Adaptive Radar Signal Detection with Integrated Learning and Knowledge Exploitation

Hongbin Li
Department of Electrical and Computer Engineering
Stevens Institute of Technology, Hoboken, NJ 07030
USA

hli@stevens.edu

Muralidhar Rangaswamy
AFRL/RYAP
Bldg 620, 2241 Avionics Circle, WPAFB, OH 45433-7132
USA

muralidhar.rangaswamy@us.af.mil

ABSTRACT

We consider the problem of weak signal detection in strong disturbance with a subspace structure. Unlike conventional subspace detection techniques relying on the availability of a large amount of training data, we consider a knowledge-aided (KA) subspace detection approach for training limited scenarios by incorporating partial prior knowledge of the subspace. A unique feature of the proposed approach is that it can identify the missing subspace bases and recover the full subspace structure by using only the test signal, thus bypassing the need for training data. The proposed approach utilizes a Bayesian hierarchical model for knowledge representation. The model is integrated within a sparse Bayesian framework, which promotes parsimonious subspace representation of the observed data. A variational Bayesian inference algorithm is developed based on the proposed model to recover parameters and subspace structures associated with the disturbance, which are then brought into a generalized likelihood ratio test (GLRT) to perform signal detection. Numerical results are presented to illustrate the performance of the proposed subspace detector in comparison with several notable existing methods.

1.0 INTRODUCTION

Detecting a weak signal in strong disturbance (noise, interference, clutter, jamming, etc.) is a fundamental problem in radar, sonar, and many other applications. 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 [1] and references therein). One limitation of these detectors is that they require heavy training to ensure the accuracy of the covariance matrix estimate. Training requirement can be reduced by exploiting structures of the disturbance. A frequently considered one is when the disturbance (approximately) has a low-rank subspace structure. 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 [2], which is optimum in the sense that it is uniformly most powerful (UMP) invariant [3]. 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 [4], which belongs to a highly successful class of reduced-rank (a.k.a. partially adaptive) detectors [5]-[8].

Exploiting prior knowledge is another way to reduce training requirement. Techniques following this direction is usually dubbed *knowledge-aided* (KA) processing [9], where the prior knowledge often refers to a prior estimate of the disturbance covariance matrix. Most these techniques often employs a KA estimate of the disturbance covariance matrix via *colored loading*, which involves linearly combining the prior estimate with the sample covariance matrix. The weighting coefficients can be determined via a Bayesian approach by treating the covariance matrix as a random matrix assigned with a conjugate prior, e.g., the inverse Wishart distribution [10], [11]. Then, the posterior estimate takes a linear combining form, with the combining coefficients determined by a parameter that represents the *reliability* of the prior knowledge. Most 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 typical set-ups, the training size of a KA detector can be reduced roughly by a factor of two compared with its non-KA counterpart (e.g., [11]). While impressive, the reduction may be insufficient for radar detection in non-homogeneous environments, where training data is scarce since the clutter is location dependent and may vary significantly around the test cell.

In this paper, we consider subspace detection with partial prior knowledge of the disturbance subspace, and develop new partially adaptive KA methods for detection with extremely limited data (i.e., no training). 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 [12, Section 2.6.2.]. Such information translates to knowledge of some of the subspace basis vectors. To incorporate such partial prior knowledge for detection, we introduce a hierarchical model for knowledge representation, which is 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 [13] and is distinctively different from the Bayesian framework used for covariance matrix based KA processing [10], [11]. We develop a variational Bayesian inference algorithm to recover parameters and structures associated with the disturbance, which are then brought into a generalized likelihood ratio test (GLRT) to perform detection.

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. $|\mathbb{S}|$ 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, $N(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Phi})$ denotes the circularly symmetric complex Gaussian probability density function (PDF) of random vector $\mathbf{x}$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Phi}$:

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

2.0 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 [5]. 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 [14].

In many cases of practical interest, the disturbance $\mathbf{d}$ may have a low-rank subspace representation [3]:

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

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 a low-rank component in $\mathbf{d}$. As noted before, jamming and clutter in radar often have a low-rank subspace structure.

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

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

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 $\mathcal{H}$, 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., [15]-[17]).
The prior knowledge can be represented as a group of columns of $\mathcal{H}$. The knowledge is incomplete in that the subset may miss some columns that are necessary to represent $\mathcal{H}$. More precisely, let

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

denote the index set that indexes the columns of $\mathcal{H}$. If $\mathbb{T} \subset \mathbb{S}$ denotes the true index set for $\mathcal{H}$, the knowledge can be denoted by a subset $\mathbb{P} \subset \mathbb{T}$. We do not assume knowledge of $|\mathbb{T}|$, the cardinality of $\mathbb{T}$ and, equivalently, the rank of $\mathbf{H}$. The problem of interest is to identify missing bases in $\mathbb{P}$ via Bayesian learning, which are used jointly with the prior knowledge to solve the hypothesis testing (3).

3.0 KA SUBSPACE SIGNAL DETECTION

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

$$p_0(\boldsymbol{\beta}, \mathbf{H}, \sigma^2; \mathbf{y}) = N_c(\mathbf{y}; \mathbf{H}\boldsymbol{\beta}, \sigma^2\mathbf{I}), \tag{7}$$

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

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 the maximum likelihood estimate (MLE) of the amplitude κ conditioned on $\mathbf{H}$ and $\boldsymbol{\beta}$ is

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

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

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

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

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

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

$$\hat{\sigma}_0^2 = \frac{1}{N} \| \mathbf{y} - \mathbf{H}\boldsymbol{\beta} \|^2, \tag{13}$$

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

$$\{\widehat{\mathbf{H}}_0, \widehat{\boldsymbol{\beta}}_0\} = \arg\min_{\mathbf{H},\boldsymbol{\beta}} \| \mathbf{y} - \mathbf{H}\boldsymbol{\beta} \|^2 . \tag{14}$$

Clearly, (12) and (14) 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 $\phi \in \mathbb{R}^{L\times 1}$ of the interference sources in a beamforming setup, in which the problem becomes to jointly estimate ϕ and $\boldsymbol{\beta}$. Equivalently, we can use the sparse representation (5) and write the cost function in (12) and (14) in the following unified form:

$$\min_{\mathbf{x}} \| \mathbf{z} - \mathbf{A}\mathbf{x} \|^2, \tag{15}$$

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 (15) 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., [18]). However, these techniques are difficult to incorporate 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 $\{\widehat{\mathbf{H}}_1, \widehat{\boldsymbol{\beta}}_1\}$ and $\{\widehat{\mathbf{H}}_0, \widehat{\boldsymbol{\beta}}_0\}$ have been obtained, they can be substituted in (10), (11) and (13) 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

$$T_{\text{GLRT}} \triangleq \frac{\hat{\sigma}_0^2}{\hat{\sigma}_1^2} \underset{H_0}{\overset{H_1}{\gtrless}} \tau \tag{16}$$

where τ denotes a test threshold.

4.0 KNOWLEDGE-AIDED SUBSPACE RECOVERY

4.1 A Bayesian Knowledge Model

As shown in Section III, the subspace estimation problems (12) and (14) under both hypotheses can be cast in one framework based on the following measurement model:

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{e}, \tag{17}$$

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 $N_c(\mathbf{0}, \gamma^{-1}\mathbf{I})$, where γ denotes the inverse variance, which is also unknown. Sparse Bayesian learning (SBL) [13] is a popular approach that can be used to recover the sparse vector $\mathbf{x}$ from (17). 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 [13] 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

$$p(\mathbf{x} \mid \boldsymbol{\alpha}) = \prod_{m=1}^{M} N_c(x_m; 0, \alpha_m^{-1}), \tag{18}$$

where x_m denotes the m -th element of $\mathbf{x}$ and α_m^{-1} its inverse variance. Meanwhile, a Gamma prior is employed for the inverse variance vector $\boldsymbol{\alpha} \triangleq [\alpha_1, \ldots, \alpha_M]^{\mathrm{T}}$:

$$p(\boldsymbol{\alpha}) = \prod_{m=1}^{M} \mathrm{Gamma}(\alpha_m; a, b), \tag{19}$$

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) [13]. 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., $m \subset \mathbb{P}$, 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. Therefore, we propose to replace (19) with a fixed b by the following prior model to incorporate prior knowledge:

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

where b_m is allowed to vary with m. Specifically, we choose a relatively larger value for b_m, e.g., $b_m \in [0.1, 1]$, if $m \subset \mathbb{P}$, so that the prior is non-sparsifying over $\mathbb{P}$, while the other b_m remain small. The resulting model is referred to as the *subspace knowledge* (SK) model:

$$\text{SK:} \quad b_m = \begin{cases} \bar{b}, & m \in \mathbb{P}, \\ 10^{-6}, & m \in \mathbb{P}^c, \end{cases} \tag{21}$$

where $\bar{b} \in [0.1, 1]$ and $\mathbb{P}^c$ denotes the complement of $\mathbb{P}$.

Finally, the inverse variance γ of the noise in (17) 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 [13]:

$$p(\gamma) = \mathrm{Gamma}(\gamma; c, d), \tag{22}$$

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

4.2 A Bayesian Knowledge Model

Let $\boldsymbol{\theta} \triangleq \{\mathbf{x}, \boldsymbol{\alpha}, \gamma\}$ 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*. A standard Bayesian inference procedure would proceed to compute the posterior

$$p(\boldsymbol{\theta} \mid \mathbf{z}) = \frac{p(\mathbf{z} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\mathbf{z})}, \tag{23}$$

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 [19] and references therein). Specifically, we write $\boldsymbol{\theta}$ in a partitioned form: $\boldsymbol{\theta} \triangleq \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_K\}$, where $K = 3$ for the SK model with: $\boldsymbol{\theta}_1 \triangleq \mathbf{x}$, $\boldsymbol{\theta}_2 \triangleq \boldsymbol{\alpha}$, and $\boldsymbol{\theta}_3 \triangleq \gamma$. A popular approximation of the posterior $p(\boldsymbol{\theta}\,|\,\mathbf{z})$ is based on the mean field approximation (e.g, [19]): $p(\boldsymbol{\theta}\,|\,\mathbf{z}) \approx \prod_{k=1}^{K} q_k(\boldsymbol{\theta}_k)$, where the component PDF is given by

$$q_k(\boldsymbol{\theta}_k) = \frac{\exp\left(\left\langle \ln p(\mathbf{z},\boldsymbol{\theta})\right\rangle_{l\neq k}\right)}{\int \exp\left(\left\langle \ln p(\mathbf{z},\boldsymbol{\theta})\right\rangle_{l\neq k}\right)\mathrm{d}\boldsymbol{\theta}_k}, \tag{24}$$

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,

$$\left\langle \ln p(\mathbf{z},\boldsymbol{\theta})\right\rangle_{l\neq k} = \int \ln p(\mathbf{z},\boldsymbol{\theta}) \prod_{l\neq k} q_l(\boldsymbol{\theta}_l)\mathrm{d}\boldsymbol{\theta}_l. \tag{25}$$

Note that (24) 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 (24). 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 by integrating the knowledge model introduced in Section IV-A.

With the SK model, we have

$$p(\mathbf{z},\mathbf{x},\boldsymbol{\alpha},\gamma) = p(\mathbf{z}\,|\,\mathbf{x},\gamma)p(\mathbf{x}\,|\,\boldsymbol{\alpha})p(\boldsymbol{\alpha})p(\gamma) \tag{26}$$

where $p(\mathbf{z}\,|\,\mathbf{x},\gamma) = N_{\mathrm{c}}(\mathbf{z};\mathbf{A}\mathbf{x},\gamma^{-1}\mathbf{I})$, $p(\mathbf{x}\,|\,\boldsymbol{\alpha})$ is given by (18), $p(\boldsymbol{\alpha})$ by (20), and $p(\gamma)$ by (22), respectively. It can be shown the factor posteriors $q_x(\mathbf{x})$, $q_\alpha(\boldsymbol{\alpha})$, and $q_\gamma(\gamma)$ are respectively Gaussian, product Gamma, and Gamma distributions: $q_x(\mathbf{x}) = \mathcal{N}_{\mathrm{c}}(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Phi})$, $q_\alpha(\boldsymbol{\alpha}) = \prod_{m=1}^{M} \mathrm{Gamma}(\alpha_m; a+1, \tilde{b}_m)$, $q_\gamma(\gamma) = \mathrm{Gamma}\left(\gamma; c+N, \tilde{d}\right)$, where $\boldsymbol{\mu} = \langle\gamma\rangle\boldsymbol{\Phi}\mathbf{A}^{\mathrm{H}}\mathbf{z}$, $\boldsymbol{\Phi} = \left(\langle\gamma\rangle\mathbf{A}^{\mathrm{H}}\mathbf{A} + \langle\mathbf{D}\rangle\right)^{-1}$, $\tilde{b}_m = b_m + \mu_m^2 + \Phi_{m,m}$, $m = 1,\ldots,M$, $\tilde{d} = d + \|\mathbf{z} - \mathbf{A}\boldsymbol{\mu}\|^2 + \mathrm{tr}\left\{\mathbf{A}\boldsymbol{\Phi}\mathbf{A}^{\mathrm{H}}\right\}$.

As such, the update of the factor posteriors boil down to the update of these parameters. We summarize the resulting estimator in Algorithm 1.

Algorithm 1 - SK-based subspace estimator

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

repeat

1) Set $t = t + 1$.

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

$$\mathbf{\Phi}^{(t)} = \left(\langle \gamma^{(t-1)} \rangle \mathbf{A}^{\mathrm{H}} \mathbf{A} + \langle \mathbf{D}^{(t-1)} \rangle \right)^{-1}, \tag{27}$$

$$\boldsymbol{\mu}^{(t)} = \langle \gamma^{(t-1)} \rangle \mathbf{\Phi}^{(t)} \mathbf{A}^{\mathrm{H}} \mathbf{z}. \tag{28}$$

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

$$\tilde{b}_m^{(t)} = b_m + \left(\mu_m^{(t)} \right)^2 + \Phi_{m,m}^{(t)}, \tag{29}$$

$$\langle \alpha_m^{(t)} \rangle = (a+1) / \tilde{b}_m^{(t)}, \quad m = 1, \ldots, M, \tag{30}$$

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

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

$$\tilde{d}^{(t)} = d + \| \mathbf{z} - \mathbf{A}\boldsymbol{\mu}^{(t)} \|^2 + \text{tr}\{\mathbf{A}\mathbf{\Phi}^{(t)}\mathbf{A}^{\mathrm{H}}\}, \tag{31}$$

$$\langle \gamma^{(t)} \rangle = (c + N) / \tilde{d}^{(t)}. \tag{32}$$

Until convergence

Convergence is reached when the difference of some parameter estimates over two consecutive iterations is sufficiently small. We use $\boldsymbol{\mu}^{(t)}$, the posterior mean of $q_x(\mathbf{x})$ and also an estimate of the sparse vector $\mathbf{x}$. 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 $\boldsymbol{\mu}^{(t)}$.

5.0 NUMERICAL RESULTS

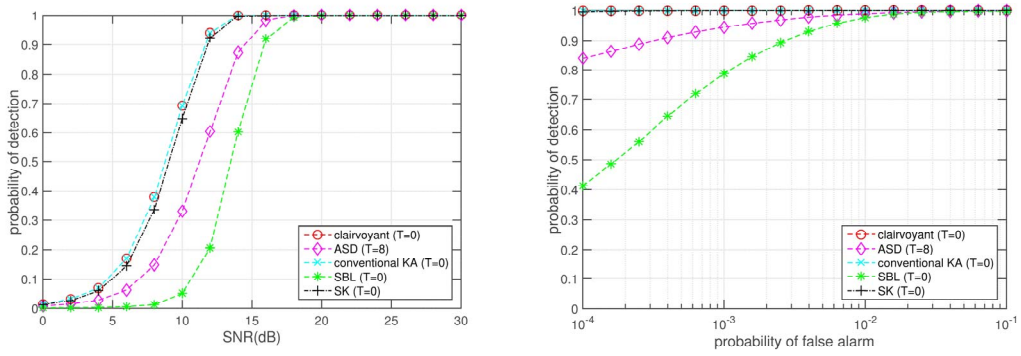
We now present simulation results to illustrate the performance of the proposed detector. The disturbance $\mathbf{d}$ has a subspace structure as in (4), where $\mathbf{H}$ is formed by L Fourier vectors with frequencies centered around the zero frequency, i.e., a lowpass narrowband interference. Specifically, let $\mathcal{H}'$ denotes an $N \times M'$ discrete Fourier transform (DFT) matrix, and the subspace matrix $\mathbf{H}$ consists of the following columns of $\mathcal{H}'$: $\left\{1, 2, \ldots, \left\lfloor \frac{L+1}{2} \right\rfloor, \cdots, M' - \left\lceil \frac{L-1}{2} \right\rceil + 1, M' - \left\lceil \frac{L-1}{2} \right\rceil + 2, \ldots, M'\right\}$, 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. 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 $\mathcal{H}'$, which cover the mainlobe of the target response, are removed to form the \$N\\times M\$ dictionary matrix $\mathcal{H}$ [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 (21) is referred to as the *SK* detector. We compare with 4 other known detectors as benchmarks, namely the *clairvoyant* subspace detector of [2], 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*), 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. The ASD is also identical to the clairvoyant detector but replaces $\mathbf{H}$ with an estimate consisting of the L principal eigenvectors of the sample covariance matrix. The latter is constructed from $T = 8$ target-free homogeneous training signals, which share the same subspace matrix $\mathbf{H}$. Note $T = 0$ for the other detectors, which require no training. For the KA detectors, including the conventional KA and SK, we consider two cases for the prior knowledge $\mathbb{P}$ including:

- **Case 1:** $\mathbb{P} = \mathbb{T}$. The prior knowledge is perfect;
- **Case 2:** $\mathbb{P} \subset \mathbb{T}$. The prior knowledge for each simulation trial contains $|\mathbb{P}| = 4$ randomly selected indices of $\mathbb{T}$.

5.1 Case 1: Full Knowledge

Case 1 is of interest to show how the proposed SK behaves in the presence of perfect prior knowledge. Figure 1(a) depicts the probability of detection P_d versus the SNR for the various detectors, where $\text{INR} = 30$ dB and the probability of false alarm $P_{\text{f}} = 10^{-3}$. The conventional KA in the current case reduces to the clairvoyant detector. The performance of the proposed SK is also nearly identical to the clairvoyant, manifesting the benefit of the prior knowledge. The near optimality also indicates that SK rarely rejects correct basis vectors in $\mathbb{P}$ or adds erroneous basses in $\mathbb{P}^{\text{c}}$. Meanwhile, without using the prior knowledge $\mathbb{P}$, SBL is the worst detector. 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. Figure 1(b) shows P_d versus P_{f}, i.e., the receiver operating characteristic (ROC) curve, where $\text{SNR} = 15$ dB and $\text{INR} = 30$ dB. The relations among the various detectors are similar to what were observed before.

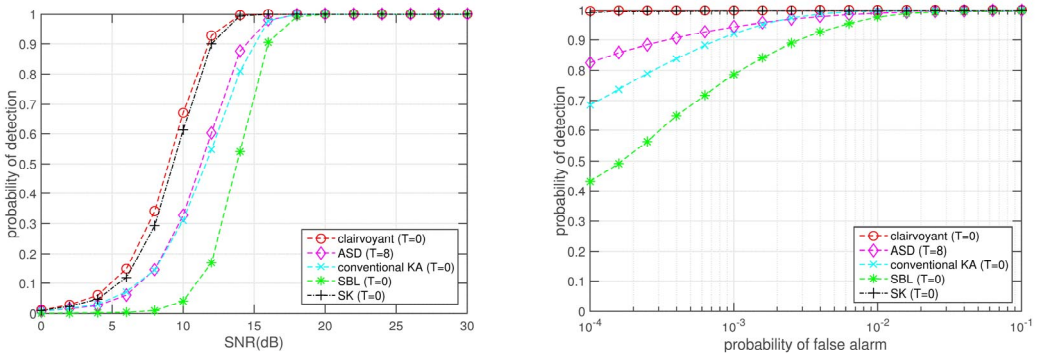


(a) (b)

Figure 1: Case 1 results. (a) P_d vs. SNR with INR = 30 dB and $P_f = 10^{-3}$. (b)ROC curve with SNR = 15 dB and INR = 30 dB.

5.2 Case 2: Partial Knowledge

We next consider a partial knowledge case with $\mathbb{P} \subset \mathbb{T}$ and $|\mathbb{P}| = 4$, that is, the prior knowledge $\mathbb{P}$ contains 4 randomly selected bases from $\mathbb{T}$ in each trial. The results are shown in Figure 2. It is seen that the SK detector is the closest to the clairvoyant and significantly outperforms the conventional KA. This is because SK is able to identify missing bases in $\mathbb{P}^c$ and therefore can better reject the interference than the conventional KA detector.



(a) (b)

Figure 2: Case 2 (partial knowledge) results. (a) P_d vs. SNR with INR = 30 dB and $P_f = 10^{-3}$. (b)ROC curve with SNR = 15 dB and INR = 30 dB.

6.0 CONCLUSIONS

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 feature is that the proposed approach accounts for the fact that the prior knowledge available in practice may be incomplete. To address such uncertainties, we introduced a Bayesian hierarchical model for knowledge representation. The model was integrated in a Bayesian learning framework, and a variational inference algorithm based on simple iteration steps was developed to solve the associated inference problem. Simulation results, which cover two cases of full and, respectively, partial prior knowledge, demonstrate that the proposed KA detectors can benefit from prior knowledge and significantly outperform conventional detectors when the available prior knowledge is incomplete.

REFERENCE

[1] A. De Maio, “Rao test for adaptive detection in Gaussian interference with unknown covariance matrix,” IEEE Transactions on Signal Processing, vol. 55, no. 7, pp. 3577–3584, 2007.

[2] P. Wang, H. Li, and B. Himed, “Moving target detection using distributed MIMO radar in clutter with nonhomogeneous power,” IEEE Transactions on Signal Processing, vol. 59, no. 10, pp. 4809–4820, Oct. 2011.

[3] L. L. Scharf and B. Friedlander, "Matched subspace detectors," IEEE Transactions on Signal Processing, vol. 42, no. 8, pp. 2146–2157, August 1994.

[4] A. Haimovich, "The eigencanceler: Adaptive radar by eigenanalysis methods," IEEE Transactions on Aerospace and Electronic Systems, vol. 32, no. 2, pp. 532–542, April 1996.

[5] J. R. Guerci, Space-Time Adaptive Processing for Radar. Artech House, 2014.

[6] J. S. Goldstein, I. S. Reed, and L. L. Scharf, "A multistage representation of the Wiener filter based on orthogonal projections," IEEE Transactions on Information Theory, vol. 44, no. 7, pp. 2943–2959, November 1998.

[7] C. Jiang, H. Li, and M. Rangaswamy, "On the conjugate gradient matched filter," IEEE Transactions on Signal Processing, vol. 60, no. 5, pp. 2660–2666, May 2012.

[8] Z. Chen, H. Li, and M. Rangaswamy, "Conjugate gradient adaptive matched filter," IEEE Transactions on Aerospace and Electronic Systems, vol. 51, no. 1, pp. 178–191, 2015.

[9] J. R. Guerci, "Cognitive radar: a knowledge-aided fully adaptive approach," in Radar Conference, 2010 IEEE. IEEE, 2010, pp. 1365–1370.

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

[11] A. De Maio, A. Farina, and G. Foglia, "Knowledge-aided Bayesian radar detectors & their application to live data," IEEE Transactions on Aerospace and Electronic Systems, vol. 46, no. 1, pp. 170–183, February 2010.

[12] J. Ward, "Space-time adaptive processing for airborne radar," Lincoln Laboratory, MIT, Technical Report 1015, December 1994.

[13] M. Tipping, "Sparse Bayesian learning and the relevance vector machine," Journal of Machine Learning Research, vol. 1, pp. 211–244, 2001.

[14] M. A. Richards, Fundamentals of Radar Signal Processing. McGrawHill, 2005.

[15] E. J. Candes and C. Fernandez-Granda, "Super-resolution from noisy data," Journal of Fourier Analysis and Applications, vol. 19, no. 6, pp. 1229–1254, 2013.

[16] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," IEEE Transactions on Information Theory, vol. 59, no. 11, pp. 7465–7490, Nov. 2013.

[17] J. Fang, F. Wang, Y. Shen, H. Li, and R. Blum, "Super-resolution compressed sensing for line spectral estimation: an iterative reweighted approach," IEEE Transactions on Signal Processing, vol. 64, no. 18, pp. 4649–4662, September 2016.

[18] S. Foucart and H. Rauhut, A Mathematical Introduction to Compressive Sensing. Springer, 2013.

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