Matched Direction Detectors and Estimators for Array Processing With Subspace Steering Vector Uncertainties

Olivier Besson, *Senior Member, IEEE*, Louis L. Scharf, *Fellow, IEEE*, and François Vincent, *Member, IEEE*

***Abstract*—In this paper, we consider the problem of estimating and detecting a signal whose associated spatial signature is known to lie in a given linear subspace but whose coordinates in this subspace are otherwise unknown, in the presence of subspace interference and broad-band noise. This situation arises when, on one hand, there exist uncertainties about the steering vector but, on the other hand, some knowledge about the steering vector errors is available. First, we derive the maximum-likelihood estimator (MLE) for the problem and compute the corresponding Cramér–Rao bound. Next, the maximum-likelihood estimates are used to derive a generalized likelihood ratio test (GLRT). The GLRT is compared and contrasted with the standard matched subspace detectors. The performances of the estimators and detectors are illustrated by means of numerical simulations.**

***Index Terms*—Cramér–Rao bound, detection, generalized likelihood ratio test, maximum-likelihood estimation, steering vector uncertainties.**

I. Problem Statement

In many applications of radar, sonar, or communications, it is desired to recover a signal of interest in the presence of interferences and noise using an array of L sensors [1], [2]. Briefly stated, the problem amounts to estimating the temporal waveform $s(t)$ in the model

$$\begin{aligned} \boldsymbol{y}(t) &= \boldsymbol{x}(t) + \boldsymbol{i}(t) + \boldsymbol{n}(t) \\ \boldsymbol{x}(t) &= \boldsymbol{a}s(t); \quad \boldsymbol{i}(t) = \boldsymbol{A}\boldsymbol{u}(t) \end{aligned} \tag{1}$$

where $\boldsymbol{a} \in \mathbb{C}^L$ stands for the signature (or steering vector) of interest, while $\boldsymbol{i}(t)$ and $\boldsymbol{n}(t)$ denote the interference and broadband noise, respectively.

In most situations, exact knowledge of $\boldsymbol{a}$ is difficult to obtain, due to uncalibrated arrays, uncertainties about the direction of arrival (DOA) of the source, local scattering, etc. Differences between the actual signature and the presumed one are known to have deleterious effects on signal waveform estimation unless some proper measures are taken. Robust adaptive beamforming [2], [3] turns out to be an effective solution to mitigate the effects of steering vector errors. This approach consists of designing a beamformer whose performance does not deteriorate dramatically when the actual signature departs from the presumed signature. Among the many methods presented in the literature, diagonal loading has proved to be one of the most efficient. Diagonal loading is able to compensate for a wide variety of errors while remaining a simple solution. However, when the errors on the signature grow large, then its performance is no longer optimal (see, e.g., [4]), and its performance departs substantially from that of a (hypothetical) clairvoyant beamformer that would know $\boldsymbol{a}$. This suggests another alternative, namely estimating $\boldsymbol{a}$ and then using this information to estimate the waveform of interest. This kind of approach was used in [5]–[9], among others. Weiss and Friedlander [7], [8] consider the case where the steering vector depends on some unknown deterministic parameters (for instance, the gains and phases of the sensors), while [5] and [9] use Bayesian approaches where the unknown parameters are random with a known probability density function. As a prerequisite to a Bayesian approach, some *a priori* knowledge is necessary. This is often available. For instance, one can know in which range the sensors gains and phases fall.

In this paper, we consider estimating the unknown steering vector of interest, assuming some knowledge about it is available. More precisely, we assume that the steering vector of interest belongs to a known linear subspace but that its coordinates within this subspace are otherwise unknown. The rationale for doing so is now briefly explained through a couple of examples. Let us assume that the actual steering vector can be written as $\boldsymbol{a} = \boldsymbol{a}(\phi)$, while the presumed direction of arrival is ϕ_0, i.e., there exists an uncertainty about the DOA. Then, assuming that the error $\phi - \phi_0$ is "small," one can write the following Taylor series expansion:

$$\boldsymbol{a}(\phi) \simeq \boldsymbol{a}(\phi_0) + \sum_{k=1}^{r} \frac{1}{r!} \left. \frac{\partial^r \boldsymbol{a}(\phi)}{\partial \phi^r} \right|_{\phi_0} (\phi - \phi_0)^r .$$

Therefore, the actual steering vector belongs to the subspace spanned by $[\,\boldsymbol{a}(\phi_0) (\partial \boldsymbol{a}(\phi)/\partial\phi)|_{\phi_0} \cdots . (\partial^r \boldsymbol{a}(\phi)/\partial\phi^r)|_{\phi_0}\,]$. Note that another intuitively appealing solution in the case of DOA uncertainties would be to assume that $\boldsymbol{a}$ belongs to the range of $[\,\boldsymbol{a}(\phi_0) \quad \boldsymbol{a}(\phi_0 - \Delta) \quad \boldsymbol{a}(\phi_0 + \Delta)\,]$, where Δ is related to the expected range of DOAs for the source of interest. Again, this results in subspace modeling of the steering vector. Another

Manuscript received May 14, 2004; revised February 6, 2005. The work of L. L. Scharf is supported by the Office of Naval Research under Contract N00014-01-1-1019. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Simon J. Godsill.

O. Besson and F. Vincent are with the Department of Avionics and Systems, ENSICA, 31056 Toulouse, France (e-mail: besson,vincent@ensica.fr).

L. L. Scharf is with the Departments of Electrical and Computer Engineering, and Statistics, Colorado State University, Fort Collins, CO 805223-1373 USA (e-mail: scharf@engr.colostate.edu).

Digital Object Identifier 10.1109/TSP.2005.859336

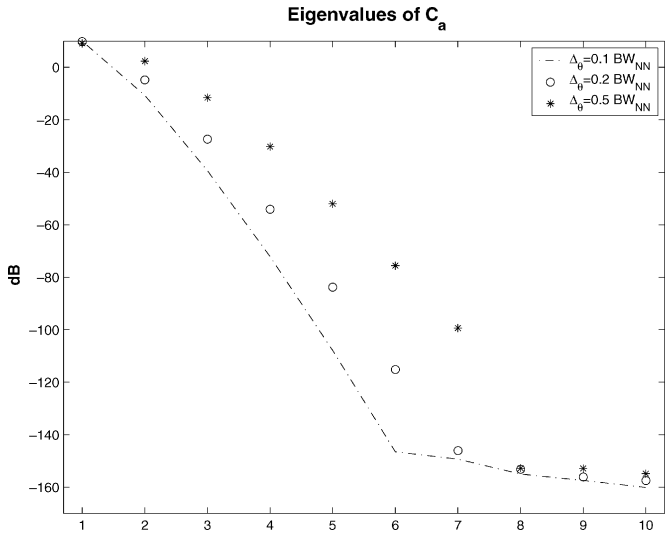

Fig. 1. Eigenvalues of $\boldsymbol{C}_a$ in (3). Scatterers uniformly distributed on $[-\Delta\theta\,;\,\Delta\theta]$. $L = 10$.

pertinent example stems from the case of a Ricean channel for which the steering vector can be written as [10]

$$\boldsymbol{a} = \boldsymbol{a}_0 + \frac{1}{\sqrt{q}} \sum_{k=1}^{q} g_k \boldsymbol{a}(\phi_k) \tag{2}$$

where $\boldsymbol{a}_0$ corresponds to the line-of-sight component. The g_k are zero-mean, independent, and identically distributed random variables with power σ_g^2, and ϕ_k are independent random variables with probability density function $p(\phi)$. The covariance matrix of the steering vector errors is given by [10]

$$\boldsymbol{C}_a = \sigma_g^2 \int \boldsymbol{a}(\phi)\boldsymbol{a}^H(\phi)p(\phi)\,d\phi. \tag{3}$$

When the standard deviation of ϕ_k (referred to as angular spread in the literature) is small, it is well known that the covariance matrix of the matrix $\boldsymbol{C}_a$ is close to rank deficient. As an illustration, Fig. 1 plots the eigenvalues of $\boldsymbol{C}_a$ in case of a uniform linear array with ten elements spaced a half-wavelength apart when ϕ is uniformly distributed on $[-\Delta\phi\,;\,\Delta\phi]$. In this figure, $\mathrm{BW_{NN}}$ refers to the null-to-null beamwidth of the array. As observed, even if $\boldsymbol{C}_a$ is full rank, the eigenvalues drop very quickly to zero. At least, one can define an "effective rank" r that contains most of the energy in the eigenvalues of $\boldsymbol{C}_a$. That is, $\boldsymbol{C}_a \simeq \boldsymbol{U}_r\boldsymbol{\Lambda}_r\boldsymbol{U}_r^H$, where $\boldsymbol{U}_r$ contains the r-dominant eigenvectors and $\boldsymbol{\Lambda}_r$ is the diagonal matrix of the r-dominant eigenvalues. Consequently, the actual steering vector approximately lies in the subspace spanned by $[\,\boldsymbol{a}_0 \quad \boldsymbol{U}_r\,]$. Note that in the Rayleigh case, $\boldsymbol{a}_0 = \boldsymbol{0}$, but the subspace modeling of $\boldsymbol{a}$ remains valid. Therefore, assuming that the steering vector lies in a given linear subspace appears to be a relevant approach to model uncertainties. Such a subspace model for the steering vector was also advocated in [11] and [12], where it is referred to as a generalized array manifold and is used to model spatial signatures in the presence of local scattering. The use of a subspace model when there exist uncertainties about the steering vector is also proposed in [13], where the space–time steering vector is modeled as the rank-one Kronecker product of a spatial and a temporal signature, each subject to uncertainties and assumed to belong to a linear subspace. Therefore, the overall space–time signature lies in a subspace. This modeling is relevant in space–time problems where, for instance, the presence of a target is detected on a grid of potential spatial and Doppler frequencies, whereas the actual spatial and Doppler frequencies lie in between the grid (see, e.g., [14]). As argued in [13], a subspace model for the steering vector is also adequate to account for multipath effects. Finally, it should be mentioned that the detection of a subspace signal from a single snapshot has been considered in the literature (see, e.g., [1], [13]–[20] and reference therein). The general theory of matched subspace detectors is developed in [1] and [15]. A robustification is presented in [14] in the case where there exists an uncertainty about the signal subspace. Methods for choosing an extended signal subspace are proposed and evaluated. In [20], robust detectors are presented to handle the case of a partially known or unknown interference subspace and generalized Gaussian noise. In [19], the assumption of known interference subspace is relaxed, and the authors show how the projection operators needed in the GLRT can be computed from the data. References [13] and [16]–[18] consider the problem of detecting a subspace signal in the presence of interference with arbitrary covariance matrix, i.e., not necessarily low rank. Bose and Steinhardt [13] utilize the theory of maximal invariants to detect uncertain rank-one waveforms, while [16]–[18] assume that secondary data is available to estimate the interference plus noise covariance matrix and the GLRT is derived. A comprehensive overview of adaptive subspace detectors along with a thorough analysis of their properties is presented in [18].

In many practical situations of interest, it seems reasonable to model noise as subspace interference plus broad-band noise [15], [19]. Herein, we consider the case where $\boldsymbol{i}(t)$ belongs to a known linear subspace. This assumption is somehow restrictive as precise knowledge of the interference subspace is, except for a few specific cases, generally not available. Hence, the detectors to be derived subsequently can be viewed as a reference to which detectors that do not need to know the interference subspace should be compared. Deriving such detectors is beyond the scope of the present paper and constitutes a topic of future research. To summarize, this paper addresses the problem of detecting and estimating a signal whose spatial signature is unknown but lies in a given subspace, and in the presence of low-rank interference. The paper is organized as follows. Our modeling assumptions are stated in Section II. In Section III, the maximum-likelihood estimator (MLE) is derived along with the associated Cramér–Rao bound (CRB). The maximum-likelihood estimates are then used in Section IV to derive a GLRT for detecting the presence of the signal of interest. The estimators and detectors we derive might reasonably be called matched *direction* estimators and detectors, as they use multiple snapshots of array data to determine a maximum-likelihood direction in an interference-free subspace. Numerical illustrations are provided in Sections V, and VI draws conclusions and perspectives.

II. Data Model

Let us consider the following model for the L-dimensional received signal

$$\begin{aligned} \boldsymbol{y}(t) &= \boldsymbol{a}s(t) + \boldsymbol{A}\boldsymbol{u}(t) + \boldsymbol{n}(t) \\ \boldsymbol{a} &= \boldsymbol{H}\boldsymbol{\theta} \end{aligned} \tag{4}$$

where:

1) $\boldsymbol{a}$ is the steering vector of interest that belongs to the p-dimensional subspace $\langle \boldsymbol{H} \rangle$;
2) $s(t)$ is the emitted signal waveform;
3) the columns of $\boldsymbol{A}$ span the J-dimensional interference subspace $\langle \boldsymbol{A} \rangle$, and $\boldsymbol{u}(t)$ denotes the interference waveforms;
4) $\boldsymbol{n}(t)$ is a zero-mean complex-valued Gaussian noise with covariance matrix $\sigma^2 \boldsymbol{I}$.

In this paper, we assume that $\boldsymbol{H}$ and $\boldsymbol{A}$ are known full-rank matrices and that the subspaces $\langle \boldsymbol{H} \rangle$ and $\langle \boldsymbol{A} \rangle$ are linearly independent. This implies that no element of $\langle \boldsymbol{H} \rangle$ can be written as a linear combination of vectors in $\langle \boldsymbol{A} \rangle$ and that the composite matrix $[\boldsymbol{H} \quad \boldsymbol{A}]$ is full rank. Furthermore, we consider the case where $J + p < L$. In "space-only" problems, this applies if the number of interferences plus sources is smaller than the number of array elements; note that p is usually very small (see below). In space–time problems, where L is the number of array elements times the number of pulses, this assumption is not restrictive. Various assumptions can be made regarding $s(t)$ and $\boldsymbol{u}(t)$, leading to different statistical models. (See [19] for a discussion.) Herein, it is assumed that $s(t)$ and $\boldsymbol{u}(t)$ are deterministic sequences such that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} \begin{bmatrix} s(t) \\ \boldsymbol{u}(t) \end{bmatrix} [\, s^*(t) \quad \boldsymbol{u}^H(t) \,] = \begin{bmatrix} P & \boldsymbol{0}^H \\ \boldsymbol{0} & \boldsymbol{\Pi} \end{bmatrix} \tag{5}$$

where $\boldsymbol{\Pi}$ is a full-rank matrix. Finally, we assume that the noise level σ^2 is known.

In the sequel, we consider the problem of estimating $\{s(t), \boldsymbol{u}(t)\}_{t=1}^{N}$ and $\boldsymbol{\theta}$ from N snapshots $\{\boldsymbol{y}(t)\}_{t=1}^{N}$ drawn from (4). Since the estimates of $\{s(t), \boldsymbol{u}(t)\}_{t=1}^{N}$ will depend on the maximum-likelihood estimate of $\boldsymbol{\theta}$, we focus on the latter and derive the corresponding CRB. Observe that there exists an inherent scaling ambiguity between $\boldsymbol{\theta}$ and $s(t)$ in (4) as multiplying $\boldsymbol{\theta}$ by any scalar α and $s(t)$ by α^{-1} results in the same model. In order to remedy this problem, a constraint may be enforced on $\boldsymbol{\theta}$. A meaningful constraint, when the first column of $\boldsymbol{H}$ is $\boldsymbol{a}_0$, is to set the first element of $\boldsymbol{\theta}$ to 1. This is the convention we adopt in the sequel, and thus we partition $\boldsymbol{\theta}$ as $\boldsymbol{\theta} = [\, \theta_1 \quad \boldsymbol{\theta}_2^T \,]^T$ and $\boldsymbol{H}$ as $\boldsymbol{H} = [\, \boldsymbol{h}_1 \quad \boldsymbol{H}_2 \,]$.

Prior to deriving the MLE, the following remarks are in order. The present model is related to that in [1], [15], and [19], where the problem of estimating and detecting a signal subspace in the presence of interference subspace is considered. However, the model in (4) considers multiple snapshots and shows a major difference with that of [1], [15], and [19]. The model in [1], [15], and [19] for multiple snapshots would be $\boldsymbol{y}(t) = \boldsymbol{H}\boldsymbol{s}(t) + \boldsymbol{A}\boldsymbol{u}(t) + \boldsymbol{n}(t)$ with $\boldsymbol{s}(t)$ a $p \times 1$ vector that wanders unconstrained in $\mathbb{C}^p$. The model herein writes $\boldsymbol{y}(t) = \boldsymbol{H}\boldsymbol{\theta}s(t) + \boldsymbol{A}\boldsymbol{u}(t) + \boldsymbol{n}(t)$, with $\boldsymbol{s}(t) = \boldsymbol{\theta}s(t)$ forced to wander along the line $\langle \boldsymbol{\theta} \rangle$. Hence, the problems are different, as will be the associated estimators. However, for $N = 1$, the problems become identical, and we will show later that the generalized likelihood ratios derived here reduce to those of [15] and [19] in the single-snapshot case.

III. Estimation

In this section, we consider the problem of estimating the unknown parameter $\boldsymbol{\theta}$ in the model of (4). First, the MLE is derived, and we show that it amounts to searching for the maximum eigenvector of a certain matrix. Then, the CRB for the problem at hand is derived. Finally, an analysis of the output signal-to-interference-and-noise ratio (SINR) obtained with the maximum-likelihood beamformer is carried out.

A. Maximum-Likelihood Estimation

For the sake of clarity in subsequent derivations, let us first define $\boldsymbol{Y} = [\, \boldsymbol{y}(1) \quad \cdots \quad \boldsymbol{y}(N) \,]$, $\boldsymbol{s} = [\, s(1) \quad \cdots \quad s(N) \,]$, and $\boldsymbol{U} = [\, \boldsymbol{u}(1) \quad \cdots \quad \boldsymbol{u}(N) \,]$. Under the hypotheses made, the observations are proper Gaussian distributed and the likelihood function is given by [1] and [21]

$$\ell(\boldsymbol{Y}) = \frac{\exp\left\{ -\frac{1}{\sigma^2} \sum_{t=1}^{N} \| \boldsymbol{y}(t) - \boldsymbol{H}\boldsymbol{\theta}s(t) - \boldsymbol{A}\boldsymbol{u}(t) \|^2 \right\}}{(\pi\sigma^2)^{mN}}. \tag{6}$$

In order to derive the MLE, $\ell(\boldsymbol{Y})$ should be maximized, or equivalently

$$\begin{aligned} \Lambda &= \sum_{t=1}^{N} \| \boldsymbol{y}(t) - \boldsymbol{H}\boldsymbol{\theta}s(t) - \boldsymbol{A}\boldsymbol{u}(t) \|^2 \\ &= \mathrm{Tr}\left\{ \left(\boldsymbol{Y} - \boldsymbol{H}\boldsymbol{\theta}\boldsymbol{s}^T - \boldsymbol{A}\boldsymbol{U} \right) \left(\boldsymbol{Y} - \boldsymbol{H}\boldsymbol{\theta}\boldsymbol{s}^T - \boldsymbol{A}\boldsymbol{U} \right)^H \right\} \end{aligned} \tag{7}$$

should be minimized with respect to $\boldsymbol{s}, \boldsymbol{U}$ and $\boldsymbol{\theta}$. It is important to note that $\boldsymbol{\theta}\boldsymbol{s}^T$ is a rank-one outer product, whereas the model in [15] and [19] would permit it to be rank p. Moreover, with $\boldsymbol{\theta}$ estimated, $\boldsymbol{H}\hat{\boldsymbol{\theta}}$ will be a one-dimensional spatial model. The model in [15] and [19] would allow for space–time variation in the model $\boldsymbol{H}\boldsymbol{s}(t)$, whereas this model allows for time variation only in $s(t)$.

For any given $\boldsymbol{s}$ and $\boldsymbol{\theta}$, the matrix $\boldsymbol{U}$, which minimizes (7), is given by [1]

$$\boldsymbol{U} = \left(\boldsymbol{A}^H \boldsymbol{A} \right)^{-1} \boldsymbol{A}^H \left(\boldsymbol{Y} - \boldsymbol{H}\boldsymbol{\theta}\boldsymbol{s}^T \right). \tag{8}$$

Inserting this value into (7), we are left with the problem of minimizing

$$\begin{aligned} \hat{\Lambda} &= \mathrm{Tr}\left\{ \left(\boldsymbol{Y} - \boldsymbol{H}\boldsymbol{\theta}\boldsymbol{s}^T \right)^H \boldsymbol{P}_{\boldsymbol{A}}^{\perp} \left(\boldsymbol{Y} - \boldsymbol{H}\boldsymbol{\theta}\boldsymbol{s}^T \right) \right\} \\ &= \mathrm{Tr}\left\{ \boldsymbol{Y}\boldsymbol{Y}^H \boldsymbol{P}_{\boldsymbol{A}}^{\perp} - \boldsymbol{Y}^H \boldsymbol{P}_{\boldsymbol{A}}^{\perp} \boldsymbol{H}\boldsymbol{\theta}\boldsymbol{s}^T - \boldsymbol{s}^* \boldsymbol{\theta}^H \boldsymbol{H}^H \boldsymbol{P}_{\boldsymbol{A}}^{\perp} \boldsymbol{Y} \right\} \\ &\quad + \mathrm{Tr}\left\{ \boldsymbol{s}^* \boldsymbol{\theta}^H \boldsymbol{H}^H \boldsymbol{P}_{\boldsymbol{A}}^{\perp} \boldsymbol{H}\boldsymbol{\theta}\boldsymbol{s}^T \right\} \\ &= \left(\boldsymbol{\theta}^H \boldsymbol{H}^H \boldsymbol{P}_{\boldsymbol{A}}^{\perp} \boldsymbol{H}\boldsymbol{\theta} \right) \left\| \boldsymbol{s}^* - \frac{\boldsymbol{Y}^H \boldsymbol{P}_{\boldsymbol{A}}^{\perp} \boldsymbol{H}\boldsymbol{\theta}}{\boldsymbol{\theta}^H \boldsymbol{H}^H \boldsymbol{P}_{\boldsymbol{A}}^{\perp} \boldsymbol{H}\boldsymbol{\theta}} \right\|^2 \\ &\quad + \mathrm{Tr}\left\{ \boldsymbol{P}_{\boldsymbol{A}}^{\perp} \boldsymbol{Y}\boldsymbol{Y}^H \right\} - \frac{\boldsymbol{\theta}^H \boldsymbol{H}^H \boldsymbol{P}_{\boldsymbol{A}}^{\perp} \boldsymbol{Y}\boldsymbol{Y}^H \boldsymbol{P}_{\boldsymbol{A}}^{\perp} \boldsymbol{H}\boldsymbol{\theta}}{\boldsymbol{\theta}^H \boldsymbol{H}^H \boldsymbol{P}_{\boldsymbol{A}}^{\perp} \boldsymbol{H}\boldsymbol{\theta}} \end{aligned} \tag{9}$$

where $\boldsymbol{P}_{\boldsymbol{A}}$ denotes the orthogonal projection onto $\langle \boldsymbol{A} \rangle$ and $\boldsymbol{P}_{\boldsymbol{A}}^{\perp} = \boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A}}$ the projection onto its orthogonal complement. Note that $\boldsymbol{P}_{\boldsymbol{A}}^{\perp} \boldsymbol{H} \neq \boldsymbol{0}$ under the hypotheses made. The maximum-likelihood estimate of $\boldsymbol{\theta}$ is thus given by the principal

generalized eigenvector of $\left(\boldsymbol{G}^H\widehat{\boldsymbol{R}}\boldsymbol{G}, \boldsymbol{G}^H\boldsymbol{G}\right)$ with $\boldsymbol{G} = \boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{H}$ and

$$\widehat{\boldsymbol{R}} = N^{-1}\boldsymbol{Y}\boldsymbol{Y}^H. \tag{10}$$

$\widehat{\boldsymbol{R}}$ is the sample covariance matrix. In other words

$$\widehat{\boldsymbol{\theta}} = \beta\mathcal{P}\left\{\left(\boldsymbol{G}^H\boldsymbol{G}\right)^{-1}\boldsymbol{G}^H\widehat{\boldsymbol{R}}\boldsymbol{G}\right\} \tag{11}$$

where $\mathcal{P}\{.\}$ stands for the principal eigenvector of the matrix between braces, and the scalar β is determined such that $\widehat{\theta}_1 = 1$. Equivalently, $\widehat{\boldsymbol{\theta}}$ can be obtained up to a scaling factor as the generalized singular vector of $\left(\boldsymbol{Y}^H\boldsymbol{G}, \boldsymbol{G}\right)$ associated with the largest generalized singular value [22]. This result shows that, when multiple snapshots are available, a preferred direction in the subspace $\langle\boldsymbol{G}\rangle$ may be determined. When $\boldsymbol{A} = \boldsymbol{0}$, this preferred direction in $\langle\boldsymbol{H}\rangle$ is $\mathcal{P}\left\{\boldsymbol{P}_{\boldsymbol{H}}\widehat{\boldsymbol{R}}\right\}$, where $\boldsymbol{P}_{\boldsymbol{H}}$ denotes the orthogonal projection onto $\langle\boldsymbol{H}\rangle$.

It is straightforward to show that the above maximum-likelihood estimate is consistent. Indeed, under the hypotheses made, the sample covariance matrix converges, as N goes to infinity, to

$$\boldsymbol{R} = P\boldsymbol{a}\boldsymbol{a}^H + \boldsymbol{A}\boldsymbol{\Pi}\boldsymbol{A}^H + \sigma^2\boldsymbol{I}. \tag{12}$$

Since $\boldsymbol{G}^H\boldsymbol{A} = \boldsymbol{0}$, it follows that

$$\begin{aligned}\left(\boldsymbol{G}^H\boldsymbol{G}\right)^{-1}\boldsymbol{G}^H\boldsymbol{R}\boldsymbol{G} &= \left(\boldsymbol{G}^H\boldsymbol{G}\right)^{-1}\left[P\boldsymbol{G}^H\boldsymbol{a}\boldsymbol{a}^H\boldsymbol{G} + \sigma^2\boldsymbol{G}^H\boldsymbol{G}\right]\\ &= \left(\boldsymbol{G}^H\boldsymbol{G}\right)^{-1}\left[P\boldsymbol{G}^H\boldsymbol{H}\boldsymbol{\theta}\boldsymbol{\theta}^H\boldsymbol{H}^H\boldsymbol{G} + \sigma^2\boldsymbol{G}^H\boldsymbol{G}\right]\\ &= \left(\boldsymbol{G}^H\boldsymbol{G}\right)^{-1}\left[P\boldsymbol{G}^H\boldsymbol{G}\boldsymbol{\theta}\boldsymbol{\theta}^H\boldsymbol{G}^H\boldsymbol{G} + \sigma^2\boldsymbol{G}^H\boldsymbol{G}\right]\\ &= P\boldsymbol{\theta}\boldsymbol{\theta}^H\boldsymbol{G}^H\boldsymbol{G} + \sigma^2\boldsymbol{I}.\end{aligned} \tag{13}$$

Obviously, $\boldsymbol{\theta}$ is an eigenvector of $\left(\boldsymbol{G}^H\boldsymbol{G}\right)^{-1}\boldsymbol{G}^H\boldsymbol{R}\boldsymbol{G}$ with eigenvalue $P\boldsymbol{\theta}^H\boldsymbol{G}^H\boldsymbol{G}\boldsymbol{\theta} + \sigma^2$. In addition, since $\left(\boldsymbol{G}^H\boldsymbol{G}\right)^{-1}\boldsymbol{G}^H\boldsymbol{R}\boldsymbol{G}$ is the sum of a rank-one matrix—whose eigenvector is $\boldsymbol{\theta}$—and a scaled identity matrix, $\boldsymbol{\theta}$ is associated with the largest eigenvalue, which proves consistency. We should also mention that when the signal-to-noise ratio grows large, i.e., when σ^2 goes to zero, $\widehat{\boldsymbol{\theta}}$ converges to $\boldsymbol{\theta}$. To see this, note that for $\sigma^2 = 0$, $\boldsymbol{Y}^H\boldsymbol{G} = \boldsymbol{s}^*\boldsymbol{\theta}^H\boldsymbol{G}^H\boldsymbol{G}$ and thus the principal generalized singular vector of $\left(\boldsymbol{Y}^H\boldsymbol{G}, \boldsymbol{G}\right)$ is $\boldsymbol{\theta}$. This proves that $\widehat{\boldsymbol{\theta}}$ converges to $\boldsymbol{\theta}$ as $N \to \infty$ or $\sigma^2 \to 0$.

Finally, note that the ML signal waveform estimate can be written as

$$\widehat{\boldsymbol{s}} = \boldsymbol{Y}^T\widehat{\boldsymbol{w}}^* \tag{14}$$

$$\widehat{\boldsymbol{w}} = \left(\widehat{\boldsymbol{\theta}}^H\boldsymbol{G}^H\boldsymbol{G}\widehat{\boldsymbol{\theta}}\right)^{-1}\boldsymbol{G}\widehat{\boldsymbol{\theta}} \tag{15}$$

and hence the beamformer weight vector belongs to the projection of $\langle\boldsymbol{H}\rangle$ onto the orthogonal complement of $\langle\boldsymbol{A}\rangle$. More precisely, the maximum-likelihood direction $\widehat{\boldsymbol{\theta}}$ places the beamformer $\widehat{\boldsymbol{w}}$ in the best direction in the subspace $\langle\boldsymbol{G}\rangle$. The weight vector hence cancels interference while being matched to the best direction in the interference-free signal subspace.

B. Cramér–Rao Bound

Before deriving the CRB, a few remarks are in order. Since $s(t)$ and $\boldsymbol{u}(t)$ are considered as deterministic sequences, the number of unknowns in the model grows with N, and thus, there does not exist a consistent estimate of the whole parameter vector. The framework considered here is in fact very similar to that of DOA estimation with deterministic signals [23], [24]. Hence, the general theory of maximum-likelihood estimation (which states that the MLE is asymptotically efficient) does not apply and one cannot claim asymptotic efficiency of $\widehat{\boldsymbol{\theta}}$ [23], [24]. However, the CRB is still a lower bound with which the MLE should be compared. In the appendix , we show that the CRB for estimation of $\boldsymbol{\theta}_2$ can be written as

$$\text{CRB}\left(\boldsymbol{\theta}_2\right) = \frac{\sigma^2}{NP_N}\left(\boldsymbol{H}_2^H\boldsymbol{P}_{\boldsymbol{B}}^{\perp}\boldsymbol{H}_2\right)^{-1} \tag{16}$$

with $P_N = N^{-1}\sum_{t=1}^{N}|s(t)|^2$, $\boldsymbol{H} = [\boldsymbol{h}_1 \quad \boldsymbol{H}_2]$, $\boldsymbol{B} = [\boldsymbol{a} \quad \boldsymbol{A}]$, and where $\boldsymbol{P}_{\boldsymbol{B}}^{\perp}$ is the orthogonal projector onto the orthogonal complement of $\boldsymbol{B}$. Note that $\boldsymbol{H}_2$ corresponds to the columns of $\boldsymbol{H}$ for which the coordinates of $\boldsymbol{a}$ in $\langle\boldsymbol{H}\rangle$—namely, $\boldsymbol{\theta}_2$—are unknown. It is interesting to note that (16) is identical in form to the CRB formula in [25]. Observe that the CRB goes to zero as $N \to \infty$ or $\sigma^2 \to 0$, which agrees with the previous discussion about consistency. It is also instructive to observe that the CRB depends on the projection of $\boldsymbol{H}_2$ onto the space orthogonal to the interferences and to the actual steering vector. Therefore, if the first column $\boldsymbol{h}_1$ of $\boldsymbol{H}$ is the nominal steering vector $\boldsymbol{a}_0$ and under the hypothesis of small steering vector errors $\boldsymbol{a} - \boldsymbol{a}_0$, it is preferable to choose the columns of $\boldsymbol{H}_2$ orthogonal to $\boldsymbol{h}_1$ so as to minimize the CRB.

C. SINR Analysis

We now turn to the analysis of the SINR provided by the maximum-likelihood beamformer (15). For any weight vector $\boldsymbol{w}$, we define the corresponding SINR as

$$\text{SINR}(\boldsymbol{w}) = \frac{P\left|\boldsymbol{w}^H\boldsymbol{a}\right|^2}{\boldsymbol{w}^H\boldsymbol{C}\boldsymbol{w}} = \frac{P\left|\boldsymbol{w}^H\boldsymbol{H}\boldsymbol{\theta}\right|^2}{\boldsymbol{w}^H\boldsymbol{C}\boldsymbol{w}} \tag{17}$$

where $\boldsymbol{C} = \boldsymbol{A}\boldsymbol{\Pi}\boldsymbol{A}^H + \sigma^2\boldsymbol{I}$ denotes the interference-plus-noise covariance matrix. It is well known [2] that the weight vector, which maximizes SINR($\boldsymbol{w}$), is given, up to a scaling factor that does not affect the SINR, by

$$\boldsymbol{w}_{\text{opt}} \propto \boldsymbol{C}^{-1}\boldsymbol{a} \tag{18}$$

which provides the optimal output SINR

$$\text{SINR}_{\text{opt}} = P\boldsymbol{a}^H\boldsymbol{C}^{-1}\boldsymbol{a}. \tag{19}$$

Note that the SINR definition in (17) is more appropriate when the signal waveforms are stochastic. Hence, the maximum-likelihood beamformer derived herein under the assumptions of deterministic waveforms may not be optimal for this definition. However, as discussed in [19], when the interference to noise ratio is large, its performance should be relatively close to that

of (18). The SINR provided by the ML beamformer (15) can be written as

$$\begin{aligned}\mathrm{SINR}(\widehat{\boldsymbol{w}}) &= \frac{P\left|\widehat{\boldsymbol{w}}^H\boldsymbol{a}\right|^2}{\widehat{\boldsymbol{w}}^H\left(\boldsymbol{A}\boldsymbol{\Pi}\boldsymbol{A}^H+\sigma^2\boldsymbol{I}\right)\widehat{\boldsymbol{w}}} \\ &= \frac{P}{\sigma^2}\frac{\widehat{\boldsymbol{\theta}}^H\boldsymbol{G}^H\boldsymbol{a}\boldsymbol{a}^H\boldsymbol{G}\widehat{\boldsymbol{\theta}}}{\widehat{\boldsymbol{\theta}}^H\boldsymbol{G}^H\boldsymbol{G}\widehat{\boldsymbol{\theta}}} \end{aligned} \tag{20}$$

where we used the fact that $\boldsymbol{G}^H\boldsymbol{A}=\boldsymbol{0}$. $\mathrm{SINR}(\widehat{\boldsymbol{w}})$ is a random quantity. It appears difficult to derive its probability density function, although $\mathrm{SINR}(\widehat{\boldsymbol{w}})$ is the ratio of two quadratic forms in $\widehat{\boldsymbol{\theta}}$, and the latter is asymptotically Gaussian distributed. However, under mild conditions, $\mathrm{SINR}(\widehat{\boldsymbol{w}})$ converges in mean square to

$$\overline{\mathrm{SINR}} = \frac{P}{\sigma^2}\frac{\left|\boldsymbol{\theta}^H\boldsymbol{G}^H\boldsymbol{H}\boldsymbol{\theta}\right|^2}{\|\boldsymbol{G}\boldsymbol{\theta}\|^2} = \frac{P}{\sigma^2}\boldsymbol{a}^H\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{a}. \tag{21}$$

This asymptotic SINR should be compared with $\mathrm{SINR}_{\mathrm{opt}}$, which can be rewritten as

$$\begin{aligned}\mathrm{SINR}_{\mathrm{opt}} &= P\boldsymbol{a}^H\left(\boldsymbol{A}\boldsymbol{\Pi}\boldsymbol{A}^H+\sigma^2\boldsymbol{I}\right)^{-1}\boldsymbol{a} \\ &= \frac{P}{\sigma^2}\boldsymbol{a}^H\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{a}+P\sum_{j=1}^{J}\frac{\left|\boldsymbol{a}^H\boldsymbol{u}_j\right|^2}{\lambda_j+\sigma^2}\end{aligned} \tag{22}$$

where $\{\boldsymbol{u}_j\}_{j=1}^J$ and $\{\lambda_j\}_{j=1}^J$ are the J eigenvectors and eigenvalues of $\boldsymbol{A}\boldsymbol{\Pi}\boldsymbol{A}^H$. Comparing (21) with (22), it can be inferred, as discussed in [19], that the MLE provides a close-to-optimal SINR. This fact will be illustrated in the numerical examples of Section V.

IV. Detection

In this section, we address the problem of detecting the presence of the signal of interest and thus of deciding between the two hypotheses

$$\begin{cases} H_0: & \boldsymbol{Y}=\boldsymbol{A}\boldsymbol{U}+\boldsymbol{N} \\ H_1: & \boldsymbol{Y}=\boldsymbol{H}\boldsymbol{\theta}\boldsymbol{s}^T+\boldsymbol{A}\boldsymbol{U}+\boldsymbol{N}. \end{cases} \tag{23}$$

Toward this end, the maximum-likelihood estimates derived in the previous section are used in a GLRT. As discussed in [15], it is convenient to replace the GLR by the logarithmic GLR

$$L_1(\boldsymbol{Y}) = \ln\frac{\widehat{\ell}(\boldsymbol{Y}\mid H_1)}{\widehat{\ell}(\boldsymbol{Y}\mid H_0)} = \frac{1}{\sigma^2}\left[\left\|\widehat{\boldsymbol{N}}_0\right\|^2-\left\|\widehat{\boldsymbol{N}}_1\right\|^2\right] \tag{24}$$

where $\widehat{\ell}(\boldsymbol{Y}|H_k)$ is the likelihood function under hypothesis k with the unknown parameters replaced by their maximum-likelihood estimates. $\widehat{\boldsymbol{N}}_0$ and $\widehat{\boldsymbol{N}}_1$ are the maximum-likelihood estimates of $\boldsymbol{N}$ under H_0 and H_1, respectively, and are given by

$$\widehat{\boldsymbol{N}}_0 = \boldsymbol{Y}-\boldsymbol{A}\widehat{\boldsymbol{U}} \tag{25}$$

$$\widehat{\boldsymbol{N}}_1 = \boldsymbol{Y}-\boldsymbol{H}\widehat{\boldsymbol{\theta}}\widehat{\boldsymbol{s}}^T-\boldsymbol{A}\widehat{\boldsymbol{U}}. \tag{26}$$

Using (9), under the null hypothesis, $\left\|\widehat{\boldsymbol{N}}_0\right\|^2$ is given by

$$\left\|\widehat{\boldsymbol{N}}_0\right\|^2 = \mathrm{Tr}\left\{\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{Y}\boldsymbol{Y}^H\right\} \tag{27}$$

whereas under H_1

$$\left\|\widehat{\boldsymbol{N}}_1\right\|^2 = \mathrm{Tr}\left\{\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{Y}\boldsymbol{Y}^H\right\}-\lambda_{\max}\left\{\left(\boldsymbol{G}^H\boldsymbol{G}\right)^{-1}\boldsymbol{G}^H\boldsymbol{Y}\boldsymbol{Y}^H\boldsymbol{G}\right\}. \tag{28}$$

Here, $\lambda_{\max}\{.\}$ is the largest eigenvalue of the matrix between braces. However, observe that

$$\begin{aligned}\lambda_{\max}\left\{\left(\boldsymbol{G}^H\boldsymbol{G}\right)^{-1}\boldsymbol{G}^H\boldsymbol{Y}\boldsymbol{Y}^H\boldsymbol{G}\right\} &= \lambda_{\max}\left\{\boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{Y}\boldsymbol{Y}^H\right\} \\ &= \lambda_{\max}\left\{\boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{Y}\boldsymbol{Y}^H\boldsymbol{P}_{\boldsymbol{G}}\right\} \\ &= \sigma^2_{\max}\left\{\boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{Y}\right\}\end{aligned} \tag{29}$$

where $\boldsymbol{P}_{\boldsymbol{G}}$ is the orthogonal projection onto $\langle\boldsymbol{G}\rangle$ and $\sigma_{\max}\{.\}$ stands for the largest singular value of the matrix between braces. Therefore, the GLR takes the form

$$\begin{aligned}L_1(\boldsymbol{Y}) &= \frac{1}{\sigma^2}\lambda_{\max}\left\{\boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{Y}\boldsymbol{Y}^H\right\} = \frac{1}{\sigma^2}\lambda_{\max}\left\{\boldsymbol{Y}^H\boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{Y}\right\} \\ &= \frac{1}{\sigma^2}\sigma^2_{\max}\left\{\boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{Y}\right\}.\end{aligned} \tag{30}$$

Briefly stated, the detector consists of searching for the *direction of maximum energy* in the subspace $\langle\boldsymbol{G}\rangle$, i.e., in the part of $\langle\boldsymbol{H}\rangle$, which is orthogonal to the interference subspace. Equivalently, it computes the covariance of $\boldsymbol{Y}$ in the subspace $\langle\boldsymbol{G}\rangle$ and tests the energy along its principal direction. Hence, the detector here can be called a *matched-direction detector*. This contrasts with the corresponding detector in [15], which would use $\mathrm{Tr}\left\{\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{Y}\boldsymbol{Y}^H\right\}$, the total energy in $\langle\boldsymbol{G}\rangle$, as the signal would be allowed to move around in the subspace $\langle\boldsymbol{H}\rangle$ from snapshot to snapshot. In the present paper, the signal is fixed in the subspace $\langle\boldsymbol{H}\rangle$, and multiple snapshots may be used to estimate its fixed location $\boldsymbol{H}\widehat{\boldsymbol{\theta}}$. Therefore, the detector searches for a single vector in $\langle\boldsymbol{G}\rangle$, namely the one that bears most energy, and then compares this energy with a threshold. Inspection of (30) also reveals that the GLR $L_1(\boldsymbol{Y})$ is invariant to transformations that rotate $\boldsymbol{Y}$ within $\langle\boldsymbol{G}\rangle$ and add a component in $\langle\boldsymbol{G}\rangle^{\perp}$.

Remark 1: In the single-snapshot case, $\boldsymbol{Y}$ boils down to a vector $\boldsymbol{y}$ and the GLR becomes

$$L_1(\boldsymbol{y}) = \frac{1}{\sigma^2}\boldsymbol{y}^H\boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{y} \tag{31}$$

where $\boldsymbol{P}_{\boldsymbol{G}}$ is the orthogonal projection onto $\langle\boldsymbol{G}\rangle$. Equation (31) coincides with [15, eq. (7.3)], i.e., the generalized likelihood ratio for detecting a subspace signal in subspace interference and noise of known level when a single snapshot is available. This agrees with the fact that the models here and in [15] are identical for $N=1$.

The detection test is thus

$$\phi(\boldsymbol{Y}) = \begin{cases} 1\sim H_1, & L_1(\boldsymbol{Y})>\eta \\ 0\sim H_0, & L_1(\boldsymbol{Y})\le\eta. \end{cases} \tag{32}$$

In order to set the threshold η of the test for a given probability of false alarm P_{FA} and to obtain the probability of detection, it is required to derive the probability density function (pdf) of $\sigma^{-2}\lambda_{\max}\left\{\boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{Y}\boldsymbol{Y}^H\right\}$ or at least its cumulative distribution function (CDF). For notational convenience, let us define $s=\min(p,N)$, $t=\max(p,N)$ and let us denote by ϕ_s the largest eigenvalue of $\sigma^{-2}\boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{Y}\boldsymbol{Y}^H$.

Under the null hypothesis

$$\begin{aligned} L_1(\boldsymbol{Y}|H_0) &= \sigma^{-2}\lambda_{\max}\left\{\boldsymbol{P_G}\boldsymbol{Y}\boldsymbol{Y}^H\right\} \\ &= \sigma^{-2}\lambda_{\max}\left\{\boldsymbol{P_G}\left[\boldsymbol{A}\boldsymbol{U}+\boldsymbol{N}\right]\left[\boldsymbol{A}\boldsymbol{U}+\boldsymbol{N}\right]^H\right\} \\ &= \lambda_{\max}\left\{\sigma^{-2}\boldsymbol{P_G}\boldsymbol{N}\boldsymbol{N}^H\right\} \end{aligned} \tag{33}$$

since $\boldsymbol{P_G}\boldsymbol{A} = \boldsymbol{0}$. Hence, we are left with the problem of deriving the pdf—or the CDF—of the largest eigenvalue of $\sigma^{-2}\boldsymbol{P_G}\boldsymbol{N}\boldsymbol{N}^H$, where $\boldsymbol{N}$ is an $L \times N$ matrix whose columns are independent L-variate complex Gaussian vectors with covariance matrix $\sigma^2\boldsymbol{I}$. Results for the pdf of the largest eigenvalue of $\sigma^{-2}\boldsymbol{N}\boldsymbol{N}^H$ are available (see, e.g., [26] and references therein). We now make use of them to derive the pdf of $\lambda_{\max}\left\{\sigma^{-2}\boldsymbol{P_G}\boldsymbol{N}\boldsymbol{N}^H\right\}$. Toward this end, let $\boldsymbol{U}_G \in \mathbb{C}^{L\times p}$ be an orthonormal basis for $\langle\boldsymbol{G}\rangle$ so that $\boldsymbol{P_G} = \boldsymbol{U}_G\boldsymbol{U}_G^H$. We can thus rewrite $\lambda_{\max}\left\{\sigma^{-2}\boldsymbol{P_G}\boldsymbol{N}\boldsymbol{N}^H\right\}$ as

$$\begin{aligned} \lambda_{\max}\left\{\sigma^{-2}\boldsymbol{P_G}\boldsymbol{N}\boldsymbol{N}^H\right\} &= \lambda_{\max}\left\{\sigma^{-2}\boldsymbol{U}_G\boldsymbol{U}_G^H\boldsymbol{N}\boldsymbol{N}^H\boldsymbol{U}_G^H\boldsymbol{U}_G\right\} \\ &= \lambda_{\max}\left\{\sigma^{-2}\boldsymbol{U}_G^H\boldsymbol{N}\boldsymbol{N}^H\boldsymbol{U}_G\right\} \\ &= \lambda_{\max}\left\{\sigma^{-2}\widetilde{\boldsymbol{N}}\widetilde{\boldsymbol{N}}^H\right\} \end{aligned} \tag{34}$$

where $\widetilde{\boldsymbol{N}} = \boldsymbol{U}_G^H\boldsymbol{N}$ is a $p \times N$ matrix whose columns are independent p-variate complex Gaussian vectors with covariance matrix $\sigma^2\boldsymbol{I}$. Then, the CDF of ϕ_s is given by [26]

$$\Pr\left(\phi_s \le \eta \mid H_0\right) = \frac{\left|\boldsymbol{\Psi}_c(\eta)\right|}{\prod_{k=1}^{s}\Gamma(t-k+1)\Gamma(s-k+1)} \tag{35}$$

where $\boldsymbol{\Psi}_c(\eta)$ is an $s\times s$ Hankel matrix function of $\eta \ge 0$ whose (k,ℓ) element is given by

$$\left[\boldsymbol{\Psi}_c(\lambda)\right]_{k,\ell} = \gamma(t-s+k+\ell-1,\lambda) \tag{36}$$

and $\gamma(n,x) = \int_0^x t^{n-1}e^{-t}\,dt$ stands for the incomplete gamma function. Note that an expression for the exact pdf of ϕ_s can also be found in [26]. Equation (35) provides the necessary material to compute the threshold η for a given $P_{\mathrm{FA}} = 1 - \Pr\left(\phi_s \le \eta \mid H_0\right)$.

Under H_1, the GLR can be written as

$$\begin{aligned} L_1(\boldsymbol{Y} \mid H_1) &= \lambda_{\max}\left\{\sigma^{-2}\boldsymbol{P_G}\left[\boldsymbol{a}\boldsymbol{s}^T+\boldsymbol{N}\right]\left[\boldsymbol{a}\boldsymbol{s}^T+\boldsymbol{N}\right]^H\right\} \\ &= \lambda_{\max}\left\{\sigma^{-2}\boldsymbol{U}_G^H\left[\boldsymbol{a}\boldsymbol{s}^T+\boldsymbol{N}\right]\left[\boldsymbol{a}\boldsymbol{s}^T+\boldsymbol{N}\right]^H\boldsymbol{U}_G\right\} \\ &= \lambda_{\max}\left\{\sigma^{-2}\widetilde{\boldsymbol{Y}}\widetilde{\boldsymbol{Y}}^H\right\} \end{aligned} \tag{37}$$

with $\widetilde{\boldsymbol{Y}} = \boldsymbol{U}_G^H\left[\boldsymbol{a}\boldsymbol{s}^T+\boldsymbol{N}\right]$. Hence, $\widetilde{\boldsymbol{Y}}$ has a multivariate normal distribution with mean $\boldsymbol{M} = \boldsymbol{U}_G^H\boldsymbol{a}\boldsymbol{s}^T$ and covariance matrix $\sigma^2\boldsymbol{I}$. The results of [26]—especially Corollary 1 and Appendix B—can again be used to obtain the pdf or CDF of $L_1(\boldsymbol{Y} \mid H_1)$. More precisely, let $\lambda_1 = \sigma^{-2}\left\|\boldsymbol{U}_G^H\boldsymbol{a}\right\|^2\|\boldsymbol{s}\|^2$ be the single nonzero eigenvalue of $\sigma^{-2}\boldsymbol{M}^H\boldsymbol{M}$. Then, the CDF of $L_1(\boldsymbol{Y} \mid H_1)$ is given by [26]

$$\Pr\left(\phi_s \le \eta \mid H_1\right) = \frac{e^{-\lambda_1}}{\Gamma(t-s+1)\lambda_1^{s-1}}\frac{\left|\boldsymbol{\Psi}_{\mathrm{i.i.d}}(\eta)\right|}{\prod_{k=1}^{s-1}\Gamma(t-k)\Gamma(s-k)} \tag{38}$$

where $\boldsymbol{\Psi}_{\mathrm{i.i.d}}(\eta)$ is a $s\times s$ matrix whose expression can be found in [26]. Equation (38) enables us to calculate the probability of detection $P_{\mathrm{D}} = 1 - \Pr\left(\phi_s \le \eta \mid H_1\right)$.

V. Numerical Examples

The aim of this section is twofold. First, we evaluate the SINR performance of the maximum-likelihood estimated beamformer. Second, the detection performance is illustrated. Throughout this section, we use a uniform linear array of $L = 10$ sensors spaced a half-wavelength apart. The source of interest impinges from broadside and we consider the case of a Ricean channel. We assume a Gaussian distribution for the scatterers with standard deviation $\sigma_\theta = 15^\circ$. The actual steering vector is generated as

$$\boldsymbol{a} = \boldsymbol{a}_0 + \boldsymbol{U}_r\boldsymbol{\Lambda}_r^{1/2}\boldsymbol{\theta}_2 = \boldsymbol{a}_0 + \boldsymbol{H}_2\boldsymbol{\theta}_2 \tag{39}$$

where $\boldsymbol{U}_r$ is the matrix formed by the r principal eigenvectors of $\boldsymbol{C}_a$, and the latter is given by (3). Unless otherwise stated, $r = 2$, and $\boldsymbol{\theta}_2$ is drawn from a proper complex-valued multivariate normal distribution with zero-mean and unit variance. We define the uncertainty ratio (UR) as

$$\mathrm{UR} = 10\log_{10}\left(\frac{\mathrm{Tr}\left\{\boldsymbol{H}_2^H\boldsymbol{H}_2\right\}}{\boldsymbol{a}_0^H\boldsymbol{a}_0}\right). \tag{40}$$

UR as defined above can be viewed as a multidimensional extension of the inverse of the conventional Rice factor for scalar signals. It measures the ratio of the average power of the non-line-of-sight component to the power of the line-of-sight component. In all simulations, we consider that the noise component consists of a proper complex white noise contribution with power σ^2 and two interferences whose DOAs are -20° and 30° and whose powers are 20 and 30 dB above the white noise level, respectively. The SNR is defined as

$$\mathrm{SNR} = 10\log_{10}\left(\frac{P\left[\boldsymbol{a}_0^H\boldsymbol{a}_0 + \mathrm{Tr}\left\{\boldsymbol{H}_2^H\boldsymbol{H}_2\right\}\right]}{\sigma^2}\right)$$

and corresponds to an *average array SNR*.

A. Performance of the Maximum-Likelihood Estimator

In order to evaluate the performance of the MLE, we first compare its mean-squared error (MSE) with the CRB. A fixed value of $\boldsymbol{\theta}$ is considered and the number of snapshots N is varied so as to test the asymptotic performance of the MLE. Fig. 2 displays both the MSE of $\widehat{\boldsymbol{\theta}}_2$ and the CRB versus N. As was conjectured above, and similarly to what is known for the deterministic ML DOA estimation, the MLE is consistent but not asymptotically efficient. However, its performance is quite close to the CRB. Next, we consider the "average" behavior of the maximum-likelihood beamformer in terms of output SINR. Toward this end, 500 Monte Carlo simulations are run with a different random $\boldsymbol{a}$ drawn from (39). For each value of $\boldsymbol{a}$, N random snapshots are generated, $\widehat{\boldsymbol{\theta}}$ is computed, and the optimal and maximum-likelihood SINRs are evaluated as in (19) and (20). They are then averaged. Hence, we display the average performance when $\boldsymbol{a}$ is drawn randomly. Figs. 3 and 4 study the influence of the number of snapshots and the signal to noise ratio, respectively. From inspection of these figures, it can be verified that the MLE has a performance very close to the optimal SINR when N increases. However, even for small values of N, the SINR obtained with the MLE is close to optimal. For instance, for $N = 10$, the SINR of the MLE is only 0.5 dB from the

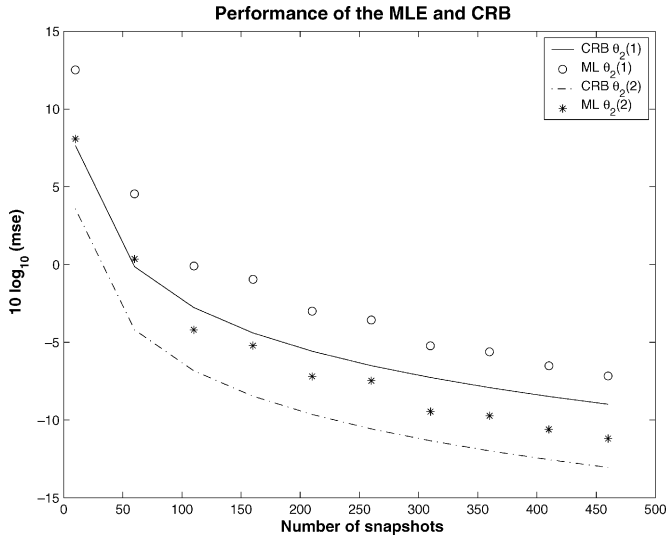

Fig. 2. Mean-square error of the MLE and CRB versus the number of snapshots. UR $= -6$ dB and SNR $= 3$ dB.

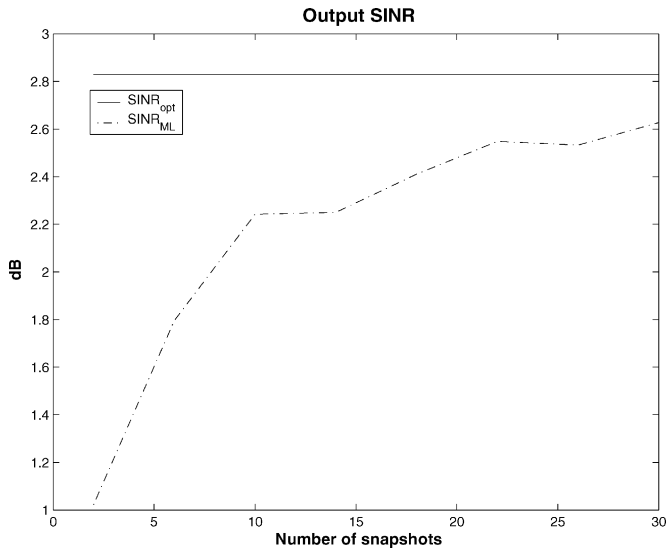

Fig. 3. Output SINR versus the number of snapshots. UR $= -6$ dB and SNR $= 3$ dB.

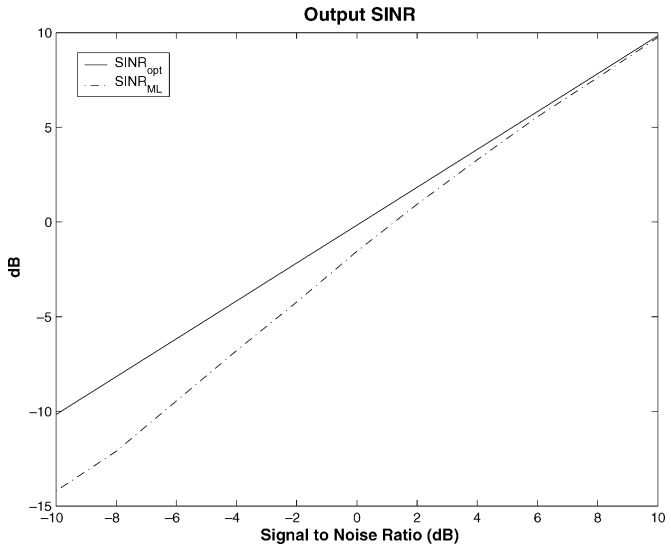

Fig. 4. Output SINR versus SNR. UR $= -6$ dB and $N = 10$.

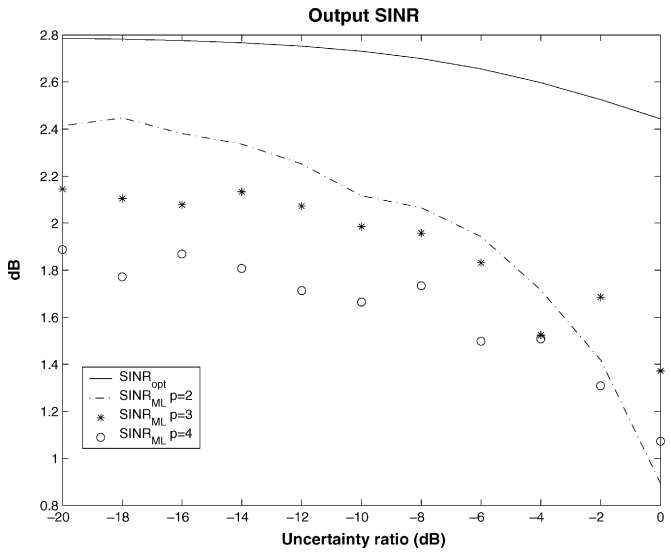

Fig. 5. Output SINR versus UR for various steering vector subspace dimensions. $N = 10$ and SNR $= 3$ dB.

optimum. Accordingly, when the noise level decreases, the performance of the MLE becomes closer and closer to the optimal performance.

Finally, we test for the *robustness* of the estimator. $\boldsymbol{a}$ is now generated according to (2) and the MLE is evaluated with different values of p. Observe that the actual steering vector no longer obeys the model (4) since $\boldsymbol{a}$ is not entirely in $\langle \boldsymbol{H} \rangle$. Fig. 5 displays the average SINR versus the UR—which is defined here as in (40) with $\boldsymbol{H}_2^H \boldsymbol{H}_2$ replaced by $\boldsymbol{C}_a$—for $p = 2, 3, 4$. As can be observed, despite the fact that $\boldsymbol{a}$ is not actually in $\langle \boldsymbol{H} \rangle$, the performance remains very good. Moreover, for low UR, selecting only $\boldsymbol{a}_0$ and the first eigenvector of $\boldsymbol{C}_a$ as the basis for the steering vector subspace provides the best performance. When the UR increases, it is preferable to include an additional eigenvector to account for the increasing power of the non line of sight component. However, choosing $p = 4$ is not useful as the performance decreases. Therefore, even if the steering vector is not entirely in a subspace, using a subspace modeling turns out to be an effective approach.

B. Detection Performance

We now consider the detection performance. First, we validate the theoretical expression of P_{D} as given by (38). Toward this end, a fixed $\boldsymbol{a}$ is drawn from (39), and 500 000 simulations are run to evaluate the empirical probability of detection. The latter is compared with the theoretical probability of detection in Fig. 6. As can be observed the empirical and theoretical results are in perfect agreement, which validates (38).

Next, we characterize the average behavior of the GLRT by changing $\boldsymbol{a}$ in each of the 500 000 runs. Fig. 7 displays the average probability of detection versus the SNR for different P_{FA}. Comparing Figs. 6 and 7 it follows that the "average" behavior is similar to that obtained with a single realization of $\boldsymbol{a}$. In Fig. 8, we investigate the influence of the number of snapshots N on the detection performance. The false alarm probability is set to

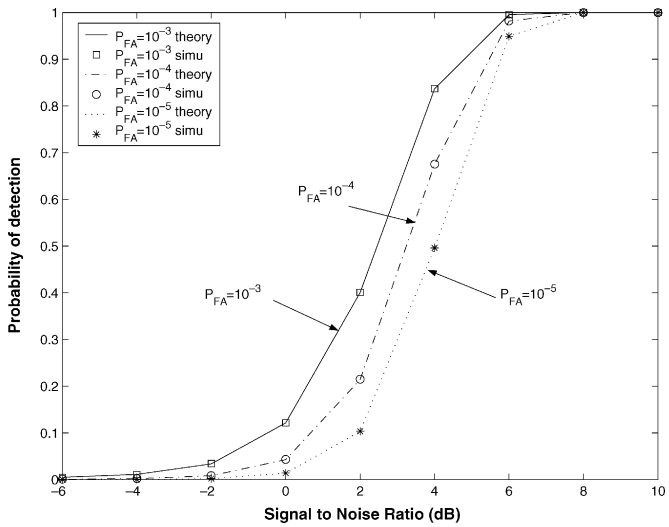

Fig. 6. Theoretical and empirical probability of detection versus SNR. UR $= -6$ dB and $N = 10$.

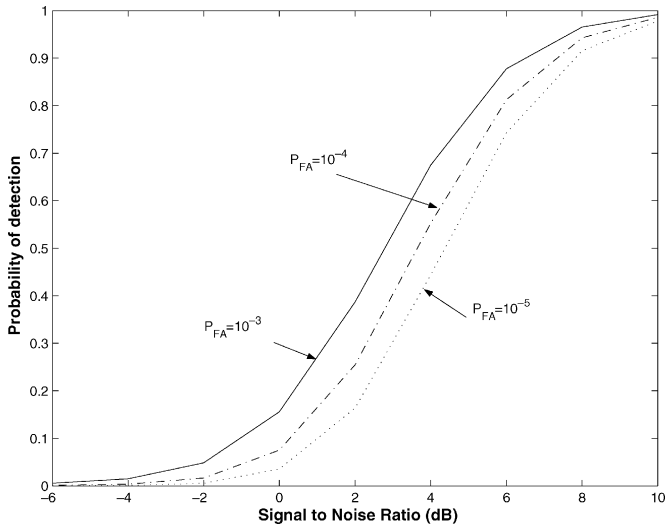

Fig. 7. Average probability of detection versus SNR. UR $= -6$ dB and $N = 10$.

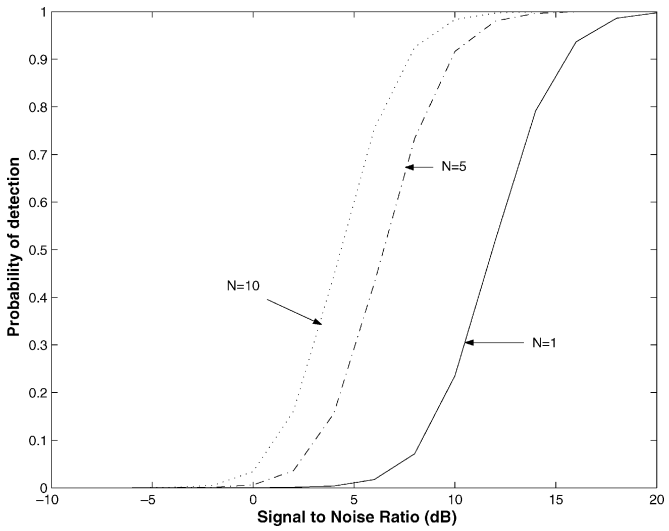

Fig. 8. Average probability of detection versus SNR for various number of snapshots. $P_{\rm FA} = 10^{-5}$ and UR $= -6$ dB.

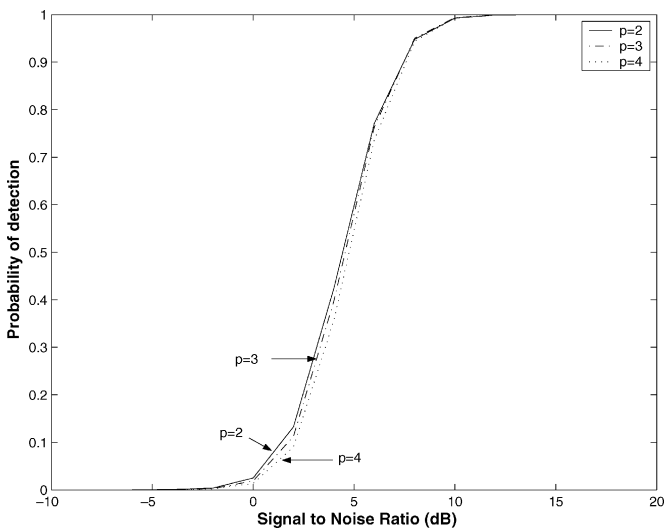

Fig. 9. Average probability of detection versus SNR for various steering vector subspace dimensions. $P_{\rm FA} = 10^{-5}$, UR $= -6$ dB and $N = 10$.

$P_{\rm FA} = 10^{-5}$. It can be observed that N has a significant influence on the detection performance, especially in moderate SNR, where $P_{\rm D}$ can be significantly improved. For instance, for $P_{\rm D} = 0.9$, a 5.4-dB SNR improvement is observed when N goes from $N = 1$ to $N = 5$, and an additional 2.1-dB improvement occurs from $N = 5$ to $N = 10$.

Finally, we test the robustness of the detector when $\boldsymbol{a}$ is no longer exactly in a subspace but is generated according to (2). Fig. 9 displays $P_{\rm D}$ for different values of the subspace dimension used in the detector. Similarly to the estimation part, it can be observed that no real improvement is achieved when increasing p; the best results are even obtained with $p = 2$; note, however, that $\boldsymbol{a}_0$ is always included in $\langle \boldsymbol{H} \rangle$. Of course, the optimal value of p depends on both the uncertainty ratio and the eigenvalue spread of $\boldsymbol{C}_a$. In conclusion, for detection purposes as well, subspace modeling of the spatial signature seems to be a robust and effective solution.

VI. Conclusion

In this paper, we have considered the problem of detecting and estimating a signal whose spatial signature lies in a given linear subspace. The work in this paper can be viewed as an extension of the matched subspace detectors of [15] to the case of multiple snapshots. We showed that the estimator and the detector can be interpreted as matched *direction* estimators–detectors as they are given by the principal eigenvector and eigenvalue of the data matrix projected onto the subspace orthogonal to the interference and matched to the spatial signature subspace. Cramér–Rao bounds were derived along with a theoretical expression for the probability of detection. Numerical examples illustrate the performance and the robustness of the proposed schemes. A future area of research consists of extending the estimators–detectors proposed herein to the case of unknown interference subspace.

APPENDIX
CRAMÉR–RAO BOUNDS

In this appendix, we derive the Cramér–Rao bounds (CRBs) for estimation of the unknown parameters of the model, and we focus more particularly on the CRB for estimation of $\boldsymbol{\theta}$. As advocated in [27]–[29], the CRB will be derived in two steps. First, an expression for the unconstrained CRB —i.e., the CRB where no constraints on $\boldsymbol{\theta}$ are enforced—is obtained. Next, constraints are taken into account and the theory developed in [27] and [28] is used to derive the constrained CRB. For the sake of clarity, let us introduce the following vectors and matrices (along with their dimensions), which will be of use in subsequent derivations:

$$\begin{aligned}
\widetilde{\boldsymbol{s}}(t) &= [\, s(t) \quad \boldsymbol{u}(t)^T \,]^T \qquad J+1 \times 1 \\
\boldsymbol{B} &= [\, \boldsymbol{a} \quad \boldsymbol{A} \,] \qquad L \times J+1 \\
\boldsymbol{\eta}_\theta &= \left[\, \mathrm{Re}\,[\boldsymbol{\theta}]^T \quad \mathrm{Im}\,[\boldsymbol{\theta}]^T \,\right]^T \qquad 2p \times 1 \\
\boldsymbol{\eta}_s &= \left[\mathrm{Re}\left[\widetilde{\boldsymbol{s}}^T(1)\right] \ \mathrm{Im}\left[\widetilde{\boldsymbol{s}}^T(1)\right] \cdots \right. \\
&\qquad \left. \cdots \mathrm{Re}\left[\widetilde{\boldsymbol{s}}^T(N)\right] \ \mathrm{Im}\left[\widetilde{\boldsymbol{s}}^T(N)\right]\right]^T \quad 2N(J+1) \times 1 \\
\boldsymbol{\eta} &= [\, \boldsymbol{\eta}_\theta^T \quad \boldsymbol{\eta}_s^T \,]^T \qquad 2N(J+1)+2p \times 1 \\
\boldsymbol{\mu}(t) &= \boldsymbol{B}\widetilde{\boldsymbol{s}}(t) \qquad L \times 1.
\end{aligned} \tag{41}$$

We also let $\boldsymbol{h}_k$ and $\boldsymbol{b}_k$ denote the kth column of $\boldsymbol{H}$ and $\boldsymbol{B}$, respectively. Accordingly, $\widetilde{s}_k(t)$ stands for the kth element of $\widetilde{\boldsymbol{s}}(t)$. Let us denote by $\boldsymbol{F}_u$ the Fisher Information Matrix (FIM) for estimation of $\boldsymbol{\eta}$, where the subscript $_u$ stands for "unconstrained." Under the stated assumptions, $\boldsymbol{y}(t)$ is Gaussian distributed with mean $\boldsymbol{\mu}(t)$ and covariance matrix $\sigma^2\boldsymbol{I}$. Hence, the (k,ℓ)th element of the FIM is given by [21]

$$\boldsymbol{F}_u(k,\ell) = \frac{2}{\sigma^2}\mathrm{Re}\left[\sum_{t=1}^{N} \frac{\partial \boldsymbol{\mu}^H(t)}{\partial \boldsymbol{\eta}_k}\frac{\partial \boldsymbol{\mu}(t)}{\partial \boldsymbol{\eta}_\ell}\right]. \tag{42}$$

The FIM will have the following partitioned form:

$$\boldsymbol{F}_u = \begin{bmatrix} \boldsymbol{F}_{\theta\theta} & \boldsymbol{F}_{\theta s} \\ \boldsymbol{F}_{s\theta} & \boldsymbol{F}_{ss} \end{bmatrix} \tag{43}$$

where the partitioning corresponds to that of $\boldsymbol{\eta}$. We now derive the FIM on a block-by-block basis. Toward this end, the following relations will be used repeatedly:

$$\begin{aligned}
\frac{\partial \boldsymbol{\mu}(t)}{\partial \mathrm{Re}\,[\theta_k]} &= \boldsymbol{h}_k s(t) \quad \frac{\partial \boldsymbol{\mu}(t)}{\partial \mathrm{Im}\,[\theta_k]} = i\boldsymbol{h}_k s(t) \\
\frac{\partial \boldsymbol{\mu}(t)}{\partial \mathrm{Re}\,[\widetilde{s}_k(t_1)]} &= \boldsymbol{b}_k \delta(t,t_1) \ \frac{\partial \boldsymbol{\mu}(t)}{\partial \mathrm{Im}\,[\widetilde{s}_k(t_1)]} = i\boldsymbol{b}_k \delta(t,t_1).
\end{aligned} \tag{44}$$

Using the previous relations, it is straightforward to establish that

$$\mathrm{Re}\left[\sum_{t=1}^{N} \frac{\partial \boldsymbol{\mu}^H(t)}{\partial \mathrm{Re}\,[\theta_k]}\frac{\partial \boldsymbol{\mu}(t)}{\partial \mathrm{Re}\,[\theta_\ell]}\right] = NP_N \mathrm{Re}\left[\boldsymbol{h}_k^H \boldsymbol{h}_\ell\right] \tag{45a}$$

$$\mathrm{Re}\left[\sum_{t=1}^{N} \frac{\partial \boldsymbol{\mu}^H(t)}{\partial \mathrm{Re}\,[\theta_k]}\frac{\partial \boldsymbol{\mu}(t)}{\partial \mathrm{Im}\,[\theta_\ell]}\right] = -NP_N \mathrm{Im}\left[\boldsymbol{h}_k^H \boldsymbol{h}_\ell\right] \tag{45b}$$

$$\mathrm{Re}\left[\sum_{t=1}^{N} \frac{\partial \boldsymbol{\mu}^H(t)}{\partial \mathrm{Im}\,[\theta_k]}\frac{\partial \boldsymbol{\mu}(t)}{\partial \mathrm{Re}\,[\theta_\ell]}\right] = NP_N \mathrm{Im}\left[\boldsymbol{h}_k^H \boldsymbol{h}_\ell\right] \tag{45c}$$

$$\mathrm{Re}\left[\sum_{t=1}^{N} \frac{\partial \boldsymbol{\mu}^H(t)}{\partial \mathrm{Im}\,[\theta_k]}\frac{\partial \boldsymbol{\mu}(t)}{\partial \mathrm{Im}\,[\theta_\ell]}\right] = NP_N \mathrm{Re}\left[\boldsymbol{h}_k^H \boldsymbol{h}_\ell\right] \tag{45d}$$

where $P_N = N^{-1}\sum_{t=1}^{N}|s(t)|^2$. Gathering the set of previous equations yields the following compact expression:

$$\boldsymbol{F}_{\theta\theta} = \frac{2}{\sigma^2} NP_N \begin{bmatrix} \mathrm{Re}\left[\boldsymbol{H}^H\boldsymbol{H}\right] & -\mathrm{Im}\left[\boldsymbol{H}^H\boldsymbol{H}\right] \\ \mathrm{Im}\left[\boldsymbol{H}^H\boldsymbol{H}\right] & \mathrm{Re}\left[\boldsymbol{H}^H\boldsymbol{H}\right] \end{bmatrix}. \tag{46}$$

Let us now turn to the derivation of $\boldsymbol{F}_{\theta s}$. Observing that

$$\mathrm{Re}\left[\sum_{t=1}^{N} \frac{\partial \boldsymbol{\mu}^H(t)}{\partial \mathrm{Re}\,[\theta_k]}\frac{\partial \boldsymbol{\mu}(t)}{\partial \mathrm{Re}\,[\widetilde{s}_\ell(t_1)]}\right] = \mathrm{Re}\left[s^*(t_1)\boldsymbol{h}_k^H \boldsymbol{b}_\ell\right] \tag{47a}$$

$$\mathrm{Re}\left[\sum_{t=1}^{N} \frac{\partial \boldsymbol{\mu}^H(t)}{\partial \mathrm{Re}\,[\theta_k]}\frac{\partial \boldsymbol{\mu}(t)}{\partial \mathrm{Im}\,[\widetilde{s}_\ell(t_1)]}\right] = -\mathrm{Im}\left[s^*(t_1)\boldsymbol{h}_k^H \boldsymbol{b}_\ell\right] \tag{47b}$$

$$\mathrm{Re}\left[\sum_{t=1}^{N} \frac{\partial \boldsymbol{\mu}^H(t)}{\partial \mathrm{Im}\,[\theta_k]}\frac{\partial \boldsymbol{\mu}(t)}{\partial \mathrm{Re}\,[\widetilde{s}_\ell(t_1)]}\right] = \mathrm{Im}\left[s^*(t_1)\boldsymbol{h}_k^H \boldsymbol{b}_\ell\right] \tag{47c}$$

$$\mathrm{Re}\left[\sum_{t=1}^{N} \frac{\partial \boldsymbol{\mu}^H(t)}{\partial \mathrm{Im}\,[\theta_k]}\frac{\partial \boldsymbol{\mu}(t)}{\partial \mathrm{Im}\,[\widetilde{s}_\ell(t_1)]}\right] = \mathrm{Re}\left[s^*(t_1)\boldsymbol{h}_k^H \boldsymbol{b}_\ell\right] \tag{47d}$$

it follows that

$$\begin{aligned}
\boldsymbol{F}_{\theta s} &= \frac{2}{\sigma^2}\,[\, \boldsymbol{Z}(1) \quad \cdots \quad \boldsymbol{Z}(N) \,] \\
\boldsymbol{Z}(t) &= \begin{bmatrix} \mathrm{Re}\left[s^*(t)\boldsymbol{H}^H\boldsymbol{B}\right] & -\mathrm{Im}\left[s^*(t)\boldsymbol{H}^H\boldsymbol{B}\right] \\ \mathrm{Im}\left[s^*(t)\boldsymbol{H}^H\boldsymbol{B}\right] & \mathrm{Re}\left[s^*(t)\boldsymbol{H}^H\boldsymbol{B}\right] \end{bmatrix}.
\end{aligned} \tag{48}$$

In order to complete the derivation of the FIM, let us write that

$$\mathrm{Re}\left[\sum_{t=1}^{N} \frac{\partial \boldsymbol{\mu}^H(t)}{\partial \mathrm{Re}\,[\widetilde{s}_\ell(t_1)]}\frac{\partial \boldsymbol{\mu}(t)}{\partial \mathrm{Re}\,[\widetilde{s}_\ell(t_2)]}\right] = \mathrm{Re}\left[\boldsymbol{b}_k^H \boldsymbol{b}_\ell\right]\delta(t_1,t_2) \tag{49a}$$

$$\mathrm{Re}\left[\sum_{t=1}^{N} \frac{\partial \boldsymbol{\mu}^H(t)}{\partial \mathrm{Re}[\widetilde{s}_\ell(t_1)]}\frac{\partial \boldsymbol{\mu}(t)}{\partial \mathrm{Im}\,[\widetilde{s}_\ell(t_2)]}\right] = -\mathrm{Im}\left[\boldsymbol{b}_k^H \boldsymbol{b}_\ell\right]\delta(t_1,t_2) \tag{49b}$$

$$\mathrm{Re}\left[\sum_{t=1}^{N} \frac{\partial \boldsymbol{\mu}^H(t)}{\partial \mathrm{Im}\,[\widetilde{s}_\ell(t_1)]}\frac{\partial \boldsymbol{\mu}(t)}{\partial \mathrm{Re}\,[\widetilde{s}_\ell(t_2)]}\right] = \mathrm{Im}\left[\boldsymbol{b}_k^H \boldsymbol{b}_\ell\right]\delta(t_1,t_2) \tag{49c}$$

$$\mathrm{Re}\left[\sum_{t=1}^{N} \frac{\partial \boldsymbol{\mu}^H(t)}{\partial \mathrm{Im}[\widetilde{s}_\ell(t_1)]}\frac{\partial \boldsymbol{\mu}(t)}{\partial \mathrm{Im}[\widetilde{s}_\ell(t_2)]}\right] = \mathrm{Re}\left[\boldsymbol{b}_k^H \boldsymbol{b}_\ell\right]\delta(t_1,t_2) \tag{49d}$$

and therefore

$$\boldsymbol{F}_{ss} = \frac{2}{\sigma^2}\boldsymbol{I}_N \otimes \begin{bmatrix} \mathrm{Re}\left[\boldsymbol{B}^H\boldsymbol{B}\right] & -\mathrm{Im}\left[\boldsymbol{B}^H\boldsymbol{B}\right] \\ \mathrm{Im}\left[\boldsymbol{B}^H\boldsymbol{B}\right] & \mathrm{Re}\left[\boldsymbol{B}^H\boldsymbol{B}\right] \end{bmatrix} \tag{50}$$

where $\otimes$ stands for the Kronecker product. This completes the derivation of the unconstrained FIM. Next, the constraints on $\boldsymbol{\theta}$ are enforced, namely, we impose that $\theta_1 = 1$. The $2 \times 2p$ gradient matrix of the constraints $\boldsymbol{f}(\boldsymbol{\eta}_\theta) = 0$ is given by

$$\boldsymbol{D} = \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{\eta}_\theta^T} = \begin{bmatrix} \boldsymbol{e}_1^T & \boldsymbol{0}^T \\ \boldsymbol{0}^T & \boldsymbol{e}_1^T \end{bmatrix} \tag{51}$$

where $\boldsymbol{e}_1$ is a $p \times 1$ vector whose first element is one, and all others are equal to zero. Let $\boldsymbol{U}_1 = \begin{bmatrix} \boldsymbol{0}^T \\ \boldsymbol{I}_{p-1} \end{bmatrix}$ be a $p \times p-1$ orthonormal matrix whose columns span the null space of $\boldsymbol{e}_1$ and define

$$\boldsymbol{U}_\theta = \begin{bmatrix} \boldsymbol{U}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{U}_1 \end{bmatrix} \in \mathbb{R}^{2p \times 2p-2}. \tag{52}$$

Then, the columns of $\boldsymbol{U}_\theta$ form an orthonormal basis for the null space of $\boldsymbol{D}$, i.e., $\boldsymbol{U}_\theta^T\boldsymbol{U}_\theta = \boldsymbol{I}_{2p-2}$ and $\boldsymbol{D}\boldsymbol{U}_\theta = \boldsymbol{0}$. Similarly, let us define the orthonormal matrix

$$\boldsymbol{U}_\eta = \begin{bmatrix} \boldsymbol{U}_\theta & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{2(J+1)N} \end{bmatrix} \tag{53}$$

$$\begin{aligned}\mathrm{CRB}(\boldsymbol{\eta}) &= \boldsymbol{U}_\eta \left(\boldsymbol{U}_\eta^T \boldsymbol{F}_u \boldsymbol{U}_\eta\right)^{-1} \boldsymbol{U}_\eta^T \\ &= \begin{bmatrix} \boldsymbol{U}_\theta & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{2(J+1)N} \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_\theta^T \boldsymbol{F}_{\theta\theta} \boldsymbol{U}_\theta & \boldsymbol{U}_\theta^T \boldsymbol{F}_{\theta s} \\ \boldsymbol{F}_{\theta s}^T \boldsymbol{U}_\theta & \boldsymbol{F}_{ss} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{U}_\theta^T & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{2(J+1)N} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{U}_\theta \left(\boldsymbol{U}_\theta^T \boldsymbol{\Psi} \boldsymbol{U}_\theta\right)^{-1} \boldsymbol{U}_\theta^T & -\boldsymbol{U}_\theta \left(\boldsymbol{U}_\theta^T \boldsymbol{\Psi} \boldsymbol{U}_\theta\right)^{-1} \boldsymbol{U}_\theta^T \boldsymbol{F}_{\theta s} \boldsymbol{F}_{ss}^{-1} \\ -\boldsymbol{F}_{ss}^{-1} \boldsymbol{F}_{\theta s}^T \boldsymbol{U}_\theta \left(\boldsymbol{U}_\theta^T \boldsymbol{\Psi} \boldsymbol{U}_\theta\right)^{-1} \boldsymbol{U}_\theta^T & \left\{\boldsymbol{F}_{ss} - \boldsymbol{F}_{\theta s}^T \boldsymbol{U}_\theta \left(\boldsymbol{U}_\theta^T \boldsymbol{F}_{\theta\theta} \boldsymbol{U}_\theta\right)^{-1} \boldsymbol{U}_\theta^T \boldsymbol{F}_{\theta s}\right\}^{-1} \end{bmatrix} \end{aligned} \tag{54}$$

which is orthogonal to the gradient of the constraints with respect to the whole parameter vector $\boldsymbol{\eta}$. Then, the CRB for estimation of $\boldsymbol{\eta}$ is given by (54) [28], shown at the top of the page with

$$\begin{aligned}\boldsymbol{\Psi} &= \boldsymbol{F}_{\theta\theta} - \boldsymbol{F}_{\theta s} \boldsymbol{F}_{ss}^{-1} \boldsymbol{F}_{\theta s}^T \\ &= \frac{2}{\sigma^2} N P_N \begin{bmatrix} \mathrm{Re}\left[\boldsymbol{H}^H \boldsymbol{P}_{\boldsymbol{B}}^{\perp} \boldsymbol{H}\right] & -\mathrm{Im}\left[\boldsymbol{H}^H \boldsymbol{P}_{\boldsymbol{B}}^{\perp} \boldsymbol{H}\right] \\ \mathrm{Im}\left[\boldsymbol{H}^H \boldsymbol{P}_{\boldsymbol{B}}^{\perp} \boldsymbol{H}\right] & \mathrm{Re}\left[\boldsymbol{H}^H \boldsymbol{P}_{\boldsymbol{B}}^{\perp} \boldsymbol{H}\right] \end{bmatrix}. \end{aligned} \tag{55}$$

To obtain the last equality, we made use of (46), (48), and (50). Let us now partition $\boldsymbol{H}$ as $\boldsymbol{H} = [\boldsymbol{h}_1 \quad \boldsymbol{H}_2]$; the $p-1$ columns of $\boldsymbol{H}_2$ correspond to the $p-1$ unknown coordinates of $\boldsymbol{a}$ in $\langle \boldsymbol{H} \rangle$ since $\theta_1 = 1$ is known. Then, one can partition $\boldsymbol{H}^H \boldsymbol{P}_{\boldsymbol{B}}^{\perp} \boldsymbol{H}$ as

$$\boldsymbol{H}^H \boldsymbol{P}_{\boldsymbol{B}}^{\perp} \boldsymbol{H} = \begin{bmatrix} - & - \\ - & \boldsymbol{H}_2^H \boldsymbol{P}_{\boldsymbol{B}}^{\perp} \boldsymbol{H}_2 \end{bmatrix} \tag{56}$$

where the "$-$" denote components that will not be of use below. Using the upper left corner of $\mathrm{CRB}(\boldsymbol{\eta})$ in (54), the CRB for estimation of $\boldsymbol{\eta}_\theta$ can be written in closed form as

$$\begin{aligned}\mathrm{CRB}(\boldsymbol{\eta}_\theta) &= \boldsymbol{U}_\theta \left(\boldsymbol{U}_\theta^T \boldsymbol{\Psi} \boldsymbol{U}_\theta\right)^{-1} \boldsymbol{U}_\theta^T = \frac{\sigma^2}{2NP_N} \\ &\times \begin{bmatrix} 0 & \boldsymbol{0}^T & 0 & \boldsymbol{0}^T \\ \boldsymbol{0} & \mathrm{Re}\left[\left(\boldsymbol{H}_2^H \boldsymbol{P}_{\boldsymbol{B}}^{\perp} \boldsymbol{H}_2\right)^{-1}\right] & \boldsymbol{0} & -\mathrm{Im}\left[\left(\boldsymbol{H}_2^H \boldsymbol{P}_{\boldsymbol{B}}^{\perp} \boldsymbol{H}_2\right)^{-1}\right] \\ 0 & \boldsymbol{0}^T & 0 & \boldsymbol{0}^T \\ \boldsymbol{0} & \mathrm{Im}\left[\left(\boldsymbol{H}_2^H \boldsymbol{P}_{\boldsymbol{B}}^{\perp} \boldsymbol{H}_2\right)^{-1}\right] & \boldsymbol{0} & \mathrm{Re}\left[\left(\boldsymbol{H}_2^H \boldsymbol{P}_{\boldsymbol{B}}^{\perp} \boldsymbol{H}_2\right)^{-1}\right] \end{bmatrix}. \end{aligned} \tag{57}$$

In order to obtain the second equality, we made use of the readily verified fact that, for any $\boldsymbol{M} \in \mathbb{R}^{p \times p}$, $\boldsymbol{U}_1^T \boldsymbol{M} \boldsymbol{U}_1$ extract the $p-1$ last lines and $p-1$ last rows, while $\boldsymbol{U}_1 \left(\boldsymbol{U}_1^T \boldsymbol{M} \boldsymbol{U}_1\right) \boldsymbol{U}_1^T$ restores the original size of $\boldsymbol{M}$ by inserting zeros in the first row and first column. Finally, the CRB for estimation of $\boldsymbol{\theta}_2$ is given by

$$\mathrm{CRB}\left(\boldsymbol{\theta}_2\right) = \frac{\sigma^2}{NP_N} \left(\boldsymbol{H}_2^H \boldsymbol{P}_{\boldsymbol{B}}^{\perp} \boldsymbol{H}_2\right)^{-1}. \tag{58}$$

Before closing this appendix, note that the CRB for estimation of $\boldsymbol{\eta}_\theta$ is always given by the first line of (57), provided that $\boldsymbol{U}_\theta$ is orthonormal and spans the null space of $\partial \boldsymbol{f} / \partial \boldsymbol{\eta}_\theta^T$.

ACKNOWLEDGMENT

The authors would like to thank M. Kang for providing them with some of the codes used to compute the cumulative distribution functions (CDFs) in (35) and (38).

REFERENCES

[1] L. Scharf, *Statistical Signal Processing: Detection, Estimation and Time Series Analysis*. Reading, MA: Addison Wesley, 1991.

[2] H. V. Trees, *Optimum Array Processing*. New York: Wiley, 2002.

[3] A. Gershman, "Robustness issues in adaptive beamforming and high-resolution direction finding," in *High Resolution and Robust Signal Processing*, Y. Hua, A. Gershman, and Q. Chen, Eds. New York: Marcel Dekker, 2003, ch. 2, pp. 63–110.

[4] O. Besson and F. Vincent, "Performance analysis of beamformers using generalized loading of the covariance matrix in the presence of random steering vector errors," *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 452–459, Feb. 2005.

[5] M. Viberg and A. Swindlehurst, "A Bayesian approach to auto-calibration for parametric array processing," *IEEE Trans. Signal Process.*, vol. 42, no. 12, pp. 3495–3507, Dec. 1994.

[6] C.-Y. Tseng, D. Feldman, and L. Griffiths, "Steering vector estimation in uncalibrated arrays," *IEEE Trans. Signal Process.*, vol. 43, no. 6, pp. 1397–1412, Jun. 1995.

[7] A. Weiss and B. Friedlander, "'Almost blind' signal estimation using second-order moments," *Proc. Inst. Elect. Eng., Radar, Sonar Navigation*, vol. 142, no. 5, pp. 213–217, Oct. 1995.

[8] ——, "Comparison of signal estimation using calibrated and uncalibrated arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 33, no. 1, pp. 241–249, Jan. 1997.

[9] A. Swindlehurst, "A maximum *a posteriori* approach to beamforming in the presence of calibration errors," in *Proc. IEEE Statistical Signal and Array Processing (SSAP) Conf.*, Corfu, Greece, Jun. 1996, pp. 82–85.

[10] G. Fuks, J. Goldberg, and H. Messer, "Bearing estimation in a Ricean channel-Part I: Inherent accuracy limitations," *IEEE Trans. Signal Process.*, vol. 49, no. 5, pp. 925–937, May 2001.

[11] D. Astély, B. Ottersten, and A. Swindlehurst, "Generalized array manifold model for wireless communication channels with local scattering," *Proc. Inst. Elect. Eng.—F Radar, Sonar Navigation*, vol. 145, no. 1, pp. 51–57, Feb. 1998.

[12] D. Astély and B. Ottersten, "The effects of local scattering on direction of arrival estimation with MUSIC," *IEEE Trans. Signal Process.*, vol. 47, no. 12, pp. 3220–3234, Dec. 1999.

[13] S. Bose and A. Steinhardt, "Adaptive array detection of uncertain rank one waveforms," *IEEE Trans. Signal Process.*, vol. 44, no. 11, pp. 2801–2809, Nov. 1996.

[14] A. Zeira and B. Friedlander, "Robust subspace detectors," in *Proc. 31st Asilomar Conf. Signals Systems Computers*, Pacific Grove, CA, Nov. 2–5, 1997, pp. 778–782.

[15] L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 2146–2157, Aug. 1994.

[16] E. Kelly, "Adaptive Detection in Non-Stationary Interference, Part III," Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA, Tech. Rep. 761, 1987.

[17] R. Raghavan, N. Pulsone, and D. McLaughlin, "Performance of the GLRT for adaptive vector subspace detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, no. 4, pp. 1473–1487, Oct. 1996.

[18] S. Kraut, L. Scharf, and L. McWhorter, "Adaptive subspace detectors," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 1–16, Jan. 2001.

[19] L. Scharf and M. McCloud, "Blind adaptation of zero forcing projections and oblique pseudo-inverses for subspace detection and estimation when interference dominates noise," *IEEE Trans. Signal Process.*, vol. 50, no. 12, pp. 2938–2946, Dec. 2002.

[20] M. Desai and R. Mangoubi, "Robust Gaussian and non-Gaussian matched subspace detection," *IEEE Trans. Signal Process.*, vol. 51, no. 12, pp. 3115–3127, Dec. 2003.

[21] S. Kay, *Fundamentals of Statistical Signal Processing : Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[22] G. Golub and C. V. Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: The John Hopkins Univ. Press, 1996.

[23] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood and Cramér–Rao bound," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 5, pp. 720–741, May 1989.

[24] B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai, "Exact and large sample maximum likelihood techniques for parameter estimation and detection in array processing," in *Radar Array Processing*, S. Haykin, J. Litva, and T. Shepherd, Eds. Berlin, Germany: Springer-Verlag, 1993, ch. 4, pp. 99–151.

[25] L. Scharf and L. McWhorter, "Geometry of the Cramér–Rao bound," *Signal Process.*, vol. 31, no. 3, pp. 301–311, Apr. 1993.

[26] M. Kang and M.-S. Alouini, "Largest eigenvalue of complex Wishart matrices and performance analysis of MIMO MRC systems," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 418–426, Apr. 2003.

[27] T. Marzetta, "A simple derivation of the constrained multiple parameter Cramér–Rao bound," *IEEE Trans. Signal Process.*, vol. 41, no. 6, pp. 2247–2249, Jun. 1993.

[28] P. Stoica and B. Ng, "On the Cramér–Rao bound under parametric constraints," *IEEE Signal Process. Lett.*, vol. 5, no. 7, pp. 177–179, Jul. 1998.

[29] B. Sadler and R. Kozick, "Bounds on bearing and symbol estimation with side information," *IEEE Trans. Signal Process.*, vol. 49, no. 4, pp. 822–834, Apr. 2001.

Olivier Besson (SM'04) received the Ph.D. degree in signal processing and the "Habilitation à Diriger des Recherches" degree from INP, Toulouse, France, in 1992 and 1998, respectively.

He is currently an Associate Professor with the Department of Avionics and Systems of ENSICA, Toulouse, France. His research interests are in the general area of statistical signal and array processing with applications in communications and radar.

Dr. Besson is a member of the IEEE SAM Technical Committee and served as the Co-Technical Chairman of the IEEE SAM 2004 workshop. He was formerly an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and currently serves as an Associate Editor for IEEE SIGNAL PROCESSING LETTERS.

Louis L. Scharf (S'67–M'69–SM'77–F'86) received the Ph.D. degree from the University of Washington, Seattle.

From 1971 to 1982, he served as Professor of electrical engineering and statistics at Colorado State University (CSU), Fort Collins. From 1982 to 1985, he was Professor and Chairman of electrical and computer engineering at the University of Rhode Island, Kingston. From 1985 to 2000, he was Professor of electrical and computer engineering at the University of Colorado, Boulder. In January 2001, he rejoined CSU as Professor of electrical and computer engineering and statistics. He has held several visiting positions in the United States and abroad: Ecole Supérieure d'Electricité, Gif-sur-Yvette, France; EURECOM, Nice, France; the University of La Plata, La Plata, Argentia; Duke University, Durham, NC; the University of Wisconsin, Madison; and the University of Tromso, Tromso, Norway. His interests are in statistical signal processing, as it applies to adaptive radar, sonar, and wireless communication. His most important contributions to date are invariance theories for detection and estimation; matched and adaptive subspace detectors and estimators for radar, sonar, and data communication; and canonical decompositions for reduced dimensional filtering and quantizing. His current interests are in rapidly adaptive receiver design for space–time and frequency–time signal processing in the wireless communication channel.

Prof. Scharf was Technical Program Chair for 1980 ICASSP in Denver, CO, Tutorials Chair for ICASSP 2001 in Salt Lake City, UT, and Technical Program Chair for Asilomar 2002, Pacific Grove, CA. He is past Chair of the Fellow Committee for the IEEE Signal Processing Society and serves on its Technical Committees for Theory and Methods and for Sensor Arrays and Multichannel Signal Processing. He has received numerous awards for his research contributions to statistical signal processing, including an IEEE Distinguished Lectureship, IEEE Third Millennium Medal, and the Technical Achievement Award from the IEEE Signal Processing Society.

François Vincent (M'04) received the Engineer and D.E.A. degrees from ENSEEIHT, Toulouse, France, in 1995 and the Ph.D. degree in radar signal processing from the University of Toulouse, Toulouse, France, in 1999.

From 1999 to 2001, he was a Research Engineer with Siemens, Toulouse. Since 2001, he has been an Assistant Professor with the Department of Avionics and Systems, ENSICA, Toulouse, France. His research interests are in array processing and radar.