Adaptive Subspace Signal Detection with Uncertain Partial Prior Knowledge

Hongbin Li, *Senior Member, IEEE*, Yuan Jiang, Jun Fang, *Member, IEEE*, and Muralidhar Rangaswamy, *Fellow, IEEE*

***Abstract*—This paper is concerned with signal detection in strong disturbance with a subspace structure. Unlike conventional subspace detection techniques relying on the availability of ample training data, we consider a knowledge-aided subspace detection approach for training limited scenarios by incorporating partial prior knowledge of the subspace. A unique advantage of the proposed approach is that it allows the prior knowledge to be incomplete and uncertain, consisting of both correct and incorrect basis vectors. However, the correct and incorrect bases cannot be identified *a priori*. Two hierarchical models are introduced for knowledge representation. One is suitable for the case when the prior knowledge is largely accurate, while the other tries to identify possible errors in the prior knowledge by checking it against and learning from the observed data. The proposed hierarchical models are integrated within a sparse Bayesian framework, which promotes parsimonious subspace representation of the observed data. Variational Bayesian inference algorithms are developed based on the proposed models to recover parameters and subspace structures associated with the disturbance, which are then used in a generalized likelihood ratio test to perform signal detection. Numerical results are presented to illustrate the performance of the proposed subspace detectors in comparison with several notable existing methods.**

***Index Terms*—Subspace signal detection, knowledge-aided processing, Bayesian interference, radar applications.**

I. Introduction

DETECTING a weak signal in strong disturbance (noise, interference, clutter, jamming, etc.) is a fundamental problem in communications, radar/sonar, and many other applications (e.g., [1]–[3]). A popular approach is based on using an estimated covariance matrix of the disturbance obtained from training data for disturbance mitigation. This has led to a family of covariance matrix based (a.k.a. fully adaptive) detectors (see [4] and references therein). One limitation of these detectors is that they require a large amount of training data to ensure the accuracy of the covariance matrix estimate. In particular, the number of training signals should at least double the problem dimension so that the average performance loss relative to the known covariance matrix case in signal-to-interference-and-noise ratio (SINR) is no more than 3 dB [5]. This requirement is prohibitively high for some applications with a large problem dimension, e.g., space-time adaptive processing (STAP) [2] and massive multi-input multi-output (MIMO) systems [6].

Training requirement can be reduced by exploiting disturbance structure. Frequently the disturbance (approximately) has a low-rank subspace structure [7], [8]. This is the case for narrowband interference, which can be expanded using a few Fourier bases, a jamming signal arriving from one or several azimuth angles, or radar clutter whose eigen-spectrum is often dominated by its principal eigenvalues [5]. When the subspace is fully known, signal detection may proceed by projecting the observation into the orthogonal complement of the subspace, followed by cross-correlating with the target signal and energy normalization. This leads to a beta test statistic [9], [10], which is optimum in the sense that it is uniformly most powerful (UMP) invariant [11]. When the subspace is unknown, it can be estimated by using, e.g., the principal eigenvectors of the sample covariance matrix constructed from training data. The resulting subspace detector is called the eigencanceler [12], which belongs to a highly successful class of reduced-rank (a.k.a. partially adaptive) methods, which have been widely used in radar detection, estimation and filtering, and wireless communications [2], [13], [14]. Additional notable reduced-rank detectors include the multi-stage Wiener filter [15] and conjugate-gradient (CG) algorithm based methods [16]–[18].

Exploiting prior knowledge is another way to reduce training requirement. Techniques following this direction are called *knowledge-aided* (KA) processing [19], where the prior knowledge often refers to a prior estimate of the disturbance covariance matrix. Most these techniques often employ a KA estimate of the disturbance covariance matrix via *colored loading*, which invovles linearly combining the prior estimate with the sample covariance matrix. The weighting coefficients can be determined from a deterministic approach [20] or from a stochastic approach (e.g., [21]–[24]). The latter employs a Bayesian framework by treating the covariance matrix as a random matrix assigned with

Manuscript received November 22, 2016; revised April 26, 2017; accepted May 19, 2017. Date of publication June 5, 2017; date of current version June 23, 2017. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Frederic Pascal. The work of H. Li and Y. Jiang was supported by the Air Force Office of Scientific Research under Award FA9550-16-1-0243. The work of J. Fang, which he completed during a visit to Sevens Institute of Technology, was supported by National Science Foundation of China under Grants 61522104 and 61428103. *(Corresponding author: Hongbin Li.)*

H. Li and Y. Jiang are with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030 USA (e-mail: hli@stevens.edu; yjiang25@stevens.edu).

J. Fang was with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030 USA. He is now with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: aquarius.fang@gmail.com).

M. Rangaswamy is with AFRL/RYAP, Wright-Patterson Air Force Base, OH 45433-7132 USA (e-mail: muralidhar.rangaswamy@us.af.mil).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2017.2712125

a conjugate prior, e.g., the inverse Wishart distribution. Then, the posterior estimate takes the aforementioned linear combining form, with the combining coefficients determined by a parameter that represents the *reliability* of the prior knowledge: the larger the parameter, the more reliable the prior estimate, in which case the posterior estimate is combined using a higher weight. Choosing the reliability parameter is a tricky problem [25]. Other than covariance matrix based KA processing, there are a number of recent works addressing exploitation of other types of prior knowledge for detection, e.g., inverse covariance matrix structures [26], spectral structures [27], [28], group symmetric/persymmetric structure [29], [30] and both array and spectral structures of the disturbance [31].

Most of the above KA based techniques can be thought of as extensions of the fully adaptive detection, since they rely on an improved estimate of the full-dimensional covariance matrix. Because of their fully adaptive nature, these KA detectors still require considerable training unless the data dimension is fairly small or the prior knowledge is sufficiently accurate. In many instances, the training size of a KA detector can be reduced approximately by a factor of two compared with its non-KA counterpart (e.g., [23]). While impressive, the reduction may be insufficient for some challenging cases with very limited training. Consider, for example, radar detection with non-homogeneous location-dependent clutter, which occurs when, e.g., radar operates in an urban environment, has a bi-static or multi-static/MIMO configuration, or is equipped with a conformal or forward-looking antenna array [5]. Radar training data is obtained from observations associated with range cells that are spatially close to the test cell. In the aforementioned cases, homogeneous training data becomes scarce since the clutter is location dependent and may vary significantly around the test cell.

In this paper, we consider subspace detection with partial and uncertain prior knowledge of the disturbance subspace, and develop new partially adaptive KA methods for detection with no training data. This effort, to the best of our knowledge, is the first of its kind. Our study is motivated by the fact that in practice, we often have some prior knowledge of the disturbance subspace, either from prior observations or established database of the environment being observed, e.g., spatial locations of dominant clutter scatterers (major natural or man-made structures) in the surveillance area, and the angle-Doppler trace of the clutter spectrum (a.k.a. clutter ridge) observed by an airborne phased-array, which can be determined by motion parameters of the moving sensing platform [5, Section 2.6.2.]. Such information translates to knowledge of some of the subspace basis vectors. However, it is not unusual for such prior knowledge to be contaminated due to measurement/estimation errors, mismatch, or outdated information which must be taken into account. The challenge lies in that it is unknown a priori which part of the prior knowledge is correct and which part is not.

To address the above challenge, we introduce two hierarchical models for knowledge representation. One is suitable for the case when the prior knowledge is largely accurate, and the other tries to identify possible errors in the prior knowledge by checking it against and learning from the observed data. The proposed hierarchical models are integrated within a Bayesian framework for inference, leading to parsimonious subspace representations of the observed data. Our Bayesian framework for KA processing is based on sparse Bayesian learning [32] and is distinctively different from the Bayesian framework used for covariance matrix based KA processing [21]–[24]. We develop variational Bayesian inference algorithms to recover parameters and structures associated with the disturbance, which are then brought into a generalized likelihood ratio test (GLRT) to perform detection.

The rest of this paper is organized as follows. The problem of interest is formulated in Section II. Subspace detection and a GLRT is presented in Section III. The proposed Bayesian models for knowledge representation as well as variational inference algorithms for subspace recovery are discussed in Section IV. Numerical results are presented in Section V, followed by conclusions in Section VI.

Notation: Vectors (matrices) are denoted by boldface lower (upper) case letters. All vectors are column vectors. Superscripts $(\cdot)^*$, $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ denote complex conjugate, transpose and complex conjugate transpose, respectively. $\mathbf{I}$ denotes an identity matrix. $\mathrm{card}(\mathbb{S})$ denotes the cardinality of a set $\mathbb{S}$. $\mathrm{Gamma}(x; a, b)$ denotes the Gamma distribution of random variable x with scale and rate parameters a and b, respectively:

$$\mathrm{Gamma}(x; a, b) = \Gamma^{-1}(a) b^a x^{a-1} \mathrm{e}^{-bx}, \tag{1}$$

where $\Gamma(a) = \int_0^\infty t^{a-1}\mathrm{e}^{-t}\mathrm{d}t$ denotes the Gamma function. Finally, $\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Phi})$ denotes the circularly symmetric complex Gaussian probability density function (PDF) of random vector $\mathbf{x}$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Phi}$:

$$\mathcal{N}_{\mathrm{c}}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Phi}) = |\pi\boldsymbol{\Phi}|^{-1} \exp\Big\{-(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{H}}\boldsymbol{\Phi}^{-1}(\mathbf{x}-\boldsymbol{\mu})\Big\}. \tag{2}$$

II. Problem Formulation

Consider the hypothesis testing problem of detecting a known multichannel signal in disturbance:

$$\begin{aligned} H_0 &: \mathbf{y} = \mathbf{d} \\ H_1 &: \mathbf{y} = \kappa\mathbf{s} + \mathbf{d} \end{aligned} \tag{3}$$

where $\mathbf{y} \in \mathbb{C}^{N\times 1}$ denotes the observation (a.k.a. *test data*), $\mathbf{s}$ the target signal which is assumed known but with an unknown complex-valued amplitude κ, and $\mathbf{d}$ the disturbance signal. The multichannel observation $\mathbf{y}$ may consist of samples taken in space (with multiple antennas), time, or jointly in both domains as in STAP [2]. In phased-array and MIMO systems, $\mathbf{s}$ is often referred to as the steering vector parameterized by the radar look angle and/or target Doppler frequency, while $\mathbf{d}$ may include clutter, jamming and noise. The steering vector $\mathbf{s}$ is known since for typical radar operation, the above hypotheses are tested for specific values of angle and Doppler frequency [33].

In many cases of practical interest, the disturbance $\mathbf{d}$ may have a low-rank subspace representation [11]:

$$\mathbf{d} = \mathbf{H}\boldsymbol{\beta} + \mathbf{n} \tag{4}$$

where $\mathbf{H} \in \mathbb{C}^{N\times L}$ consists of $L < N$ linearly independent basis vectors of the subspace, $\boldsymbol{\beta} \in \mathbb{C}^{L\times 1}$ contains the subspace coefficients, and $\mathbf{n}$ is complex white Gaussian noise vector with

zero mean and covariance $\sigma^2\mathbf{I}$. For the hypothesis testing (3) to be meaningful, it is assumed that $\mathbf{s} \notin \text{span}(\mathbf{H})$. Clearly, $\mathbf{H}\boldsymbol{\beta}$ is the low-rank component in $\mathbf{d}$. As noted before, jamming and clutter in radar often have a low-rank subspace structure. For example, the superposition of clutter echoes can be modeled in terms of the direction containing the maximum energy in the space of observations and often has a subspace structure [5].

While the detection problem (3) has been well studied under the condition $\mathbf{H}$ is exactly known (e.g., [11]), we consider a more interesting and also practically motivated case where only partial and possibly contaminated prior knowledge of $\mathbf{H}$ is available. To model such prior knowledge, we employ an overcomplete dictionary matrix $\boldsymbol{\mathcal{H}} \in \mathbb{C}^{N\times M}$, $M \gg N$, such that

$$\mathbf{H}\boldsymbol{\beta} = \boldsymbol{\mathcal{H}}\mathbf{x} \tag{5}$$

where $\mathbf{x}$ is an $M \times 1$ sparse vector with sparsity L. The dictionary matrix can be formed on a fine grid covering the entire parameter space that parameterizes $\boldsymbol{\mathcal{H}}$, e.g., 1-dimensional (1D) direction-or-arrival (DOA) in beamforming or 2-dimensional (2D) angle-and-Doppler plane in STAP. To focus on the main problem (i.e., subspace detection with inaccurate prior knowledge) without causing excessive ramifications, we assume M is sufficiently large and will not consider the grid-mismatch problem due to finite discretization on the parameter space, which can be addressed by a number of recent techniques (e.g., [34]–[38]).

The prior knowledge can be represented as a group of columns of $\boldsymbol{\mathcal{H}}$ which are believed to span the column space of $\mathbf{H}$; however, the knowledge is inaccurate in that the subset may miss some columns that are necessary to represent $\mathbf{H}$ or contain erroneous columns that do not belong to span($\mathbf{H}$). More precisely, let

$$\mathbb{S} \triangleq \{1, 2, \ldots, M\}, \tag{6}$$

denote the index set that indexes the columns of $\boldsymbol{\mathcal{H}}$. The knowledge can be denoted by a subset $\mathbb{P} \subset \mathbb{S}$. We can write $\mathbb{P} = \mathbb{C} \cup \mathbb{E}$, where $\mathbb{C}$ denotes the subset containing the correct knowledge and $\mathbb{E}$ the erroneous subset. If $\mathbb{T} \subset \mathbb{S}$ denotes the true index set for $\mathbf{H}$, then $\mathbb{C} \subset \mathbb{T}$ and $\mathbb{E} \subset \mathbb{T}^{\text{c}}$. Note that only $\mathbb{P}$ is known; the partition of $\mathbb{C}$ and $\mathbb{E}$ is unknown. In addition, we do not assume knowledge of card($\mathbb{T}$), the cardinality of $\mathbb{T}$ and, equivalently, the rank of $\mathbf{H}$. The problem of interest is to solve the hypothesis testing (3) using observation $\mathbf{y}$, knowledge of the steering vector $\mathbf{s}$, and the partial/uncertain subspace information $\mathbb{P}$.

III. GLRT

We consider a generalized likelihood ratio test (GLRT) approach to solve (3) by incorporating knowledge of the subspace. The likelihood functions under the H_0 and H_1 hypotheses given observation $\mathbf{y}$ are

$$p_0(\boldsymbol{\beta}, \mathbf{H}, \sigma^2; \mathbf{y}) = \mathcal{N}_{\text{c}}(\mathbf{y}; \mathbf{H}\boldsymbol{\beta}, \sigma^2\mathbf{I}), \tag{7}$$

$$p_1(\kappa, \boldsymbol{\beta}, \mathbf{H}, \sigma^2; \mathbf{y}) = \mathcal{N}_{\text{c}}(\mathbf{y}; \kappa\mathbf{s} + \mathbf{H}\boldsymbol{\beta}, \sigma^2\mathbf{I}). \tag{8}$$

The test variable of the GLRT, given by

$$\frac{\max_{\{\kappa,\boldsymbol{\beta},\mathbf{H},\sigma^2\}} p_1(\kappa, \boldsymbol{\beta}, \mathbf{H}, \sigma^2; \mathbf{y})}{\max_{\{\boldsymbol{\beta},\mathbf{H},\sigma^2\}} p_0(\boldsymbol{\beta}, \mathbf{H}, \sigma^2; \mathbf{y})}, \tag{9}$$

requires finding estimates of the unknown parameters under both hypotheses, which are discussed next.

Under H_1, it is easy to see from (8) and (2) that the maximum likelihood estimate (MLE) of the amplitude κ conditioned on $\mathbf{H}$ and $\boldsymbol{\beta}$ is

$$\hat{\kappa} = \frac{\mathbf{s}^{\text{H}}(\mathbf{y} - \mathbf{H}\boldsymbol{\beta})}{\mathbf{s}^{\text{H}}\mathbf{s}} \tag{10}$$

Substituting $\hat{\kappa}$ into (8) and maximizing the resulting likelihood with respect to (w.r.t.) σ^2 gives the MLE of the noise variance as

$$\hat{\sigma}_1^2 = \frac{1}{N}\left\|\mathbf{P}_{\mathbf{s}}^{\perp}\mathbf{y} - \mathbf{P}_{\mathbf{s}}^{\perp}\mathbf{H}\boldsymbol{\beta}\right\|^2, \tag{11}$$

where the subscript 1 indicates the estimate is obtained under the hypothesis H_1 and $\mathbf{P}_{\mathbf{s}}^{\perp} \triangleq \mathbf{I} - \mathbf{s}(\mathbf{s}^{\text{H}}\mathbf{s})^{-1}\mathbf{s}^{\text{H}}$ denotes the projection matrix that projects to the orthogonal complement of $\mathbf{s}$. Substituting (11) and (10) back into (8), we see that the MLEs of the subspace matrix $\mathbf{H}$ and coefficient $\boldsymbol{\beta}$ can be obtained by

$$\{\hat{\mathbf{H}}_1, \hat{\boldsymbol{\beta}}_1\} = \arg\min_{\mathbf{H},\boldsymbol{\beta}} \left\|\mathbf{P}_{\mathbf{s}}^{\perp}\mathbf{y} - \mathbf{P}_{\mathbf{s}}^{\perp}\mathbf{H}\boldsymbol{\beta}\right\|^2. \tag{12}$$

The above least-square (LS) fitting implies the following interpretation for the estimation. Specifically, after concentrating out κ and σ^2 from the likelihood function, the parameter estimation problem under H_1 reduces to an equivalent and simplified one that involves estimating only $\mathbf{H}$ and $\boldsymbol{\beta}$ by using the transformed data $\mathbf{P}_{\mathbf{s}}^{\perp}\mathbf{y}$:

$$\mathbf{P}_{\mathbf{s}}^{\perp}\mathbf{y} = \mathbf{P}_{\mathbf{s}}^{\perp}\mathbf{H}\boldsymbol{\beta} + \mathbf{e} \tag{13}$$

where the $N \times 1$ noise vector $\mathbf{e}$ consists of independent and identically distributed (i.i.d.) zero-mean complex Gaussian entries. Note that in the original estimation problem, the real fitting errors $\mathbf{P}_{\mathbf{s}}^{\perp}\mathbf{y} - \mathbf{P}_{\mathbf{s}}^{\perp}\mathbf{H}\boldsymbol{\beta}$ computed at the true values of $\mathbf{H}$ and $\boldsymbol{\beta}$ are slightly correlated. This interpretation will be employed in Section IV.

The estimation under H_0 proceeds in a similar manner by using (7). Specifically, the MLE of the noise variance conditioned on $\mathbf{H}$ and $\boldsymbol{\beta}$ is

$$\hat{\sigma}_0^2 = \frac{1}{N}\left\|\mathbf{y} - \mathbf{H}\boldsymbol{\beta}\right\|^2, \tag{14}$$

where the subscript 0 signifies the estimate is obtained under hypothesis H_0. In turn, the MLEs of $\mathbf{H}$ and $\boldsymbol{\beta}$ are given by

$$\{\hat{\mathbf{H}}_0, \hat{\boldsymbol{\beta}}_0\} = \arg\min_{\mathbf{H},\boldsymbol{\beta}} \left\|\mathbf{y} - \mathbf{H}\boldsymbol{\beta}\right\|^2. \tag{15}$$

Clearly, (12) and (15) are similar, but neither can be uniquely solved without additional information of the unknowns, which are too many relative to the data size. One approach is to exploit a parametric model for the subspace matrix $\mathbf{H}$, e.g., the DOA $\boldsymbol{\phi} \in \mathbb{R}^{L\times 1}$ of the interference sources in a beamforming setup, in which the problem becomes to jointly estimate $\boldsymbol{\phi}$ and $\boldsymbol{\beta}$. Equivalently, we can use the sparse representation (5) and write the cost function in (12) and (15) in the following unified form:

$$\min_{\mathbf{x}} \left\|\mathbf{z} - \mathbf{A}\mathbf{x}\right\|^2, \tag{16}$$

where $\mathbf{z} \triangleq \mathbf{P}_{\mathbf{s}}^{\perp}\mathbf{y}$ and $\mathbf{A} \triangleq \mathbf{P}_{\mathbf{s}}^{\perp}\boldsymbol{\mathcal{H}}$ under H_1, while under H_0, $\mathbf{z} \triangleq \mathbf{y}$ and $\mathbf{A} \triangleq \boldsymbol{\mathcal{H}}$. The minimization in (16) has to be performed with a sparsity constraint on $\mathbf{x}$. The sparsity recovery

problem can be solved by using a wealth of techniques from greedy methods to ℓ_1-norm based procedures (e.g., [39]). However, these techniques do not allow for a simple integration of prior and potentially contaminated knowledge for subspace recovery. The problem is deferred to Section IV, where we develop new techniques to incorporate uncertain prior knowledge of the subspace structure to estimate $\mathbf{H}$ and $\boldsymbol{\beta}$.

Once $\{\hat{\mathbf{H}}_1, \hat{\boldsymbol{\beta}}_1\}$ and $\{\hat{\mathbf{H}}_0, \hat{\boldsymbol{\beta}}_0\}$ have been obtained, they can be substituted in (10), (11) and (14) to compute the estimates of the signal amplitude κ and noise variance σ^2. Using these parameter estimates in (8) and (7), it is easy to show that the test variable (9) can be simplified to a ratio of the noise variance estimates, and the GLRT is given by

$$T_{\text{GLRT}} \triangleq \frac{\hat{\sigma}_0^2}{\hat{\sigma}_1^2} \underset{H_0}{\overset{H_1}{\gtrless}} \tau \tag{17}$$

where τ denotes a test threshold.

IV. Knowledge-Aided Subspace Estimation

We discuss a Bayesian approach that can incorporate uncertain partial prior knowledge for subspace estimation. We first introduce two Bayesian models for knowledge representation, one corresponding to the case when the partial prior knowledge is believed to contain few errors and the other accounts for possible errors in the prior knowledge. We then develop subspace estimation algorithms for each case.

A. Bayesian Models for Knowledge Representation

Following the discussions in Section III, we have shown the subspace estimation problems (12) and (15) under both hypotheses can be cast in one framework based on the following measurement model [see (16)]:

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{e}, \tag{18}$$

where $\mathbf{z} \in \mathbb{C}^{N\times 1}$ denotes the observation, $\mathbf{A} \in \mathbb{C}^{N\times M}$ a known dictionary matrix, $\mathbf{x} \in \mathbb{C}^{M\times 1}$ an unknown sparse vector with unknown sparsity L, and $\mathbf{e}$ the measurement noise with distribution $\mathcal{N}_{\text{c}}(\mathbf{0}, \gamma^{-1}\mathbf{I})$, where γ denotes the inverse variance, which is also unknown. Sparse Bayesian learning (SBL) [32] is a popular approach that can be used to recover the sparse vector $\mathbf{x}$ from (18). However, SBL does not impose any prior knowledge on the sparsity pattern of $\mathbf{x}$. We need some extensions to incorporate prior knowledge to recover $\mathbf{x}$.

To facilitate discussions, a brief review of SBL (see [32] for more details) is useful. The approach uses a Gaussian inverse Gamma hierarchical model. Specifically, the sparse vector $\mathbf{x}$ is modeled as conditional Gaussian with PDF given by

$$p(\mathbf{x}|\boldsymbol{\alpha}) = \prod_{m=1}^{M} \mathcal{N}_{\text{c}}(x_m; 0, \alpha_m^{-1}), \tag{19}$$

where x_m denotes the m-th element of $\mathbf{x}$ and α_m its inverse variance. Meanwhile, a Gamma prior is employed for the inverse variance vector $\boldsymbol{\alpha} \triangleq [\alpha_1, \ldots, \alpha_M]^{\text{T}}$:

$$p(\boldsymbol{\alpha}) = \prod_{m=1}^{M} \text{Gamma}(\alpha_m; a, b), \tag{20}$$

where suggested choices for hyperparameters a and b are very small values, e.g., 10^{-6}, such that the prior is uniform (over a logarithmic scale) [32]. Such a broad prior over the hyperparameters allows the posterior probability mass to concentrate at very large values of some of α_m, which effectively drives the corresponding x_m (deemed irrelevant to data) to zero, thus leading to a sparse solution.

With prior knowledge on the support of $\mathbf{x}$, it is no longer meaningful to set the prior $p(\alpha_m)$ to be identically non-informative across different m. For subspace coefficients x_m belonging to the knowledge set, i.e., $m \in \mathbb{P}$, where $\mathbb{P}$ is defined in Section II, we should avoid using broad and sparsifying prior $p(\alpha_m)$, which causes the posterior mean to become unbounded. The spread of the Gamma distribution can be reduced by choosing a larger value for the rate parameter b. Based on this observation, we propose to replace (20) with a fixed b by the following prior model to incorporate prior knowledge:

$$p(\boldsymbol{\alpha}) = \prod_{m=1}^{M} \text{Gamma}(\alpha_m; a, b_m), \tag{21}$$

where b_m is allowed to vary with m.

We consider two cases for b_m. The first is a *subspace knowledge* (SK) model which assumes the prior knowledge $\mathbb{P}$ is accurate, while the second is a *subspace knowledge with learning* (SKL) model which takes into account possible errors in $\mathbb{P}$. For the SK model, we choose a relatively large value for b_m, e.g., $b_m \in [0.1, 1]$, if $m \in \mathbb{P}$, so that the prior is non-sparsifying over $\mathbb{P}$, while the other b_m remain small:

$$\text{SK:} \quad b_m = \begin{cases} \bar{b}, & m \in \mathbb{P}, \\ 10^{-6}, & m \in \mathbb{P}^{\text{c}}, \end{cases} \tag{22}$$

where $\bar{b} \in [0.1, 1]$ and $\mathbb{P}^{\text{c}}$ denotes the complement of $\mathbb{P}$. For the SKL model, $\{b_m, m \in \mathbb{P}\}$ are treated as latent variables that are to be learned from the data. Since it is unknown a priori which part of $\mathbb{P}$ is accurate and which is not, we employ a hyperprior for $\{b_m, m \in \mathbb{P}\}$. Assuming errors are not dominant in $\mathbb{P}$, we can use Gamma distributions with moderate shape and rate parameters:

$$\begin{aligned} \text{SKL:} \quad & p(b_1, \ldots, b_M) = \prod_{m=1}^{M} p(b_m), \\ & p(b_m) = \begin{cases} \text{Gamma}(b_m; u, v), & m \in \mathbb{P}, \\ \delta(b_m - 10^{-6}), & m \in \mathbb{P}^{\text{c}}, \end{cases} \end{aligned} \tag{23}$$

where u and v are some fixed values in, e.g., $[0.1, 1]$, and $\delta(\cdot)$ denotes the Dirac delta function, which effectively sets $b_m = 10^{-6}$ for $m \in \mathbb{P}^{\text{c}}$. It has been found that the results are not very sensitive to the values of u, v and $\bar{b}$ in (22). For the simulation results of Section V, they are set identically to 0.3.

Finally, the inverse variance γ of the noise in (18) can be jointly estimated in the Bayesian approach by employing a prior

for γ. We use a non-informative Gamma prior for γ as in [32]:

$$p(\gamma) = \text{Gamma}(\gamma; c, d), \tag{24}$$

where $c = d = 10^{-6}$.

B. Bayesian Inference

Let $\boldsymbol{\theta}$ denote a vector containing all parameters to be estimated. With the probabilistic modeling discussed in Section IV-A, these parameters are treated as *latent variables*. We have two cases for which the composition of $\boldsymbol{\theta}$ is slightly different:

$$\boldsymbol{\theta}_{\text{SK}} \triangleq \{\mathbf{x}, \boldsymbol{\alpha}, \gamma\}, \tag{25}$$

$$\boldsymbol{\theta}_{\text{SKL}} \triangleq \{\mathbf{x}, \boldsymbol{\alpha}, \gamma, \mathbf{b}\}. \quad \mathbf{b} \triangleq \{b_m\}, \forall m \in \mathbb{P}. \tag{26}$$

For notational simplicity, we will suppress the subscript "SK" and "SKL" wherever there is no ambiguity. A standard Bayesian inference procedure would proceed to compute the posterior

$$p(\boldsymbol{\theta}|\mathbf{z}) = \frac{p(\mathbf{z}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{z})}, \tag{27}$$

which is however infeasible for the considered problem since the marginal distribution $p(\mathbf{z}) = \int p(\mathbf{z}|\boldsymbol{\theta})p(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}$ cannot be computed analytically.

To circumvent the difficulty, we employ variational Bayesian inference that utilizes an approximation of the posterior $p(\boldsymbol{\theta}|\mathbf{z})$. Variational Bayesian methods have been used with great success in various applications (see [40]–[42] and references therein). Specifically, we write $\boldsymbol{\theta}$ in a partitioned form:

$$\boldsymbol{\theta} \triangleq \big\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_K\big\}, \tag{28}$$

where $K = 3$ for the SK model and $K = 4$ for the SKL model along with the mapping between parameters: $\boldsymbol{\theta}_1 \triangleq \mathbf{x}$, $\boldsymbol{\theta}_2 \triangleq \boldsymbol{\alpha}$, $\boldsymbol{\theta}_3 \triangleq \gamma$, and $\boldsymbol{\theta}_4 \triangleq \mathbf{b}$. A popular approximation of the posterior $p(\boldsymbol{\theta}|\mathbf{z})$ is based on the mean field approximation (e.g., [42]):

$$p(\boldsymbol{\theta}|\mathbf{z}) \approx \prod_{k=1}^{K} q_k(\boldsymbol{\theta}_k), \tag{29}$$

where the component PDF is given by

$$q_k(\boldsymbol{\theta}_k) = \frac{\exp\Big(\big\langle \ln p(\mathbf{z}, \boldsymbol{\theta})\big\rangle_{l\neq k}\Big)}{\displaystyle\int \exp\Big(\big\langle \ln p(\mathbf{z}, \boldsymbol{\theta})\big\rangle_{l\neq k}\Big)\mathrm{d}\boldsymbol{\theta}_k}, \tag{30}$$

where $p(\mathbf{z}, \boldsymbol{\theta}) = p(\mathbf{z}|\boldsymbol{\theta})p(\boldsymbol{\theta})$ denotes the joint distribution of $\mathbf{z}$ and $\boldsymbol{\theta}$, while $\langle\cdot\rangle_{l\neq k}$ denotes the statistical expectation w.r.t. distributions $q_l(\boldsymbol{\theta}_l)$, $\forall l \neq k$, that is,

$$\big\langle \ln p(\mathbf{z}, \boldsymbol{\theta})\big\rangle_{l\neq k} = \int \ln p(\mathbf{z}, \boldsymbol{\theta}) \prod_{l\neq k} q_l(\boldsymbol{\theta}_l)\mathrm{d}\boldsymbol{\theta}_l. \tag{31}$$

Note that (30) is not an explicit solution since the factor posterior $q_k(\boldsymbol{\theta}_k)$ depends on the other factors $q_l(\boldsymbol{\theta}_l)$, $l \neq k$. However, it points naturally to an iterative procedure for finding the factors. Specifically, we can start by initializing $q_k^{(t)}(\boldsymbol{\theta}_k) = p(\boldsymbol{\theta}_k)$, for $t = 0$, where t is the iteration index; that is, the factor posteriors are initialized by their corresponding prior distributions. Then, for the t-th iteration, we can update $q_k^{(t)}(\boldsymbol{\theta}_k)$ by using $q_l^{(t-1)}(\boldsymbol{\theta}_l)$, $l \neq k$, in the right-hand side of (30). Each iteration cycles through all factors from $k = 1$ to $k = K$, and the iterative process stops till a practical convergence criterion has been met. Next, we consider variational Bayesian inference based subspace estimation based on two knowledge models introduced in Section IV-A.

1) With the SK Model: For the SK problem, we have

$$p(\mathbf{z}, \mathbf{x}, \boldsymbol{\alpha}, \gamma) = p(\mathbf{z}|\mathbf{x}, \gamma)p(\mathbf{x}|\boldsymbol{\alpha})p(\boldsymbol{\alpha})p(\gamma) \tag{32}$$

where $p(\mathbf{z}|\mathbf{x}, \gamma) = \mathcal{N}_{\text{c}}(\mathbf{z}; \mathbf{A}\mathbf{x}, \gamma^{-1}\mathbf{I})$, $p(\mathbf{x}|\boldsymbol{\alpha})$ is given by (19), $p(\boldsymbol{\alpha})$ by (21), and $p(\gamma)$ by (24), respectively. In Appendix A, we show that the factor posteriors $q_x(\mathbf{x})$, $q_\alpha(\boldsymbol{\alpha})$, and $q_\gamma(\gamma)$ are respectively Gaussian, product Gamma, and Gamma distributions:

$$q_x(\mathbf{x}) = \mathcal{N}_{\text{c}}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Phi}), \tag{33}$$

$$q_\alpha(\boldsymbol{\alpha}) = \prod_{m=1}^{M} \text{Gamma}(\alpha_m; a+1, \tilde{b}_m), \tag{34}$$

$$q_\gamma(\gamma) = \text{Gamma}\big(\gamma; c+N, \tilde{d}\big), \tag{35}$$

where $\boldsymbol{\mu}$, $\boldsymbol{\Phi}$, $\tilde{b}_m$ and $\tilde{d}$ are respectively specified in (52), (53), (56), and (60) of Appendix A. As such, the update of the factor posteriors boil down to the update of these parameters. We summarize the resulting estimator in Algorithm 1 (see Appendix A for details).

Algorithm 1: SK-based Subspace Estimator.

Initialize Let $\big\langle \alpha_m^{(0)} \big\rangle = a/b_m$, $m = 1, \ldots, M$, $\big\langle \gamma^{(0)} \big\rangle = c/d$, $\mathbf{D}^{(0)} \triangleq \text{diag}\big\{\big\langle \alpha_1^{(0)} \big\rangle, \ldots, \big\langle \alpha_M^{(0)} \big\rangle\big\}$, and $t = 0$.

repeat

1) Set $t = t + 1$.

2) Update the covariance and mean of $q_x^{(t)}(\mathbf{x})$:

$$\boldsymbol{\Phi}^{(t)} = \Big(\big\langle \gamma^{(t-1)} \big\rangle \mathbf{A}^{\text{H}}\mathbf{A} + \big\langle \mathbf{D}^{(t-1)} \big\rangle\Big)^{-1}, \tag{36}$$

$$\boldsymbol{\mu}^{(t)} = \big\langle \gamma^{(t-1)} \big\rangle \boldsymbol{\Phi}^{(t)}\mathbf{A}^{\text{H}}\mathbf{z}. \tag{37}$$

3) Update the rate and mean of $q_\alpha^{(t)}(\boldsymbol{\alpha})$:

$$\tilde{b}_m^{(t)} = b_m + \big|\mu_m^{(t)}\big|^2 + \Phi_{m,m}^{(t)}, \tag{38}$$

$$\big\langle \alpha_m^{(t)} \big\rangle = (a+1)/\tilde{b}_m^{(t)}, \quad m = 1, \ldots, M, \tag{39}$$

where $\mu_m^{(t)}$ denotes the m-th element of $\boldsymbol{\mu}^{(t)}$ and $\Phi_{m,m}^{(t)}$ the m-th diagonal element of $\boldsymbol{\Phi}^{(t)}$.

4) Update the rate and mean of $q_\gamma^{(t)}(\gamma)$:

$$\tilde{d}^{(t)} = d + \big\|\mathbf{z} - \mathbf{A}\boldsymbol{\mu}^{(t)}\big\|^2 + \text{tr}\big\{\mathbf{A}\boldsymbol{\Phi}^{(t)}\mathbf{A}^{\text{H}}\big\}, \tag{40}$$

$$\big\langle \gamma^{(t)} \big\rangle = (c+N)/\tilde{d}^{(t)}. \tag{41}$$

until convergence

output $\hat{\mathbf{H}} =$ columns of $\mathbf{A}$ corresponding to the support of $\boldsymbol{\mu}^{(t)}$.

We will discuss the stopping rule for the iteration shortly after presenting a similar iterative procedure for subspace recovery based on the SKL model.

2) With the SKL Model: For the SKL problem, we have

$$p(\mathbf{z},\mathbf{x},\boldsymbol{\alpha},\gamma,\mathbf{b}) = p(\mathbf{z}|\mathbf{x},\gamma)p(\mathbf{x}|\boldsymbol{\alpha})p(\boldsymbol{\alpha}|\mathbf{b})p(\mathbf{b})p(\gamma), \quad (42)$$

where the factor distributions $p(\mathbf{z}|\mathbf{x},\gamma)$, $p(\mathbf{x}|\boldsymbol{\alpha})$, and $p(\gamma)$ are identical to those in (32) for the SK case, while $p(\boldsymbol{\alpha}|\mathbf{b})$ is given by (21) with $\mathbf{b} = \{b_m, m \in \mathbb{P}\}$ modeled as a latent variable with distribution $p(\mathbf{b})$ given by (23). Note that since the latent variables $\mathbf{x}$ and γ are not directly related to $\mathbf{b}$, their update remains unchanged as in Algorithm 1. Hence, we only need to consider the update of $q_\alpha(\boldsymbol{\alpha})$ and $q_b(\mathbf{b})$. In Appendix B, we show that the two posteriors are given by

$$q_\alpha(\boldsymbol{\alpha}) = \prod_{m=1}^{M} \text{Gamma}(\alpha_m; a+1, \tilde{b}_m) \quad (43)$$

$$q_b(\mathbf{b}) = \prod_{m\in\mathbb{P}} \text{Gamma}(b_m; u+a, \tilde{v}_m) \quad (44)$$

where $\tilde{b}_m$ and $\tilde{v}$ are respectively given by (63) and (65) in Appendix B. We summarize the steps that can be used to recover the subspace matrix $\mathbf{H}$ based on the SKL model in Algorithm 2.

Algorithm 2: SKL-based Subspace Estimator.

Initialize Let $\langle \alpha_m^{(0)} \rangle = a/b_m$, $m = 1,\ldots,M$, $\langle \gamma^{(0)} \rangle = c/d$, $\mathbf{D}^{(0)} \triangleq \text{diag}\{\langle \alpha_1^{(0)} \rangle, \ldots, \langle \alpha_M^{(0)} \rangle\}$, and $t = 0$.

repeat

1) Set $t = t+1$.

2) Update the covariance and mean of $q_x^{(t)}(\mathbf{x})$ by (36) and (37).

3) Update the rate and mean of $q_\alpha^{(t)}(\boldsymbol{\alpha})$:

$$\tilde{b}_m^{(t)} = \begin{cases} \langle b_m^{(t-1)} \rangle + \left|\mu_m^{(t-1)}\right|^2 + \Phi_{m,m}^{(t-1)}, & m \in \mathbb{P}, \\ b_m + \left|\mu_m^{(t-1)}\right|^2 + \Phi_{m,m}^{(t-1)}, & m \in \mathbb{P}^{\text{c}}, \end{cases} \quad (45)$$

$$\langle \alpha_m^{(t)} \rangle = (a+1)/\tilde{b}_m^{(t)}. \quad (46)$$

4) Update the rate and mean of $q_\gamma^{(t)}(\gamma)$ by (40) and (41).

5) Update the rate and mean of $q_b^{(t)}(\mathbf{b})$:

$$\tilde{v}_m^{(t)} = v + \langle \alpha_m^{(t)} \rangle, \quad (47)$$

$$\langle b_m^{(t)} \rangle = (u+a)/\tilde{v}_m^{(t)}. \quad (48)$$

until convergence

output $\hat{\mathbf{H}} =$ columns of $\mathbf{A}$ corresponding to the support of $\boldsymbol{\mu}^{(t)}$.

Convergence is reached when the difference of some parameter estimates over two consecutive iterations is sufficiently small, e.g., $\|\boldsymbol{\mu}^{(t)} - \boldsymbol{\mu}^{(t-1)}\| / \|\boldsymbol{\mu}^{(t)}\| \le 10^{-3}$. We use $\boldsymbol{\mu}^{(t)}$, the posterior mean of $q_x(\mathbf{x})$ and also an estimate of the sparse vector $\mathbf{x}$. In addition, we can employ a pruning process to speed up the convergence, which is standard in sparse Bayesian learning methods [32], [43]. Specifically, after each iteration, we can set those $\alpha_m^{(t)}$ that are sufficiently large to infinity (or a very large number), which effectively sets the corresponding sparse coefficient x_m to zero. An estimate of the subspace matrix (i.e., $\mathbf{H}$ under H_0 or $\mathbf{P}_{\mathbf{s}}^{\perp}\mathbf{H}$ under H_1) can be obtained as the columns of $\mathbf{A}$ corresponding to the support of $\boldsymbol{\mu}^{(t)}$.

The computational complexity of both algorithms is dominated by the update of the covariance matrix $\boldsymbol{\Phi}^{(t)}$ via (36), which involves inverting an $M \times M$ matrix with a complexity of $O(M^3)$ per iteration. If M is larger than N, the complexity can be reduced to $O(NM^2)$ by using the matrix inversion lemma and rewriting (36) as

$$\boldsymbol{\Phi}^{(t)} = \langle \mathbf{D}^{(t-1)} \rangle^{-1} - \langle \mathbf{D}^{(t-1)} \rangle^{-1} \mathbf{A}^{\text{H}} \times \left(\langle \gamma^{(t-1)} \rangle^{-1} \mathbf{I} + \mathbf{A} \langle \mathbf{D}^{(t-1)} \rangle^{-1} \mathbf{A}^{\text{H}} \right)^{-1} \mathbf{A} \langle \mathbf{D}^{(t-1)} \rangle^{-1}$$

The complexity can further be reduced (by an order of magnitude) by using a message passing approach to compute (36) and (37), following similar procedures reported in [44], [45].

V. Numerical Results

In this section, we present simulation results to illustrate the performance of the proposed detectors. The disturbance $\mathbf{d}$ has a subspace structure as in (4), where $\mathbf{H}$ is formed by L Fourier vectors with frequencies centered around the zero frequency, i.e., it corresponds to a lowpass narrowband interference. Specifically, let $\boldsymbol{\mathcal{H}}'$ denotes an $N \times M'$ discrete Fourier transform (DFT) matrix, and the subspace matrix $\mathbf{H}$ consists of the following columns of $\boldsymbol{\mathcal{H}}'$: $\{1, 2, \ldots, \lfloor \frac{L+1}{2} \rfloor, M' - \lceil \frac{L-1}{2} \rceil + 1, M' - \lceil \frac{L-1}{2} \rceil + 2, \ldots, M'\}$, where $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ denote the floor and, respectively, ceiling operators. We set $N = 32$, $M' = 64$, and $L = 7$ in simulation. The target signal $\mathbf{s}$ is a Fourier vector with a normalized frequency 0.3 Hz. The set-up may correspond to detecting the presence/absence of a moving target with Doppler frequency 0.3 Hz in narrowband interference (i.e., clutter) and noise [33]. A standard assumption is that the interference bandwidth does not overlap with the mainlobe of the target response (otherwise, the interference will be detected as target). Therefore, 4 columns of $\boldsymbol{\mathcal{H}}'$, which cover the mainlobe of the target response, are removed to form the $N \times M$ dictionary matrix $\boldsymbol{\mathcal{H}}$ [cf. Section II], where $M = 60$. The signal-to-noise ratio (SNR) and interference-to-noise ratio (INR) are defined as: $\text{SNR} = N|\kappa|^2/\sigma^2$ and $\text{INR} = N\|\boldsymbol{\beta}\|^2/\sigma^2$.

For brevity, the proposed detector based on model (22) is referred to as the *SK* detector, whereas the one based on model (23) is referred to as the *SKL* detector. We compare with 5 other known detectors as benchmarks, namely the *clairvoyant* subspace detector of [10], which assumes perfect knowledge of the subspace matrix $\mathbf{H}$, the *conventional KA* detector, which takes the same form as the clairvoyant detector except that $\mathbf{H}$ is replaced by the prior knowledge of $\mathbf{H}$, the adaptive subspace detector (*ASD*), the orthogonal matching pursuit algorithm based detector (*OMP*) [46], and a non-informative *SBL* detector, which employs a similar Bayesian inference framework as the

proposed ones but is not provided with any prior knowledge of the subspace.

Specifically, the test variable of the clairvoyant detector is given by [10]

$$T_{\text{clairvoyant}} = \frac{|\mathbf{s}^{\mathrm{H}}\mathbf{P}_{\mathbf{H}}^{\perp}\mathbf{y}|^2}{\mathbf{s}^{\mathrm{H}}\mathbf{P}_{\mathbf{H}}^{\perp}\mathbf{s}} \tag{49}$$

where $\mathbf{P}_{\mathbf{H}}^{\perp} \triangleq \mathbf{I} - \mathbf{H}(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{H}}$ denotes the projection matrix projecting to the orthogonal complement of the true interference subspace span$(\mathbf{H})$. The test variable of the conventional KA detector has the same form as in (49), except that the interference subspace matrix $\mathbf{H}$ is replaced by a prior knowledge, which may be incomplete or contain errors (detailed later). The ASD also uses the above test variable except that the project matrix $\mathbf{P}_{\mathbf{H}}^{\perp}$ is replaced by an estimate obtained from training data. In particular, let $\hat{\mathbf{R}} \in \mathbb{C}^{N\times N}$ denote the sample covariance matrix obtained from T training signals $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T$, and $\mathbf{U} \in \mathbb{C}^{N\times(N-L)}$ consist of the eigenvectors associated with the smallest $N-L$ eigenvalues of $\hat{\mathbf{R}}$. The projection matrix used by the ASD is given by $\mathbf{U}\mathbf{U}^{\mathrm{H}}$ [12]. Meanwhile, the test variable of the SBL and OMP takes the same form as that of the proposed GLRT (17), except that the subspace matrix $\mathbf{H}$ is replaced by the corresponding SBL and, respectively, OMP algorithm. Clearly, the clairvoyant detector indicates the best achievable performance of all, while the SBL provides a reference to illustrate the benefit of knowledge exploitation.

Using the set definitions of Section II, we can write the true subspace index set $\mathbb{T} = \left\{1, 2, \ldots, \lfloor\frac{L+1}{2}\rfloor, M - \lceil\frac{L-1}{2}\rceil + 1, M - \lceil\frac{L-1}{2}\rceil + 2, \ldots, M\right\}$, which is a subset of $\mathbb{S}$ in (6). For the KA detectors, including the conventional KA, SK, and SKL, we consider 4 different cases for the prior knowledge $\mathbb{P}$ to examine the effects of correct vs. erroneous information for KA detection.

- *Case 1:* $\mathbb{P} = \mathbb{T}$. The prior knowledge is perfect and contains no errors.
- *Case 2:* $\mathbb{P} \supset \mathbb{T}$ and $\mathbb{E} \neq \varnothing$. The prior knowledge has not only all basis vectors in $\mathbb{S}$ but also card$(\mathbb{E}) = 3$ additional erroneous ones.
- *Case 3:* $\mathbb{P} \subset \mathbb{T}$ and $\mathbb{E} = \varnothing$. The prior knowledge for each simulation trial contains card$(\mathbb{P}) = 4$ randomly selected indices of $\mathbb{T}$ but no erroneous ones.
- *Case 4:* $\mathbb{P} \subset \mathbb{T}$ and $\mathbb{E} \neq \varnothing$. The prior knowledge for each trial contains card$(\mathbb{P}) = 4$ randomly selected indices of $\mathbb{T}$ and card$(\mathbb{E}) = 3$ erroneous ones.

A. Case 1: Perfect Knowledge

Case 1 represents an ideal case with perfect prior knowledge $\mathbb{P} = \mathbb{T}$. The case is of interest to show the proposed SK and SKL, behaves in the presence of perfect prior knowledge. Note that due to the different prior models, Bayesian learning functions differently for the above 2 detectors and SBL. Specifically, SBL employs a non-informative prior, it screens non-discriminatorily all columns of the dictionary matrix $\boldsymbol{\mathcal{H}}$ in search of basis vectors that can describe the data. For SK and SKL, the columns of $\boldsymbol{\mathcal{H}}$ are split into 2 non-overlapping subsets $\mathbb{P}$ and $\mathbb{P}^{\mathrm{c}}$. Within $\mathbb{P}^{\mathrm{c}}$, SK and SKL also employ non-informative Bayesian learning to search for missing basis vectors (aside from the prior knowledge $\mathbb{P}$). Within $\mathbb{P}$, SK is likely to accept it as is, whereas SKL screens for possible mistakes and may reject some bases in $\mathbb{P}$.

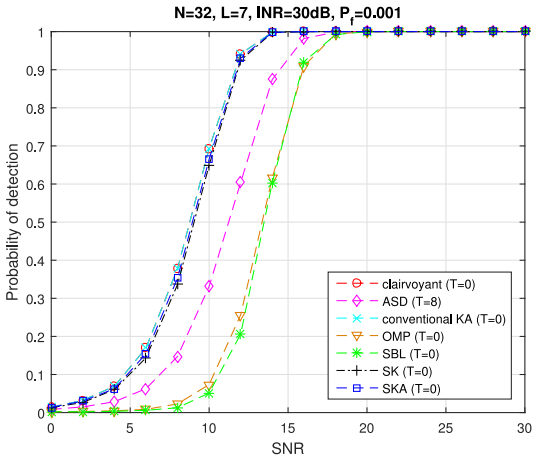
Fig. 1(a) depicts the probability of detection P_{d} versus the SNR for the various detectors, where INR $= 30$ dB and the probability of false alarm $P_{\mathrm{f}} = 10^{-3}$. The conventional KA in the current case reduces to the clairvoyant detector. The performance of the proposed SK and SKL is also nearly identical to the clairvoyant, manifesting the benefit of the prior knowledge. Their near optimality also indicates that they rarely reject correct basis vectors in $\mathbb{P}$ or add erroneous basses in $\mathbb{P}^{\mathrm{c}}$. Meanwhile, without using the prior knowledge $\mathbb{P}$, OMP and SBL are similar to each other and inferior to the other detectors. The ASD is the only detector that requires training. With $T = 8$ target-free i.i.d. training signals, its performance is still notably worse than the KA detectors. Fig. 1(b) shows P_{d} versus P_{f}, i.e., the receiver operating characteristic (ROC) curve, where SNR $=$ 15 dB and INR $= 30$ dB. The relations among the various detectors are similar to what were observed before. Fig. 1(c) shows the *recovery rate* versus the column index, converted into the normalized frequency, for the 3 Bayesian learning based detectors. The recovery rate at any column index is defined as the percentage that column is identified as a basis vector of the subspace matrix $\mathbf{H}$ by the Bayesian inference algorithm. The recovery behaviors of the SK and SKL over $\mathbb{P}$ and $\mathbb{P}^{\mathrm{c}}$ match our prediction, while the SBL misses some of the correct basis vectors with low recovery rate, which causes its performance degradation.

B. Case 2: Full Knowledge Plus Errors

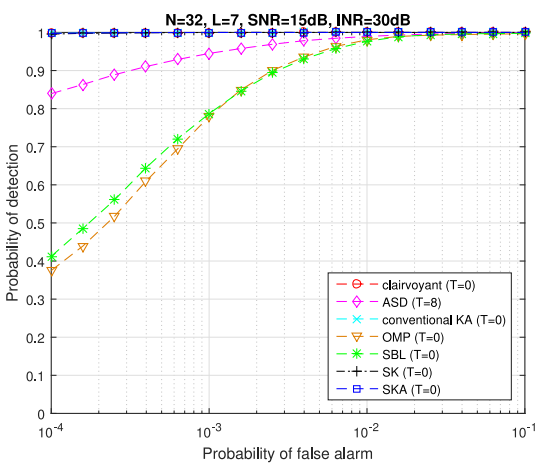
In this case, $\mathbb{P} \supset \mathbb{T}$ and card$(\mathbb{E}) = 3$, which implies the prior knowledge $\mathbb{P}$ include all basis vectors of the subspace matrix and, in addition, 3 erroneous basis vectors. The erroneous basis vectors are assumed to near the target frequency 0.3, which has more detrimental effects due to partial signal cancellation. The results are depicted in Fig. 2. Figs. 2(a) and (b) show that the proposed SKL is the one that is closest to the clairvoyant detector. In fact, its performance appears to be unaffected by the presence of erroneous knowledge and nearly identical to that in Case 1. This is corroborated by the recovery rate result in Fig. 2(c). However, the errors in $\mathbb{P}$ have an impact on the conventional KA and SK detectors, causing a considerable performance degradation for both. SBL and ASD are independent of the prior knowledge $\mathbb{P}$ and their performance remains unchanged.

C. Case 3: Partial Knowledge

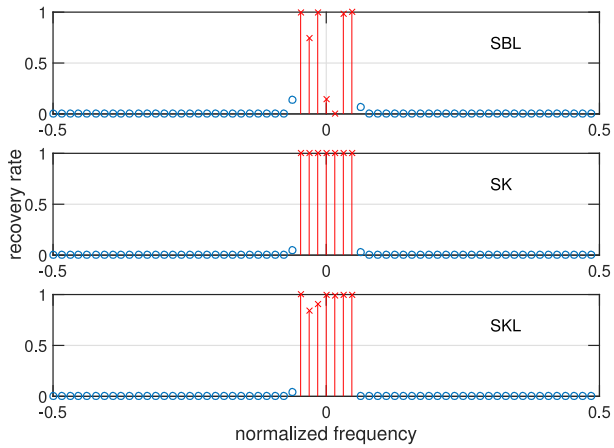
We now consider a partial knowledge case with $\mathbb{P} \subset \mathbb{T}$, card$(\mathbb{P}) = 4$, and $\mathbb{E} = \varnothing$, that is, the prior knowledge $\mathbb{P}$ contains 4 randomly selected bases from $\mathbb{T}$ in each trial but no erroneous ones. The results are shown in Fig. 3. It is seen that the SK detector is the closest to the clairvoyant in this case, also slightly better than the SKL. This is expected, since both SK and SKL use the same approach to find missing bases in $\mathbb{P}^{\mathrm{c}}$, but in $\mathbb{P}$, the two detectors behave differently. Specifically, there is a small probability with which SKL may reject some of the correct bases in $\mathbb{P}$, while SK is more likely to accept them and thus benefits more from the prior knowledge.



(a)

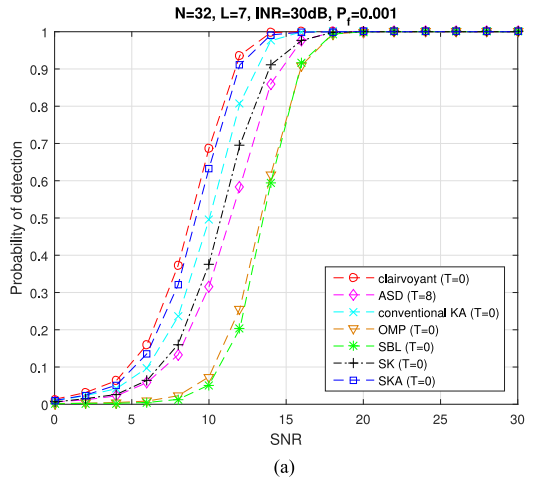


(b)

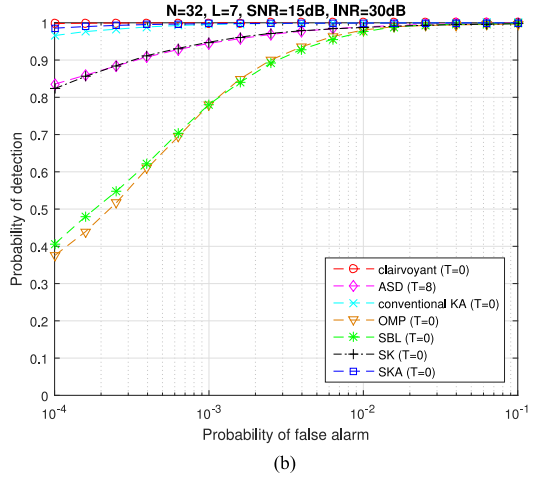


(c)

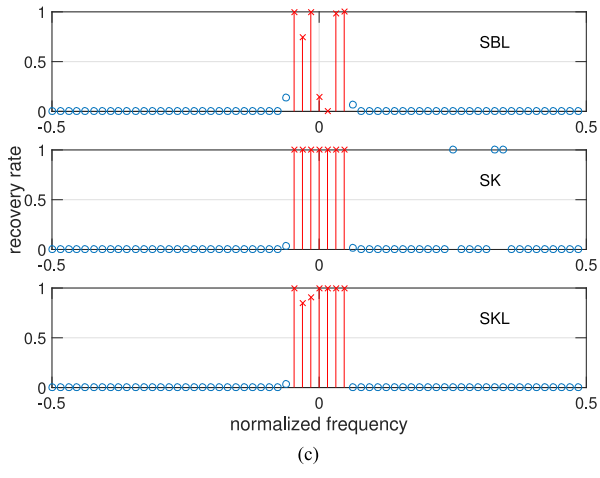
Fig. 1. Case 1 results. (a) P_{d} vs. SNR with INR $= 30$ dB and $P_{\mathrm{f}} = 10^{-3}$. (b) ROC curve with SNR $= 15$ dB and INR $= 30$ dB. (c) Recovery rate vs. normalized frequency, with red 'x' markers indicating the locations of the true subspace bases and blue 'o' markers for other bases.



(a)



(b)



(c)

Fig. 2. Case 2 (full knowledge plus errors) results. (a) P_{d} vs. SNR with INR $= 30$ dB and $P_{\mathrm{f}} = 10^{-3}$. (b) ROC curve with SNR $= 15$ dB and INR $= 30$ dB. (c) Recovery rate vs. normalized frequency.

D. Case 4: Partial Knowledge With Errors

Finally, we consider the most challenging case involving both partial and erroneous knowledge with $\mathbb{P} \subset \mathbb{T}$, $\mathrm{card}(\mathbb{P}) = 4$, and $\mathrm{card}(\mathbb{E}) = 3$, that is, the prior knowledge $\mathbb{P}$ contains 4 correct bases randomly selected from $\mathbb{T}$ and 3 erroneous ones. The results are shown in Fig. 4. Like Case 2, the proposed SKL is able to identify the erroneous bases and, as a result, outperforms all but the clairvoyant detector. Unlike Case 2, however, SKL has less correct knowledge and needs to recover missing bases from $\mathbb{P}^{\mathrm{c}}$; the overall recovery rate is lower in Case 4 [cf. Figs. 4(c) and 2(c)], which contributes to some performance loss. Among the 3 KA detectors, the conventional KA is the worse, suffering

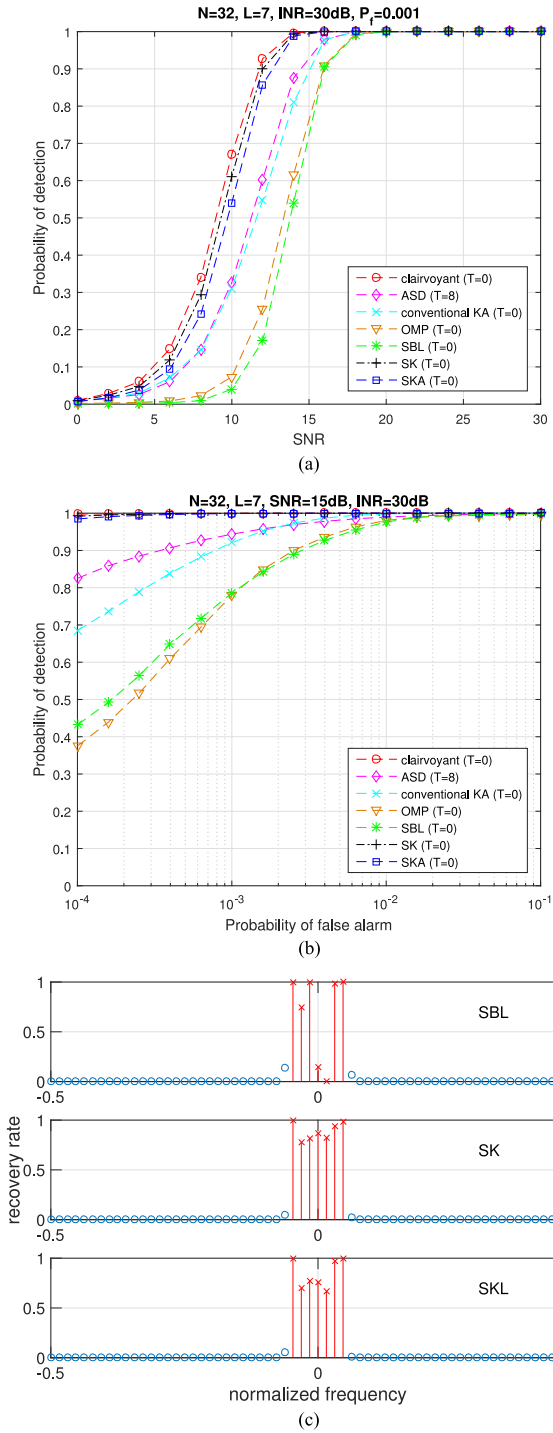


Fig. 3. Case 3 (partial knowledge) results. (a) P_{d} vs. SNR with INR $=$ 30 dB and $P_{\mathrm{f}} = 10^{-3}$. (b) ROC curve with SNR $=$ 15 dB and INR $=$ 30 dB. (c) Recovery rate vs. normalized frequency.

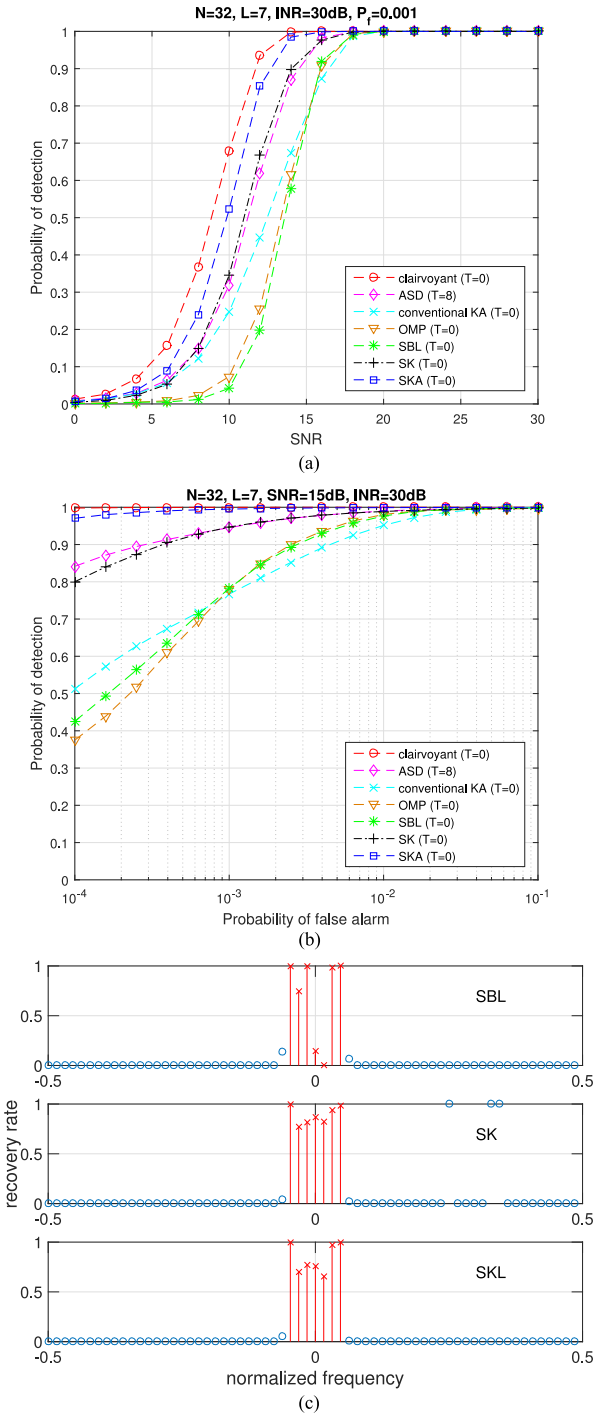


Fig. 4. Case 4 (partial knowledge with errors): (a) P_{d} vs. SNR with INR $=$ 30 dB and $P_{\mathrm{f}} = 10^{-3}$. (b) ROC curve with SNR $=$ 15 dB and INR $=$ 30 dB. (c) Recovery rate vs. normalized frequency.

from both missing bases and wrong bases in the knowledge set $\mathbb{P}$. The SK detector is able to recover most of the missing bases, as indicated by Fig. 4(c), and is therefore considerably better than the conventional KA. It is mainly affected by the erroneous bases. The results here and in Case 2 confirm the need to identify errors in prior knowledge with uncertainty.

VI. Conclusion

We presented a new knowledge-aided (KA) approach for signal detection in strong disturbance by exploiting prior knowledge of the subspace structure of the disturbance. A unique contribution is that the proposed approach accounts for the fact that the prior knowledge available in practice is often incomplete and subject to possible errors. To address such uncertainties, we introduced two Bayesian hierarchical models for

knowledge representation. One is suitable for the case when the prior knowledge is incomplete but primarily accurate, while the other tries to identify errors in the prior knowledge by learning from the observed data. These models were integrated in a Bayesian learning framework, and variational inference algorithms based on simple iteration steps were developed to solve the associated inference problem. Extensive simulation results, which cover a range of prior knowledge scenarios, demonstrate that the proposed KA detectors can benefit from uncertain prior knowledge and significantly outperform conventional detectors in such cases.

APPENDIX

A. Derivation of SK-Based Estimator

The factor posteriors $q_x(\mathbf{x})$, $q_\alpha(\boldsymbol{\alpha})$, and $q_\gamma(\gamma)$ can be computed by using the mean field approximation (30) along with the joint distribution (32). For $q_x(\mathbf{x})$, we have (keeping only the terms dependent on $\mathbf{x}$)

$$\begin{aligned}\ln q_x(\mathbf{x}) &\propto \left\langle \ln p(\mathbf{z}|\mathbf{x},\gamma) + \ln p(\mathbf{x}|\boldsymbol{\alpha}) \right\rangle_{q_\alpha(\boldsymbol{\alpha}) q_\gamma(\gamma)} \\ &\propto -\langle\gamma\rangle (\mathbf{z}-\mathbf{A}\mathbf{x})^{\mathrm{H}}(\mathbf{z}-\mathbf{A}\mathbf{x}) - \mathbf{x}^{\mathrm{H}}\langle\mathbf{D}\rangle\mathbf{x},\end{aligned} \tag{50}$$

where $\langle\gamma\rangle$ denotes the mean of γ computed w.r.t. $q_\gamma(\gamma)$ and

$$\langle\mathbf{D}\rangle \triangleq \operatorname{diag}\left\{\langle\alpha_1\rangle, \ldots, \langle\alpha_M\rangle\right\} \tag{51}$$

with $\langle\alpha_m\rangle$ denoting the mean of α_m w.r.t. $q_\alpha(\boldsymbol{\alpha})$. Both $\langle\gamma\rangle$ and $\langle\alpha_m\rangle$ are determined shortly in (61) and (57), respectively. It can be readily verified from (50) that $q_x(\mathbf{x})$ is a complex Gaussian distribution, as expressed in (33), where the mean and covariance are

$$\boldsymbol{\mu} = \langle\gamma\rangle \boldsymbol{\Phi}\mathbf{A}^{\mathrm{H}}\mathbf{z}, \tag{52}$$

$$\boldsymbol{\Phi} = \left(\langle\gamma\rangle \mathbf{A}^{\mathrm{H}}\mathbf{A} + \langle\mathbf{D}\rangle\right)^{-1}. \tag{53}$$

Likewise, the logarithmic $q_\alpha(\boldsymbol{\alpha})$ can be written as

$$\begin{aligned}\ln q_\alpha(\boldsymbol{\alpha}) &\propto \left\langle \ln p(\mathbf{x}|\boldsymbol{\alpha}) + \ln p(\boldsymbol{\alpha}) \right\rangle_{q_x(\mathbf{x})} \\ &\propto \sum_{m=1}^{M} \left[a\ln\alpha_m - \left(\left\langle |x_m|^2 \right\rangle + b_m\right)\alpha_m \right],\end{aligned} \tag{54}$$

where $\left\langle |x_m|^2 \right\rangle$ denotes the expectation of x_m w.r.t. $q_x(\mathbf{x})$:

$$\left\langle |x_m|^2 \right\rangle = |\mu_m|^2 + \Phi_{m,m}, \tag{55}$$

with μ_m and $\Phi_{m,m}$ denoting the m-th element of $\boldsymbol{\mu}$ and the m-th diagonal element of $\boldsymbol{\Phi}$, respectively. It follows from (54) that $q_\alpha(\boldsymbol{\alpha})$ is a product Gamma distribution as expressed in (34) with shape and rate parameters given by $a+1$ and, respectively,

$$\tilde{b}_m = b_m + |\mu_m|^2 + \Phi_{m,m}, \quad m = 1, \ldots, M. \tag{56}$$

The mean of α_m, which is required to compute (53), is

$$\langle\alpha_m\rangle = \frac{a+1}{\tilde{b}_m}. \tag{57}$$

Finally, for $q_\gamma(\gamma)$, we have

$$\begin{aligned}\ln q_\gamma(\gamma) &\propto \left\langle \ln p(\mathbf{z}|\mathbf{x},\gamma) + \ln p(\gamma) \right\rangle_{q_x(\mathbf{x})} \\ &\propto (c+N-1)\ln\gamma - \left(d + \left\langle \|\mathbf{z}-\mathbf{A}\mathbf{x}\|^2 \right\rangle_{q_x(\mathbf{x})}\right)\gamma.\end{aligned} \tag{58}$$

By using (52) and (53), it is easy to show

$$\left\langle \|\mathbf{z}-\mathbf{A}\mathbf{x}\|^2 \right\rangle_{q_x(\mathbf{x})} = \|\mathbf{z}-\mathbf{A}\boldsymbol{\mu}\|^2 + \operatorname{tr}\left\{\mathbf{A}\boldsymbol{\Phi}\mathbf{A}^{\mathrm{H}}\right\}. \tag{59}$$

It follows from (58) that $q_\gamma(\gamma)$ is a Gamma distribution as represented in (35) with shape $c+N$, rate

$$\tilde{d} = d + \|\mathbf{z}-\mathbf{A}\boldsymbol{\mu}\|^2 + \operatorname{tr}\left\{\mathbf{A}\boldsymbol{\Phi}\mathbf{A}^{\mathrm{H}}\right\}, \tag{60}$$

and, in turn, mean

$$\langle\gamma\rangle = \frac{c+N}{\tilde{d}}. \tag{61}$$

Noted that some parameters of the factor posteriors are related to one another and cannot be determined in closed form. For example, the mean (52) and covariance (53) of $q_x(\mathbf{x})$ depend on $\langle\gamma\rangle$ and $\{\langle\alpha_m\rangle\}$, while the latter also depend on the former, as shown in (57) and (61). A cyclic approach can be used to update these parameters iteratively, by updating one parameter at a time and fixing the others. This leads to the cyclic updating steps (36)–(41) of Algorithm 1, which are based on (52), (53), (56), (57), (60) and (61), respectively.

B. Derivation of SKL-Based Estimator

As noted in Section IV-B2, the factor posteriors $q_x(\mathbf{x})$ and $q_\gamma(\gamma)$ based on the SKL model remain identical to their SK counterparts (33) and (35), respectively. We only need to determine $q_\alpha(\boldsymbol{\alpha})$ and $q_b(\mathbf{b})$ as follows. Specifically, by using (42) in the mean field approximation (30), $q_\alpha(\boldsymbol{\alpha})$ can be written as

$$\begin{aligned}\ln q_\alpha(\boldsymbol{\alpha}) &\propto \left\langle \ln p(\mathbf{x}|\boldsymbol{\alpha}) + \ln p(\boldsymbol{\alpha}) \right\rangle_{q_x(\mathbf{x}) q_b(\mathbf{b})} \\ &\propto \sum_{m=1}^{M} \left\langle a\ln\alpha_m - \left(|x_m|^2 + b_m\right)\alpha_m \right\rangle_{q_x(\mathbf{x}) q_b(\mathbf{b})} \\ &= \sum_{m\in\mathbb{P}} \left[a\ln\alpha_m - \left(\left\langle |x_m|^2 \right\rangle + \langle b_m \rangle\right)\alpha_m \right] \\ &\quad + \sum_{m\in\mathbb{P}^{\mathrm{c}}} \left[a\ln\alpha_m - \left(\left\langle |x_m|^2 \right\rangle + b_m\right)\alpha_m \right]\end{aligned} \tag{62}$$

where $\left\langle |x_m|^2 \right\rangle$ is given by (55) and $\langle b_m \rangle$ denotes the mean of b_m w.r.t. $q_b(b_m)$ that is determined shortly. Clearly, $q_\alpha(\boldsymbol{\alpha})$ is a product Gamma distribution as expressed in (43), with shape $a+1$ and rate

$$\tilde{b}_m = \begin{cases} \left\langle |x_m|^2 \right\rangle + \langle b_m \rangle, & m \in \mathbb{P}, \\ \left\langle |x_m|^2 \right\rangle + b_m, & m \in \mathbb{P}^{\mathrm{c}}. \end{cases} \tag{63}$$

The mean of α_m has the same form as (57).

For $q_b(\mathbf{b})$, we have

$$\ln q_b(\mathbf{b}) \propto \left\langle \ln p(\boldsymbol{\alpha}|\mathbf{b}) + \ln p(\mathbf{b}) \right\rangle_{q_\alpha(\boldsymbol{\alpha})} \propto \sum_{m \in \mathbb{P}} \Big[(u+a-1)\ln b_m - \big(\langle \alpha_m \rangle + v\big) b_m \Big], \quad (64)$$

which implies that $q_b(\mathbf{b})$ is a product Gamma distribution as shown in (44), with shape $u+a$, rate

$$\tilde{v}_m = v + \langle \alpha_m \rangle, \quad (65)$$

and mean

$$\langle b_m \rangle = \frac{u+a}{\tilde{v}_m}. \quad (66)$$

Equations (63), (65), and (66), along with the expressions for the parameters associated with $q_x(\mathbf{x})$ and $q_\gamma(\gamma)$ in Appendix A, can be cycled through to provide updates of all factor posteriors. Details of the iterative steps are listed in Algorithm 2.

REFERENCES

[1] A. Paulraj and C. B. Papadias, "Space-time processing for wireless communications," *IEEE Signal Process. Mag.*, vol. 14, no. 4, pp. 49–83, Nov. 1997.

[2] J. R. Guerci, *Space-Time Adaptive Processing for Radar*. Norwood, MA, USA: Artech House, 2014.

[3] A. De Maio and M. S. Greco, Eds., *Modern Radar Detection Theory*, 1st ed. Edison, NJ, USA: SciTech Publ., 2016.

[4] A. De Maio, "Rao test for adaptive detection in Gaussian interference with unknown covariance matrix," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3577–3584, Jul. 2007.

[5] J. Ward, "Space-time adaptive processing for airborne radar," Lincoln Lab., Massachusetts Inst. Technol., Cambridge, MA, USA, Tech. Rep. 1015, Dec. 1994.

[6] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[7] A. De Maio and D. Orlando, "Adaptive radar detection of a subspace signal embedded in subspace structured plus Gaussian interference via invariance," *IEEE Trans. Signal Process.*, vol. 64, no. 8, pp. 2156–2167, Apr. 2016.

[8] F. Bandiera, D. Orlando, and G. Ricci, "Advanced radar detection schemes under mismatched signal models," in *Synthesis Lectures on Signal Processing*, J. Moura, Ed. Williston, VT, USA: Morgan & Claypool Publ., 2009, ch. 8.

[9] M. Rangaswamy, F. C. Lin, and K. R. Gerlach, "Robust adaptive signal processing methods for heterogeneous radar clutter scenarios," *Signal Process.*, vol. 84, no. 9, pp. 1653–1665, 2004.

[10] P. Wang, H. Li, and B. Himed, "Moving target detection using distributed MIMO radar in clutter with nonhomogeneous power," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4809–4820, Oct. 2011.

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

[12] A. Haimovich, "The eigencanceler: Adaptive radar by eigenanalysis methods," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, no. 2, pp. 532–542, Apr. 1996.

[13] Y. Hua, M. Nikpour, and P. Stoica, "Optimal reduced-rank estimation and filtering," *IEEE Trans. Signal Process.*, vol. 49, no. 3, pp. 457–469, Mar. 2001.

[14] W. Chen, U. Mitra, and P. Schniter, "On the equivalence of three reduced rank linear estimators with applications to DS-CDMA," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2609–2614, Sep. 2002.

[15] J. S. Goldstein, I. S. Reed, and L. L. Scharf, "A multistage representation of the Wiener filter based on orthogonal projections," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2943–2959, Nov. 1998.

[16] L. L. Scharf, E. K. P. Chong, M. D. Zoltowski, J. S. Goldstein, and I. S. Reed, "Subspace expansion and the equivalence of conjugate direction and multistage wiener filters," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5013–5019, Oct. 2008.

[17] C. Jiang, H. Li, and M. Rangaswamy, "On the conjugate gradient matched filter," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2660–2666, May 2012.

[18] Z. Chen, H. Li, and M. Rangaswamy, "Conjugate gradient adaptive matched filter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, no. 1, pp. 178–191, Jan. 2015.

[19] J. R. Guerci, "Cognitive radar: A knowledge-aided fully adaptive approach," in *Proc. IEEE Radar Conf.*, 2010, pp. 1365–1370.

[20] P. Stoica, J. Li, X. Zhu, and J. R. Guerci, "On using a priori knowledge in space-time adaptive processing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2598–2602, Jun. 2008.

[21] A. De Maio and A. Farina, "Adaptive radar detection: A Bayesian approach," in *Proc. Int. Radar Symp.*, Krakow, Poland, May 2006, pp. 85–88.

[22] O. Besson, J.-Y. Tourneret, and S. Bidon, "Knowledge-aided Bayesian detection in heterogeneous environment," *IEEE Signal Process. Lett.*, vol. 14, no. 5, pp. 355–358, May 2007.

[23] A. De Maio, A. Farina, and G. Foglia, "Knowledge-aided Bayesian radar detectors & their application to live data," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46, no. 1, pp. 170–183, Feb. 2010.

[24] P. Wang, Z. Sahinoglu, M.-O. Pun, H. Li, and B. Himed, "Knowledge-aided adaptive coherence estimator in stochastic partially homogeneous environments," *IEEE Signal Process. Lett.*, vol. 18, no. 3, pp. 193–196, Mar. 2011.

[25] S. Bidon, O. Besson, and J. Y. Tourneret, "Knowledge-aided STAP in heterogeneous clutter using a hierarchical Bayesian algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, no. 3, pp. 1863–1879, Jul. 2011.

[26] A. Aubry, A. De Maio, L. Pallotta, and A. Farina, "Radar detection of distributed targets in homogeneous interference whose inverse covariance structure is defined via unitary invariant functions," *IEEE Trans. Signal Process.*, vol. 61, no. 20, pp. 4949–4961, Oct. 2013.

[27] A. De Maio, D. Orlando, C. Hao, and G. Foglia, "Adaptive detection of point-like targets in spectrally symmetric interference," *IEEE Trans. Signal Process.*, vol. 64, no. 12, pp. 3207–3220, Jun. 2016.

[28] A. Aubry, V. Carotenuto, A. De Maio, and G. Foglia, "Exploiting multiple a priori spectral models for adaptive radar detection," *IET Radar, Sonar Navig.*, vol. 8, no. 7, pp. 695–707, Jul. 2013.

[29] A. De Maio, D. Orlando, I. Soloveychik, and A. Wiesel, "Invariance theory for adaptive detection in interference with group symmetric covariance matrix," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6299–6312, Dec. 2016.

[30] A. De Maio and D. Orlando, "An invariant approach to adaptive radar detection under covariance persymmetry," *IEEE Trans. Signal Process.*, vol. 63, no. 5, pp. 1297–1309, Mar. 2015.

[31] C. Hao, D. Orlando, G. Foglia, and G. Giunta, "Knowledge-based adaptive detection: Joint exploitation of clutter and system symmetry properties," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1489–1493, Oct. 2016.

[32] M. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, 2001.

[33] M. A. Richards, *Fundamentals of Radar Signal Processing*. New York, NY, USA: McGraw-Hill, 2005.

[34] E. J. Candès and C. Fernandez-Granda, "Super-resolution from noisy data," *J. Fourier Anal. Appl.*, vol. 19, no. 6, pp. 1229–1254, 2013.

[35] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7465–7490, Nov. 2013.

[36] Z. Yang, L. Xie, and C. Zhang, "Off-grid direction of arrival estimation using sparse Bayesian inference," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 38–43, Jan. 2013.

[37] L. Hu, J. Zhou, Z. Shi, and Q. Fu, "A fast and accurate reconstruction algorithm for compressed sensing of complex sinusoids," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5744–5754, Nov. 2013.

[38] J. Fang, F. Wang, Y. Shen, H. Li, and R. Blum, "Super-resolution compressed sensing for line spectral estimation: An iterative reweighted approach," *IEEE Trans. Signal Process.*, vol. 64, no. 18, pp. 4649–4662, Sep. 2016.

[39] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Berlin, Germany: Springer, 2013.

[40] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul, "An introduction to variational methods for graphical models," *Mach. Learn.*, vol. 37, no. 2, pp. 183–233, 1999.

[41] V. Šmídl and A. Quinn, *The Variational Bayes Method in Signal Processing*. Berlin, Germany: Springer, 2006.

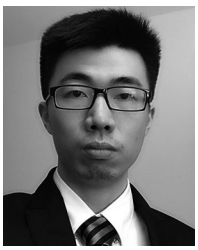
[42] D. G. Tzikas, C. Likas, and N. P. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 131–146, 2008, pp. 131–146.

[43] J. Fang, Y. Shen, H. Li, and P. Wang, "Pattern-coupled sparse Bayesian learning for recovery of block-sparse signals," *IEEE Trans. Signal Process.*, vol. 63, no. 2, pp. 360–372, Jan. 2015.
[44] M. Al-Shoukairi and B. Rao, "Sparse Bayesian learning using approximate message passing," in *Proc. 48th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2014, pp. 1957–1961.
[45] F. Jun, L. Zhang, and H. Li, "Two-dimensional pattern-coupled sparse Bayesian learning via generalized approximate message passing," *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2920–2930, Jun. 2016.
[46] J. A. Tropp and A. C. Gilber, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

Hongbin Li (M'99–SM'08) received the B.S. and M.S. degrees from the University of Electronic Science and Technology of China, in 1991 and 1994, respectively, and the Ph.D. degree from the University of Florida, Gainesville, FL, USA, in 1999, all in electrical engineering.

From July 1996 to May 1999, he was a Research Assistant in the Department of Electrical and Computer Engineering at the University of Florida. Since July 1999, he has been with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, USA, where he became a Professor in 2010. He was a Summer Visiting Faculty Member at the Air Force Research Laboratory in the summers of 2003, 2004, and 2009. His general research interests include statistical signal processing, wireless communications, and radars.

Dr. Li received the IEEE Jack Neubauer Memorial Award in 2013 from the IEEE Vehicular Technology Soceity, the Outstanding Paper Award from the IEEE AFICON Conference in 2011, the Harvey N. Davis Teaching Award in 2003, the Jess H. Davis Memorial Award for excellence in research in 2001 from Stevens Institute of Technology, and the Sigma Xi Graduate Research Award from the University of Florida in 1999. He was a member of the IEEE SPS Signal Processing Theory and Methods Technical Committee (TC) and the IEEE SPS Sensor Array and Multichannel TC, an Associate Editor for *Signal Processing* (Elsevier), the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE SIGNAL PROCESSING LETTERS, and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, as well as a Guest Editor for the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, and *EURASIP Journal on Applied Signal Processing*. He has been involved in various conference organization activities and was a General Co-Chair for the 7th IEEE Sensor Array and Multichannel Signal Processing Workshop, Hoboken, NJ, USA, June 17–20, 2012. He is currently a member of Tau Beta Pi and Phi Kappa Phi.



Yuan Jiang received the B.Eng. degree from in electrical engineering the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2015. He received the M.S degree, in 2016, from Stevens Institute of Technology, Hoboken, NJ, USA, where he is currently working toward the Ph.D. degree, both in electrical engineering.

Since 2016, he has been a Research Assistant in the Department of Electrical and Computer Engineering, Stevens Institute of Technology. His research interests include statistical signal processing, Bayesian inference, and sparse signal recovery.

Jun Fang (M'08) received the B.S. and M.S. degrees from the Xidian University, Xi'an, China, in 1998 and 2001, respectively, and the Ph.D. degree from the National University of Singapore, Singapore, in 2006, all in electrical engineering.

During 2006, he was a Postdoctoral Research Associate in the Department of Electrical and Computer Engineering, Duke University. From January 2007 to December 2010, he was a Research Associate in the Department of Electrical and Computer Engineering, Stevens Institute of Technology. Since 2011, he has been with the University of Electronic of Science and Technology of China. His research interests include compressed sensing and sparse theory, massive MIMO/mmWave communications, and statistical inference.

Dr. Fang received the IEEE Jack Neubauer Memorial Award in 2013 for the best systems paper published in the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He is currently an Associate Technical Editor of the IEEE COMMUNICATIONS MAGAZINE and an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS.

Muralidhar Rangaswamy (S'89–M'93–SM'98–F'06) received the B.E. degree from Bangalore University, Bangalore, India, in 1985, and the M.S. and Ph.D. degrees from Syracuse University, Syracuse, NY, USA, in 1992, all in electrical engineering.

He is currently the Senior Advisor for Radar Research at the RF Exploitation Branch within the Sensors Directorate of the Air Force Research Laboratory, Wright-Patterson Air Force Base (AFRL), OH, USA. Prior to this, he has held industrial and academic appointments. His research interests include radar signal processing, spectrum estimation, modeling non-Gaussian interference phenomena, and statistical communication theory. A common theme of his research in these areas is that of big data. A large class of radar signal processing problems incurs the curse of dimensionality, i.e., the training data requirement and computational cost grow explosively with increasing dimensionality. He has coauthored more than 200 refereed journal and conference record papers in the areas of his research interests. Additionally, he is a contributor to eight books and is a co-inventor on three U.S. patents. He has developed innovative techniques to address these challenges by using physics based approaches and exploiting underlying structure in the data gained from a priori information.

Dr. Rangaswamy was the Technical Editor (Associate Editor-in-Chief) for radar systems in the IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS (IEEE-TAES) from 2007–2015 and an Associate Editor for radar systems within the IEEE-TAES from 2004–2007. He was the Co-Editor-in-Chief for the *Digital Signal Processing Journal* between 2005 and 2011. He was a member of the Senior Editorial Board of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING (January 2012–December 2014). He was a two term elected member of the Sensor Array and Multichannel Processing Technical Committee (SAM-TC) of the IEEE Signal Processing Society between January 2005 and December 2010. He was the General Chairman for the 4th IEEE Workshop on Sensor Array and Multichannel Processing (SAM-2006), Waltham, MA, USA, July 2006. He was on the Technical Committee of the IEEE Radar Conference series in a myriad of roles (Track Chair, Session Chair, Special Session Organizer and Chair, Paper Selection Committee Member, Tutorial Lecturer). He was the Publicity Chair for the 1st IEEE International Conference on Waveform Diversity and Design, Edinburgh, U.K., November 2004. He was also the Technical Program Chairman for the 2014 IEEE Radar Conference. He is a member of the Radar Systems Panel in the IEEE-AES Society. He also serves on the conference subcommittee of the RSP.

He received the IEEE Warren White Radar Award in 2013, the 2013 Affiliate Societies Council Dayton Outstanding Scientist and Engineer Award, the 2007 IEEE Region 1 Award, the 2006 IEEE Boston Section Distinguished Member Award, and the 2005 IEEE-AESS Fred Nathanson Memorial Outstanding Young Radar Engineer Award. He also received the 2012 and 2005 Charles Ryan Basic Research Award from the Sensors Directorate of AFRL, in addition to more than 40 scientific achievement awards. He was elected as a Fellow of the IEEE in January 2006 with the citation "for contributions to mathematical techniques for radar space-time adaptive processing."