Knowledge-Aided Parametric Adaptive Matched Filter With Automatic Combining for Covariance Estimation

Pu Wang, *Member, IEEE*, Zhe Wang, *Student Member, IEEE*, Hongbin Li, *Senior Member, IEEE*, and Braham Himed, *Fellow, IEEE*

***Abstract*—In this paper, a knowledge-aided parametric adaptive matched filter (KA-PAMF) is proposed that utilizing both observations (including the test and training signals) and *a priori* knowledge of the spatial covariance matrix. Unlike existing KA-PAMF methods, the proposed KA-PAMF is able to automatically adjust the combining weight of *a priori* covariance matrix, thus gaining enhanced robustness against uncertainty in the prior knowledge. Meanwhile, the proposed KA-PAMF is significantly more efficient than its KA nonparametric counterparts when the amount of training signals is limited. One distinct feature of the proposed KA-PAMF is the inclusion of both the test and training signals for automatic determination of the combining weights for the prior spatial covariance matrix and observations. Numerical results are presented to demonstrate the effectiveness of the proposed KA-PAMF, especially in the limited training scenarios.**

***Index Terms*—Knowledge-aided processing, multi-channel autoregressive process, parametric adaptive matched filter, space-time adaptive processing (STAP).**

I. Introduction

TRADITIONAL space-time adaptive processing (STAP) methods such as the Kelly's generalized likelihood ratio test (GLRT) [1] and the adaptive matched filter (AMF) [2] usually require excessive homogeneous training (secondary) data to obtain an accurate estimate of the disturbance covariance matrix for adaptive detection of targets. For example, it is known that for these methods at least $K \geq JN$ homogeneous training signals are required for a full-rank covariance matrix estimator, where J is the number of channels and N is the number of temporal observations.

Knowledge-aided detectors have been introduced to reduce the demanding need of training signals by fusing some prior knowledge in the estimation of the disturbance covariance matrix [3]. One approach toward this goal is based on the Bayesian framework, which embeds the *a priori* knowledge via a prior distribution of the disturbance covariance matrix [4]–[12]. Another approach was based on the regularized method [13]–[16], which usually linearly shrinks the eigenvalues of the sample covariance matrix towards a targeted covariance matrix, e.g., the identity matrix up to a scaling factor [13], a diagonal matrix consisting of the diagonal entries of the sample covariance matrix [14], or the *a priori* covariance matrix [16]. Interestingly, both approaches result in a colored loading form between the *a priori* matrix and the sample covariance matrix. While the weights in the Bayesian approach are determined by the hyper-parameters of the statistical model, the regularized method uses the (training) signals to determine the amount of regularization.

The regularized method has been considered for STAP detection, which employs the loaded covariance matrix for signal whitening and test statistic calculation. Specifically, [16] introduces the knowledge-aided AMF (KA-AMF) which first linearly combines the sample covariance matrix and the *a priori* covariance matrix [13], which is then used in the conventional AMF for adaptive detection. The linear combining weights are determined from the training signals. Results obtained with the high-fidelity site-specific radar simulation (KASSPER) data [17] show that with $J = 11$ channels and $N = 32$ pulses, the proposed KA-AMF offers good detection performance by using $K = 50$ training signals. Still, it may be difficult to obtain $K = 50$ homogeneous training signals in a non-homogeneous environment, where a more efficient solution with less training data is desirable. Moreover, the computational complexity of the KA-AMF is still high, since it needs to compute the inverse of the $JN \times JN$ covariance matrix.

In this paper, we aim to address both issues of limited homogeneous training signals and the complexity by extending the parametric adaptive matched filter (PAMF) [18], [19] and integrate knowledge-aided processing. As shown with numerous simulated and measured STAP datasets [19], [20], the parametric framework using a multichannel auto-regressive (AR) process can effectively and efficiently capture the correlation structure of the disturbance in STAP. Furthermore, we develop

Manuscript received November 19, 2013; revised March 27, 2014 and June 16, 2014; accepted July 08, 2014. Date of publication July 11, 2014; date of current version August 14, 2014. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ami Wiesel. The work of Z. Wang and H. Li was supported in part by a subcontract with Dynetics, Inc. for research sponsored by the Air Force Research Laboratory (AFRL) under Contract FA8650-08-D-1303. This work has been performed while P. Wang was with the Stevens Institute of Technology.

P. Wang was with the Stevens Institute of Technology, Hoboken, NJ 07030 USA. He is now with the Schlumberger-Doll Research, Cambridge, MA 01239 USA (e-mail: pwang@ieee.org).

Z. Wang and H. Li are with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030 USA (e-mail: zwang23@stevens.edu; Hongbin.Li@stevens.edu, hongbin.li@stevens.edu).

B. Himed is with RF Technology Branch, Air Force Research Laboratory, AFRL/RYMD, WPAFB, Dayton, OH 45433 USA (e-mail: braham.himed@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.2014.2338838

a regularized method which automatically determines the combining weights *jointly* from the test signal and training signals. Our scheme is different from other regularized methods, such as the KA-AMF [16], which uses only the training signals to determine the combining weights. It appears that the inclusion of the test signal for weight calculation is critical to achieving a robust performance in scenarios where the number of training data is limited. Our proposed knowledge-aided PAMF with automatic combining (referred to as the KA-AC-PAMF hereafter) is derived in a three-step approach. First, conditioned on the given AR temporal correlation matrices, a partially adaptive detector is derived according to the GLRT principle which yields an estimate of the spatial covariance matrix. Then, the estimate of the spatial covariance matrix is linearly combined with the prior knowledge in an adaptive way from the test and training signals. Finally, the fully adaptive KA-AC-PAMF is obtained by replacing the AR temporal correlation matrices in the partially adaptive detector by its maximum likelihood (ML) estimate.

There are previous efforts proposed to extend the parametric detectors with knowledge-aided processing. In [21], [22], the knowledge-aided PAMF was derived within the Bayesian framework. The resulting Bayesian PAMF (B-PAMF) relies on the hyper-parameters which may not be known in advance. Hence, the B-PAMF may be vulnerable to uncertainties in the prior knowledge. In Section IV, the sensitivity of the B-PAMF to the prior uncertainty is numerically demonstrated, while the proposed KA-AC-PAMF exhibits enhanced robustness to such prior uncertainty.

The remainder of this paper is organized as follows. Section II describes the multichannel AR model of the parametric framework. The proposed KA-AC-PAMF is derived in Section III, including details on using both the test and training signals to determine the linear combining weights. Section IV provides simulation results with the synthetic AR and KASSPER datasets. Finally, conclusions are drawn in Section V.

II. Signal Model

Consider the problem of detecting a *known* multi-channel signal with *unknown* amplitude in the presence of spatially and temporally correlated disturbance (e.g., [23]):

$$\begin{aligned} H_0: &\quad \mathbf{x}_0(n) = \mathbf{d}_0(n), \\ H_1: &\quad \mathbf{x}_0(n) = \alpha\mathbf{s}(n) + \mathbf{d}_0(n), \\ &\quad n = 0, 1, \ldots, N-1, \end{aligned} \tag{1}$$

where all vectors are of dimension $J\times 1$ obtained from J spatial channels/receivers, and N is the number of temporal observations/snapshots. The subindex of $\mathbf{x}_0(n)$ is referred to the range bin of interest, and $\{\mathbf{x}_0(n)\}_{n=1}^{N}$ forms the test signal from J receivers and N pulses. The steering vector $\{\mathbf{s}(n)\}_{n=1}^{N}$ takes into account of the array geometry with spatial frequency ω_s and the Doppler frequency ω_d. For a uniformly equi-spaced linear array, the (normalized) steering vector is given as

$$\mathbf{s}(n) = \frac{1}{\sqrt{JN}} e^{j\omega_d(n-1)}[1, e^{j\omega_s}, \ldots, e^{j\omega_s(J-1)}]^T. \tag{2}$$

In addition, α denotes the *unknown*, deterministic and complex-valued signal amplitude, and $\mathbf{d}_0(n)$ is the disturbance signal that is correlated in space and time. Besides the test signal $\mathbf{x}_0(n)$, there may be a set of *target-free* training signals $\mathbf{x}_k(n)$:

$$\mathbf{x}_k(n) = \mathbf{d}_k(n), \quad k = 1, \ldots, K. \tag{3}$$

Denote the $JN\times 1$ space-time vectors of the steering vector, disturbance signals, and received signals as

$$\begin{aligned} \mathbf{s} &\triangleq \left[\mathbf{s}^T(0), \mathbf{s}^T(1), \ldots, \mathbf{s}^T(N-1)\right]^T, \\ \mathbf{d}_k &\triangleq \left[\mathbf{d}_k^T(0), \mathbf{d}_k^T(1), \ldots, \mathbf{d}_k^T(N-1)\right]^T, \\ \mathbf{x}_k &\triangleq \left[\mathbf{x}_k^T(0), \mathbf{x}_k^T(1), \ldots, \mathbf{x}_k^T(N-1)\right]^T. \end{aligned} \tag{4}$$

The hypothesis testing problem in (1) can be rewritten as

$$\begin{aligned} H_0: &\quad \mathbf{x}_0 = \mathbf{d}_0, \\ &\quad \mathbf{x}_k = \mathbf{d}_k, \quad k = 1, \ldots, K \\ H_1: &\quad \mathbf{x}_0 = \alpha\mathbf{s} + \mathbf{d}_0, \\ &\quad \mathbf{x}_k = \mathbf{d}_k, \quad k = 1, \ldots, K. \end{aligned} \tag{5}$$

It is assumed that the disturbance signals $\mathbf{d}_k$, $k = 0, 1, \ldots, K$, are independent and identically distributed (i.i.d.) with the complex Gaussian distribution $\mathbf{d}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{R})$, where $\mathbf{R}$ is the unknown space-time covariance matrix [1]. The parametric framework further assumes that the disturbance signals $\{\mathbf{d}_k\}_{k=0}^{K}$ in the test and training signals follow the assumption below [19]:

- **AS — Multi-Channel AR Model:** The disturbance signals $\mathbf{d}_k(n)$, $k = 0, \ldots, K$, in the test and training signals are modeled as J-channel $\mathrm{AR}(P)$ processes with model order P:

$$\mathbf{d}_k(n) = -\sum_{i=1}^{P} \mathbf{A}^H(i)\mathbf{d}_k(n-i) + \boldsymbol{\varepsilon}_k(n), \tag{6}$$

  where $\{\mathbf{A}(i)\}_{i=1}^{P}$ denote the *unknown* $J\times J$ AR coefficient matrices, $\mathbf{A}^H$ denotes the conjugate transpose of $\mathbf{A}$, $\boldsymbol{\varepsilon}_k(n)$ denote the $J\times 1$ spatial noise vectors that are temporally white but spatially colored Gaussian noise: $\{\boldsymbol{\varepsilon}_k(n)\}_{k=0}^{K} \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q})$, and $\mathbf{Q}$ denotes the *unknown* $J \times J$ spatial covariance matrix.

In other words, the disturbance covariance matrix $\mathbf{R}$ is parametrized in **AS** with P AR coefficient matrices $\mathbf{A}(p)$ and the spatial covariance matrix $\mathbf{Q}$. In many cases, it is possible to have some *a priori* knowledge on $\mathbf{R}$ which can be utilized for improved detection performance. Such knowledge can be obtained from previously acquired database, e.g., digital terrain maps, synthetic aperture radar (SAR) images, as well as from real-time information including the transmit/receive array configurations, beampatterns, etc. [3]. In this paper, we only assume *a priori* knowledge of the spatial covariance matrix, denoted as $\bar{\mathbf{Q}}$, which can be obtained from several different ways [22]: one can generate $\bar{\mathbf{Q}}$ directly from prior spatial information such as platform height, or performing a block Lower-triangle-Diagonal-Upper-triangle (LDU) matrix decomposition of spatial-temporal covariance $\bar{\mathbf{R}}$, or learn $\bar{\mathbf{Q}}$ by solving a multi-channel levinson algorithm by using $\bar{\mathbf{R}}$ as the covariance matrix for the observation.

With $\bar{\mathbf{Q}}$ and **AS**, the problem of interest is to develop a knowledge-aided parametric detector for the binary hypothesis testing problem (5).

III. Proposed Approach

In this section, we consider a three-step approach to develop the KA-AC-PAMF detector. First, assuming that the AR coefficient matrix

$$\mathbf{A} \triangleq [\mathbf{A}^T(1), \mathbf{A}^T(2), \ldots, \mathbf{A}^T(P)]^T, \tag{7}$$

is known, a partially adaptive PAMF is derived by finding the ML estimates of unknown parameters $\mathbf{Q}$ and α that maximize the joint likelihood function of the test signal $\mathbf{x}_0$ and the training signals $\mathbf{x}_k$. Then, the ML estimate of $\mathbf{Q}$ is regularized with the prior $\bar{\mathbf{Q}}$ according to the minimum mean squared error (MMSE) criterion. Third, the AR coefficient matrix $\mathbf{A}$ in the partially adaptive PAMF is replaced by its ML estimate of $\mathbf{A}$ leading to the fully adaptive KA-AC-PAMF.

A. Partially Adaptive PAMF

Assuming **AS** with a known $\mathbf{A}$, the partially adaptive PAMF takes the form of a likelihood ratio test

$$T = \frac{\max_{\alpha, \mathbf{Q}} f_1(\alpha, \mathbf{Q})}{\max_{\mathbf{Q}} f_0(\mathbf{Q})}, \tag{8}$$

where $f_i(\alpha, \mathbf{Q}), \ \ i = 0, 1$ ($\mathbf{Q} \in^{J \times J}$) is the joint asymptotic ($N \gg P$) likelihood function of $\mathbf{x}_0$ and $\mathbf{x}_k$ under $H_i, \ i = 0, 1$,

$$f_i(\alpha, \mathbf{Q}) = \left[\frac{1}{\pi^J |\mathbf{Q}|} e^{-\operatorname{tr}(\mathbf{Q}^{-1}\mathbf{\Gamma}(\alpha))} \right]^{(K+1)(N-P)}, \tag{9}$$

with $\alpha = 0$ when $i = 0$,

$$\mathbf{\Gamma}(\alpha) = \left(\tilde{\mathbf{X}}_0 - \alpha \tilde{\mathbf{S}}\right)\left(\tilde{\mathbf{X}}_0 - \alpha \tilde{\mathbf{S}}\right)^H + \sum_{k=1}^{K} \tilde{\mathbf{X}}_k \tilde{\mathbf{X}}_k^H, \tag{10}$$

and

$$\tilde{\mathbf{S}} = [\tilde{\mathbf{s}}(P), \ldots, \tilde{\mathbf{s}}(N-1)] \in^{J \times (N-P)}, \tag{11}$$

$$\tilde{\mathbf{X}}_k = [\tilde{\mathbf{x}}_k(P), \ldots, \tilde{\mathbf{x}}_k(N-1)] \in^{J \times (N-P)}, \tag{12}$$

$$\tilde{\mathbf{x}}_k(n) = \mathbf{x}_k(n) + \mathbf{A}^H \mathbf{y}_k(n), \tag{13}$$

$$\tilde{\mathbf{s}}(n) = \mathbf{s}(n) + \mathbf{A}^H \mathbf{t}(n), \tag{14}$$

$$\mathbf{y}_k(n) = \left[\mathbf{x}_k^T(n-1), \ldots, \mathbf{x}_k^T(n-P)\right]^T, \tag{15}$$

$$\mathbf{t}(n) = \left[\mathbf{s}^T(n-1), \ldots, \mathbf{s}^T(n-P)\right]^T, \tag{16}$$

for $k = 0, 1, \ldots, K$. Note that (13) and (14) perform the temporal whitening for $\mathbf{x}_k(n)$ and $\mathbf{s}(n)$. The ML estimate of the unknown parameter $\mathbf{Q}$ can be obtained by taking the derivative of (9) and equating it to zero

$$\hat{\mathbf{Q}}_{1,\mathrm{ML}}(\alpha) = \mathbf{\Gamma}(\alpha), \quad \hat{\mathbf{Q}}_{0,\mathrm{ML}} = \mathbf{\Gamma}(\mathbf{0}). \tag{17}$$

Taking it back to (9), the partially adaptive PAMF is equivalent to

$$T \propto \frac{|\mathbf{\Gamma}(\mathbf{0})|}{\left|\max_{\alpha} \mathbf{\Gamma}(\alpha)\right|} = \frac{|\mathbf{\Gamma}(\mathbf{0})|}{|\mathbf{\Gamma}(\hat{\alpha}_{\mathrm{ML}})|}. \tag{18}$$

The (asymptotic) ML estimate of α under H_1 can be obtained asymptotically under a first-order approximation as [22]

$$\hat{\alpha}_{\mathrm{AML}} = \frac{\operatorname{tr}\left(\tilde{\mathbf{S}}^H \mathbf{\Psi}^{-1} \tilde{\mathbf{X}}_0\right)}{\operatorname{tr}\left(\tilde{\mathbf{S}}^H \mathbf{\Psi}^{-1} \tilde{\mathbf{S}}\right)}, \tag{19}$$

where

$$\mathbf{\Psi} = \frac{\tilde{\mathbf{X}}_0 \mathbf{P}^{\perp} \tilde{\mathbf{X}}_0^H + \sum_{k=1}^{K} \tilde{\mathbf{X}}_k \tilde{\mathbf{X}}_k^H}{L}, \tag{20}$$

with $L = K(N-P) + N - P - 1$ and $\mathbf{P}^{\perp} = \mathbf{I} - \mathbf{P} = \mathbf{I} - \tilde{\mathbf{S}}^H (\tilde{\mathbf{S}}^H)^{\dagger}$ denoting the projection matrix projecting to the orthogonal complement of the range of $\tilde{\mathbf{S}}^H$.

Finally, taking $\hat{\alpha}_{\mathrm{AML}}$ back to (18), we have

$$T = \frac{\left|\operatorname{tr}\left\{\tilde{\mathbf{S}}^H \mathbf{\Psi}^{-1} \tilde{\mathbf{X}}_0\right\}\right|^2}{\operatorname{tr}\left\{\tilde{\mathbf{S}}^H \mathbf{\Psi}^{-1} \tilde{\mathbf{S}}\right\}} = \frac{\left|\sum_{n=P}^{N-1} \tilde{\mathbf{s}}^H(n) \mathbf{\Psi}^{-1} \tilde{\mathbf{x}}_0(n)\right|^2}{\sum_{n=P}^{N-1} \tilde{\mathbf{s}}^H(n) \mathbf{\Psi}^{-1} \tilde{\mathbf{s}}(n)}. \tag{21}$$

It is clear that the partially adaptive PAMF first performs the temporal whitening process to obtain the test and training signals $\{\tilde{\mathbf{x}}_k(n)\}_{k=0}^K$ and the steering vector $\tilde{\mathbf{s}}(n)$ via (13) and (14), and then performs the spatial whitening process in (21) with $\mathbf{\Psi}$ of (20). More importantly, the spatial whitening matrix $\mathbf{\Psi}$ of (20) includes contribution from the temporally whitened test signal $\tilde{\mathbf{x}}_0(n)$ (after projecting onto the orthogonal complement of the range space of $\tilde{\mathbf{S}}^H$) and the temporally whitened training signals $\tilde{\mathbf{x}}_k(n)$.

B. Unbiasedness of $\mathbf{\Psi}$

Before proceeding to address the linear combination of $\mathbf{\Psi}$ and the prior $\bar{\mathbf{Q}}$, we show in the following that $\mathbf{\Psi}$ of (20) is an unbiased estimator of $\mathbf{Q}$ under the two hypotheses. This property will be utilized in the next section to determine the optimal combining weights according to the minimum mean-squared error (MMSE) criterion.

Proposition: Given the signal model of $\{\mathbf{x}_k(n)\}_{k=0}^K$ and $\mathbf{s}(n)$ and assuming the multichannel AR model in **AS**, the estimate $\mathbf{\Psi}$ of (20) is an unbiased estimate of $\mathbf{Q}$ under the two hypotheses

$$E\{\mathbf{\Psi}\} = \mathbf{Q}, \quad \text{under } H_0 \text{ and } H_1. \tag{22}$$

Proof: From **AS** the temporally whitened test signal $\tilde{\mathbf{x}}_0(n)$ is statistically equivalent to the spatial noise vector $\boldsymbol{\epsilon}_0(n)$ plus $\alpha \tilde{\mathbf{s}}(n)$, while the training signals $\tilde{\mathbf{x}}_k(n)$ are statistically equivalent to $\boldsymbol{\epsilon}_k(n)$:

$$\begin{aligned}\tilde{\mathbf{x}}_0(n) &= \alpha_i \tilde{\mathbf{s}}(n) + \boldsymbol{\epsilon}_0(n) \sim \mathcal{CN}(\alpha_i \tilde{\mathbf{s}}(n), \mathbf{Q}),\\ &\quad i = 0, 1,\\ \tilde{\mathbf{x}}_k(n) &= \boldsymbol{\epsilon}_k(n) \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q}),\\ &\quad k = 1, \ldots, K,\end{aligned} \tag{23}$$

where $\alpha_0 = 0$ under H_0 and $\alpha_1 = \alpha$ under H_1. Therefore, we have

$$\tilde{\mathbf{X}}_0 = \alpha_i \tilde{\mathbf{S}} + \mathbf{E}_0, \quad \tilde{\mathbf{X}}_k = \mathbf{E}_k, \tag{24}$$

where

$$\mathbf{E}_k = [\boldsymbol{\epsilon}_k(P), \ldots, \boldsymbol{\epsilon}_k(N-1)], \quad k = 0, 1, \ldots, K \tag{25}$$

with columns $\boldsymbol{\epsilon}_k(n)$ distributed as i.i.d. complex Gaussian vectors with zero mean and covariance matrix $\mathbf{Q}$. Thus we have $E\{\boldsymbol{\epsilon}_k(i)\boldsymbol{\epsilon}_k(l)^H\} = 0 \in^{J\times J}$ for all $i \neq l$. Due to the orthogonal projection, it is straightforward to show that

$$\mathbf{P}^\perp \tilde{\mathbf{X}}_0 = \alpha_i(\tilde{\mathbf{S}} - \mathbf{P}\tilde{\mathbf{S}}) + \mathbf{P}^\perp \mathbf{E}_0 = \mathbf{P}^\perp \mathbf{E}_0, \tag{26}$$

which holds under the two hypotheses H_0 and H_1. As result, we have

$$L\boldsymbol{\Psi} = \mathbf{E}_0 \mathbf{P}^\perp \mathbf{E}_0^H + \mathbf{E}_k \mathbf{E}_k^H, \tag{27}$$

under H_0 and H_1. Denoting $\mathbf{P}^\perp_{i,l}$ as the element of i-th row and l-th column in matrix $\mathbf{P}^\perp$, then the proof is completed by taking the expectation on $\boldsymbol{\Psi}$:

$$\begin{aligned} E\{L\boldsymbol{\Psi}\} &= E\{\mathbf{E}_0\mathbf{P}^\perp\mathbf{E}_0^H\} + E\left\{\sum_{k=1}^{K}\mathbf{E}_k\mathbf{E}_k^H\right\} \\ &= \sum_{i=1}^{N-P}\sum_{l=1}^{N-P}\mathbf{P}^\perp_{i,l}E\{\boldsymbol{\epsilon}_0(i-1+P)\boldsymbol{\epsilon}_0(l-1+P)^H\} \\ &\quad + \sum_{k=1}^{K}\sum_{i=1}^{N-P}\sum_{l=1}^{N-P}E\{\boldsymbol{\epsilon}_k(i)\boldsymbol{\epsilon}_k(l)^H\} \\ &= \sum_{l=1}^{N-P}\mathbf{P}^\perp_{l,l}E\{\boldsymbol{\epsilon}_0(l-1+P)\boldsymbol{\epsilon}_0^H(l-1+P)\} \\ &\quad + \sum_{k=1}^{K}\sum_{n=P}^{N-1}E\{\boldsymbol{\epsilon}_k(n)\boldsymbol{\epsilon}_k^H(n)\} \\ &= [\mathrm{tr}\{\mathbf{P}^\perp\} + K(N-P)]\mathbf{Q} \\ &= [(N-P-1) + K(N-P)]\mathbf{Q}, \end{aligned} \tag{28}$$

where the second equality is due to the i.i.d. columns of $\mathbf{E}_k, \quad k = 0, 1, \ldots, K$. Note that since $\tilde{\mathbf{S}}$ is a rank-1 matrix. The rank of its projection matrix $\mathbf{P}^\perp = \mathbf{I} - \tilde{\mathbf{S}}^H(\tilde{\mathbf{S}}^H)^\dagger$ which projects to the orthogonal space of range of $\tilde{\mathbf{S}}$ is $N-P-1$. Therefore the last equality holds since the orthogonal projection matrix has $N-P-1$ unit eigenvalues and one eigenvalue to be zero. ∎

C. Automatic Weighting Between $\boldsymbol{\Psi}$ and $\bar{\mathbf{Q}}$

We consider a linear combination scheme between the unbiased $\boldsymbol{\Psi}$ and the prior $\bar{\mathbf{Q}}$ according to the MMSE criterion [13], [16]. Consistent with the fact that the estimate $\boldsymbol{\Psi}$ consists of both the test and training signals, we propose to use the test and training signals for automatic determination of the linear combining weights, thus extending the regularized method in [13] and [16] which uses only the training signals. As shown in later numerical examples, the inclusion of the test signal leads to improved performance when the number of training signals is limited.

Specifically, we consider a convex combination between $\bar{\mathbf{Q}}$ and the estimate $\boldsymbol{\Psi}$[1]

$$\tilde{\mathbf{Q}} = \beta\bar{\mathbf{Q}} + (1-\beta)\boldsymbol{\Psi}, \tag{29}$$

where $\beta \in [0,1]$ is the combining weight to be determined. This scheme is to balance the contribution from the prior knowledge and the observed signals. One can replace $\bar{\mathbf{Q}}$ by an identify matrix in the case that $\bar{\mathbf{Q}}$ is unavailable or has a large amount of uncertainty.

The optimal $\tilde{\mathbf{Q}}$ is determined by minimizing the MSE defined as

$$\begin{aligned} \mathrm{MSE} &= E\{\|\tilde{\mathbf{Q}} - \mathbf{Q}\|^2\} \\ &= E\{\|(\boldsymbol{\Psi} - \mathbf{Q}) - \beta(\boldsymbol{\Psi} - \bar{\mathbf{Q}})\|^2\} \\ &= E\{\|\boldsymbol{\Psi} - \mathbf{Q}\|^2\} + \beta^2 E\{\|\boldsymbol{\Psi} - \bar{\mathbf{Q}}\|^2\} \\ &\quad - 2\beta\Re\{\mathrm{tr}\{E\{(\boldsymbol{\Psi} - \mathbf{Q})(\boldsymbol{\Psi} - \bar{\mathbf{Q}})^H\}\}\}. \end{aligned} \tag{30}$$

with $\|\cdot\|^2$ defining the Frobenius norm of a matrix, under the linear equality constraint of (29). Since $E\{\boldsymbol{\Psi}\} = \mathbf{Q}$ and $E\{\boldsymbol{\Psi}^H\} = \mathbf{Q}^H$, the last term in the above equation can be simplified as

$$\begin{aligned} &\Re\{\mathrm{tr}\{E\{(\boldsymbol{\Psi} - \mathbf{Q})(\boldsymbol{\Psi} - \bar{\mathbf{Q}})^H\}\}\} \\ &= \Re\{\mathrm{tr}\{\mathbf{Q}\bar{\mathbf{Q}}^H - E\{\boldsymbol{\Psi}\}\bar{\mathbf{Q}}^H \\ &\quad + E\{\boldsymbol{\Psi}\boldsymbol{\Psi}^H\} - \mathbf{Q}E\{\boldsymbol{\Psi}^H\}\}\} \\ &= \Re\{\mathrm{tr}\{E\{\boldsymbol{\Psi}\boldsymbol{\Psi}^H\} - \mathbf{Q}\mathbf{Q}^H\}\} \\ &= \Re\{\mathrm{tr}\{-E\{\boldsymbol{\Psi}\}\mathbf{Q}^H + \mathbf{Q}E\{\boldsymbol{\Psi}^H\} \\ &\quad E\{\boldsymbol{\Psi}\boldsymbol{\Psi}^H\} - \mathbf{Q}\mathbf{Q}^H\}\} \\ &= \Re\{\mathrm{tr}\{E\{(\boldsymbol{\Psi} - \mathbf{Q})(\boldsymbol{\Psi} - \mathbf{Q})^H\}\}\} \\ &= E\{\|\boldsymbol{\Psi} - \mathbf{Q}\|^2\}. \end{aligned} \tag{31}$$

which leads to

$$\mathrm{MSE} = (1-2\beta)E\{\|\boldsymbol{\Psi} - \mathbf{Q}\|^2\} + \beta^2 E\{\|\boldsymbol{\Psi} - \bar{\mathbf{Q}}\|^2\}. \tag{32}$$

Taking the derivative of the MSE and equating it to zero, the optimal combining weight β is given by

$$\beta = \frac{E\{\|\boldsymbol{\Psi} - \mathbf{Q}\|^2\}}{E\{\|\boldsymbol{\Psi} - \bar{\mathbf{Q}}\|^2\}} = \frac{E\{\|\boldsymbol{\Psi} - \mathbf{Q}\|^2\}}{E\{\|\boldsymbol{\Psi} - \mathbf{Q}\|^2\} + \|\bar{\mathbf{Q}} - \mathbf{Q}\|^2}. \tag{33}$$

Define $\rho \triangleq E\{\|\boldsymbol{\Psi} - \mathbf{Q}\|^2\}$ and $\nu \triangleq \|\bar{\mathbf{Q}} - \mathbf{Q}\|^2$. Since ρ and ν depend on the true but unobservable $\mathbf{Q}$, the optimal combining weight β has to be estimated from the observations. In our case, we use both the test and training signals to achieve this purpose.

First, regarding the estimate of ρ, we show in the following that $\boldsymbol{\Psi}$ can be considered as equivalently the sample covariance matrix from a set of L i.i.d. Gaussian vectors, among which $N-P-1$ vectors are obtained from the test signal and the

[1]Other linear combination such as the generalized linear combination (GLC) can be derived similarly, which has a similar performance.

remaining $K(N-P)$ vectors are from the training signals. From (24) and (26), $\boldsymbol{\Psi}$ consists of two components:

$$\begin{aligned} L\boldsymbol{\Psi} &= \tilde{\mathbf{X}}_0\mathbf{P}^{\perp}\tilde{\mathbf{X}}_0^H + \sum_{k=1}^{K}\tilde{\mathbf{X}}_k\tilde{\mathbf{X}}_k^H \\ &= \mathbf{E}_0\mathbf{P}^{\perp}\mathbf{E}_0^H + \sum_{k=1}^{K}\mathbf{E}_k\mathbf{E}_k^H. \end{aligned} \tag{34}$$

Then the $(N-P)\times(N-P)$ orthogonal projection matrix $\mathbf{P}^{\perp}$ can be decomposed to

$$\mathbf{P}^{\perp} = \mathbf{U}_{\mathbf{P}}\mathbf{U}_{\mathbf{P}}^H, \tag{35}$$

where $\mathbf{U}_{\mathbf{P}}$ is an $(N-P)\times(N-P-1)$ matrix with $N-P-1$ orthonormal columns. Together with the $N-P$ i.i.d. Gaussian columns in $\mathbf{E}_0$ and the $(N-P-1)$ orthonormal columns of $\mathbf{U}_{\mathbf{P}}$, the resulting $N-P-1$ columns of

$$\begin{aligned} \mathbf{Z}_0 &= \tilde{\mathbf{X}}_0\mathbf{U}_{\mathbf{P}} = \mathbf{E}_0\mathbf{U}_{\mathbf{P}} \\ &= [\mathbf{z}_0(1), \mathbf{z}_0(2), \ldots, \mathbf{z}_0(N-P-1)] \end{aligned} \tag{36}$$

are i.i.d. Gaussian vectors with zero mean and covariance matrix $\mathbf{Q}$, i.e., $\mathbf{z}_0(n) \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q})$ and follow the same steps in (28). As a result,

$$\begin{aligned} \tilde{\mathbf{X}}_0\mathbf{P}^{\perp}\tilde{\mathbf{X}}_0^H &= \mathbf{E}_0\mathbf{P}^{\perp}\mathbf{E}_0^H = \mathbf{Z}_0\mathbf{Z}_0^H \\ &= \sum_{n=1}^{N-P-1}\mathbf{z}_0(n)\mathbf{z}_0^H(n) \end{aligned} \tag{37}$$

For the training signal component, we define $\mathbf{z}_k(n) = \tilde{\mathbf{x}}_k(n)$ and

$$\sum_{k=1}^{K}\tilde{\mathbf{X}}_k\tilde{\mathbf{X}}_k^H = \sum_{k=1}^{K}\sum_{n=P}^{N-1}\mathbf{z}_k(n)\mathbf{z}_k^H(n), \tag{38}$$

where, again, $\mathbf{z}_k(n)$ are i.i.d. Gaussian vectors with zero mean and covariance matrix $\mathbf{Q}$. Stacking all $\mathbf{z}_k(n)$, we have a set of L i.i.d. Gaussian vectors

$$\mathbf{z}(n) = \begin{cases} \mathbf{z}_0(l), & n = 1, 2, \ldots, N-P-1, \\ & l = n + P - 1 \\ \mathbf{z}_k(l), & n = N-P, \ldots, L, \\ & k = \left\lfloor \frac{n-(N-P)}{N-P} \right\rfloor, \\ & l = n - (k+1)(N-P) + P. \end{cases} \tag{39}$$

Therefore,

$$\boldsymbol{\Psi} = \frac{1}{L}\sum_{n=1}^{L}\mathbf{z}(n)\mathbf{z}^H(n), \tag{40}$$

can be considered as the sample covariance matrix from L i.i.d. Gaussian vectors $\mathbf{z}(n)$ with zero mean and covariance matrix $\mathbf{Q}$.

With the above results, the estimate of ρ reduces to the estimation of β from L i.i.d. random vectors with zero mean and covariance matrix $\mathbf{Q}$. As shown in the Appendix, the coefficient ρ can be adaptively estimated as

$$\tilde{\rho} = \frac{1}{(L-1)L}\sum_{n=1}^{L}\|\mathbf{z}(n)\mathbf{z}^H(n) - \boldsymbol{\Psi}\|^2. \tag{41}$$

For the second quantity ν, we simply replace the true $\mathbf{Q}$ by the unbiased estimate $\boldsymbol{\Psi}$

$$\tilde{\nu} = \|\bar{\mathbf{Q}} - \boldsymbol{\Psi}\|^2, \tag{42}$$

which leads to the knowledge-aided spatial covariance matrix estimate

$$\tilde{\mathbf{Q}} = \frac{\tilde{\rho}}{\tilde{\rho}+\tilde{\nu}}\bar{\mathbf{Q}} + \frac{\tilde{\nu}}{\tilde{\rho}+\tilde{\nu}}\boldsymbol{\Psi}. \tag{43}$$

D. *Fully Adaptive KA-PAMF*

Finally, an adaptive estimate of $\mathbf{A}$ is needed to enable a fully adaptive KA-AC-PAMF. In the following, the ML estimate of $\mathbf{A}$ from the training signals, derived in [24], is used. Specifically, the ML estimate of $\mathbf{A}$ can be computed as follows

$$\hat{\mathbf{A}}_{\text{ML}} = -\hat{\mathbf{R}}_{yx,K}^H\hat{\mathbf{R}}_{yy,K}^{-1}, \tag{44}$$

where

$$\hat{\mathbf{R}}_{yy,K} = \sum_{k=1}^{K}\sum_{n=P}^{N-1}\mathbf{y}_k(n)\mathbf{y}_k^H(n) \tag{45}$$

$$\hat{\mathbf{R}}_{yx,K} = \sum_{k=1}^{K}\sum_{n=P}^{N-1}\mathbf{y}_k(n)\mathbf{x}_k^H(n), \tag{46}$$

with $\mathbf{y}_k(n)$ is defined in (15).

Using the ML estimate of $\mathbf{A}$ in (21), the proposed KA-AC-PAMF detector takes the form of

$$T_{\text{KA-PAMF}} = \frac{\left|\sum_{n=P}^{N-1}\hat{\tilde{\mathbf{s}}}^H(n)\hat{\tilde{\mathbf{Q}}}^{-1}\hat{\tilde{\mathbf{x}}}_0(n)\right|^2}{\sum_{n=P}^{N-1}\hat{\tilde{\mathbf{s}}}^H(n)\hat{\tilde{\mathbf{Q}}}^{-1}\hat{\tilde{\mathbf{s}}}(n)}, \tag{47}$$

where the fully adaptively temporally whitened vectors

$$\hat{\tilde{\mathbf{x}}}_k(n) = \mathbf{x}_k(n) + \sum_{p=1}^{P}\hat{\mathbf{A}}_{\text{ML}}^H(p)\mathbf{x}_k(n-p), \tag{48}$$

$$\hat{\tilde{\mathbf{s}}}(n) = \mathbf{s}(n) + \sum_{p=1}^{P}\hat{\mathbf{A}}_{\text{ML}}^H(p)\mathbf{s}(n-p), \tag{49}$$

and, correspondingly,

$$\hat{\tilde{\mathbf{X}}}_k = \left[\hat{\tilde{\mathbf{x}}}_k(P), \hat{\tilde{\mathbf{x}}}_k(P+1), \ldots, \hat{\tilde{\mathbf{x}}}_k(N-1)\right], \tag{50}$$

$$\hat{\tilde{\mathbf{S}}} = \left[\hat{\tilde{\mathbf{s}}}(P), \hat{\tilde{\mathbf{s}}}(P+1), \ldots, \hat{\tilde{\mathbf{s}}}(N-1)\right]. \tag{51}$$

Then, the vectors $\hat{\mathbf{z}}(n)$ is formed from the columns of the following matrices

$$\hat{\mathbf{Z}}_0 = \hat{\tilde{\mathbf{X}}}_0\hat{\mathbf{U}}_{\hat{\mathbf{P}}}, \quad \hat{\mathbf{Z}}_k = \hat{\tilde{\mathbf{X}}}_k, \quad k = 1, 2, \ldots, K, \tag{52}$$

where $\hat{\mathbf{U}}_{\hat{\mathbf{P}}}$ given as $\hat{\mathbf{U}}_{\hat{\mathbf{P}}} = \texttt{null}(\hat{\mathbf{P}})$ with $\hat{\mathbf{P}} = \hat{\tilde{\mathbf{S}}}^H(\hat{\tilde{\mathbf{S}}}^H)^{\dagger}$. Finally, the spatial covariance matrix is estimated as

$$\hat{\tilde{\mathbf{Q}}} = \frac{\hat{\tilde{\rho}}}{\hat{\tilde{\rho}}+\hat{\tilde{\nu}}}\bar{\mathbf{Q}} + \frac{\hat{\tilde{\nu}}}{\hat{\tilde{\rho}}+\hat{\tilde{\nu}}}\hat{\boldsymbol{\Psi}}, \tag{53}$$

where

$$\hat{\boldsymbol{\Psi}} = \frac{1}{L}\sum_{n=1}^{L}\hat{\mathbf{z}}(n)\hat{\mathbf{z}}^H(n), \tag{54}$$

$$\hat{\tilde{\rho}} = \frac{1}{(L-1)L}\sum_{n=1}^{L}\|\hat{\mathbf{z}}(n)\hat{\mathbf{z}}^H(n) - \hat{\boldsymbol{\Psi}}\|^2, \tag{55}$$

$$\hat{\tilde{\nu}} = \|\bar{\mathbf{Q}} - \hat{\boldsymbol{\Psi}}\|^2. \tag{56}$$

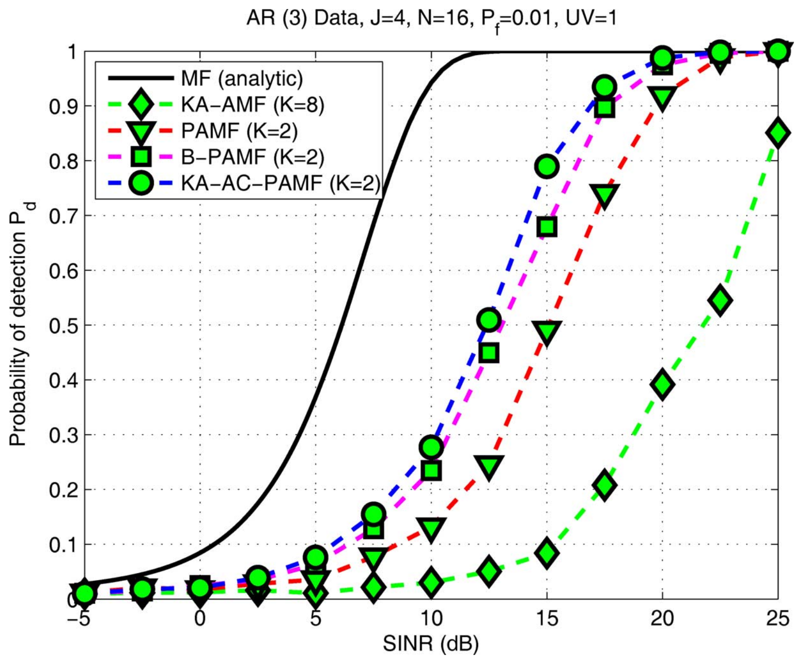

Fig. 1. Probability of detection with a reliable prior ($\mathrm{UV} = 1$) when $K = 2$, $J = 4$, $N = 16$, and $P_f = 0.01$.

IV. Numerical Results

In this section, numerical results are provided to compare the proposed KA-AC-PAMF with other conventional and knowledge-aided parametric detectors in terms of the detection performance versus the signal-to-interference-plus-noise ratio (SINR) for a probability of false alarm $P_f = 0.01$. Specifically, we consider 1) the conventional PAMF [19], 2) the Bayesian KA-PAMF [22] with a random guess of the hyper-prior parameter, and 3) the KA-AMF [16]. The optimal matched filter is also shown to provide the benchmark.

In the first set of simulations, the disturbance signals $\mathbf{d}_k$ are generated from the $\mathrm{AR}(3)$ process with a given $\mathbf{A}$ and $\mathbf{Q}$. To account for the uncertainty of prior knowledge about $\mathbf{Q}$, a perturbed version of $\mathbf{Q}$ is used as $\bar{\mathbf{Q}}$ [16]

$$\bar{\mathbf{Q}} = \mathbf{Q} \odot \mathbf{t}_s \mathbf{t}_s^H, \tag{57}$$

where $\mathbf{t}_s$ is a $J \times 1$ vector of i.i.d. Gaussian random vectors with mean 1 and variance σ_t^2, and $\odot$ denotes the Hadamard matrix product. Specifically, we refer to the σ_t^2 as the uncertainty variance (UV). As the UV increases, the prior knowledge $\bar{\mathbf{Q}}$ is on average away from the true $\mathbf{Q}$. For the KA-AMF method which relies on a prior knowledge $\bar{\mathbf{R}}$ of the space-time covariance matrix $\mathbf{R}$, $\bar{\mathbf{R}}$ is computed from the given $\bar{\mathbf{Q}}$ and the AR coefficient matrix $\mathbf{A}$ by using the multichannel Levinson algorithm. This ensures the methods are subject to the same level of uncertainty in their prior knowledge. The SINR is defined by $\mathbf{R}$ as

$$\mathrm{SINR} = |\alpha^2| \mathbf{s}^H \mathbf{R}^{-1} \mathbf{s}, \tag{58}$$

For the synthesized AR dataset, the number of channel is $J = 4$ and the number of temporal observations is $N = 16$.

We first consider the case of $\mathrm{UV} = 1$, e.g., a case with a relatively reliable prior ($\bar{\mathbf{Q}}$ for the parametric detectors and $\bar{\mathbf{R}}$ for the KA-AMF) and the number of training signals $K = 2$. Without using the prior knowledge, the conventional PAMF fully relies on $K = 2$ training signals and, as shown in Fig. 1, its performance is the worst among the parameter detectors. Although utilizing the prior knowledge, the KA-AMF shows worse performance than the parametric detectors since the limited $K = 8$ training signals, compared with the total dimension $JN = 64$, are unable to obtain a good covariance matrix estimate without exploiting the structural information of $\mathbf{R}$. In contrast, by using the prior $\bar{\mathbf{Q}}$ and exploiting the multi-channel AR structure, the knowledge-aided parametric detectors, i.e., the B-PAMF and the proposed KA-AC-PAMF, give improved performance than the conventional PAMF and the KA-AMF. In addition, the proposed KA-AC-PAMF has an SINR improvement of about 1 dB over the B-PAMF.

In the second example, we increase the prior uncertainty to $\mathrm{UV} = 5$. Since the conventional PAMF uses no prior knowledge, its performance is independent of the UV as shown in Fig. 2. Comparison between Figs. 1 and 2 reveals that for a larger UV, a noticeable performance degradation happens for the B-PAMF which uses a non-adaptive weight on $\bar{\mathbf{Q}}$, regardless of the prior uncertainty. On the other hand, the proposed KA-AC-PAMF is less sensitive. It should be noted that as the UV further increases so that the prior knowledge $\bar{\mathbf{Q}}$ becomes less and less accurate, the combining coefficient for $\bar{\mathbf{Q}}$ in (53) will decrease to zero, and the proposed method will degrade to its non-knowledge-aided counterpart.

Next, we test the detectors by using the KASSPER dataset (see [17] for a detailed description of the KASSPER dataset) when the disturbance is not exactly a multi-channel AR process. Specifically, the ground clutter covariance matrix $\mathbf{R}$ at the range bin 100 is used to generate disturbances in the test and training signals. For the knowledge-aided detectors, we first learn the corresponding $\mathbf{Q}$ from the covariance matrix $\mathbf{R}$ for the

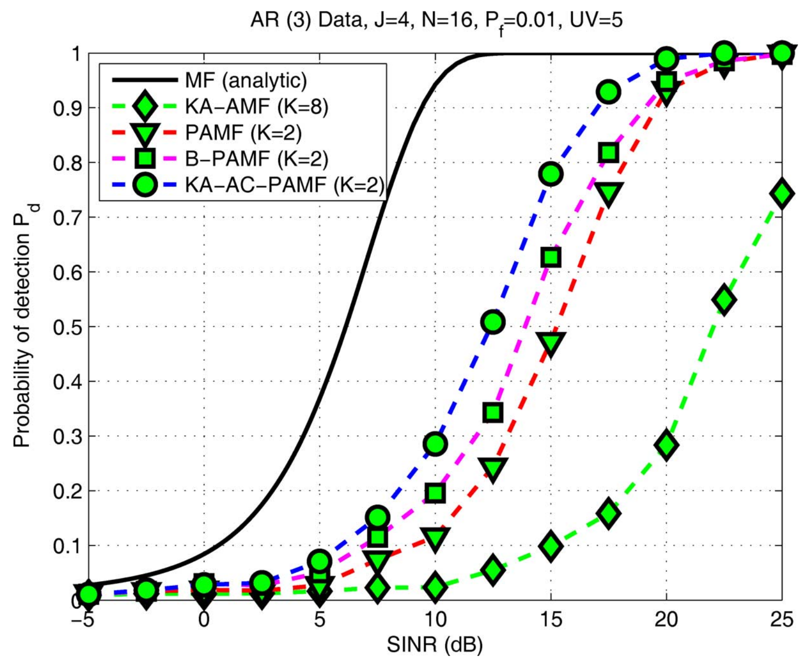

Fig. 2. Probability of detection with a less reliable prior $(\mathrm{UV}=5)$ when $K=2$, $J=4$, $N=16$, and $P_f=0.01$.

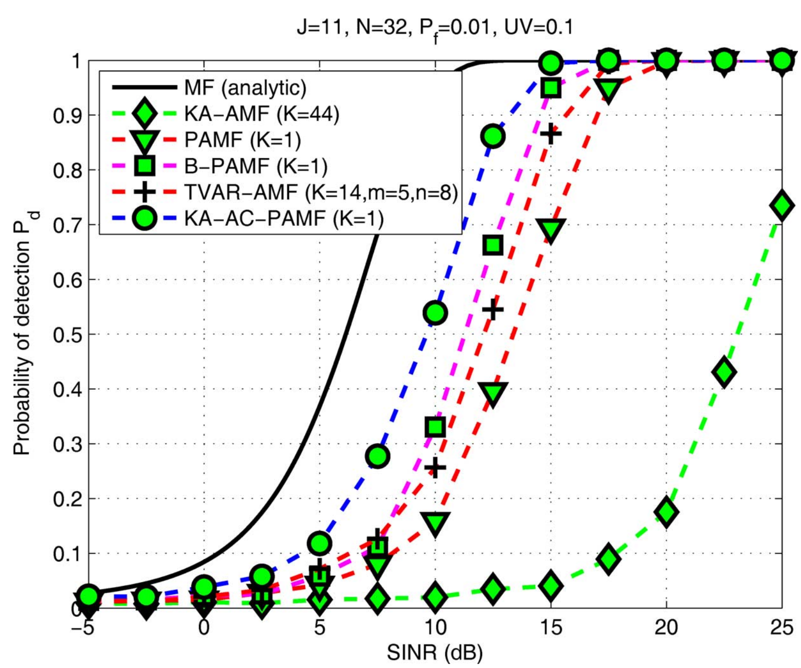

Fig. 3. Probability of detection for the KASSPER dataset when $K=1$, $J=11$, $N=32$, $P_f=0.01$, and $\mathrm{UV}=0.1$.

observation by the multi-channel Levinson algorithm, and then a perturbed prior $\bar{\mathbf{Q}}$ is generated according to (57). In addition to the methods considered above, we include another detector denoted as TVAR-AMF, which employs the time-varying auto regressive (TVAR) covariance matrix estimation method [25] in the AMF. Although the TVAR estimator is not knowledge-aided (in the sense that it does not require a prior covariance matrix estimate), it does exploit the Toeplitz-block-Toeplitz structure of the space-time covariance matrix for covariance estimation. The performance of the TVAR estimator was examined by using KASSPER data in [26], which shows the estimator is a strong competitor for applications in training-limited scenarios. For the TVAR-AMF detector, we use $m=5, n=8$ as the model orders in the spatial and temporal domain, respectively (same as in [26]). The KASSPER data has $J=11$ spatial channels and $N=32$ pulses. We consider an extremely limited training case of $K=1$ for the parametric detectors while $K=14$ for TVAR-AMF and $K=44$ for KA-AMF. For the parametric detectors, we use a range of possible AR model order $P=\{1,2,3\}$ and choose the one that yields the best detection performance ($P=1$ in our simulation). As shown in Fig. 3, when the $\mathrm{UV}=0.1$, the proposed KA-AC-PAMF attains the best per-

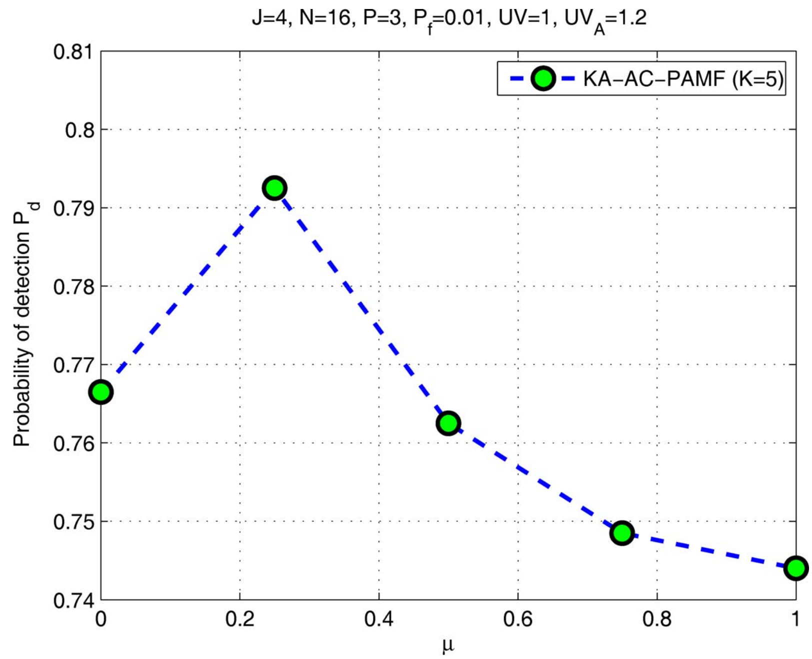

Fig. 4. Impact of using a prior knowledge $\bar{\mathbf{A}}$ of the AR coefficient $\mathbf{A}$ on the detection probability.

formance among all considered detectors. It provides about 2-dB SINR gain than the B-PAMF which also benefits from the accurate prior knowledge, and 3-dB SINR gain than the TVAR method. Among the parametric detectors, the conventional PAMF shows worse performance than the knowledge-aided parametric detectors. For the KA-AMF, $K = 14$ training signals are not sufficient to support the estimate of the 352×352 covariance matrix $\mathbf{R}$ and hence a big performance degradation is observed when it is compared to the parametric detectors. It is notable that the non-knowledge-aided TVAR-AMF outperforms the knowledge-aided KA-AMF.

So far, we have only examined the benefit of using a prior knowledge of the spatial covariance matrix $\bar{\mathbf{Q}}$ in proposed detector. One would naturally ask if this can be extended to the AR coefficient matrix $\mathbf{A}$. Here, we examine via computer simulation the impact of using a prior knowledge, denoted by $\bar{\mathbf{A}}$, of $\mathbf{A}$. Similar to what is applied to spatial covariance matrix, we consider a convex combination of the prior knowledge $\bar{\mathbf{A}}$ and the adaptive estimate $\hat{\mathbf{A}}$ given by (44), namely $\tilde{\mathbf{A}} = \mu\bar{\mathbf{A}} + (1-\mu)\hat{\mathbf{A}}$, where $0 \leq \mu \leq 1$. To simulate the fact that $\bar{\mathbf{A}}$ is inaccurate, we follow the similar step as in (57):

$$\bar{\mathbf{A}} = \mathbf{A} \odot \mathbf{t}_a \mathbf{t}_a^H, \tag{59}$$

where $\mathbf{t}_a$ is a $J \times 1$ vector of i.i.d. Gaussian random variables with mean 1 and variance UV_A. Fig. 4 depicts the detection performance of the proposed detector by using a convex combination of $\bar{\mathbf{A}}$ and $\hat{\mathbf{A}}$ with μ varying between 0 and 1, where we set uncertainty parameters $\mathrm{UV} = 1$ for $\mathbf{Q}$ and $\mathrm{UV}_A = 1.2$ for $\bar{\mathbf{A}}$. It is noted that with $\mu = 1$, only the prior knowledge $\bar{\mathbf{A}}$ is used for the detector, whereas with $\mu = 0$, only the estimate $\hat{\mathbf{A}}$ is used. Fig. 4 shows that for the 5 different combining coefficient considered, the best detection probability is attached at $\mu = 0.25$, which indicates that it is beneficial to employ both the prior knowledge $\bar{\mathbf{A}}$ and the adaptive estimate $\hat{\mathbf{A}}$ for detection. However, how to optimumly combining $\bar{\mathbf{A}}$ and $\hat{\mathbf{A}}$ remains a future work.

V. Conclusion

In this paper, a new knowledge-aided PAMF is proposed which automatically determines the linear combining weights between the prior covariance matrix and the conventional covariance estimate using both the test and training signals. On one hand, the proposed detector is more robust against uncertainty in the prior knowledge than the existing knowledge-aided PAMF. On the other hand, it also outperforms the knowledge-aided non-parametric detector in scenarios with limited training signals. Simulation results confirm the effectiveness of the proposed detector.

Appendix

From (40), $\mathbf{\Psi}$ can be estimated from L i.i.d. signals $\mathbf{z}(n)$ stacked from both the test and training signals. As a result, an adaptive estimate of ρ can be derived in a similar way to [13], [16].

The quantity ρ can be rewritten as

$$\rho = E\{\|\mathbf{\Psi} - \mathbf{Q}\|^2\} = \sum_{j=1}^{J} E\{\|\tilde{\mathbf{q}}_j - \mathbf{q}_j\|^2\}, \tag{60}$$

where $\tilde{\mathbf{q}}_j$ and $\mathbf{q}_j$ are the j-th columns of $\mathbf{\Psi}$ and $\mathbf{Q}$, respectively

$$\tilde{\mathbf{q}}_j = \frac{1}{L}\sum_{n=1}^{L} \mathbf{z}(n) z_j^*(n) \triangleq \frac{1}{L}\sum_{n=1}^{L} \mathbf{w}_j(n), \tag{61}$$

$$\mathbf{q}_j = E\{\mathbf{z}(n) z_j^*(n)\}, \tag{62}$$

with $z_j(n)$ denoting the j-th element of $\mathbf{z}(n)$. Since $\mathbf{z}(n)$ are i.i.d. vectors, we have $\mathbf{w}_j(n)$ are i.i.d. vector across n with mean $\mathbf{q}_j$. This leads to

$$\begin{aligned} E\{\|\tilde{\mathbf{q}}_j - \mathbf{q}_j\|^2\} &= E\left\{\left\|\frac{1}{L}\sum_{n=1}^{L}\mathbf{w}_j(n) - \mathbf{q}_j\right\|^2\right\} \\ &= E\left\{\left\|\frac{1}{L}\sum_{n=1}^{L}(\mathbf{w}_j(n) - \mathbf{q}_j)\right\|^2\right\} \\ &= \frac{1}{L}E\left\{\|\mathbf{w}_j(n) - \mathbf{q}_j\|^2\right\}, \end{aligned} \tag{63}$$

where the last equality is due to, again, i.i.d. $\mathbf{w}_j(n)$. Then the variance $\mathbf{Q}_j \triangleq \|\mathbf{w}_j(n) - \mathbf{q}_j\|^2$ can be estimated from the sample variance estimate from L i.i.d. $\mathbf{z}(n)$, i.e.,

$$\begin{aligned} \tilde{\mathbf{Q}}_j &= \frac{\sum_{n=1}^{L}\|\mathbf{w}_j(n) - \hat{\mathbf{q}}_j\|^2}{L-1} \\ &= \frac{\sum_{n=1}^{L}\|\mathbf{z}(n)z_j^*(n) - \hat{\mathbf{q}}_j\|^2}{L-1}, \end{aligned} \tag{64}$$

where $\hat{\mathbf{q}}_j$ is the sample mean estimate

$$\hat{\mathbf{q}}_j = \frac{1}{L}\sum_{n=1}^{L}\mathbf{z}(n)z_j^*(n). \tag{65}$$

Taking the sample variance estimate back to (63), we have

$$E\{\|\tilde{\mathbf{q}}_j - \mathbf{q}_j\|^2\} = \frac{\sum_{n=1}^{L}\|\mathbf{z}(n)z_j^*(n) - \hat{\mathbf{q}}_j\|^2}{(L-1)L}. \tag{66}$$

Finally, taking the above quantity back to (60), we have

$$\begin{aligned} \rho &= \sum_{j=1}^{J}E\{\|\tilde{\mathbf{q}}_j - \mathbf{q}_j\|^2\} \\ &= \frac{\sum_{j=1}^{J}\sum_{n=1}^{L}\|\mathbf{z}(n)z_j^*(n) - \hat{\mathbf{q}}_j\|^2}{(L-1)L} \\ &= \frac{\sum_{n=1}^{L}\|\mathbf{z}(n)\mathbf{z}^H(n) - \mathbf{\Psi}\|^2}{(L-1)L}. \end{aligned} \tag{67}$$

REFERENCES

[1] E. J. Kelly, “An adaptive detection algorithm,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 22, no. 1, pp. 115–127, Mar. 1986.

[2] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, “A CFAR adaptive matched filter detector,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 1, pp. 208–216, Jan. 1992.

[3] M. C. Wicks, M. Rangaswamy, R. Adve, and T. B. Hale, “Space-time adaptive processing: A knowledge-based perspective for airborne radar,” *IEEE Signal Process. Mag.*, vol. 23, no. 1, pp. 51–65, Jan. 2006.

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

[5] A. De Maio, A. Farina, and G. Foglia, “Adaptive radar detection: A Bayesian approach,” in *Proc. IEEE Int. Conf. Radar*, Waltham, MA, USA, Apr. 2007, pp. 624–629.

[6] O. Besson, J. Y. Tourneret, and S. Bidon, “Knowledge-aided Bayesian detection in heterogeneous environment,” *IEEE Signal Process. Lett.*, pp. 355–358, May 2007.

[7] 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.

[8] S. Bidon, O. Besson, and J.-Y. Tourneret, “A Bayesian approach to adaptive detection in non-homogeneous environments,” *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 205–217, Jan. 2008.

[9] 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.

[10] 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.

[11] A. Aubry, A. D. Maio, L. Pallotta, and A. Farina, “Maximum likelihood estimation of a structured covariacne matrix with a condition number constraint,” *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 3004–3021, Jun. 2012.

[12] A. Aubry, A. D. 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.

[13] O. Ledoit and M. Wolf, “A well-conditioned estimator for large-dimensional covariance matrices,” *J. Multivar. Anal.*, vol. 88, no. 2, pp. 365–411, 2004.

[14] J. P. Hoffbeck and D. A. Landgrebe, “Covariance matrix estimation and classification with limited training data,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 7, pp. 763–767, Jul. 1996.

[15] P. J. Bickel and E. Levina, “Regularized estimation of large covariance matrices,” *Ann. Statist.*, vol. 36, no. 1, pp. 199–227, 2008.

[16] 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.

[17] J. S. Bergin and P. M. Techau, “High-fidelity site-specific radar simulation: KASSPER’02 Workshop Datacube,” Information Systems Labs., Inc., Vienna, VA, USA, Tech. Rep. ISL-SCRD-TR-02-105, 2002.

[18] M. Rangaswamy and J. H. Michels, “A parametric multichannel detection algorithm for correlated non-Gaussian random processes,” in *Proc. IEEE Nat. Radar Conf.*, Syracuse, NY, USA, May 1997, pp. 349–354.

[19] J. R. Román, M. Rangaswamy, D. W. Davis, Q. Zhang, B. Himed, and J. H. Michels, “Parametric adaptive matched filter for airborne radar applications,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 2, pp. 677–692, Apr. 2000.

[20] J. H. Michels, B. Himed, and M. Rangaswamy, “Robust STAP detection in a dense signal airborne radar environment,” *Signal Process., Special Issue on New Trends and Findings in Antenna Array Processing for Radar*, vol. 84, pp. 1625–1636, 2004.

[21] P. Wang, H. Li, and B. Himed, “A Bayesian parametric test for multichannel adaptive signal detection in non-homogeneous environments,” *IEEE Signal Process. Lett.*, vol. 17, no. 4, pp. 351–354, Apr. 2010.

[22] P. Wang, H. Li, and B. Himed, “Knowledge-aided parametric tests for multichannel adaptive signal detection,” *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 5970–5982, Dec. 2011.

[23] J. Ward, “Space-time adaptive processing for airborne radar,” MIT, Lincoln Lab., Lexington, MA, USA, Tech. Rep. 1015, 1994.

[24] K. J. Sohn, H. Li, and B. Himed, “Parametric Rao test for multichannel adaptive signal detection,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, no. 3, pp. 920–933, Jul. 2007.

[25] Y. I. Abramovich, B. A. Johnson, and N. K. Spencer, “Two-dimensional multivariate parametric models for radar applications—Part I: Maximum-entropy extensions for Toeplitz-block matrices,” *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5509–5526, Nov. 2008.

[26] Y. I. Abramovich, M. Rangaswamy, B. A. Johnson, P. M. Corbell, and N. K. Spencer, “Performance of 2-D mixed autoregressive models for airborne radar STAP: KASSPER-aided analysis,” in *Proc. IEEE Radar Conf.*, Rome, May 2008, pp. 1–5.

Pu Wang (S'05–M'12) received the B.Eng. and M.Eng. degrees from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2003 and 2006, respectively, and the Ph.D. degree from the Stevens Institute of Technology, Hoboken, NJ, USA, in 2011, all in electrical engineering.

He was an intern at the Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA, in summer 2010, and a Research Assistant Professor at the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, from May 2011 to September 2012. Since October 2012, he has been with the Schlumberger-Doll Research, Cambridge, MA. His current research interests include statistical signal processing, acoustics and ultrasonics in geophysical prospecting, Bayesian inference, time-frequency analysis, and sparse signal recovery. He currently holds four U.S. patents (applications).

Dr. Wang is the recipient of the IEEE Jack Neubauer Memorial Award in 2013 for the best systems paper published in the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY and the Outstanding Paper Award from the IEEE AFRICON Conference in 2011. He received the Outstanding Doctoral Thesis in EE Award in 2011, the Edward Peskin Award in 2011, the Francis T. Boesch Award in 2008, and the Outstanding Research Assistant Award in 2007, all from the Stevens Institute of Technology, the Excellent Master Thesis in Sichuan Province (of China) Award in 2007, and the Excellent Master Thesis Award from UESTC in 2006. He was an organization committee member of the Seventh IEEE Sensor Array and Multichannel Signal Processing (SAM) Workshop, Hoboken, New Jersey, June 2012, and the Fifth IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), Saint Martin, December 2013.

Zhe Wang (S'12) received the B.S. degree in 2008 and M.S. degree in 2010 from the Dalian University of Technology (DLUT), Dalian, all in electrical engineering. He is currently pursuing the Ph.D degree in electrical engineering at the Stevens Institute of Technology, Hoboken, NJ.

Since 2012 he has been a Research Assistant in the Department of Electrical and Computer Engineering at Stevens. His current research interests include digital signal processing, multichannel signal processing, adaptive detection and parameter estimation, and sparse signal processing.

Hongbin Li (M'99–SM'08) received the B.S. and M.S. degrees from the University of Electronic Science and Technology of China, Chengdu, in 1991 and 1994, respectively, and the Ph.D. degree from the University of Florida, Gainesville, FL, 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. He was a Summer Visiting Faculty Member at the Air Force Research Laboratory in the summers of 2003, 2004 and 2009. Since July 1999, he has been with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, where he became a Professor in 2010. His current research interests include statistical signal processing, wireless communications, and radars.

Dr. Li received the IEEE Jack Neubauer Memorial Award in 2013, the Harvey N. Davis Teaching Award in 2003 and 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 is presently a member of the Signal Processing Theory and Methods (SPTM) Technical Committee and served on the Sensor Array and Multichannel (SAM) Technical Committee of the IEEE Signal Processing Society. He has been an Associate Editor for *EURASIP Signal Processing*, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE SIGNAL PROCESSING LETTERS, and IEEE TRANSACTIONS ON SIGNAL PROCESSING, and a Guest Editor for *EURASIP Journal on Applied Signal Processing*. He was a General Co-Chair for the 7th IEEE Sensor Array and Multichannel Signal Processing Workshop, Hoboken, NJ, June 17–20, 2012. Dr. Li is a member of Tau Beta Pi and Phi Kappa Phi.

Braham Himed (S'88–M'90–SM'01–F'07) received his B.S. degree in electrical engineering from Ecole Nationale Polytechnique of Algiers in 1984, and his M.S. and Ph.D. degrees both in electrical engineering, from Syracuse University, Syracuse, NY, in 1987 and 1990, respectively. Dr. Himed is a Technical Advisor with the Air Force Research Laboratory, Sensors Directorate, RF Technology Branch, in Dayton Ohio, where he is involved with several aspects of radar developments. His research interests include detection, estimation, multichannel adaptive signal processing, time series analyses, array processing, space-time adaptive processing, waveform diversity, MIMO radar, passive radar, and over the horizon radar. Dr. Himed is the recipient of the 2001 IEEE region I award for his work on bistatic radar systems, algorithm development, and phenomenology. Dr. Himed is a Fellow of the IEEE and a member of the AES Radar Systems Panel. Dr. Himed is the recipient of the 2012 IEEE Warren White award for excellence in radar engineering. Dr. Himed is also a Fellow of AFRL (Class of 2013).