{\boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{a}}{2N|b|^2\left[(\boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{a})(\boldsymbol{d}^*\boldsymbol{Q}^{-1}\boldsymbol{d})-\left|\boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{d}\right|^2\right]} \tag{23}$$

$$\mathcal{E}\{(\hat{\omega}_s-\omega_0)^2\} = \frac{1}{2N|b|^2(\gamma_N-\beta_N^2)\boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{a}} \simeq \frac{6}{N^3|b|^2\boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{a}}. \tag{24}$$

Proof: See Appendix A. ■

B. Unstructured Model

The ML solution for the unstructured model may be similarly derived. The negative log-likelihood function for $\tilde{\boldsymbol{\eta}}$ is of the same form as (11) but with $b\boldsymbol{a}(\theta)$ replaced by $\boldsymbol{\alpha}$. Minimization with respect to $\boldsymbol{Q}$ gives

$$\hat{\boldsymbol{Q}}_u = \frac{1}{N}\sum_{t=1}^{N}\left(\boldsymbol{x}(t)-\boldsymbol{\alpha}e^{j\omega t}\right)\left(\boldsymbol{x}(t)-\boldsymbol{\alpha}e^{j\omega t}\right)^* \tag{25}$$

where the subscript u is used to explicitly denote an estimate obtained using the unstructured model. The criterion resulting from substitution of (25) into (11) may be simplified as follows:

$$\tilde{V}_N(\boldsymbol{\alpha},\omega) = \log\left|\frac{1}{N}\sum_{t=1}^{N}(\boldsymbol{x}(t)-\boldsymbol{\alpha}e^{j\omega t})(\boldsymbol{x}(t)-\boldsymbol{\alpha}e^{j\omega t})^*\right| \tag{26}$$

$$= \log\left|\hat{\boldsymbol{R}}-\boldsymbol{y}(\omega)\boldsymbol{\alpha}^*-\boldsymbol{\alpha}\boldsymbol{y}^*(\omega)+\boldsymbol{\alpha}\boldsymbol{\alpha}^*\right| \tag{27}$$

$$= \log\left|(\boldsymbol{\alpha}-\boldsymbol{y}(\omega))(\boldsymbol{\alpha}-\boldsymbol{y}(\omega))^*+\hat{\boldsymbol{R}}-\boldsymbol{y}(\omega)\boldsymbol{y}^*(\omega)\right| \tag{28}$$

from which it is clear that the solution for $\boldsymbol{\alpha}$ is simply

$$\hat{\boldsymbol{\alpha}}_u = \boldsymbol{y}(\omega). \tag{29}$$

The remaining term in (28) is a function of ω only and can be written as

$$\log\left|\hat{\boldsymbol{R}}-\boldsymbol{y}(\omega)\boldsymbol{y}^*(\omega)\right| = \log\left(\left|\hat{\boldsymbol{R}}\right|\left|\boldsymbol{I}-\hat{\boldsymbol{R}}^{-1}\boldsymbol{y}(\omega)\boldsymbol{y}^*(\omega)\right|\right) \tag{30}$$

$$= \left(1-\boldsymbol{y}^*(\omega)\hat{\boldsymbol{R}}^{-1}\boldsymbol{y}(\omega)\right)+\log\left|\hat{\boldsymbol{R}}\right|. \tag{31}$$

The unstructured ML estimate of ω is thus given by

$$\hat{\omega}_u = \arg\max_{\omega}\boldsymbol{y}^*(\omega)\hat{\boldsymbol{R}}^{-1}\boldsymbol{y}(\omega). \tag{32}$$

Since it appears in the ML criteria for both the structured and unstructured models, we make a note here regarding the quantity $1-z_{\hat{R}}(\omega)$, where

$$z_{\hat{R}}(\omega) \triangleq \boldsymbol{y}^*(\omega)\hat{\boldsymbol{R}}^{-1}\boldsymbol{y}(\omega). \tag{33}$$

Neither (16) nor (31) makes sense if $z_{\hat{R}}(\omega)\geq 1$. To show that this cannot occur, define

$$\boldsymbol{X} = [\boldsymbol{x}(1)\ \cdots\ \boldsymbol{x}(N)]$$
$$\boldsymbol{w}(\omega) = [e^{-j\omega}\ e^{-j2\omega}\ \cdots\ e^{-jN\omega}]^T$$

so that

$$\hat{\boldsymbol{R}} = \frac{1}{N}\boldsymbol{X}\boldsymbol{X}^*$$
$$\boldsymbol{y}(\omega) = \frac{1}{N}\boldsymbol{X}\boldsymbol{w}(\omega).$$

Then

$$z_{\hat{R}}(\omega) = \frac{1}{N^2}\boldsymbol{w}^*(\omega)\boldsymbol{X}^*\left(\frac{1}{N}\boldsymbol{X}\boldsymbol{X}^*\right)^{-1}\boldsymbol{X}\boldsymbol{w}(\omega) \tag{34}$$

$$= \frac{1}{N}\boldsymbol{w}^*(\omega)\boldsymbol{X}^*(\boldsymbol{X}\boldsymbol{X}^*)^{-1}\boldsymbol{X}\boldsymbol{w}(\omega) \tag{35}$$

$$\triangleq \frac{1}{N}\boldsymbol{w}^*(\omega)\boldsymbol{P}_{X^*}\boldsymbol{w}(\omega) \tag{36}$$

$$\leq \frac{1}{N}\boldsymbol{w}^*(\omega)\boldsymbol{w}(\omega) = 1 \tag{37}$$

since $\boldsymbol{P}_{X^*}$ is a projection matrix. Equality in (37) can hold only if $\boldsymbol{w}(\omega)$ belongs to the range of $\boldsymbol{X}^*$. Due to the presence of noise and interference in $\boldsymbol{X}$, $z_{\hat{R}}(\omega)<1$ with probability one and the ML criterion is well defined.

An interpretation for (32) can be obtained if the problem is viewed as one of estimating the frequency of a single sinewave observed in m distinct but dependent channels. Equation (32) shows that the ML frequency estimate for this problem is calculated by first spatially prewhitening the data and then finding the largest peak in the magnitude of the Fourier transform averaged over the m channels. The derivation above also implies that in principle, array calibration data is not needed in detecting the presence or absence of a target in the radar data set. Given the estimate $\hat{\omega}_u$ from (32), one could develop a statistical test for detection based on the size of

$$\left\|\hat{\boldsymbol{R}}^{-(1/2)}\hat{\boldsymbol{\alpha}}_u\right\| = \left\|\hat{\boldsymbol{R}}^{-(1/2)}\boldsymbol{y}(\hat{\omega}_u)\right\|$$

where $\hat{\boldsymbol{R}}^{-(1/2)}$ is a symmetric square root of $\hat{\boldsymbol{R}}^{-1}$. If the array is imprecisely calibrated, such a test may outperform

$$\mathrm{CRB}(\boldsymbol{\eta}_s) = \frac{1}{2N}\begin{bmatrix} \boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{a} & 0 & \mathrm{Re}\{b\boldsymbol{a}\boldsymbol{Q}^{-1}\boldsymbol{d}\} & -\beta_N\mathrm{Im}\{b\}\boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{a} \\ 0 & \boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{a} & \mathrm{Im}\{b\boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{d}\} & \beta_N\mathrm{Re}\{b\}\boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{a} \\ \mathrm{Re}\{b\boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{d}\} & \mathrm{Im}\{b\boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{d}\} & |b|^2\boldsymbol{d}^*\boldsymbol{Q}^{-1}\boldsymbol{d} & \beta_N|b|^2\mathrm{Im}\{\boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{d}\} \\ -\beta_N\mathrm{Im}\{b\}\boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{a} & \beta_N\mathrm{Re}\{b\}\boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{a} & \beta_N|b|^2\mathrm{Im}\{\boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{d}\} & \gamma_N|b|^2\boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{a} \end{bmatrix}^{-1} \tag{20}$$

the GLRT based on estimates of the target DOA. This issue is further studied in Section V and the simulations.

The statistical properties of the ML estimator for the unstructured model are given in the following theorem.

Theorem 3: The unstructured ML estimator defined by (25), (29), and (32) is consistent and statistically efficient for the model described by (3). The CRB for the target parameters

$$\tilde{\boldsymbol{\eta}}_s = \begin{bmatrix} \mathrm{Re}\{\boldsymbol{\alpha}\} \\ \mathrm{Im}\{\boldsymbol{\alpha}\} \\ \omega \end{bmatrix} \tag{38}$$

is given by

$$\mathbf{CRB}(\tilde{\boldsymbol{\eta}}_s) = \frac{1}{2N}\begin{bmatrix} \boldsymbol{Q}_r + \frac{\beta_N^2}{\delta_N}\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i^T & -\boldsymbol{Q}_i - \frac{\beta_N^2}{\delta_N}\boldsymbol{\alpha}_i\boldsymbol{\alpha}_r^T & \frac{\beta_N}{\delta_N}\boldsymbol{\alpha}_i \\ \boldsymbol{Q}_i - \frac{\beta_N^2}{\delta_N}\boldsymbol{\alpha}_r\boldsymbol{\alpha}_i^T & \boldsymbol{Q}_r + \frac{\beta_N^2}{\delta_N}\boldsymbol{\alpha}_r\boldsymbol{\alpha}_r^T & -\frac{\beta_N}{\delta_N}\boldsymbol{\alpha}_r \\ \frac{\beta_N}{\delta_N}\boldsymbol{\alpha}_i^T & -\frac{\beta_N}{\delta_N}\boldsymbol{\alpha}_r^T & \frac{1}{\delta_N} \end{bmatrix} \tag{39}$$

where

$$\boldsymbol{Q}_r = \mathrm{Re}\{\boldsymbol{Q}\} \tag{40}$$

$$\boldsymbol{Q}_i = \mathrm{Im}\{\boldsymbol{Q}\} \tag{41}$$

$$\boldsymbol{\alpha}_r = \mathrm{Re}\{\boldsymbol{\alpha}_0\} \tag{42}$$

$$\boldsymbol{\alpha}_i = \mathrm{Im}\{\boldsymbol{\alpha}_0\} \tag{43}$$

$$\delta_N = [\gamma_N - \beta_N^2]\boldsymbol{\alpha}_0^*\boldsymbol{Q}^{-1}\boldsymbol{\alpha}_0 \tag{44}$$

and β_N, γ_N are defined in (21) and (22).

Proof: See Appendix B. ■

It is interesting to observe that the asymptotic variance of $\hat{\omega}_u$ is equivalent to that of the Doppler estimate obtained using the structured ML approach. From (39), we have that

$$\mathcal{E}\{(\hat{\omega}_u - \omega_0)^2\} = \frac{1}{2N(\gamma_N - \beta_N^2)\boldsymbol{\alpha}_0^*\boldsymbol{Q}^{-1}\boldsymbol{\alpha}_0} = \frac{1}{2N|b_0|^2(\gamma_N - \beta_N^2)\boldsymbol{a}^*(\theta_0)\boldsymbol{Q}^{-1}\boldsymbol{a}(\theta_0)}$$

which is identical to the expression given in (24). This fact will have important ramifications in the development of the EXIP approach below. In particular, because $\hat{\omega}_s$ and $\hat{\omega}_u$ are asymptotically equivalent, we can replace $\hat{\omega}_s$ by $\hat{\omega}_u$ in (15) and (16). The resulting estimates for b and θ

$$\hat{b}_u = \frac{\boldsymbol{a}^*(\hat{\theta}_u)\tilde{\boldsymbol{R}}_{\hat{\omega}_u}^{-1}\hat{\boldsymbol{\alpha}}_u}{\boldsymbol{a}^*(\hat{\theta}_u)\tilde{\boldsymbol{R}}_{\hat{\omega}_u}^{-1}\boldsymbol{a}(\hat{\theta}_u)} \tag{45}$$

$$\hat{\theta}_u = \arg\max_{\theta} \frac{\left|\boldsymbol{a}^*(\theta)\hat{\boldsymbol{R}}^{-1}\hat{\boldsymbol{\alpha}}_u\right|^2}{\boldsymbol{a}^*(\theta)\hat{\boldsymbol{R}}^{-1}\boldsymbol{a}(\theta)} \tag{46}$$

should have the same asymptotic variance as the structured ML estimates in (15) and (16). This observation is formalized using the EXIP in what follows.

The main result regarding the use of EXIP to find estimates of b, θ, ω, and $\boldsymbol{Q}$ from the unstructured estimates $\hat{\boldsymbol{\alpha}}_u, \hat{\omega}_u$, and $\hat{\boldsymbol{Q}}_u$ is summarized by the following theorem.

Theorem 4: Given $\hat{\boldsymbol{Q}}_u, \hat{\boldsymbol{\alpha}}_u$, and $\hat{\omega}_u$ obtained from (25), (29), and (32), asymptotically (in N) efficient estimates of the radar target parameters in (2) can be found as follows:

$$\hat{\boldsymbol{Q}}_e = \hat{\boldsymbol{Q}}_u \tag{47}$$

$$\hat{\omega}_e = \hat{\omega}_u \tag{48}$$

$$\hat{b}_e, \hat{\theta}_e = \arg\min_{\theta,b} \left(\begin{bmatrix} (\hat{\boldsymbol{\alpha}}_u)_r \\ (\hat{\boldsymbol{\alpha}}_u)_i \end{bmatrix} - \boldsymbol{A}(\theta)\begin{bmatrix} b_r \\ b_i \end{bmatrix}\right)^T \cdot \boldsymbol{W}_Q\left(\begin{bmatrix} (\hat{\boldsymbol{\alpha}}_u)_r \\ (\hat{\boldsymbol{\alpha}}_u)_i \end{bmatrix} - \boldsymbol{A}(\theta)\begin{bmatrix} b_r \\ b_i \end{bmatrix}\right) \tag{49}$$

$$= \arg\min_{\theta,b} (\hat{\boldsymbol{\alpha}}_u - b\boldsymbol{a}(\theta))^*\hat{\boldsymbol{Q}}_u^{-1}(\hat{\boldsymbol{\alpha}}_u - b\boldsymbol{a}(\theta)) \tag{50}$$

where the subscript e denotes an EXIP-based estimate, the subscripts r and i indicate real and imaginary parts, respectively, and

$$\boldsymbol{W}_Q = 2N\begin{bmatrix} \mathrm{Re}\left\{\hat{\boldsymbol{Q}}_u^{-1}\right\} & -\mathrm{Im}\left\{\hat{\boldsymbol{Q}}_u^{-1}\right\} \\ \mathrm{Im}\left\{\hat{\boldsymbol{Q}}_u^{-1}\right\} & \mathrm{Re}\left\{\hat{\boldsymbol{Q}}_u^{-1}\right\} \end{bmatrix} \tag{51}$$

$$\boldsymbol{A}(\theta) = \begin{bmatrix} \mathrm{Re}\{\boldsymbol{a}(\theta)\} & -\mathrm{Im}\{\boldsymbol{a}(\theta)\} \\ \mathrm{Im}\{\boldsymbol{a}(\theta)\} & \mathrm{Re}\{\boldsymbol{a}(\theta)\} \end{bmatrix}. \tag{52}$$

The estimate $\hat{b}_e$ may be solved for explicitly, leading to a criterion that depends only on θ

$$\hat{b}_e = \frac{\boldsymbol{a}^*(\hat{\theta}_e)\hat{\boldsymbol{Q}}_u^{-1}\hat{\boldsymbol{\alpha}}_u}{\boldsymbol{a}^*(\hat{\theta}_e)\hat{\boldsymbol{Q}}_u^{-1}\boldsymbol{a}(\hat{\theta}_e)} \tag{53}$$

$$\hat{\theta}_e = \arg\max_{\theta} \frac{\left|\boldsymbol{a}^*(\theta)\hat{\boldsymbol{Q}}_u^{-1}\hat{\boldsymbol{\alpha}}_u\right|^2}{\boldsymbol{a}^*(\theta)\hat{\boldsymbol{Q}}_u^{-1}\boldsymbol{a}(\theta)}. \tag{54}$$

Proof: See Appendix C. ■

The simplicity of the final result is quite remarkable, especially the fact that the EXIP estimate of ω is exactly equivalent to the unstructured estimate $\hat{\omega}_u$. Although it was shown earlier that both have the same asymptotic variance (indeed, we showed that $\hat{\omega}_s$ and $\hat{\omega}_u$ have the same asymptotic variance, which should also coincide with the asymptotic variance of $\hat{\omega}_e$ by the EXIP), the weighting matrix for EXIP is not block diagonal with respect to $\boldsymbol{\alpha}$ and ω. The off-block diagonal terms just happen to combine in a fortuitous way for (48) to result. It is also interesting to observe that the estimators of b and θ in (53) and (54) are not just asymptotically equivalent to those predicted by (45) and (46) but are in fact identical. To see this, note that (25)–(28) imply

$$\hat{\boldsymbol{Q}}_u = \hat{\boldsymbol{R}} - \hat{\boldsymbol{\alpha}}_u\hat{\boldsymbol{\alpha}}_u^* = \hat{\boldsymbol{R}} - \boldsymbol{y}(\hat{\omega}_u)\boldsymbol{y}^*(\hat{\omega}_u) = \tilde{\boldsymbol{R}}_{\hat{\omega}_u}$$

which demonstrates immediately that, provided (46) and (54) are identical, (45) and (53) are as well. Using the matrix inversion lemma [e.g., see (121) in Appendix B],

we have

$$\hat{\boldsymbol{Q}}_u^{-1}\hat{\boldsymbol{\alpha}}_u = \left(\hat{\boldsymbol{R}}^{-1} - \frac{\hat{\boldsymbol{R}}^{-1}\hat{\boldsymbol{\alpha}}_u\hat{\boldsymbol{\alpha}}_u^*\hat{\boldsymbol{R}}^{-1}}{1+\hat{\boldsymbol{\alpha}}_u^*\hat{\boldsymbol{R}}^{-1}\hat{\boldsymbol{\alpha}}_u}\right)\hat{\boldsymbol{\alpha}}_u = \frac{\hat{\boldsymbol{R}}^{-1}\hat{\boldsymbol{\alpha}}_u}{1+\hat{\boldsymbol{\alpha}}_u^*\hat{\boldsymbol{R}}^{-1}\hat{\boldsymbol{\alpha}}}$$

$$\boldsymbol{a}^*(\theta)\hat{\boldsymbol{Q}}_u^{-1}\boldsymbol{a}(\theta) = \boldsymbol{a}^*(\theta)\hat{\boldsymbol{R}}^{-1}\boldsymbol{a}(\theta) - \frac{\left|\boldsymbol{a}^*(\theta)\hat{\boldsymbol{R}}^{-1}\hat{\boldsymbol{\alpha}}_u\right|^2}{1+\hat{\boldsymbol{\alpha}}_u^*\hat{\boldsymbol{R}}^{-1}\hat{\boldsymbol{\alpha}}_u}.$$

Thus, the equation at the bottom of the page proves that (46) and (54) are equivalent as well. Despite the fact that these simplified criteria can be obtained without using the EXIP, the result of Theorem 4 is still of interest. For example, the weighting of (49) can be regularized to make the algorithm more robust to array calibration errors. This idea is explained in more detail below.

For convenience, we summarize here the two steps of the proposed EXIP approach.

a) Estimate the target Doppler frequency from

$$\hat{\omega}_u = \arg\max_{\omega} \boldsymbol{y}^*(\omega)\hat{\boldsymbol{R}}^{-1}\boldsymbol{y}(\omega)$$

and then solve for the interference covariance

$$\hat{\boldsymbol{Q}}_u = \hat{\boldsymbol{R}} - \boldsymbol{y}(\hat{\omega}_u)\boldsymbol{y}^*(\hat{\omega}_u).$$

b) Estimate the target amplitude and DOA using

$$\hat{b}_e = \frac{\boldsymbol{a}^*(\hat{\theta}_e)\hat{\boldsymbol{Q}}_u^{-1}\boldsymbol{y}(\hat{\omega}_u)}{\boldsymbol{a}^*(\hat{\theta}_e)\hat{\boldsymbol{Q}}_u^{-1}\boldsymbol{a}(\hat{\theta}_e)}$$

$$\hat{\theta}_e = \arg\max_{\theta} \frac{|\boldsymbol{a}^*(\theta)\hat{\boldsymbol{Q}}_u^{-1}\boldsymbol{y}(\hat{\omega}_u)|^2}{\boldsymbol{a}^*(\theta)\hat{\boldsymbol{Q}}_u^{-1}\boldsymbol{a}(\theta)}.$$

1) Initial Estimates: The two 1-D searches listed above are rendered even more computationally efficient by the availability of simple initial estimates. For example, as mentioned above, it is likely that the target lies in the main beam of the transmit pattern, and hence that θ_0 is near the known nominal transmit direction θ_t. Consequently, θ_t is a good choice for initializing step b).

Since $\boldsymbol{y}(\omega)$ is just the discrete Fourier transform (DFT) of $\boldsymbol{x}(t)$, the criterion of (32) may be efficiently evaluated at the frequencies $\omega_k = 2\pi k/N$ for $k = 0, \cdots, N-1$ using a bank of fast (F)FT's. For sufficiently large N, the value of ω_k on the FFT grid that maximizes (32) is an excellent initial estimate for the search in step a). One can then locally maximize (32) around this initial estimate using the computationally efficient zoom or chirp FFT algorithms. In Section V, we will also see that the development of the detection criterion relies on using (32) evaluated on a grid with spacing equal to $2\pi/N$.

2) Solution Via Polynomial Rooting: If the data are collected with a uniform linear array (ULA), a further asymptotic argument can be used to show that the search for θ in step b) may be eliminated by reparameterizing the problem in terms of a simple polynomial [14]. In this approach, the polynomial coefficients are solved for explicitly in closed form, and $\hat{\theta}_e$ is found by rooting the resulting first-order polynomial equation. The availability of such a procedure is a further advantage of the EXIP approach, since a corresponding simplification is apparently not possible for the criterion in (16) due to the need for a 2-D search in both θ and ω.

To see how such a reparameterization can be used here, note that for a ULA

$$\boldsymbol{a}(\theta) = [1 \quad e^{-j\phi} \quad \cdots \quad e^{-j(m-1)\phi}]^T$$

where $\phi = 2\pi\Delta\sin\theta$ and Δ is the distance in wavelengths between adjacent antenna elements. Because of its structure, $\boldsymbol{a}(\theta)$ above lies in the nullspace of the following $(m-1)\times m$ matrix:

$$\boldsymbol{B} = \begin{bmatrix} p & -p^c & & & \\ & p & -p^c & & \\ & & \ddots & \ddots & \\ & & & p & -p^c \end{bmatrix} \tag{55}$$

$$\begin{aligned}
\hat{\theta}_e &= \arg\max_{\theta} \frac{\left|\boldsymbol{a}^*(\theta)\hat{\boldsymbol{Q}}_u^{-1}\hat{\boldsymbol{\alpha}}_u\right|^2}{\boldsymbol{a}^*(\theta)\hat{\boldsymbol{Q}}_u^{-1}\boldsymbol{a}(\theta)} \\
&= \arg\max_{\theta} \frac{\left|\boldsymbol{a}^*(\theta)\hat{\boldsymbol{R}}^{-1}\hat{\boldsymbol{\alpha}}_u\right|^2}{\left(1+\hat{\boldsymbol{\alpha}}_u^*\hat{\boldsymbol{R}}^{-1}\hat{\boldsymbol{\alpha}}_u\right)^2\left(\boldsymbol{a}^*(\theta)\hat{\boldsymbol{R}}^{-1}\boldsymbol{a}(\theta) - \frac{\left|\boldsymbol{a}^*(\theta)\hat{\boldsymbol{R}}^{-1}\hat{\boldsymbol{\alpha}}_u\right|^2}{1+\hat{\boldsymbol{\alpha}}_u^*\hat{\boldsymbol{R}}\hat{\boldsymbol{\alpha}}_u}\right)} \\
&= \arg\min_{\theta} \left\{ \frac{\left(1+\hat{\boldsymbol{\alpha}}_u^*\hat{\boldsymbol{R}}^{-1}\hat{\boldsymbol{\alpha}}_u\right)^2\boldsymbol{a}^*(\theta)\hat{\boldsymbol{R}}^{-1}\boldsymbol{a}(\theta)}{\left|\boldsymbol{a}^*(\theta)\hat{\boldsymbol{R}}^{-1}\hat{\boldsymbol{\alpha}}_u\right|^2} - \frac{1}{1+\hat{\boldsymbol{\alpha}}_u^*\hat{\boldsymbol{R}}^{-1}\hat{\boldsymbol{\alpha}}_u} \right\} \\
&= \arg\max_{\theta} \frac{\left|\boldsymbol{a}^*(\theta)\hat{\boldsymbol{R}}^{-1}\hat{\boldsymbol{\alpha}}_u\right|^2}{\boldsymbol{a}^*(\theta)\hat{\boldsymbol{R}}^{-1}\boldsymbol{a}(\theta)} = \hat{\theta}_u,
\end{aligned}$$

where p is a θ-dependent variable and p^c its conjugate. The variables p and p^c are the coefficients of a first-order polynomial $p-p^c z$ whose root is $z = e^{-j\phi}$; the polynomial coefficients are conjugates of one another since the root lies on the unit circle: $p = \kappa e^{-j\phi/2}, p^c = \kappa e^{j\phi/2}$, where κ is an arbitrary scaling. Since κ does not affect the equation $\boldsymbol{B}\boldsymbol{a}(\theta) = 0$, the matrix $\boldsymbol{B}$ can be thought of as depending on a single real-valued scalar. For example, scale p such that its real part is unity $p = 1 + jp_i$; then $\boldsymbol{B}$ depends explicitly only on p_i, which is uniquely related to[4] θ. To reparameterize (54), define

$$\boldsymbol{\Pi}(\theta) = \hat{\boldsymbol{Q}}_u^{-(1/2)}\boldsymbol{a}(\theta)[\boldsymbol{a}^*(\theta)\hat{\boldsymbol{Q}}_u^{-1}\boldsymbol{a}(\theta)]^{-1}\boldsymbol{a}^*(\theta)\hat{\boldsymbol{Q}}_u^{-(1/2)}$$

and note that

$$\begin{aligned}\boldsymbol{\Pi}^{\perp}(\theta) &= \boldsymbol{I} - \boldsymbol{\Pi}(\theta)\\ &= \hat{\boldsymbol{Q}}_u^{1/2}\boldsymbol{B}^*(p_i)[\boldsymbol{B}(p_i)\hat{\boldsymbol{Q}}_u\boldsymbol{B}^*(p_i)]^{-1}\boldsymbol{B}(p_i)\hat{\boldsymbol{Q}}_u^{1/2}\\ &= \boldsymbol{\Pi}(p_i).\end{aligned}$$

The criterion of (54) may thus be rewritten as

$$\max_{\theta}\frac{\left|\boldsymbol{a}^*(\theta)\hat{\boldsymbol{Q}}_u^{-1}\hat{\boldsymbol{\alpha}}_u\right|^2}{\boldsymbol{a}^*(\theta)\hat{\boldsymbol{Q}}_u^{-1}\boldsymbol{a}(\theta)} = \max_{\theta}\left\|\boldsymbol{\Pi}(\theta)\hat{\boldsymbol{Q}}_u^{-(1/2)}\hat{\boldsymbol{\alpha}}_u\right\|^2 \quad (56)$$

$$= \min_{\theta}\left\|\boldsymbol{\Pi}^{\perp}(\theta)\hat{\boldsymbol{Q}}_u^{-(1/2)}\hat{\boldsymbol{\alpha}}_u\right\|^2 \quad (57)$$

$$= \min_{p_i}\left\|\boldsymbol{\Pi}(p_i)\hat{\boldsymbol{Q}}_u^{-(1/2)}\hat{\boldsymbol{\alpha}}_u\right\|^2. \quad (58)$$

The key to the advantage in parameterizing by p_i lies in the fact that, as shown in [14], $\boldsymbol{\Pi}(p_i)$ in (58) may be replaced by

$$\boldsymbol{\Pi}(p_i) \rightarrow \hat{\boldsymbol{Q}}_u^{1/2}\boldsymbol{B}^*(p_i)\left[\hat{\boldsymbol{B}}\hat{\boldsymbol{Q}}_u\hat{\boldsymbol{B}}^*\right]^{-1}\boldsymbol{B}(p_i)\hat{\boldsymbol{Q}}_u^{1/2} \quad (59)$$

where $\hat{\boldsymbol{B}}$ is a consistent estimate of $\boldsymbol{B}$, without affecting the asymptotic properties of the estimate of p_i (and hence θ). With the bracketed part of (59) fixed, the resulting criterion is simply quadratic in p_i, and a closed-form solution can be obtained. The consistent estimate of $\boldsymbol{B}$ required in (59) can be found by solving

$$\min_{p_i}\hat{\boldsymbol{\alpha}}_u^*\boldsymbol{B}^*(p_i)\boldsymbol{B}(p_i)\hat{\boldsymbol{\alpha}}_u \quad (60)$$

which can also be achieved without a search. Given an estimate of p_i (and hence p), θ is estimated as

$$\hat{\theta}_e = \sin^{-1}\left[\frac{-\angle\hat{p}^2}{2\pi\Delta}\right].$$

[4]In certain cases—for example, if $\Delta = 1/2$ and the target DOA is at end fire (90° relative to broadside)—the real part of p is zero and cannot be scaled to be unity. For this reason, it is often preferable to use the constraint $|p|^2 = 1$. In our discussion, however, we choose to set $p = 1 + jp_i$ since it simplifies the description of the algorithm.

3) Robustness to Calibration Errors: Although (49) reduces to (53) and (54), it provides a direct and very natural way of taking array calibration errors into account and thus may be preferred in some cases. To see this, suppose that

$$\boldsymbol{\alpha}_0 = b_0(\boldsymbol{a}(\theta_0) + \tilde{\boldsymbol{a}}) \quad (61)$$

where $\tilde{\boldsymbol{a}}$ is some unknown perturbation to the array response. A number of different models could be chosen for $\tilde{\boldsymbol{\alpha}}$ (see, for example, [15]–[17]), but for the sake of discussion, assume that $\tilde{\boldsymbol{a}}$ is a zero-mean random vector with known covariance

$$\boldsymbol{\Omega} = \mathcal{E}\left\{\begin{bmatrix}\mathrm{Re}\{\tilde{\boldsymbol{a}}\}\\ \mathrm{Im}\{\tilde{\boldsymbol{a}}\}\end{bmatrix}\begin{bmatrix}\mathrm{Re}^T\{\tilde{\boldsymbol{a}}\} & \mathrm{Im}^T\{\tilde{\boldsymbol{a}}\}\end{bmatrix}\right\}.$$

Such information may be available from the manufacturer of the array (e.g., given in terms of gain and phase tolerances) or from the results of several calibration measurements.

Under this assumption, the difference between $\boldsymbol{\alpha}_0$ and the nominal model $b_0\boldsymbol{a}(\theta_0)$ is the term $b_0\tilde{\boldsymbol{a}}$, whose covariance is given by

$$\begin{aligned}\boldsymbol{\Gamma}(b_0) &= \mathrm{cov}\left(\begin{bmatrix}\mathrm{Re}\{b_0\tilde{\boldsymbol{a}}\}\\ \mathrm{Im}\{b_0\tilde{\boldsymbol{a}}\}\end{bmatrix}\right)\\ &= \begin{bmatrix}\mathrm{Re}\{b_0\}\boldsymbol{I} & -\mathrm{Im}\{b_0\}\boldsymbol{I}\\ \mathrm{Im}\{b_0\}\boldsymbol{I} & \mathrm{Re}\{b_0\}\boldsymbol{I}\end{bmatrix}\\ &\quad\cdot\boldsymbol{\Omega}\begin{bmatrix}\mathrm{Re}\{b_0\}\boldsymbol{I} & \mathrm{Im}\{b_0\}\boldsymbol{I}\\ -\mathrm{Im}\{b_0\}\boldsymbol{I} & \mathrm{Re}\{b_0\}\boldsymbol{I}\end{bmatrix}.\end{aligned} \quad (62)$$

Consequently, the performance of the EXIP minimization in (49) could be improved by using a "regularized" weighting matrix

$$\boldsymbol{W}_r = \left(\boldsymbol{W}_Q^{-1} + \boldsymbol{\Gamma}(\hat{b})\right)^{-1} \quad (63)$$

where $\boldsymbol{W}_Q$ is the standard weighting of Theorem 4 (51) and $\hat{b}$ is an initial estimate of b_0. If the array errors are circular, then by definition

$$\boldsymbol{\Omega} = \begin{bmatrix}\boldsymbol{\Omega}_1 & -\boldsymbol{\Omega}_2\\ \boldsymbol{\Omega}_2 & \boldsymbol{\Omega}_1\end{bmatrix}$$

and $\boldsymbol{\Gamma}(\hat{b}) = |\hat{b}|^2\boldsymbol{\Omega}$. It is easy to show that for this special case, simplified criteria like (53) and (54) result, except with $\hat{\boldsymbol{Q}}_u^{-1}$ replaced by

$$\left(\frac{1}{2N}\hat{\boldsymbol{Q}}_u + |\hat{b}|^2(\boldsymbol{\Omega}_1 + j\boldsymbol{\Omega}_2)\right)^{-1}.$$

The initial estimate of b_0 needed for (63) could be obtained in one of the following ways:

- by an initial application of the EXIP algorithm without the regularized weighting;
- using (15) evaluated with the Doppler estimate from (32) and the known transmit direction θ_t;
- setting $\hat{b} = [\hat{\boldsymbol{\alpha}}_u]_1$, the first element of $\hat{\boldsymbol{\alpha}}_u$, which relies on the assumption previously made that the nominal array response $\boldsymbol{a}(\theta)$ is scaled such that its first element is unity.

In our experience with simulated data, the estimate of b_0 obtained from the first method is significantly more accurate than the other two but requires somewhat more computation. If θ_t is within a few degrees of θ_0, the second method will outperform the third, but usually only in difficult situations with a weak target where the variance of the estimate $[\hat{\boldsymbol{\alpha}}_u]_1$ is quite large. Our simulations show that the improvement gained from the regularization is not very sensitive to the exact value of $\hat{b}$; a change in $\hat{b}$ by a factor of three or even more causes little performance degradation. This also implies that one need not know the precise value of $\boldsymbol{\Omega}$, and hence that a simple estimate such as $\boldsymbol{\Omega} = \sigma_a^2 \boldsymbol{I}$ will probably suffice.

It is shown in Section VI that the use of the regularized weighting offers some improvement in estimation performance. A similar procedure for making the estimates of the structured algorithm (16) robust to array errors appears to be much more difficult.

C. Extension to Nonwhite Interference

As discussed earlier, when significant clutter is present in the data, it is unlikely that $\boldsymbol{e}(t)$ will be temporally white. In this section, we present a simple technique for applying the methods described thus far to the case where $\boldsymbol{e}(t)$ is temporally colored. The technique relies on the availability of secondary data taken from adjacent range bins and CPI's that do not contain target energy. Assuming that the statistics of the interference in the primary and secondary data are the same, the secondary data is used to derive a VAR filter that whitens $\boldsymbol{e}(t)$. As shown below, the use of the whitening filter has almost no effect on the implementation of the unstructured ML estimator, but it must be taken into account when using the EXIP approach to find the target DOA and amplitude.

1) A VAR Prewhitening Filter: To begin, suppose we have at our disposal a VAR filter of the form

$$\mathcal{H}(z^{-1}) = \boldsymbol{I} + \sum_{l=1}^{L} \boldsymbol{H}_l z^{-l} \tag{64}$$

such that

$$\mathcal{H}(z^{-1})\boldsymbol{e}(t) = \boldsymbol{\epsilon}(t) \tag{65}$$

is temporally white. Here, z^{-1} denotes the unit delay operator. The estimation of $\mathcal{H}(z^{-1})$ from the secondary data will be addressed below. If we preprocess the array data with $\mathcal{H}(z^{-1})$ prior to detection and estimation of the target parameters, the filtered array output will be of the form

$$\boldsymbol{x}_f(t) = \mathcal{H}(z^{-1})\boldsymbol{x}(t) = b_0 \mathcal{H}(e^{-j\omega_0})\boldsymbol{a}(\theta_0)e^{j\omega_0 t} + \boldsymbol{\epsilon}(t). \tag{66}$$

Defining the filtered spatial signature

$$\boldsymbol{\alpha}_f = b_0 \mathcal{H}(e^{-j\omega_0})\boldsymbol{a}(\theta_0)$$

we have

$$\boldsymbol{x}_f(t) = \boldsymbol{\alpha}_f e^{j\omega_0 t} + \boldsymbol{\epsilon}(t) \tag{67}$$

and we see that the unstructured ML estimator can be directly applied to $\boldsymbol{x}_f(t)$ with no change.

The unstructured algorithm will provide an estimate of the Doppler $\hat{\omega}_f$ and the filtered spatial signature $\hat{\boldsymbol{\alpha}}_f$, which must then be fit to the structured model to obtain the target DOA and amplitude. A complication arises here due to the fact that after filtering, the spatial signature depends on the target Doppler frequency. This fact induces a nonlinear dependence of the EXIP criterion on both θ and ω, which makes it unlikely that the parameters can be estimated without a 2-D search. Hence, in this case, the EXIP approach would offer no advantage over the structured ML solution. However, given the excellent accuracy of the unstructured Doppler estimate (its variance goes as N^{-3}), reasonable estimates for b and θ analogous to (53) and (54) can be obtained using

$$\hat{b}_f = \frac{\boldsymbol{a}^*(\hat{\theta}_f)\mathcal{H}^*(e^{-j\hat{\omega}_f})\hat{\boldsymbol{Q}}_f^{-1}\hat{\boldsymbol{\alpha}}_f}{\boldsymbol{a}^*(\hat{\theta}_f)\mathcal{H}^*(e^{-j\hat{\omega}_f})\hat{\boldsymbol{Q}}_f^{-1}\mathcal{H}(e^{-j\hat{\omega}_f})\boldsymbol{a}(\hat{\theta}_f)} \tag{68}$$

$$\hat{\theta}_f = \arg\max_{\theta} \frac{\left|\boldsymbol{a}^*(\theta)\mathcal{H}^*(e^{-j\hat{\omega}_f})\hat{\boldsymbol{Q}}_f^{-1}\hat{\boldsymbol{\alpha}}_f\right|^2}{\boldsymbol{a}^*(\theta)\mathcal{H}^*(e^{-j\hat{\omega}_f})\hat{\boldsymbol{Q}}_f^{-1}\mathcal{H}(e^{-j\hat{\omega}_f})\boldsymbol{a}(\theta)} \tag{69}$$

where $\hat{\boldsymbol{Q}}_f$ is the estimate of the spatial covariance of $\boldsymbol{\epsilon}(t)$ obtained from (25) using the filtered data and the estimates $\hat{\boldsymbol{\alpha}}_f$ and $\hat{\omega}_f$. Once again, only a 1-D search is required to obtain the target parameter estimates.

2) Practical Issues: Below, we outline the computation of the VAR filter $\mathcal{H}(z^{-1})$ using target-free secondary data. Suppose we have N samples of data from each of N_s range/CPI cells adjacent to the one under study. We will denote the secondary data by

$$\{\boldsymbol{e}(r_k, \phi_k, t)\}_{t=1}^{N}, \qquad k = 1, \cdots, N_s$$

where r_k and ϕ_k represent the range and CPI of the kth data set. We seek a VAR filter such that

$$\begin{aligned}\mathcal{H}(z^{-1})\boldsymbol{e}(r_k, \phi_k, t) &= \boldsymbol{e}(r_k, \phi_k, t) + \boldsymbol{H}_1\boldsymbol{e}(r_k, \phi_k, t-1) \\ &\quad + \cdots + \boldsymbol{H}_L\boldsymbol{e}(r_k, \phi_k, t-L) \\ &= \boldsymbol{\epsilon}(r_k, \phi_k, t)\end{aligned} \tag{70}$$

is temporally white for each k. This approach assumes that the statistics of the clutter and interference are the same in the primary and in all secondary data sets, so clearly, only range/CPI cells adjacent to the primary data cell should be used. The order L of the filter can be determined using standard methods such as those in [18]–[20]. Note that the process of filtering the array data as in (66) leads to the loss of L data samples at the beginning of the observation window, so a penalty term should be applied to keep L fairly small.

The standard approach to the problem of estimating the VAR filter coefficients is based on the following least-squares error criterion [20]:

$$\begin{aligned}\hat{\boldsymbol{H}}_1, \cdots, \hat{\boldsymbol{H}}_L = \arg\min_{\boldsymbol{H}_1, \cdots, \boldsymbol{H}_L} &\sum_{k=1}^{N_s}\sum_{t=L+1}^{N} \\ &\cdot \left\| \boldsymbol{e}(r_k, \phi_k, t) - \sum_{l=1}^{L} \boldsymbol{H}_l \boldsymbol{e}(r_k, \phi_k, t-l) \right\|_F^2.\end{aligned} \tag{71}$$

If we define

$$\boldsymbol{H} = [\boldsymbol{H}_1 \cdots \boldsymbol{H}_L] \tag{72}$$

$$\boldsymbol{\psi}_k(t) = -\begin{bmatrix} \boldsymbol{e}(r_k, \phi_k, t-1) \\ \vdots \\ \boldsymbol{e}(r_k, \phi_k, t-L) \end{bmatrix} \tag{73}$$

$$\boldsymbol{\Psi}_k = [\boldsymbol{\psi}_k(L+1) \cdots \boldsymbol{\psi}_k(N)] \tag{74}$$

$$\boldsymbol{\Psi} = [\boldsymbol{\Psi}_1 \cdots \boldsymbol{\Psi}_{N_s}] \tag{75}$$

$$\boldsymbol{E}_k = [\boldsymbol{e}(r_k, \phi_k, L+1) \cdots \boldsymbol{e}(r_k, \phi_k, N)] \tag{76}$$

$$\boldsymbol{E} = [\boldsymbol{E}_1 \cdots \boldsymbol{E}_{N_s}] \tag{77}$$

the minimization of (71) can be rewritten in a more compact form

$$\hat{\boldsymbol{H}} = \arg\min_{H} ||\boldsymbol{E} - \boldsymbol{H}\boldsymbol{\Psi}||_F^2 \tag{78}$$

and, provided that the matrix $\boldsymbol{\Psi}\boldsymbol{\Psi}^*$ is full rank, the solution for $\hat{\boldsymbol{H}}$ is simply

$$\hat{\boldsymbol{H}} = \boldsymbol{E}\boldsymbol{\Psi}^*(\boldsymbol{\Psi}\boldsymbol{\Psi}^*)^{-1}. \tag{79}$$

A necessary condition for the existence of the matrix inverse in (79) is that $\boldsymbol{\Psi}$ have at least as many columns as rows: $N_s(N-L) \geq mL$, a condition that can easily be satisfied by choosing a large enough N_s. Of course, an upper bound for N_s exists, effectively determined by how many adjacent range/CPI cells can be considered to have identical clutter and interference statistics.

Some comments regarding the VAR prewhitening approach are given below.

- As mentioned, the standard "fully adaptive" approaches for dealing with spatially and temporally correlated clutter involve estimating and inverting the $mN \times mN$ covariance associated with all the interference terms in an observation window [5]. In essence, these techniques attempt to whiten the data in both space and time, and they are computationally more involved than the above VAR approach.
- The spatial covariance of the whitened interference data could also be estimated using the secondary data and then used to whiten the interference spatially as well. However, this whitening will of course be imperfect owing to finite-sample effects and the fact that the statistics of the interference in the primary and secondary data may (at least slightly) differ from one another. This observation, along with the fact that the estimation methods introduced in this paper can handle spatially correlated interference in the primary data without any problem, suggests that spatial prewhitening is not necessary or even desirable.
- In principle, it is possible to compute the covariance of the error in $\hat{\boldsymbol{\alpha}}_f$ and $\hat{\omega}_f$ incurred by using an estimated $\boldsymbol{H}$ to prewhiten the data. This error covariance could be used to regularize the EXIP least-squares fit in the same way that information about the covariance of the calibration errors was used in Section IV-C3.

V. Detection

In this section, we discuss the binary hypothesis testing problem that arises when determining whether or not a target is present in a given data set. In particular, it is necessary to decide which of the following two hypotheses is valid for $\boldsymbol{x}(t), t = 1, \cdots, N$:

$$H_0\text{: } \boldsymbol{x}(t) = \boldsymbol{e}(t)$$
$$H_1\text{: } \boldsymbol{x}(t) = b_0\boldsymbol{a}(\theta_0)e^{j\omega_0 t} + \boldsymbol{e}(t).$$

The most common approach to solving this type of problem involves first estimating the signal parameters as if a target were present and then comparing the likelihood of H_1 (with the true parameters replaced by estimates) to that of H_0. This is referred to as a GLRT.

A. GLRT for Structured Model

If p_{H_0} and p_{H_1} represent the probability densities of the data corresponding to the two hypotheses, the GLRT for the structured ML model can be formulated as

$$\lambda' = \frac{p_{H_1}(\boldsymbol{x}(1), \cdots, \boldsymbol{x}(N) | \hat{b}_s, \hat{\theta}_s, \hat{\omega}_s, \hat{\boldsymbol{Q}}_s)}{p_{H_0}(\boldsymbol{x}(1), \cdots, \boldsymbol{x}(N) | \hat{\boldsymbol{Q}}_s)} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda'_T \tag{80}$$

where $\hat{b}_s, \hat{\theta}_s, \hat{\omega}_s, \hat{\boldsymbol{Q}}_s$ are the structured ML estimates and λ'_T is some threshold. The threshold is chosen to achieve some desired performance, most often specified in terms of a fixed probability of false alarm (PFA). Since p_{H_0} and p_{H_1} are Gaussian, the logarithm of both sides of (80) is usually taken, in which case the GLRT takes the form

$$\lambda = 2\log\lambda' = 2N(\log|\hat{\boldsymbol{R}}| - \log|\boldsymbol{C}(\hat{b}_s, \hat{\theta}_s, \hat{\omega}_s)|) \underset{H_0}{\overset{H_1}{\gtrless}} 2\log\lambda'_T = \lambda_T \tag{81}$$

where we have replaced $\hat{\boldsymbol{Q}}_s$ with its ML estimate under each of the hypotheses. The parameter estimates in (81) can be replaced by the EXIP estimates $\hat{b}_e, \hat{\theta}_e$, and $\hat{\omega}_e$, which have the same asymptotic variance but are computationally simpler to obtain.

If we use PFA as the criterion for choosing λ_T, we must be able to find the probability that λ exceeds the threshold when it should not, i.e., when H_0 holds. This in turn requires that the distribution of λ under H_0 be determined, a very difficult problem due to the complicated dependence of the test statistic on the data. However, it can be shown that for general models (not necessarily of the form considered here) and under quite general conditions, the *asymptotic* distribution of the GLRT quantity is chi-squared (χ^2) with degrees of freedom equal to the difference between the number of free parameters under H_1 and H_0 [20].

In the structured ML problem, the number of extra free parameters is four, corresponding to two parameters for b (a real and imaginary part) and one each for θ and ω. A complication arises here due to the fact that under H_0, the parameters θ and ω are undefined, which renders the general result of [20] inapplicable. The approach taken in [13] to overcome this problem is to compute the distribution of λ *conditioned* on θ and ω, in which case the number of extra

parameters is two. Thus, conditioned on θ and ω, λ is a $\chi^2(2)$ random variable. Since the test statistic is computed by maximizing λ over θ and ω, order statistics must be used to find the unconditional density for λ and the desired PFA threshold. This too is quite a difficult problem in the general case. A reasonable heuristical approximation was proposed in [13], where θ was treated as a known quantity equal to θ_t and order statistics applied only to samples of λ taken in ω. It was shown in [13] that under these assumptions, a given PFA P_f could be obtained by choosing

$$\lambda_T = -2\log\left(1-(1-P_f)^{1/N}\right). \tag{82}$$

In the discussion below, we apply the GLRT to the unstructured model and derive a threshold test for a given PFA similar to the one above. The derivation is considerably easier due to the simplicity of the unstructured model, however, and the resulting test is much more robust in cases where $\theta_0 \neq \theta_t$ and calibration errors are present. Additionally, the unstructured test obtained below can be readily extended to the case of temporally correlated noise, unlike the structured GLRT, which in that case requires a 2-D search (see Section IV-C2).

B. GLRT for Unstructured Model

For the unstructured model, the two competing hypotheses are

$$\begin{aligned} H_0&\colon \boldsymbol{x}(t) = \boldsymbol{e}(t) \\ H_{1u}&\colon \boldsymbol{x}(t) = \boldsymbol{\alpha}_0 e^{j\omega_0 t} + \boldsymbol{e}(t). \end{aligned}$$

The resulting GLRT is similar in form to (81), with $\hat{b}_s, \hat{\theta}_s$ replaced by $\hat{\boldsymbol{\alpha}}_u$

$$\lambda_u = 2N(\log|\hat{\boldsymbol{R}}| - \log|\boldsymbol{C}(\hat{\boldsymbol{\alpha}}_u, \hat{\omega}_u)|) \underset{H_0}{\overset{H_1}{\gtrless}} \lambda_{T_u}. \tag{83}$$

From (31) and (33), we have that

$$\log|\boldsymbol{C}(\hat{\boldsymbol{\alpha}}_u, \hat{\omega}_u)| = \log(1 - z_{\hat{R}}(\hat{\omega}_u)) + \log|\hat{\boldsymbol{R}}|$$

and thus the problem is to determine λ_{T_u} such that

$$\lambda_u = -2N\log(1 - z_{\hat{R}}(\hat{\omega}_u)) > \lambda_{T_u} \tag{84}$$

occurs under H_0 with a given PFA P_f. Since the logarithm is monotonic, the inequality of (84) is equivalent to the condition

$$2Nz_{\hat{R}}(\hat{\omega}_u) > 2N(1 - \exp(-\lambda_{T_u}/2N)) \tag{85}$$

which for large N can further be approximated as

$$2Nz_{\hat{R}}(\hat{\omega}_u) > \lambda_{T_u} \tag{86}$$

since λ_{T_u}/N approaches zero as N increases.

Under H_0, we have that

$$\lim_{N\to\infty} \hat{\boldsymbol{R}}^{-1} = \boldsymbol{Q}^{-1} \quad \text{w.p.1} \tag{87}$$

$$\begin{aligned} \mathcal{E}\{\boldsymbol{y}(\omega)\boldsymbol{y}^*(\omega)\} &= \frac{1}{N^2}\sum_{t=1}^{N}\sum_{s=1}^{N}\mathcal{E}\{\boldsymbol{e}(t)\boldsymbol{e}^*(s)\}e^{j\omega(s-t)} \\ &= \frac{1}{N}\boldsymbol{Q}. \end{aligned} \tag{88}$$

With

$$z_Q(\omega) = \boldsymbol{y}^*(\omega)\boldsymbol{Q}^{-1}\boldsymbol{y}(\omega) \tag{89}$$

(87) implies that the asymptotic distributions of $z_{\hat{R}}(\omega)$ and $z_Q(\omega)$ coincide and hence that (84) is asymptotically equivalent to

$$2Nz_Q(\hat{\omega}_u) > \lambda_{T_u}. \tag{90}$$

Since $\boldsymbol{y}(\omega)$ is an m-element zero-mean circular Gaussian random vector with covariance $\boldsymbol{Q}/N$, $Nz_Q(\omega)$ is the sum of the squares of $2m$ zero-mean real-valued independent Gaussian random variables with variance $1/2$. Thus

$$2Nz_Q(\omega) \sim \chi^2(2m) \tag{91}$$

for every ω, and hence λ_u is (asymptotically) the maximum value of a collection of $\chi^2(2m)$ random variables. Note that, conditioned on ω, the number of extra free parameters under H_{1u} for the unstructured model is $2m$ (a real and imaginary part for each element of $\boldsymbol{\alpha}$), which means that (91) is in agreement with the general result of [20] mentioned earlier.

Define $\boldsymbol{\epsilon}_k(t)$ to be the kth element of the vector $\boldsymbol{\epsilon}(t) = \boldsymbol{Q}^{-(1/2)}\boldsymbol{e}(t)$ and observe that under H_0

$$\begin{aligned} Nz_Q(\omega) &= \frac{1}{N}\left\|\sum_{t=1}^{N}\boldsymbol{\epsilon}(t)e^{-j\omega t}\right\|^2 \\ &= \frac{1}{N}\sum_{i=1}^{m}\left|\sum_{t=1}^{N}\boldsymbol{\epsilon}_i(t)e^{-j\omega t}\right|^2 \\ &\triangleq \sum_{i=1}^{m}\Phi_i(\omega) \end{aligned}$$

which is simply the sum of the periodograms $\Phi_i(\omega)$ corresponding to each element of $\boldsymbol{\epsilon}(t)$. Since the elements of $\boldsymbol{\epsilon}(t)$ are independent circular white Gaussian processes, $\Phi_i(\omega_k)$ and $\Phi_j(\omega_l)$ can be shown to be (asymptotically) independent random variables whenever $i \neq j$ or $|\omega_k - \omega_l| \geq 2\pi/N$ [21]. If we restrict the evaluation of our detection variable to the DFT frequencies $\omega_k = 2\pi k/N, k = 0, \cdots, N-1$, then $z_Q(\omega_k)$ and $z_Q(\omega_l)$ are independent for $k \neq l$ and λ_u is asymptotically distributed as the maximum value of a collection of *independent* $\chi^2(2m)$ random variables. In what follows, we let $\hat{\bar{\omega}}_u$ denote the maximizer of the criterion $z_{\hat{R}}(\omega)$ evaluated on the aforementioned DFT frequency grid.

To find the threshold λ_{T_u} needed to achieve a given P_f, define

$$F(\zeta) = P\{\rho \leq \zeta | \rho \sim \chi^2(2m)\} \tag{92}$$

where $P\{\cdot\}$ denotes the probability of a generic event. For λ_u maximized over the N DFT frequencies, we have

$$\begin{aligned}P_f &\triangleq P\{\lambda_u > \lambda_{T_u}|H_0\}\\ &= P\{2Nz_{\hat{R}}(\hat{\bar{\omega}}_u) > 2N(1-\exp(-\lambda_{T_u}/2N))|H_0\}\\ &\simeq P\{2Nz_{\hat{R}}(\hat{\bar{\omega}}_u) > \lambda_{T_u}|H_0\}\\ &\simeq P\{2Nz_Q(\hat{\bar{\omega}}_u) > \lambda_{T_u}|H_0\}\\ &= 1-P\{2Nz_Q(\hat{\bar{\omega}}_u) \le \lambda_{T_u}|H_0\}\\ &= 1-P\{2Nz_Q(\omega_k) \le \lambda_{T_u}|H_0\},\ \forall\, k=0,\cdots,N-1\\ &= 1-[F(\lambda_{T_u})]^N.\end{aligned}$$

Thus, given P_f, we can calculate

$$F(\lambda_{T_u}) = (1-P_f)^{1/N}$$

and determine λ_{T_u} using (92) and a table of χ^2 probabilities.

The GLRT detection procedure derived in this section is summarized in the following.

Given a collection of data $\boldsymbol{x}(1),\cdots,\boldsymbol{x}(N)$, suppose it is desired to decide between the two hypotheses

$$\begin{aligned}H_0&\colon \boldsymbol{x}(t) = \boldsymbol{e}(t)\\ H_{1u}&\colon \boldsymbol{x}(t) = \boldsymbol{\alpha}_0 e^{j\omega_0 t} + \boldsymbol{e}(t)\end{aligned}$$

with probability of false alarm P_f, where $\boldsymbol{e}(t)$ is temporally white with unknown spatial covariance. A GLRT that asymptotically achieves this goal is outlined below.

1) Compute the threshold λ_{T_u} by solving
$$F(\lambda_{T_u}) = (1-P_f)^{(1/N)}$$
where $F(\zeta)$ is defined by (92).
2) Compute the sample covariance $\hat{\boldsymbol{R}}$ and calculate the DFT $\boldsymbol{y}(\omega_k)$ at the frequencies $\omega_k = 2\pi k/N$ for $k=0,\cdots,N-1$.
3) Find the DFT frequency ω_k that maximizes the vector periodogram $\boldsymbol{y}^*(\omega)\hat{\boldsymbol{R}}^{-1}\boldsymbol{y}(\omega)$; call this frequency $\hat{\bar{\omega}}_u$.
4) Decide that H_{1u} is true if
$$\lambda_u = -2N\log\left(1-\boldsymbol{y}^*(\hat{\bar{\omega}}_u)\hat{\boldsymbol{R}}^{-1}\boldsymbol{y}(\hat{\bar{\omega}}_u)\right) > \lambda_{T_u}$$
otherwise assume that H_0 holds.

The case of temporally correlated clutter and interference can be handled using the "whitening approach" outlined in Section IV-C. The performance degradation induced by the imperfections in the estimated whitening filter is expected to be minor.

1) Finite Sample Behavior of the GLRT: As emphasized already, the detection tests developed in Sections V-A and V-B are valid asymptotically in N. However, the application of the tests in situations where N is relatively small bears somewhat further discussion. We raise two issues here. First, the FFT grid used to evaluate the detection statistics is quite coarse when N is small. Consequently, if a target is present, its Doppler frequency is less likely to be near the center of any FFT bin, and hence the peak in the detection criterion due to the target may be artificially reduced. This in turn decreases the ability of the tests to detect a weak target. One way to overcome this problem is to obtain a finer frequency grid by zero-padding the data before taking the FFT. The asymptotic distribution of the resulting test statistic is more difficult to obtain (since the samples of the periodogram are no longer independent) and the asymptotic probability of false alarm is no longer guaranteed, but this may be a reasonable price to pay, especially considering that the asymptotic arguments are less accurate for small N anyway.

The second small-N issue we discuss has to do with the asymptotic approximation

$$2N(1-\exp(-\lambda_{T_u}/2N)) \simeq \lambda_{T_u} \tag{93}$$

used in going from (85) to (86). In principle, such a step might seem unnecessary. Indeed, as explained above, it is possible to choose a threshold

$$\lambda_z = 2N(1-\exp(-\lambda_{T_u}/2N)) \tag{94}$$

such that

$$2Nz_{\hat{R}}(\hat{\bar{\omega}}_u) > \lambda_z$$

with a certain desired probability P_f under H_0. The threshold λ_{T_u} could then be found exactly by inverting (94)

$$\lambda'_{T_u} = -2N\log\left(1-\frac{\lambda_z}{2N}\right). \tag{95}$$

We use the notation λ'_{T_u} to differentiate this threshold from the one obtained using the approximation of (93), which results in $\lambda_{T_u} = \lambda_z$. For a fixed P_f, λ_z increases very slowly with N [22], and it can indeed be shown that

$$\lim_{N\to\infty} \lambda'_{T_u} = \lambda_z = \lambda_{T_u}.$$

Despite the fact that the use of (95) would reduce our reliance on asymptotic arguments, there are some subtle reasons to prefer the test developed using the approximation of (93). First, when N is relatively small, the solution for λ'_{T_u} in (95) may not be defined for common values of P_f. In particular, the desired $\chi^2(2m)$ threshold λ_z is not guaranteed to be less than $2N$. In such cases, the equivalence between the tests in (84) and (85) breaks down, and the derivation of a threshold for (84) would require the finite sample (nonasymptotic) distribution of λ_u to be found. Using λ_{T_u} instead of (95) avoids this difficulty and also provides a well-defined test in all cases, although one could question its performance for small N. As demonstrated in Section VI, however, simulations show that the simpler threshold λ_{T_u} performs quite well for small N, essentially as well as the GLRT developed for the structured array model. But more than that, and quite surprisingly, it turns out that λ_{T_u} yields much better results than the more "exact" threshold derived using (95)—at least when N is large enough that λ'_{T_u} is well defined. The reason behind this interesting result can be explained as follows.

- Since $\log(z) \le z-1$ for all $z>0$, it can easily be shown that $\lambda'_{T_u} \ge \lambda_z = \lambda_{T_u}$.

- Although $2Nz_{\hat{R}}(\omega)$ is asymptotically $\chi^2(2m)$, it was shown in (37) that $2Nz_{\hat{R}}(\omega) \leq 2N$, implying that in finite samples, the random variables $\{2Nz_{\hat{R}}(\omega_k)\}$ tend to be smaller than $\chi^2(2m)$ random variables (which have a nonzero probability of exceeding $2N$).

Consequently, the lower threshold provided by λ_{T_u} tends to compensate for the fact that $2Nz_{\hat{R}}(\hat{\bar{\omega}}_u)$, and hence λ_u, are smaller than predicted by their asymptotic distributions.

VI. Simulation Results

We present below the results of several simulations comparing the detection and estimation performance of the structured and unstructured ML estimators. The following parameters were common to all of the simulations.

- The array was composed of $m = 12$ nominally identical, uniformly spaced, unit-gain antennas separated by one-half wavelength.
- If present, the calibration errors were generated as Gaussian random variables with covariance $\boldsymbol{\Omega} = \sigma_a^2 \boldsymbol{I}$, which amounts to assuming that the calibration errors are independent and identically distributed from antenna to antenna. The term *calibration error* is used below to refer to the quantity $\sqrt{2}\sigma_a$, which is the standard deviation of the total perturbation (both real and imaginary parts) for a given sensor.
- The target was located at $\theta_0 = 0°$ (broadside), with amplitude $b_0 = 1$ and Doppler shift $\omega_0 = 2\pi/10$.
- The covariance of the interference was set at $\boldsymbol{Q} = \sigma^2 \tilde{\boldsymbol{Q}}$ for various σ, where $\tilde{\boldsymbol{Q}}$ is a randomly chosen positive definite matrix scaled and regularized so that its determinant is unity. The same $\tilde{\boldsymbol{Q}}$ was used in all of the simulations; its maximum and minimum eigenvalues were 2.36 and 0.095, corresponding to a condition number of about 25. The signal-to-noise ratio (SNR) was defined to be $-20\log\sigma$ dB, as the target amplitude and nominal antenna gains were both unity.
- In some of the simulations, a point source "jammer" with a DOA of 5° was introduced into the data. The jammer signal was assumed to be temporally white (circular) Gaussian noise.
- The PFA was set to $P_f = 10^{-6}$ for the detection tests.

In what follows, the term "unstructured GLRT" (or simply UGLRT) will be used to refer to the GLRT obtained for the unstructured array model. UML will refer to the process of estimating the target parameters via the unstructured ML algorithm and the EXIP. A corresponding meaning is to be attached to the acronyms SGLRT and SML.

A. Verification of Asymptotics

The purpose of the first example we consider is to "validate" the asymptotic arguments presented earlier. In particular, we provide insight into the size of N required to achieve the asymptotic regime. The results presented here are indicative of numerous simulation runs. Fig. 1 shows the PFA for the structured and unstructured GLRT's versus the number of snapshots taken from the array with no signal present. The results are based on 1000 trials conducted for each value of N. The number of trials was not enough to verify if the desired $P_f = 10^{-6}$ was reached, but our main purpose was to compare the threshold behavior of the structured and unstructured GLRT's. While both tests approach a low value of P_f with increasing N, the "threshold N" for the structured GLRT appears to be a few (two to three) snapshots smaller than that for the unstructured GLRT.

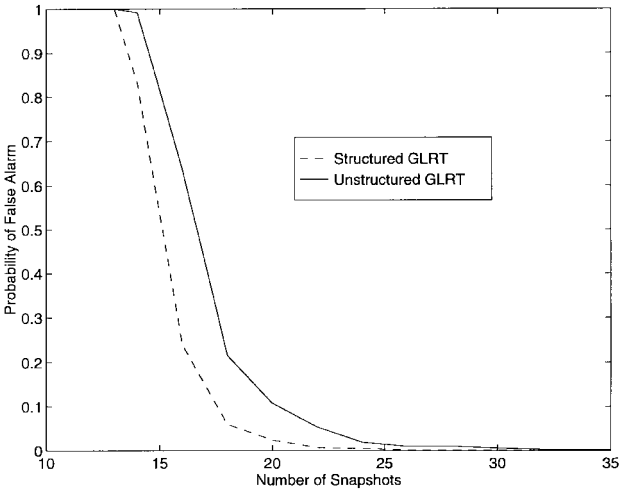

Fig. 1. PFA for structured and unstructured GLRT's versus number of snapshots.

The empirical behavior of the SML and UML parameter estimates is compared with the CRB in the next example. Figs. 2–4 show the root mean square (rms) error of the DOA, Doppler, and amplitude estimates, respectively, versus N for a case with 0-dB SNR and an ideal array (no calibration errors). Each data point on the plots is calculated from the results of 1000 independent experiments. It is clear from the figures that there is essentially no difference between the quality of the SML and UML estimates for all values of N. Thus, for these cases at least, the asymptotic equivalence of the two methods is apparent with very little data. Somewhat more data are required for their performance to reach the CRB, but there is little excess error for $N \geq 30$. Similar results have also been obtained at lower SNR's, though plots for these cases are not included here since they are almost identical to Figs. 2–4.

B. Robust Detection

Fig. 5 compares the detection probabilities of the SGLRT and UGLRT at various SNR (500 trials at each SNR) for a case involving an array with a calibration error of 0.1. Since the unperturbed array is assumed to have unit gain sensors, this corresponds to standard deviations in the array's gain and phase of -20 dB and 0.1 rad, respectively. A different perturbation was applied to the array at each trial, so the curves in Fig. 5 represent the detection probability averaged over uniform linear arrays with this level of calibration error. The three curves for the SGLRT show its performance for "pointing" errors of 0°, 1°, and 2°. Recall that the

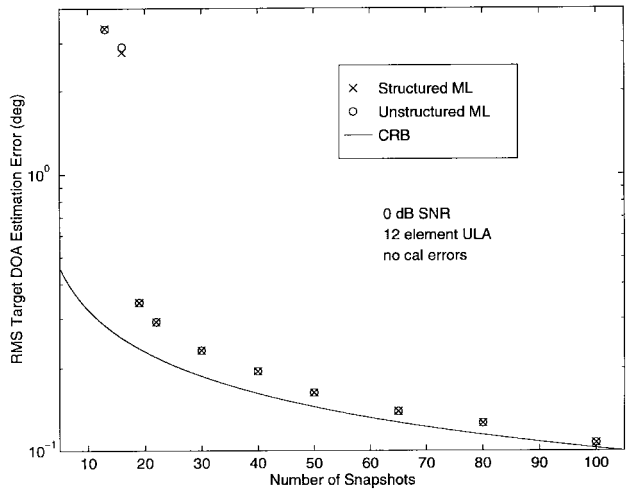

Fig. 2. DOA estimation performance and CRB versus number of snapshots, 0-dB SNR.

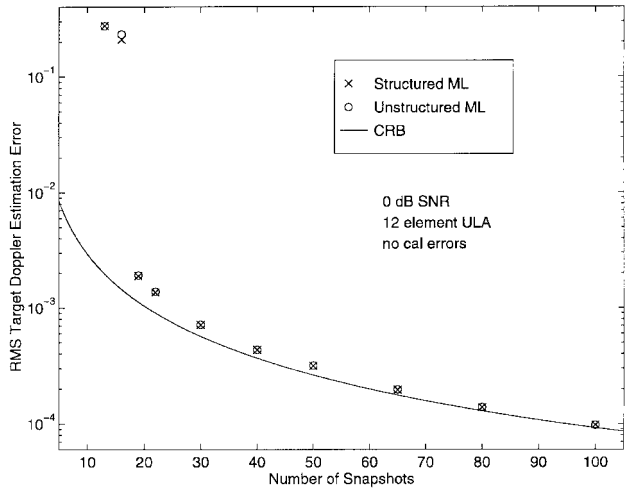

Fig. 3. Doppler estimation performance and CRB versus number of snapshots, 0-dB SNR.

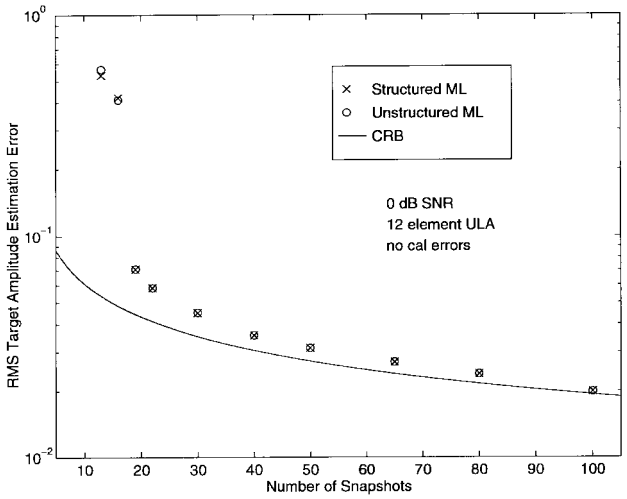

Fig. 4. Amplitude estimation performance and CRB versus number of snapshots, 0-dB SNR.

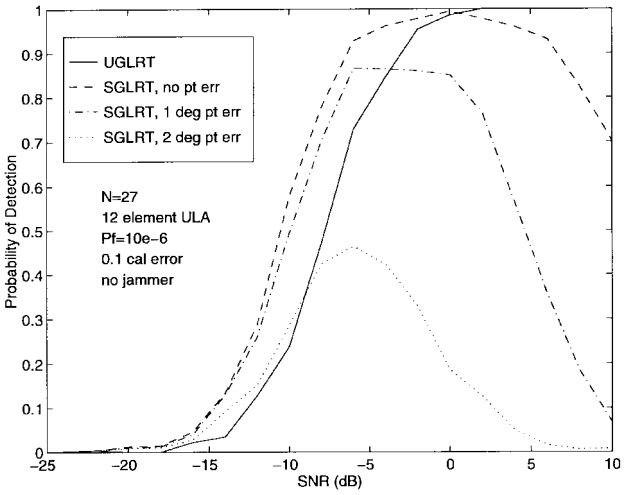

Fig. 5. Probability of detection versus SNR for various pointing errors.

SGLRT is implemented by evaluating the test statistic λ at $\hat{\theta}_s = \theta_t$, the pointing angle of the transmit beam. A pointing error occurs if the actual target DOA is slightly different from θ_t. Fig. 5 demonstrates that the SGLRT is very sensitive to the assumption that $\theta_t = \theta_0$. No such difficulty is observed with the UGLRT, of course, since it does not use DOA information. Even with no pointing error, we see that the performance of the SGLRT deteriorates at high SNR, since the presence of the calibration errors reduces the relative height of the peak value of the test statistic λ evaluated over ω. At higher SNR, there is less of a chance that the noise will artificially push the peak over the threshold. This deterioration is more evident with pointing errors, since the peak has been reduced even further.

The above experiment was repeated with a strong jammer added to the data, and the resulting probabilities of detection are plotted in Fig. 6. The jammer was located at a DOA of 5°, and the jammer-to-signal power ratio (JSR) was held fixed at 20 dB. Since the jammer is assumed to broadcast temporally white noise, its net statistical effect is to make a large rank one contribution to $\boldsymbol{Q}$. The presence of the jammer causes a slight reduction in probability of detection at low SNR but appears to lead to an "improvement" in performance for the SGLRT at higher SNR. This improvement is only artificial, however, since it is the random contribution of the jammer that is pushing the peak of the detection criterion above the threshold. We will see below that the estimation performance of SML in these high SNR regions is very poor.

C. Robust Estimation

We study here the estimation performance of SML and UML for the same scenario described in the previous section for the detection tests. The only difference in this example is that the calibration error for the array was increased to 0.2, corresponding to errors of -14 dB in gain and 0.2 rad in phase. The target parameters were estimated by maximizing the SML and UML criteria on a grid with spacing $2\pi/N_F$ in ω and 0.05° in θ over the interval [$-7°$,

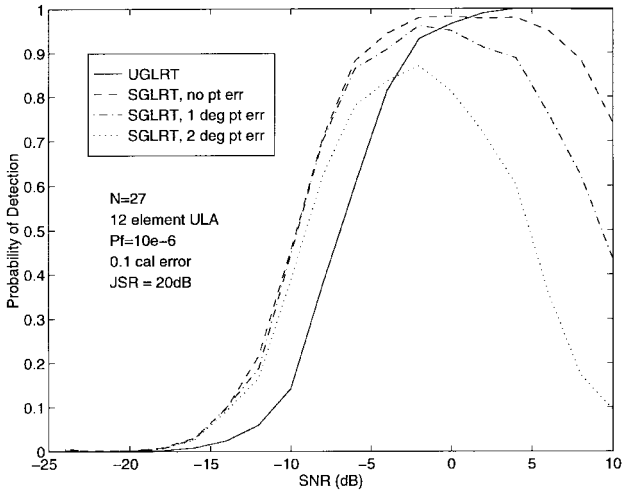

Fig. 6. Probability of detection versus SNR for various pointing errors, strong jammer included.

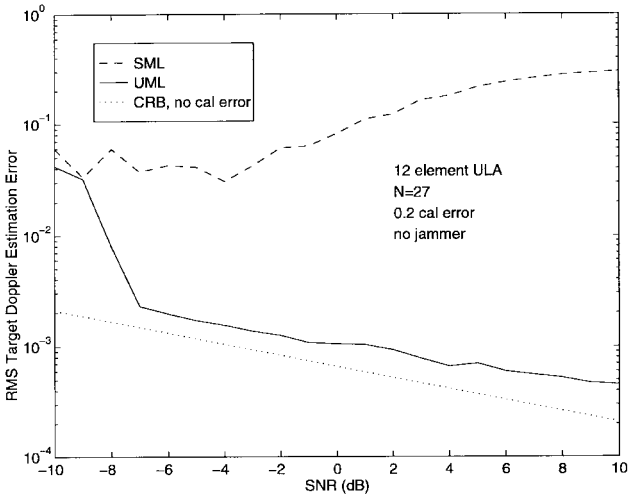

Fig. 8. Doppler estimation performance versus SNR, no jammer present.

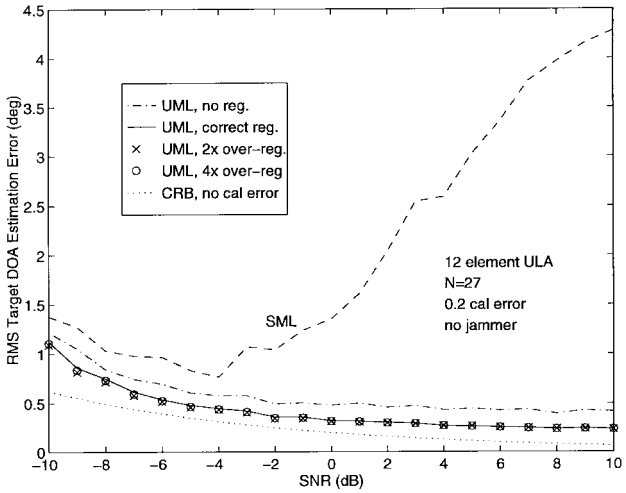

Fig. 7. DOA estimation performance versus SNR, no jammer present.

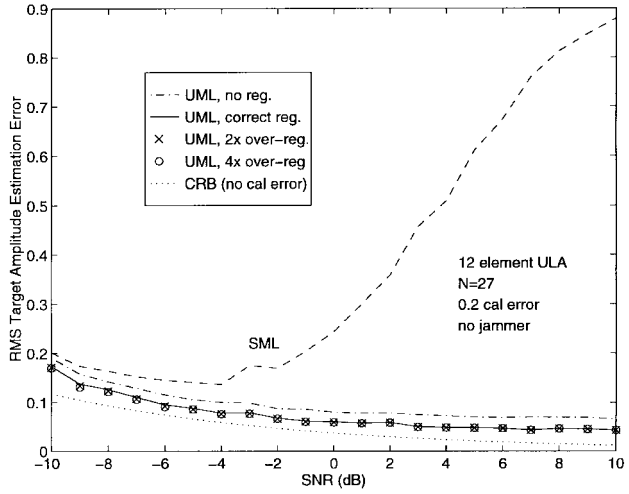

Fig. 9. Amplitude estimation performance versus SNR, no jammer present.

7°]. The frequency resolution was increased by choosing $N_F > N$ and zero-padding the data to the required length. In each case, N_F was chosen large enough so that the error due to the grid quantization would be several times smaller than the error predicted by the CRB. This was also the criterion used in choosing the grid spacing for the DOA parameter.

The regularized weighting of (63) was used for UML, but the value of σ_a used for the weighting was varied to study the sensitivity of the method to knowledge of the calibration error statistics. The initial estimate of b_0 required for the regularization was obtained by first applying UML with the standard weighting. Figs. 7–9 show the rms error versus SNR with no jammer present for the target DOA, Doppler, and amplitude estimates, respectively. Figs. 10–12 show the results when a jammer at 5° is present with JSR $= 20$ dB. The dash-dot line indicates the performance of UML with no regularization, while the solid line gives the results for the correct weighting. The symbols "x" and "o" on the plots correspond to the results for regularized UML implemented with σ_a overestimated by a factor of two and four, respectively. It is seen that the performance improvement offered by the regularization is very insensitive to exact knowledge of this parameter.

With or without regularization, the UML method is several times more accurate than SML at high SNR. The presence of the calibration error reduces the peak of the 2-D SML criterion at the true θ and ω, and produces a local maximum that is far removed from them. The UML technique does not suffer from this problem since at high SNR, it is able accurately to estimate ω without being affected by the array perturbation. With an accurate estimate of ω, the error in the DOA parameter is correspondingly reduced. The difference between the methods is smaller at low SNR but not insignificant. If the calibration error is reduced to 0.1, as in the detection examples above, all of the algorithms perform essentially identical at low SNR but SML still degrades rapidly as the SNR is increased above 0 dB. Last, note that the CRB is included in each plot to indicate the best performance that would be possible if no calibration errors were present.

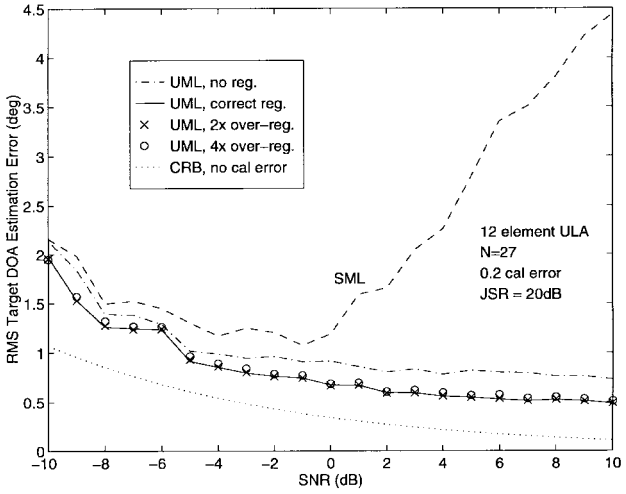

Fig. 10. DOA estimation performance versus SNR, strong jammer present.

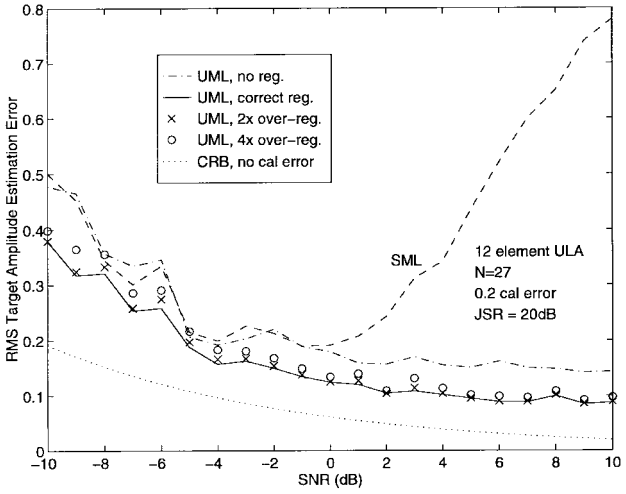

Fig. 12. Amplitude estimation performance versus SNR, strong jammer present.

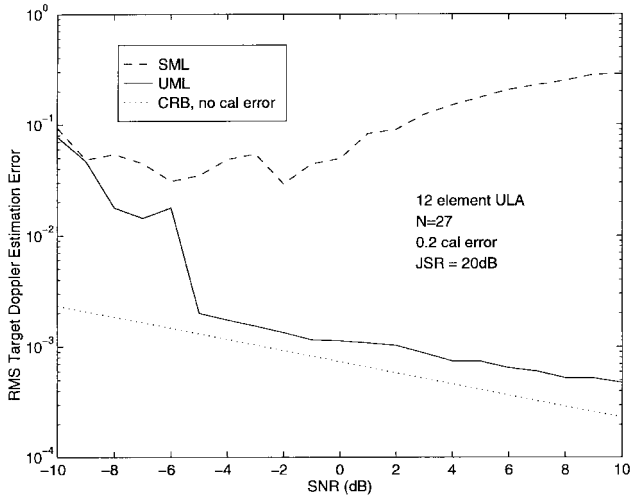

Fig. 11. Doppler estimation performance versus SNR, strong jammer present.

VII. Conclusions

We have presented several techniques for detection and estimation in radar problems where a target is observed by an antenna array in noise with unknown spatial covariance. Two data models were introduced: one that uses a structured model with the target signal amplitude and DOA as parameters and one that is a simpler, unstructured model for the array response. ML solutions were derived for each case and the corresponding CRB's established. The EXIP was used to show how the simpler unstructured solution could be processed to estimate the target parameters with the same asymptotic accuracy as the more complicated structured algorithm.

Several advantages of the EXIP approach were observed. First, statistically efficient estimates of the target parameters can be obtained using two 1-D searches rather than one 2-D search. Since the estimates are often determined by evaluating the given criterion on a grid, this can provide substantial computational savings. Furthermore, if a uniform linear array is used, one of the 1-D searches can be eliminated for further reduction of the computational load. Second, the two-step approach using the unstructured model provides a simple and effective way of taking array calibration errors into account in the estimation of the target DOA. The resulting DOA estimates can be significantly more accurate and much less sensitive to such errors. Last, the use of an unstructured model for the array response allows for the development of a likelihood ratio detection test that does not depend on the accuracy or the availability of array calibration data, nor on the availability of prior information about the target DOA. Consequently, the test is robust in situations where such information is imprecisely known. Our simulations indicate that when the array is perfectly calibrated, little degradation results from using the simpler two-step algorithm over the direct structured approach. When calibration or pointing errors are present, the simpler algorithm offers much better performance.

In summary, the use of an unstructured array model together with the EXIP provides a very simple procedure for applying ML methods to radar signal processing problems. The computational simplicity and robustness of the proposed estimation and detection algorithms make them worthy of serious consideration for real radar applications.

Appendix A
Properties of the Structured ML Estimator

A. Consistency

We first establish the consistency of the structured ML algorithm. To begin with, we have

$$\lim_{N\to\infty} \boldsymbol{y}(\omega) = \lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N}\left[b_0\boldsymbol{a}(\theta_0)e^{j(\omega_0-\omega)t} + \boldsymbol{e}(t)e^{-j\omega t}\right] \tag{96}$$

$$= \lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} b_0\boldsymbol{a}(\theta_0)e^{j(\omega_0-\omega)t} + \lim_{N\to\infty}\frac{1}{N} \cdot \sum_{t=1}^{N}\boldsymbol{e}(t)e^{-j\omega t}. \tag{97}$$

If we let $\delta(\cdot)$ denote the Kronecker delta, the limit of the first term in (97) is simply $b_0\boldsymbol{a}(\theta_0)\delta(\omega_0-\omega)$ since

$$\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}e^{j(\omega_0-\omega)t}=\begin{cases}1 & \text{if } \omega=\omega_0\\ 0 & \text{otherwise.}\end{cases}$$

By the law of large numbers [23], the second term converges uniformly to zero with probability one since $\boldsymbol{e}(t)$ is zero mean. Consequently

$$\lim_{N\to\infty}\boldsymbol{y}(\omega)=\begin{cases}b_0\boldsymbol{a}(\theta_0) & \text{if } \omega=\omega_0\\ 0 & \text{otherwise}\end{cases}\quad \text{w.p.1.} \tag{98}$$

Letting $\boldsymbol{R}=\lim_{N\to\infty}\hat{\boldsymbol{R}}$, the ML criterion in (16) is thus asymptotically given by

$$\lim_{N\to\infty}\mathcal{V}_N(\theta,\omega)=\begin{cases}\dfrac{|b_0|^2}{1-|b_0|^2\boldsymbol{a}^*(\theta_0)\boldsymbol{R}^{-1}\boldsymbol{a}(\theta_0)}\dfrac{|\boldsymbol{a}^*(\theta)\boldsymbol{R}^{-1}\boldsymbol{a}(\theta_0)|^2}{\boldsymbol{a}^*(\theta)\boldsymbol{R}^{-1}\boldsymbol{a}(\theta)} & \text{if } \omega=\omega_0\\ 0 & \text{otherwise}\end{cases}$$

with probability one. At $\omega=\omega_0$, the Cauchy–Schwartz inequality may be used to obtain the following upper bound for the limiting ML criterion:

$$\lim_{N\to\infty}\mathcal{V}_N(\theta,\omega_0)=\frac{|b_0|^2}{c}\frac{|\boldsymbol{a}^*(\theta)\boldsymbol{R}^{-1}\boldsymbol{a}(\theta_0)|^2}{\boldsymbol{a}^*(\theta)\boldsymbol{R}^{-1}\boldsymbol{a}(\theta)}\le\frac{|b_0|^2\boldsymbol{a}^*(\theta_0)\boldsymbol{R}^{-1}\boldsymbol{a}(\theta_0)}{c}$$

where $c=1-|b_0|^2\boldsymbol{a}^*(\theta_0)\boldsymbol{R}^{-1}\boldsymbol{a}(\theta_0)$. Equality is achieved iff $\boldsymbol{a}(\theta)=\boldsymbol{a}(\theta_0)$, which in turn is possible iff $\theta=\theta_0$ since we implicitly assume an "unambiguous" array.

Although the limiting ML criterion has a global maximum at the true parameter values, establishing consistency is complicated by the fact that the criterion is discontinuous in ω. This problem can be solved using arguments similar to those in Lemma 1 and Theorem 1 of [24]. It was shown in [24] that consistency is achieved provided that $\mathcal{V}_N(\theta,\omega)$ can be written as the sum of two components

$$\mathcal{V}_N(\theta,\omega)=f_N(\theta,\omega)+\varepsilon_N(\theta,\omega)$$

that satisfy the following conditions:

1) $\lim_{N\to\infty}\varepsilon_N(\theta,\omega)=0$ uniformly in ω;
2) for any $\delta>0$, there exist $\epsilon>0$ and N_M such that
$$f_N(\theta,\omega)\le f_N(\theta,\omega_0)+\epsilon$$
for all $|\omega-\omega_0|>\delta$ and $N>N_M$.

Using (16) and the definition for $\boldsymbol{y}(\omega)$, we may write $\mathcal{V}_N(\theta,\omega)$ as (99) and (100), shown at the bottom of the page, where $\boldsymbol{Z}_\theta=\hat{\boldsymbol{R}}^{-1}\boldsymbol{a}(\theta)\boldsymbol{a}^*(\theta)\hat{\boldsymbol{R}}^{-1}$. If we associate $f_N(\theta,\omega)$ with (99) and $\varepsilon_N(\theta,\omega)$ with (100), it can be shown that conditions 1) and 2) listed above are satisfied (a proof of this result will not be given here, however). Using reasoning similar to that for the proof of Theorem 1 in [24], it follows that

$$\lim_{N\to\infty}\hat{\theta}_s=\theta_0 \tag{101}$$

$$\lim_{N\to\infty}\hat{\omega}_s=\omega_0. \tag{102}$$

The consistency of the estimates for b and $\boldsymbol{Q}$ is easily established by substituting θ_0 and ω_0 into (15) and (12) and (13).

B. Cramér–Rao Bound

Since the structured ML estimates are consistent, they will be statistically efficient, and their asymptotic variance will be given by the CRB. Under the structured model, the observations $\boldsymbol{x}(t),t=1,\cdots,N$ are independent circular Gaussian random variables with mean $\boldsymbol{\mu}(t)=b\boldsymbol{a}(\theta)e^{j\omega t}$ and covariance $\boldsymbol{Q}$, where we have dropped the subscript 0 for brevity. In this case, element k,l of the FIM for the N observations can be shown to be equal to [21], [25], [26]

$$\mathbf{FIM}_{kl}=N\,\mathrm{Tr}\left(\boldsymbol{Q}^{-1}\frac{\partial\boldsymbol{Q}}{\partial\boldsymbol{\eta}_k}\boldsymbol{Q}^{-1}\frac{\partial\boldsymbol{Q}}{\partial\boldsymbol{\eta}_l}\right)+2\,\mathrm{Re}\sum_{t=1}^{N}\left\{\left(\frac{\partial\boldsymbol{\mu}(t)}{\partial\boldsymbol{\eta}_k}\right)^*\boldsymbol{Q}^{-1}\left(\frac{\partial\boldsymbol{\mu}(t)}{\partial\boldsymbol{\eta}_l}\right)\right\}. \tag{103}$$

Since $\boldsymbol{\mu}(t)$ and $\boldsymbol{Q}$ depend on different elements of $\boldsymbol{\eta}$, it is clear that $\mathbf{FIM}$ will be block diagonal with respect to the signal $(\boldsymbol{\eta}_s=[\mathrm{Re}\{b\},\mathrm{Im}\{b\},\theta,\omega]^T)$ and noise $(\boldsymbol{Q})$ parameters. In particular, the first term of (103) will give a nonzero result only for the noise block, while the second term is only nonzero for the signal block. Since we are concerned with the CRB only for the signal parameters, we need only consider the second term

$$\mathbf{FIM}_{kl}(\boldsymbol{\eta}_s)=2\sum_{t=1}^{N}\mathrm{Re}\left\{\left(\frac{\partial\boldsymbol{\mu}(t)}{\partial\boldsymbol{\eta}_{s,k}}\right)^*\boldsymbol{Q}^{-1}\left(\frac{\partial\boldsymbol{\mu}(t)}{\partial\boldsymbol{\eta}_{s,l}}\right)\right\}. \tag{104}$$

$$\mathcal{V}_N(\theta,\omega)=\frac{\displaystyle\sum_{t,s=1}^{N}\left|b_0\boldsymbol{a}^*(\theta)\hat{\boldsymbol{R}}^{-1}\boldsymbol{a}(\theta_0)\right|^2e^{j(\omega_0-\omega)(s-t)}}{N^2\Big(1-\boldsymbol{y}^*(\omega)\hat{\boldsymbol{R}}^{-1}\boldsymbol{y}(\omega)\Big)\boldsymbol{a}^*(\theta)\hat{\boldsymbol{R}}^{-1}\boldsymbol{a}(\theta)} \tag{99}$$

$$+\frac{\displaystyle\sum_{t,s=1}^{N}(b_0^*\boldsymbol{a}^*(\theta_0)\boldsymbol{Z}_\theta\boldsymbol{e}(t)e^{-j\omega_0 s}+b_0\boldsymbol{e}^*(s)\boldsymbol{Z}_\theta\boldsymbol{a}(\theta_0)e^{j\omega_0 t}+\boldsymbol{e}^*(s)\boldsymbol{Z}_\theta\boldsymbol{e}(t))e^{j\omega(s-t)}}{N^2\Big(1-\boldsymbol{y}^*(\omega)\hat{\boldsymbol{R}}^{-1}\boldsymbol{y}(\omega)\Big)\boldsymbol{a}^*(\theta)\hat{\boldsymbol{R}}^{-1}\boldsymbol{a}(\theta)} \tag{100}$$

The partial derivatives needed to compute (104) are given below

$$\frac{\partial \boldsymbol{\mu}(t)}{\partial \mathrm{Re}\{b\}} = \boldsymbol{a}(\theta)e^{j\omega t} \tag{105}$$

$$\frac{\partial \boldsymbol{\mu}(t)}{\partial \mathrm{Im}\{b\}} = j\boldsymbol{a}(\theta)e^{j\omega t} \tag{106}$$

$$\frac{\partial \boldsymbol{\mu}(t)}{\partial \theta} = b\boldsymbol{d}(\theta)e^{j\omega t} \tag{107}$$

$$\frac{\partial \boldsymbol{\mu}(t)}{\partial \omega} = jtb\boldsymbol{a}(\theta)e^{j\omega t} \tag{108}$$

where

$$\boldsymbol{d}(\theta) = \frac{\partial \boldsymbol{a}(\theta)}{\partial \theta}.$$

The expression for the CRB, shown in (109) at the bottom of the page, is obtained by substituting (105)–(108) into (104) where

$$\beta_N = \frac{(N+1)}{2} \tag{110}$$

$$\gamma_N = \frac{(N+1)(N+\frac{1}{2})}{3} \tag{111}$$

and $\boldsymbol{a} = \boldsymbol{a}(\theta_0), \boldsymbol{d} = \boldsymbol{d}(\theta_0)$.

To find the variance of the ML Doppler and DOA estimates, we use the following formula for the inverse of a partitioned symmetric matrix:

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}^T & \boldsymbol{C} \end{bmatrix}^{-1} = \begin{bmatrix} (\boldsymbol{A}-\boldsymbol{B}\boldsymbol{C}^{-1}\boldsymbol{B}^T)^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{D} \\ -\boldsymbol{D}\boldsymbol{B}^T\boldsymbol{A}^{-1} & \boldsymbol{D} \end{bmatrix} \tag{112}$$

where $\boldsymbol{D} = (\boldsymbol{C}-\boldsymbol{B}^T\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}$ is the lower 2×2 block of $\mathbf{CRB}(\boldsymbol{\eta}_s)$. For convenience, define $c = \boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{d}, d = \boldsymbol{d}^*\boldsymbol{Q}^{-1}\boldsymbol{d}, f = \boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{a}$ and let the subscripts r and i denote real and imaginary parts, respectively. Then

$$\begin{aligned}
&\frac{1}{2N}\boldsymbol{D}^{-1} \\
&\quad = \boldsymbol{C} - \frac{1}{f}\boldsymbol{B}^T\boldsymbol{B} \\
&\quad = |b|^2\begin{bmatrix} d & \beta_N c_i \\ \beta_N c_i & \gamma_N f \end{bmatrix} - \frac{1}{f} \\
&\qquad \cdot \begin{bmatrix} |bc|^2 & \beta_N f[b_r(bc)_i - b_i(bc)_r] \\ \beta_N f[b_r(bc)_i - b_i(bc)_r] & \beta_N^2 f^2(b_i^2+b_r^2) \end{bmatrix} \\
&\quad = |b|^2\begin{bmatrix} d & \beta_N c_i \\ \beta_N c_i & \gamma_N f \end{bmatrix} - \frac{|b|^2}{f}\begin{bmatrix} |c|^2 & \beta_N f c_i \\ \beta_N f c_i & \beta_N^2 f^2 \end{bmatrix} \\
&\quad = \begin{bmatrix} |b|^2\left(d - \frac{|c|^2}{f}\right) & 0 \\ 0 & |b|^2 f(\gamma_N - \beta_N^2) \end{bmatrix}.
\end{aligned}$$

The estimate variances are thus given by

$$\begin{aligned}
&\mathcal{E}\{(\hat{\theta}_s - \theta_0)^2\} \\
&\qquad = \frac{\boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{a}}{2N|b|^2[(\boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{a})(\boldsymbol{d}^*\boldsymbol{Q}^{-1}\boldsymbol{d}) - |\boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{d}|^2]}
\end{aligned} \tag{113}$$

$$\begin{aligned}
&\mathcal{E}\{(\hat{\omega}_s - \omega_0)^2\} \\
&\qquad = \frac{1}{2N|b|^2(\gamma_N - \beta_N^2)\boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{a}} \simeq \frac{6}{N^3|b|^2\boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{a}}.
\end{aligned} \tag{114}$$

Appendix B
Properties of the Unstructured ML Estimator

A. Consistency

The consistency of the unstructured ML algorithm is established in a manner similar to the structured case. Following (96)–(98), we have

$$\lim_{N\to\infty} \boldsymbol{y}(\omega) = \begin{cases} \boldsymbol{\alpha}_0 & \text{if } \omega = \omega_0 \\ 0 & \text{otherwise} \end{cases}$$

and hence

$$\lim_{N\to\infty} \boldsymbol{y}^*(\omega)\hat{\boldsymbol{R}}^{-1}\boldsymbol{y}(\omega) = \begin{cases} \boldsymbol{\alpha}_0^*\boldsymbol{R}^{-1}\boldsymbol{\alpha}_0 & \text{if } \omega = \omega_0 \\ 0 & \text{otherwise.} \end{cases}$$

Splitting the criterion into two terms as

$$\begin{aligned}
&\boldsymbol{y}^*(\omega)\hat{\boldsymbol{R}}^{-1}\boldsymbol{y}(\omega) \\
&\quad = \frac{1}{N^2}\sum_{t,s=1}^{N} \boldsymbol{\alpha}_0^*\hat{\boldsymbol{R}}^{-1}\boldsymbol{\alpha}_0 e^{-j(\omega_0-\omega)(t-s)} \qquad (115) \\
&\qquad + \frac{1}{N^2}\sum_{t,s=1}^{N}\Big[\boldsymbol{\alpha}_0^*\hat{\boldsymbol{R}}^{-1}\boldsymbol{e}(t)e^{-j\omega_0 s} + \boldsymbol{e}^*(s) \\
&\qquad\qquad \cdot\hat{\boldsymbol{R}}^{-1}\boldsymbol{\alpha}_0 e^{j\omega_0 t} + \boldsymbol{e}^*(s)\hat{\boldsymbol{R}}^{-1}\boldsymbol{e}(t)\Big] \\
&\qquad \cdot e^{-j\omega(t-s)} \qquad (116)
\end{aligned}$$

we again have two functions that satisfy the consistency conditions of [24] given in Appendix A. Thus, $\lim_{N\to\infty}\hat{\omega}_u = \omega_0$ and furthermore

$$\lim_{N\to\infty}\hat{\boldsymbol{\alpha}}_u = \lim_{N\to\infty}\boldsymbol{y}(\hat{\omega}_u) = \boldsymbol{\alpha}_0.$$

The consistency of $\hat{\boldsymbol{Q}}_u$ follows from substituting $\boldsymbol{\alpha}_0, \omega_0$ into (25).

$$\mathbf{CRB}(\boldsymbol{\eta}_s) = \frac{1}{2N}\begin{bmatrix} \boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{a} & 0 & \mathrm{Re}\{b\boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{d}\} & -\beta_N \mathrm{Im}\{b\}\boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{a} \\ 0 & \boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{a} & \mathrm{Im}\{b\boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{d}\} & \beta_N \mathrm{Re}\{b\}\boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{a} \\ \mathrm{Re}\{b\boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{d}\} & \mathrm{Im}\{b\boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{d}\} & |b|^2\boldsymbol{d}^*\boldsymbol{Q}^{-1}\boldsymbol{d} & \beta_N|b|^2 \mathrm{Im}\{\boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{d}\} \\ -\beta_N \mathrm{Im}\{b\}\boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{a} & \beta_N \mathrm{Re}\{b\}\boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{a} & \beta_N|b|^2 \mathrm{Im}\{\boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{d}\} & \gamma_N|b|^2\boldsymbol{a}^*\boldsymbol{Q}^{-1}\boldsymbol{a} \end{bmatrix}^{-1} \tag{109}$$

B. Cramér–Rao Bound

The data $\boldsymbol{x}(t), t = 1, \cdots, N$ in the unstructured model are independent circular Gaussian random variables with mean $\boldsymbol{\mu}(t) = \boldsymbol{\alpha} e^{j\omega t}$ and covariance $\boldsymbol{Q}$. As for the structured case studied in Appendix A, the unstructured CRB is block diagonal with respect to the signal $(\tilde{\boldsymbol{\eta}}_s = [\mathrm{Re}\{\boldsymbol{\alpha}^T\}, \mathrm{Im}\{\boldsymbol{\alpha}^T\}, \omega]^T)$ and noise $(\boldsymbol{Q})$ parameters. The FIM is computed exactly as in (104), with $\boldsymbol{\eta}_s$ replaced by $\tilde{\boldsymbol{\eta}}_s$, and the required partial derivatives are given by

$$\frac{\partial \boldsymbol{\mu}(t)}{\partial \mathrm{Re}\{\boldsymbol{\alpha}\}} = e^{j\omega t}\boldsymbol{I} \tag{117}$$

$$\frac{\partial \boldsymbol{\mu}(t)}{\partial \mathrm{Im}\{\boldsymbol{\alpha}\}} = je^{j\omega t}\boldsymbol{I} \tag{118}$$

$$\frac{\partial \boldsymbol{\mu}(t)}{\partial \omega} = jt\boldsymbol{\alpha} e^{j\omega t}. \tag{119}$$

Substituting (117)–(119) into (104) yields the expression for the FIM shown in (120) at the bottom of the page.

An explicit expression for the CRB can be obtained by inverting (120). To this end, we make use of (112) and the matrix inversion lemma

$$(\boldsymbol{A} - \boldsymbol{B}\boldsymbol{C}^{-1}\boldsymbol{B}^T)^{-1} = \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{D}\boldsymbol{B}^T\boldsymbol{A}^{-1} \tag{121}$$

by associating $\boldsymbol{A}$ with the upper 2×2 block of $\mathbf{FIM}(\tilde{\boldsymbol{\eta}}_s)$ and $\boldsymbol{C}$ with the scalar in the lower right corner. We begin by noting that

$$\begin{aligned}\boldsymbol{A}^{-1} &= \frac{1}{2N}\begin{bmatrix}\mathrm{Re}\{\boldsymbol{Q}^{-1}\} & -\mathrm{Im}\{\boldsymbol{Q}^{-1}\} \\ \mathrm{Im}\{\boldsymbol{Q}^{-1}\} & \mathrm{Re}\{\boldsymbol{Q}^{-1}\}\end{bmatrix}^{-1} \\ &= \frac{1}{2N}\begin{bmatrix}\boldsymbol{Q}_r & -\boldsymbol{Q}_i \\ \boldsymbol{Q}_i & \boldsymbol{Q}_r\end{bmatrix}\end{aligned}$$

where we have used the notation $\boldsymbol{Q}_r = \mathrm{Re}\{\boldsymbol{Q}\}, \boldsymbol{Q}_i = \mathrm{Im}\{\boldsymbol{Q}\}$. Furthermore

$$\boldsymbol{A}^{-1}\boldsymbol{B} = \begin{bmatrix}\boldsymbol{Q}_r & -\boldsymbol{Q}_i \\ \boldsymbol{Q}_i & \boldsymbol{Q}_r\end{bmatrix}\begin{bmatrix}-\beta_N \mathrm{Im}\{\boldsymbol{Q}^{-1}\boldsymbol{\alpha}\} \\ \beta_N \mathrm{Re}\{\boldsymbol{Q}^{-1}\boldsymbol{\alpha}\}\end{bmatrix} \tag{122}$$

$$= \beta_N\begin{bmatrix}-\boldsymbol{Q}_r \mathrm{Im}\{\boldsymbol{Q}^{-1}\boldsymbol{\alpha}\} - \boldsymbol{Q}_i \mathrm{Re}\{\boldsymbol{Q}^{-1}\boldsymbol{\alpha}\} \\ -\boldsymbol{Q}_i \mathrm{Im}\{\boldsymbol{Q}^{-1}\boldsymbol{\alpha}\} + \boldsymbol{Q}_r \mathrm{Re}\{\boldsymbol{Q}^{-1}\boldsymbol{\alpha}\}\end{bmatrix} \tag{123}$$

$$= \beta_N\begin{bmatrix}-\boldsymbol{\alpha}_i \\ \boldsymbol{\alpha}_r\end{bmatrix} \tag{124}$$

where $\boldsymbol{\alpha}_r = \mathrm{Re}\{\boldsymbol{\alpha}\}, \boldsymbol{\alpha}_i = \mathrm{Im}\{\boldsymbol{\alpha}\}$. Thus

$$\boldsymbol{D} = \frac{1}{2N}\Bigg(\gamma_N \boldsymbol{\alpha}^* \boldsymbol{Q}^{-1}\boldsymbol{\alpha} - \beta_N^2\begin{bmatrix}\mathrm{Im}\{\boldsymbol{\alpha}^*\boldsymbol{Q}^{-1}\} & \mathrm{Re}\{\boldsymbol{\alpha}^*\boldsymbol{Q}^{-1}\}\end{bmatrix}\begin{bmatrix}-\boldsymbol{\alpha}_i \\ \boldsymbol{\alpha}_r\end{bmatrix}\Bigg)^{-1} \tag{125}$$

$$= \frac{1}{2N}(\gamma_N \boldsymbol{\alpha}^*\boldsymbol{Q}^{-1}\boldsymbol{\alpha} - \beta_N^2[-\mathrm{Im}\{\boldsymbol{\alpha}^*\boldsymbol{Q}^{-1}\}\boldsymbol{\alpha}_i + \mathrm{Re}\{\boldsymbol{\alpha}^*\boldsymbol{Q}^{-1}\}\boldsymbol{\alpha}_r])^{-1} \tag{126}$$

$$= \frac{1}{2N(\gamma_N - \beta_N^2)\boldsymbol{\alpha}^*\boldsymbol{Q}^{-1}\boldsymbol{\alpha}} = \frac{1}{2N\delta_N} \tag{127}$$

which is the lower right element of (39). Together, (124) and (127) also establish the validity of the lower left and upper right blocks of (39). The remaining block also easily follows

$$\begin{aligned}&\boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{D}\boldsymbol{B}^T\boldsymbol{A}^{-1} \\ &\quad= \frac{1}{2N}\left(\begin{bmatrix}\boldsymbol{Q}_r & -\boldsymbol{Q}_i \\ \boldsymbol{Q}_i & \boldsymbol{Q}_r\end{bmatrix} + \frac{\beta_N^2}{\delta_N}\begin{bmatrix}-\boldsymbol{\alpha}_i \\ \boldsymbol{\alpha}_r\end{bmatrix}[-\boldsymbol{\alpha}_i^T \quad \boldsymbol{\alpha}_r^T]\right) \\ &\quad= \frac{1}{2N}\begin{bmatrix}\boldsymbol{Q}_r + \frac{\beta_N^2}{\delta_N}\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i^T & -\boldsymbol{Q}_i - \frac{\beta_N^2}{\delta_N}\boldsymbol{\alpha}_i\boldsymbol{\alpha}_r^T \\ \boldsymbol{Q}_i - \frac{\beta_N^2}{\delta_N}\boldsymbol{\alpha}_r\boldsymbol{\alpha}_i^T & \boldsymbol{Q}_r + \frac{\beta_N^2}{\delta_N}\boldsymbol{\alpha}_r\boldsymbol{\alpha}_r^T\end{bmatrix}.\end{aligned}$$

Appendix C
Proof of Theorem 4

The theorem is proved by applying the general EXIP result in Theorem 1 to the criteria of the structured $(V_N(\boldsymbol{\eta}))$ and unstructured $(\tilde{V}_N(\tilde{\boldsymbol{\eta}}))$ ML radar target parameter estimators. In particular, a function $f(\boldsymbol{\eta})$ will be defined that satisfies (7) and relates the SML and UML parameters, and it will be shown that the WLS minimization of (9) reduces to (47)–(50), (53), and (54). Thus, Theorem 1 states that the estimates of $\boldsymbol{Q}, \omega, b$, and θ, obtained respectively by (47), (48), (53), and (54), are asymptotically as accurate as those obtained by the structured ML algorithm and hence are statistically efficient.

The function f relating the elements of D_η and $D_{\tilde{\eta}}$ is defined implicitly by

$$\boldsymbol{Q}_u = \boldsymbol{Q} \tag{128}$$

$$\omega_u = \omega \tag{129}$$

$$\begin{bmatrix}\mathrm{Re}\{\boldsymbol{\alpha}_u\} \\ \mathrm{Im}\{\boldsymbol{\alpha}_u\}\end{bmatrix} = \boldsymbol{A}(\theta)\begin{bmatrix}\mathrm{Re}\{b\} \\ \mathrm{Im}\{b\}\end{bmatrix} \tag{130}$$

where

$$\boldsymbol{A}(\theta) = \begin{bmatrix}\mathrm{Re}\{\boldsymbol{a}(\theta)\} & -\mathrm{Im}\{\boldsymbol{a}(\theta)\} \\ \mathrm{Im}\{\boldsymbol{a}(\theta)\} & \mathrm{Re}\{\boldsymbol{a}(\theta)\}\end{bmatrix}.$$

These equations, together with the consistency of both algorithms established in Appendixes A and B, satisfy conditions (7) and (8) of Theorem 1.

Since ML criteria are used in this problem, the optimal weighting in (10) is seen to be equal to the FIM, evaluated in this case for the unstructured model. It is shown in Appendix B that the FIM for the unstructured model is block diagonal with respect to the signal $(\tilde{\boldsymbol{\eta}}_s)$ and noise $(\boldsymbol{Q})$ parameters, so the criterion of (9) will be composed of two terms, one involving only $\boldsymbol{\eta}_s$ and the other only the elements of $\boldsymbol{Q}$. These two terms may thus be minimized separately. Since the parameterization for $\boldsymbol{Q}$ is the same under both the structured and unstructured models, the term involving $\boldsymbol{Q}$ can be made to be zero by simply setting $\hat{\boldsymbol{Q}}_e = \hat{\boldsymbol{Q}}_u$. This establishes (47).

$$\mathbf{FIM}(\tilde{\boldsymbol{\eta}}_s) = 2N\begin{bmatrix}\mathrm{Re}\{\boldsymbol{Q}^{-1}\} & -\mathrm{Im}\{\boldsymbol{Q}^{-1}\} & -\beta_N \mathrm{Im}\{\boldsymbol{Q}^{-1}\boldsymbol{\alpha}\} \\ \mathrm{Im}\{\boldsymbol{Q}^{-1}\} & \mathrm{Re}\{\boldsymbol{Q}^{-1}\} & \beta_N \mathrm{Re}\{\boldsymbol{Q}^{-1}\boldsymbol{\alpha}\} \\ \beta_N \mathrm{Im}\{\boldsymbol{\alpha}^*\boldsymbol{Q}^{-1}\} & \beta_N \mathrm{Re}\{\boldsymbol{\alpha}^*\boldsymbol{Q}^{-1}\} & \gamma_N \boldsymbol{\alpha}^*\boldsymbol{Q}^{-1}\boldsymbol{\alpha}\end{bmatrix}. \tag{120}$$

The remaining term involving $\boldsymbol{\eta}_s$ is found by substituting (129) and (130) into (9)

$$\begin{bmatrix}\hat{b}_{er}\\ \hat{b}_{ei}\\ \hat{\theta}_e\\ \hat{\omega}_e\end{bmatrix} = \arg\min_{b,\theta,\omega}\left(\begin{bmatrix}(\hat{\boldsymbol{\alpha}}h_u)_r\\ (\hat{\boldsymbol{\alpha}}h_u)_i\end{bmatrix} - \boldsymbol{A}(\theta)\begin{bmatrix}b_r\\ b_i\end{bmatrix}\right)^T \cdot \boldsymbol{W}\left(\begin{bmatrix}(\hat{\boldsymbol{\alpha}}h_u)_r\\ (\hat{\boldsymbol{\alpha}}h_u)_i\end{bmatrix} - \boldsymbol{A}(\theta)\begin{bmatrix}b_r\\ b_i\end{bmatrix}\right) \quad (131)$$

where the weighting $\boldsymbol{W}$ is given by the FIM in (120) evaluated at the UML parameter estimates and as before, the subscripts r and i denote real and imaginary parts, respectively.

Define

$$\boldsymbol{g} = \begin{bmatrix}-\beta_N(\hat{\boldsymbol{Q}}_u^{-1}\hat{\boldsymbol{\alpha}}h_u)_i\\ \beta_N(\hat{\boldsymbol{Q}}_u^{-1}\hat{\boldsymbol{\alpha}}h_u)_r\end{bmatrix}$$

so that

$$\boldsymbol{W} = \begin{bmatrix}\boldsymbol{W}_Q & \boldsymbol{g}\\ \boldsymbol{g}^T & \gamma_N\hat{\boldsymbol{\alpha}}_u^*\hat{\boldsymbol{Q}}_u^{-1}\hat{\boldsymbol{\alpha}}_u\end{bmatrix}$$

where $\boldsymbol{W}_Q$ is the upper left $2m\times 2m$ block of the FIM evaluated at $\hat{\boldsymbol{Q}}_u$. If we minimize (131) with respect to ω, a simple calculation yields

$$\hat{\omega}_u - \hat{\omega}_e = \frac{\boldsymbol{g}^T\boldsymbol{A}(\theta)\begin{bmatrix}b_r\\ b_i\end{bmatrix}}{\gamma_N\hat{\boldsymbol{\alpha}}h_u^*\hat{\boldsymbol{Q}}_u^{-1}\hat{\boldsymbol{\alpha}}_u} = \frac{\beta_N\left(b\hat{\boldsymbol{\alpha}}_u^*\hat{\boldsymbol{Q}}_u^{-1}\boldsymbol{a}(\theta)\right)_i}{\gamma_N\hat{\boldsymbol{\alpha}}_u^*\hat{\boldsymbol{Q}}_u^{-1}\hat{\boldsymbol{\alpha}}_u} \quad (132)$$

which when substituted into (131) gives the following criterion involving only b and θ:

$$\begin{bmatrix}\hat{b}_{er}\\ \hat{b}_{ei}\\ \hat{\theta}_e\end{bmatrix} = \arg\min_{b,\theta}\left(\begin{bmatrix}(\hat{\boldsymbol{\alpha}}_u)_r\\ (\hat{\boldsymbol{\alpha}}_u)_i\end{bmatrix} - \boldsymbol{A}(\theta)\begin{bmatrix}b_r\\ b_i\end{bmatrix}\right)^T \cdot \boldsymbol{W}_Q\left(\begin{bmatrix}(\hat{\boldsymbol{\alpha}}_u)_r\\ (\hat{\boldsymbol{\alpha}}_u)_i\end{bmatrix} - \boldsymbol{A}(\theta)\begin{bmatrix}b_r\\ b_i\end{bmatrix}\right) \quad (133)$$

$$+ \frac{-[b_r \quad b_i]\boldsymbol{A}^T(\theta)\boldsymbol{g}\boldsymbol{g}^T\boldsymbol{A}(\theta)\begin{bmatrix}b_r\\ b_i\end{bmatrix}}{\gamma_N\hat{\boldsymbol{\alpha}}_u^*\hat{\boldsymbol{Q}}_u^{-1}\hat{\boldsymbol{\alpha}}_u}. \quad (134)$$

Define $f = \boldsymbol{a}^*(\theta)\hat{\boldsymbol{Q}}_u^{-1}\boldsymbol{a}(\theta), h = \gamma_N\hat{\boldsymbol{\alpha}}_u^*\hat{\boldsymbol{Q}}_u^{-1}\hat{\boldsymbol{\alpha}}h_u, p = \hat{\boldsymbol{\alpha}}_u^*\hat{\boldsymbol{Q}}_u^{-1}\boldsymbol{a}(\theta)$ and set the derivative with respect to $[b_r \quad b_i]^T$ equal to zero to obtain

$$\begin{aligned}0 &= \boldsymbol{A}^T(\theta)\left[\boldsymbol{W}_Q - \frac{1}{h}\boldsymbol{g}\boldsymbol{g}^T\right]\boldsymbol{A}(\theta)\begin{bmatrix}\hat{b}_{er}\\ \hat{b}_{ei}\end{bmatrix} - \boldsymbol{A}^T(\theta)\boldsymbol{W}_Q\begin{bmatrix}(\hat{\boldsymbol{\alpha}}_u)_r\\ (\hat{\boldsymbol{\alpha}}h_u)_i\end{bmatrix}\\ &= f\begin{bmatrix}\hat{b}_{er}\\ \hat{b}_{ei}\end{bmatrix} - \frac{\beta_N^2}{h}\begin{bmatrix}p_i^2 & p_ip_r\\ p_ip_r & p_r^2\end{bmatrix}\begin{bmatrix}\hat{b}_{er}\\ \hat{b}_{ei}\end{bmatrix} - \begin{bmatrix}p_r\\ -p_i\end{bmatrix}.\end{aligned}$$

Solving for $[\hat{b}_{er} \quad \hat{b}_{ei}]^T$ yields

$$\begin{aligned}\begin{bmatrix}\hat{b}_{er}\\ \hat{b}_{ei}\end{bmatrix} &= \begin{bmatrix}f-\frac{\beta_N^2}{h}p_i^2 & -\frac{\beta_N^2}{h}p_ip_r\\ -\frac{\beta_N^2}{h}p_ip_r & f-\frac{\beta_N^2}{h}p_r^2\end{bmatrix}^{-1}\begin{bmatrix}p_r\\ -p_i\end{bmatrix}\\ &= \frac{1}{f\left(f-\frac{\beta_N^2}{h}|p|^2\right)}\begin{bmatrix}f-\frac{\beta_N^2}{h}p_r^2 & \frac{\beta_N^2}{h}p_ip_r\\ \frac{\beta_N^2}{h}p_ip_r & f-\frac{\beta_N^2}{h}p_i^2\end{bmatrix}\cdot\begin{bmatrix}p_r\\ -p_i\end{bmatrix}\\ &= \frac{1}{f\left(f-\frac{\beta_N^2}{h}|p|^2\right)}\begin{bmatrix}fp_r-\frac{\beta_N^2}{h}p_r|p|^2\\ -fp_i+\frac{\beta_N^2}{h}p_i|p|^2\end{bmatrix}\\ &= \frac{1}{f}\begin{bmatrix}p_r\\ -p_i\end{bmatrix}\\ &= \frac{1}{\boldsymbol{a}^*(\theta)\hat{\boldsymbol{Q}}_u^{-1}\boldsymbol{a}(\theta)}\begin{bmatrix}\mathrm{Re}\{\hat{\boldsymbol{\alpha}}_u^*\hat{\boldsymbol{Q}}_u^{-1}\boldsymbol{a}(\theta)\}\\ -\mathrm{Im}\{\hat{\boldsymbol{\alpha}}_u^*\hat{\boldsymbol{Q}}_u^{-1}\boldsymbol{a}(\theta)\}\end{bmatrix}\end{aligned}$$

and thus

$$\hat{b}_e = \frac{\boldsymbol{a}^*(\theta)\hat{\boldsymbol{Q}}_u^{-1}\hat{\boldsymbol{\alpha}}_u}{\boldsymbol{a}^*(\theta)\hat{\boldsymbol{Q}}_u^{-1}\boldsymbol{a}(\theta)} \quad (135)$$

which is (53).

Since $(\hat{b}_e\hat{\boldsymbol{\alpha}}_u^*\hat{\boldsymbol{Q}}_u^{-1}\boldsymbol{a}(\theta))_i = 0$, we see from (132) that $\hat{\omega}_e = \hat{\omega}_u$, which verifies (48). The fact that $\hat{\omega}_u - \hat{\omega}_e = 0$ also proves that (131) is equal to (49). Equation (50) follows from (49) if we combine real and imaginary parts, and (54) is obtained by substituting (53) into (50). Thus, all of the statements in Theorem 4 are verified and the proof is complete.

REFERENCES

[1] B. Van Veen and K. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE Acoust., Speech, Signal Processing Mag.*, vol. 5, pp. 4–24, Apr. 1988.

[2] S. Haykin and A. Steinhardt, Eds., *Adaptive Radar Detection and Estimation*. New York: Wiley, 1992.

[3] A. Farina, *Antenna Based Signal Processing Techniques for Radar Systems*. Norwood, MA: Artech House, 1992.

[4] S. Haykin, J. Litva, and T. Shepherd, Eds, *Radar Array Processing*. Berlin: Springer-Verlag, 1993.

[5] J. Ward, "Space-time adaptive processing for airborne radar," Massachusetts Institute of Technology, Lincoln Labs, Cambridge, MA, Tech. Rep. TR-1015, Dec. 1994.

[6] A. Steinhardt, "Adaptive multisensor detection and estimation," in *Adaptive Radar Detection and Estimation*, S. Haykin and A. Steinhardt, Eds. New York: Wiley, 1992, pp. 91–160.

[7] M. Viberg, B Ottersten, and Ö. Erikmats, "A comparison of model-based detection and adaptive sidelobe cancelling for radar array processing," Chalmers University of Technology, Gothenburg, Sweden, Tech. Rep. CTH-TE-13, Apr. 1994.

[8] P. Stoica and T. Söderström, "On reparametrization of loss functions used in estimation and the invariance principle," *Signal Process.*, vol. 17, pp. 383–387, Aug. 1989.

[9] S. Zacks, *Parametric Statistical Inference: Basic Theory and Modern Approaches*. Oxford, England: Pergamon, 1981.

[10] T. Söderström, P. Stoica, and B. Friedlander, "An indirect prediction error method for system identification," *Automatica*, vol. 27, pp. 183–188, Jan. 1991.

[11] A. Graham, *Kronecker Products and Matrix Calculus with Applications*. Chichester, England: Ellis Horwood, 1981.

[12] J. Li and P. Stoica, "Adaptive filtering approach to spectral estimation and SAR imaging," *IEEE Trans. Signal Processing*, vol. 44, pp. 1469–1484, June 1996.
[13] J. Eriksson, "Aspects on parameter estimation and data reduction in Doppler radar," Licentiate thesis, Chalmers University of Technology, Gothenburg, Sweden, 1996.
[14] P. Stoica and K. Sharman, "Maximum likelihood methods for direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1132–1143, July 1990.
[15] A. Swindlehurst and T. Kailath, "A performance analysis of subspace-based methods in the presence of model errors—Part 1: The MUSIC algorithm," *IEEE Trans. Signal Processing*, vol. 40, pp. 1758–1774, July 1992.
[16] M. Viberg and A. Swindlehurst, "Analysis of the combined effects of finite samples and model errors on array processing performance," *IEEE Trans. Signal Processing*, vol. 42, pp. 3073–3083, Nov. 1994.
[17] ——, "A Bayesian approach to auto-calibration for parametric array signal processing," *IEEE Trans. Signal Processing*, vol. 42, pp. 3495–3507, Dec. 1994.
[18] H. Akaike, "Fitting autoregressive models for prediction," *Ann. Inst. Statist Math.*, vol. 21, pp. 243–347, 1969.
[19] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.
[20] T. Söderström and P. Stoica, *System Identification*. London, England: Prentice-Hall, 1989.
[21] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Upper Saddle River, NJ: Prentice-Hall, 1997.
[22] S. Berman, "A law of large numbers for the maximum in a stationary Gaussian sequence," *Ann. Math. Statist.*, vol. 33, pp. 93–97, 1962.
[23] K. L. Chung, *A Course in Probability Theory*. New York: Harcourt Brace World, 1968.
[24] M. Viberg, B. Ottersten, and A. Nehorai, "Performance analysis of direction finding with large arrays and finite data," *IEEE Trans. Signal Processing*, vol. 43, pp. 469–477, Feb. 1995.
[25] D. Slepian, "Estimation of signal parameters in the presence of noise," *IRE Trans. Prof. Group Inform. Theory*, vol. 3, pp. 68–89, 1954.
[26] W. J. Bangs, "Array processing with generalized beamformers," Ph.D. dissertation, Yale University, New Haven, CT, 1971.

A. Lee Swindlehurst (Member, IEEE) received the B.S. *(summa cum laude)* and M.S. degrees in electrical engineering from Brigham Young University, Provo, UT, in 1985 and 1986, respectively, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1991.

From 1983 to 1984, he was with Eyring Research Institute, Provo, as a Scientific Programmer writing software for a Minuteman missile simulation system. During 1984–1986, he was a Research Assistant in the Department of Electrical Engineering, Brigham Young University. During 1985–1988, he was with the Information Systems Laboratory at Stanford University. From 1986 to 1990, he also was with ESL, Inc., Sunnyvale, CA, where he was involved in the design of algorithms and architectures for a variety of radar and sonar signal-processing systems. He joined the Faculty of the Department of Electrical and Computer Engineering, Brigham Young University, in 1990, where he currently is an Associate Professor. During 1996–1997, he held a joint appointment as a Visiting Scholar at both Uppsala University, Uppsala, Sweden, and the Royal Institute of Technology, Stockholm, Sweden. His research interests include sensor array signal processing, detection, and estimation theory, system identification, and control.

Dr. Swindlehurst was an Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING. He received an Office of Naval Research Graduate Fellowship for 1985–1988.

Petre Stoica (Fellow, IEEE) received the M.Sc. and D.Sc. degrees in automatic control from the Bucharest Polytechnic Institute, Romania, in 1972 and 1979, respectively.

Since 1972, he has been with the Department of Automatic Control and Computers of the Bucharest Polytechnic Institute, where he is a Professor of system identification and signal processing. He spent 1992, 1993, and the first half of 1994 with the Systems and Control Group of Uppsala University, Sweden, as a Guest Professor. In the second half of 1994, he held a Chalmers 150th Anniversary Visiting Professorship with the Applied Electronics Department, Chalmers University of Technology, Gothenburg, Sweden. Currently, he is affiliated with the Systems and Control Group of Uppsala University as a Docent. He has published more than 310 papers in journals and conference records on these topics. He also has published six book chapters and seven books. Most recently, he is the coauthor (with R. Moses) of *Introduction to Spectral Analysis* (Englewood Cliffs, NJ: Prentice-Hall, 1997). He is on the editorial boards of four journals in the field. His scientific interests include system identification, time-series analysis and prediction, statistical signal and array processing, spectral analysis, and radar signal processing.

Dr. Stoica is a corresponding member of the Romanian Academy and a fellow of the Royal Statistical Society. In 1993, he received an honorary doctorate degree from Uppsala University. He was the corecipient (with A. Nehorai) of the 1989 IEEE ASSP Senior Award for contributions to array processing. He received the 1996 Technical Achievement Award from the IEEE Signal Processing Society for fundamental contributions to statistical signal processing as well as several other awards and prizes. He is listed in *Who's Who in the World*.