Knowledge-Aided Parametric Tests for Multichannel Adaptive Signal Detection

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

***Abstract*—In this paper, the problem of detecting a multichannel signal in the presence of spatially and temporally colored disturbance is considered. By modeling the disturbance as a multi-channel auto-regressive (AR) process with a random cross-channel (spatial) covariance matrix, two knowledge-aided parametric adaptive detectors are developed within a Bayesian framework. The first knowledge-aided parametric detector is developed using an *ad hoc* two-step procedure for the estimation of the signal and disturbance parameters, which leads to a successive spatio-temporal whitening process. The second knowledge-aided parametric detector takes a joint approach for the estimation of the signal and disturbance parameters, which leads to a joint spatio-temporal whitening process. Both knowledge-aided parametric detectors are able to utilize prior knowledge about the spatial correlation through colored-loading that combines the sample covariance matrix with a prior covariance matrix. Computer simulation using various data sets, including the KASPPER dataset, show that the knowledge-aided parametric adaptive detectors yield improved detection performance over existing parametric solutions, especially in the case of limited data.**

***Index Terms*—Bayesian inference, generalized likelihood ratio test, knowledge-aided process, multi-channel auto-regressive model, space-time adaptive processing.**

I. Introduction

SPACE-TIME adaptive processing (STAP) has been successfully utilized to mitigate the effect of clutter and other interfering signals in radar, remote sensing, and communication systems [1], [2]. Traditional STAP detectors, such as the Reed, Mallett, and Brennan (RMB) detector [3], Kelly's generalized likelihood ratio test (GLRT) [4], the adaptive matched filter (AMF) [5], [6], the adaptive coherence estimator (ACE) [7] (also known as the normalized AMF [8]), the Rao test [9], etc., involve estimating and inverting the full-dimensional space-time covariance matrix, denoted by $\mathbf{R}$, of the disturbance signal. At a minimum, $K \geq JN$ training signals are needed to ensure a full-rank estimate of the $JN \times JN$ space-time covariance matrix $\mathbf{R}$, where J denotes the number of spatial channels and N the number of temporal observations. Meanwhile, the available training data is often limited in practice because of nonhomogeneous clutter background. Thus, it is necessary to reduce the training data requirements and computational complexity associated with traditional STAP detectors for practical applications.

In addition to the partially adaptive STAP techniques [1, ch. 4], reduced-rank STAP [10]–[13] and reduced-dimensional STAP [14], a class of parametric STAP detectors have recently been introduced and are shown to offer effective means of mitigating the requirement of excessive training signals and considerably reduce the associated computational complexity ([15]–[25] and reference therein). The parametric STAP detectors assume a multi-channel auto-regressive (AR) model for the disturbances in both test and training signals. Multi-channel AR modeling has extensively been studied, in particular for airborne radar signal detection, and its effectiveness has been shown using both simulated and measured datasets [15], [16], [18]–[20], [26]. Examples of the parametric STAP detectors are the parametric AMF (PAMF) [15], the parametric Rao (P-Rao) test [19], the parametric GLRT (P-GLRT) [20], and the simplified P-GLRT [25]. Meanwhile, extensions of the multi-channel AR modeling to nonstationary cases for STAP detection are investigated in [21]–[24].

Recently, knowledge-aided STAP (KA-STAP) has received much attention as it has been found that STAP detection can significantly be improved by exploiting prior knowledge of the disturbance signal [27]–[32]. A natural way to incorporate prior knowledge in solving the detection problem is a Bayesian approach that models the disturbance covariance matrix as a random matrix with some prior [33]–[38]. Some of these detectors are modified versions of the standard AMF or GLRT through diagonal or colored-loading [34]–[36], while others take more sophisticated forms such as the minimum mean-square estimate (MMSE) via the Gibbs sampling [36], [38]. These Bayesian STAP detectors, when compared with traditional STAP detectors, offer enhanced detection performance in the case of limited training data (due to contribution from prior knowledge) and are shown to be more robust in some nonhomogeneous background cases. However, the computational complexity of these detectors is still high, especially when the joint space-time dimension JN is large.

In this paper, we move a step further by considering a knowledge-aided parametric approach for the STAP detection. Specifically, we introduce a multi-channel AR process with a random cross-channel or spatial covariance matrix for the spatially and

Manuscript received March 22, 2011; revised July 14, 2011; accepted August 29, 2011. Date of publication September 15, 2011; date of current version November 16, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Biao Chen. This work 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.

P. Wang and H. Li are with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030 USA (e-mail: pwang4@stevens.edu; hli@stevens.edu).

B. Himed is with AFRL/RYMD, Dayton, OH 45433 USA (e-mail: braham.himed@wpafb.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.2011.2168220

temporally colored disturbance. Following a Bayesian framework, two knowledge-aided parametric adaptive detectors are developed, which are shown to be capable of utilizing prior knowledge as well as both the test and training signals for detection. The first knowledge-aided parametric detector, which is referred to as the knowledge-aided parametric AMF (KA-PAMF), is derived in a two-step manner. First, we develop a partially adaptive detector by assuming the AR coefficient matrices to be known. Second, the partially adaptive detector is modified to arrive at a fully adaptive detector by replacing the AR coefficient matrices with their estimates obtained from training signals. The closed-form test statistic reveals that the KA-PAMF employs a successively spatial whitening followed by temporal whitening, similar to the conventional PAMF detector [15] and the P-Rao test [19]. However, the spatial whitening matrix of the KA-PAMF is different and employs the prior knowledge. The second knowledge-aided parametric detector, referred to as the knowledge-aided parametric GLRT (KA-PGLRT), is obtained via a one-step joint estimation/optimization approach for both the signal and disturbance parameters. Since the exact maximum likelihood (ML) estimate of the target amplitude is intractable, we use the Schur complements and find an asymptotic but closed-form ML estimate of the amplitude. In turn, it enables us to find a simple and closed-form detection variable for the KA-PGLRT. The results show that the KA-PGLRT applies a joint spatio-and-*sub*temporal whitening with the ability to utilize the prior knowledge.

The remainder of this paper is organized as follows. Section II describes the proposed data model. Existing parametric STAP solutions are briefly reviewed in Section III. Section IV contains the derivation of the KA-PAMF detector obtained in a two-step procedure, while Section V contains the derivation of the KA-PGLRT detector involving joint estimation. Numerical results and conclusions are provided in Sections VI and VII, respectively.

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., [1]):

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

where all vectors are $J\times 1$ vectors, J denotes the number of spatial channels, and N is the number of temporal observations. In the sequel, $\mathbf{x}_0(n)$ is referred to as the test signal, $\mathbf{s}(n)$ the steering vector that is known to the detector, α the *unknown*, deterministic and complex-valued signal amplitude, and $\mathbf{d}_0(n)$ the disturbance signal that is correlated in space and time. For a uniform equi-spaced linear array, the steering vector is given as

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

where ω_d and ω_s denote the normalized target Doppler and spatial frequencies, respectively. It is also noted that the parametric detectors including the proposed detectors are able to handle other array configurations.

Besides the test signal $\mathbf{x}_0(n)$, there may be a set of training signals $\mathbf{x}_k(n), k=1,2,\ldots,K$, to assist in the signal detection:

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

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

$$\begin{aligned} H_0:\ & \mathbf{x}_0=\mathbf{d}_0 \\ H_1:\ & \mathbf{x}_0=\alpha\mathbf{s}+\mathbf{d}_0. \end{aligned} \tag{5}$$

We follow the standard assumptions [3]–[5], [7], [9] 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.

In this paper, a parametric approach is adopted, which uses the following assumption for the disturbance signal [15], [16], [19], [20], [25].

- **AS1—Multi-channel AR Model**: The disturbance signals $\mathbf{d}_k(n)$, $k=0,\ldots,K$, in the test and training signals are modeled as a J-channel $\mathrm{AR}(P)$ process 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 $\{\mathbf{A}^H(i)\}_{i=1}^P$ denote the *unknown* $J\times J$ AR coefficient matrices, $\boldsymbol{\varepsilon}_k(n)$ 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 addition to the above assumption, the spatial covariance matrix $\mathbf{Q}$ is assumed to be a random matrix with mean $\bar{\mathbf{Q}}$ that is known *a priori*:

- **AS2—Random Spatial Covariance Matrix:** The random spatial covariance matrix $\mathbf{Q}$ follows an inverse complex Wishart distribution with degrees of freedom μ and mean $\bar{\mathbf{Q}}$, i.e., $\mathbf{Q}\sim\mathcal{CW}^{-1}((\mu-J)\bar{\mathbf{Q}},\mu)$, (cf. [35], [36]):

$$p(\mathbf{Q})=\frac{|(\mu-J)\bar{\mathbf{Q}}|^{\mu}}{\tilde{\Gamma}(J,\mu)|\mathbf{Q}|^{(\mu+J)}}e^{-(\mu-J)\,\mathrm{tr}\,(\mathbf{Q}^{-1}\bar{\mathbf{Q}})} \tag{7}$$

  where

$$\tilde{\Gamma}(J,\mu)=\pi^{J(J-1)/2}\prod_{k=1}^{J}\Gamma(\mu-J+k), \tag{8}$$

  with Γ denoting the Gamma function, and $\mathrm{tr}(\cdot)$ denotes the matrix trace operator. The matrix $\bar{\mathbf{Q}}$ quantifies the prior knowledge about the disturbance from the probing environment. The parameter μ reflects the reliability of the prior knowledge $\bar{\mathbf{Q}}$. The larger μ is, the more reliable $\bar{\mathbf{Q}}$ is.

Some comments on the above model are in order. Multichannel AR processes have been extensively studied, using various experimentally collected radar data sets, and shown to provide effective modeling and characterization of radar clutter [15], [16], [18], [19]. In this paper, we introduce a stochastic AR modeling approach, whereby the AR coefficient matrices $\mathbf{A}(p)$ are deterministic, as in standard AR processes, but the spatial covariance matrix $\mathbf{Q}$ is random. Our approach is motivated by the following considerations. First, in several STAP scenarios, e.g., the coherent processing interval (CPI) is very short, the temporal clutter correlation, induced by the radar platform mobility, is often (approximately) stationary over the CPI, while the spatial clutter fluctuation could vary dramatically due to the heterogeneity of the ground terrain [2, Sec. 2.4.1], e.g., the boundary of distinct terrain, urban area, etc. Therefore, the random spatial parameters are needed to describe discrepancies between the "prior knowledge" and the received test/training data. Second, the model of a random spatial covariance matrix allows us to incorporate some *a priori* knowledge at the algorithmic design stage and develop knowledge-aided versions of the parametric STAP detector. In a multitude of recent studies, knowledge-aided processing has been proved as an efficient way to improve the detection performance over conventional STAP approaches [27]–[31]. Third, the prior knowledge on the overall space-time covariance matrix $\bar{\mathbf{R}}$ can be usually 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.; refer to [29], [31], [35], and [36] for more details. More specifically, in this paper, we may obtain the prior knowledge $\bar{\mathbf{Q}}$ in the following ways: 1) one can generate $\bar{\mathbf{Q}}$ directly from prior spatial information such as platform height, antenna look angle, transmit and receive beampatterns and past measurements of ground clutter [35]; 2) obtaining $\bar{\mathbf{Q}}$ by first performing a block Lower-triangle-Diagonal-Upper-triangle (LDU) matrix decomposition of the prior overall spatial-temporal covariance matrix $\bar{\mathbf{R}}$ and then averaging the last P block-diagonal matrices [39]; and 3) learning $\bar{\mathbf{Q}}$ by solving a multi-channel Levinson algorithm by using $\bar{\mathbf{R}}$ as the covariance matrix for the observation.

III. Existing Parametric Detectors

In this section, we provide a brief review of existing parametric detectors for easy reference and to facilitate latter discussions and comparisons. If the AR coefficient matrices $\mathbf{A} \triangleq [\mathbf{A}^T(1), \ldots, \mathbf{A}^T(P)]^T$ and the spatial covariance matrix $\mathbf{Q}$ are both known exactly, the clairvoyant parametric matched filter (PMF) takes the following test statistic [15]

$$T_{\mathrm{PMF}} = \frac{\left| \sum_{n=P}^{N-1} \tilde{\mathbf{s}}^H(n) \mathbf{Q}^{-1} \tilde{\mathbf{x}}_0(n) \right|^2}{\sum_{n=P}^{N-1} \tilde{\mathbf{s}}^H(n) \mathbf{Q}^{-1} \tilde{\mathbf{s}}(n)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_{\mathrm{PMF}} \tag{9}$$

where γ_{PMF} denotes the PMF threshold subject to a selected probability of false alarm, and $\tilde{\mathbf{s}}(n)$ and $\tilde{\mathbf{x}}_0(n)$ denote, respectively, the *temporally whitened* steering vector and the *temporally whitened* test signal using the *true* AR coefficient matrices $\mathbf{A}$; see [19, eqs. (9) and (10)].

The PAMF detector replaces $\mathbf{A}$ and $\mathbf{Q}$ in the above statistic by their maximum likelihood (ML) estimates $\hat{\mathbf{A}}$ and $\hat{\mathbf{Q}}$ obtained only from training signals $\mathbf{x}_k(n), k = 1, \ldots, K$. The PAMF takes the following test statistic [15]:

$$T_{\mathrm{PAMF}} = \frac{\left| \sum_{n=P}^{N-1} \hat{\tilde{\mathbf{s}}}^H(n) \hat{\mathbf{Q}}^{-1} \hat{\tilde{\mathbf{x}}}_0(n) \right|^2}{\sum_{n=P}^{N-1} \hat{\tilde{\mathbf{s}}}^H(n) \hat{\mathbf{Q}}^{-1} \hat{\tilde{\mathbf{s}}}(n)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_{\mathrm{PAMF}} \tag{10}$$

where the temporally whitened steering vector $\hat{\tilde{\mathbf{s}}}(n)$ and test signal $\hat{\tilde{\mathbf{x}}}_0(n)$ are obtained *adaptively* by using the ML estimates $\hat{\mathbf{A}}$ and $\hat{\mathbf{Q}}$ *adaptively* from training signals only (see [15, Sec. V]). Meanwhile, the P-Rao test was developed from the general Rao test principle in [19] and coincides with the PAMF (10) except that it uses both the test and training signals for the ML estimates of $\mathbf{A}$ and $\mathbf{Q}$; see [19, eqs. (23)–(25)]. From (10), we conclude that the PAMF and the P-Rao test perform successively temporal whitening (i.e., $\hat{\tilde{\mathbf{x}}}_0(n)$ and $\hat{\tilde{\mathbf{s}}}(n)$) followed by a spatial whitening (i.e., $\hat{\mathbf{Q}}^{-1}$).

The P-GLRT was developed in [20] as an improved detector with respect to the P-Rao test. However, the P-GLRT involves a highly nonlinear cost function with respect to the target amplitude α. Newton-like iterative nonlinear two-dimensional searches were employed to find the ML amplitude estimate, which is computationally intensive and also suffers from local convergence. To overcome this difficulty, the simplified P-GLRT was proposed in [25], and its closed-form test statistic sheds more insights on how it suppresses the disturbance. The observation in [25] concludes that the P-GLRT performs a spatio-and-*sub*temporal whitening across the $J(P+1)$ dimensions by using a $J(P+1) \times J(P+1)$ whitening matrix.

A Bayesian PAMF (B-PAMF) detector has recently been introduced in [39] to handle nonhomogeneous environments, where the covariance matrices of the training and test signals are assumed to be different from each other. The B-PAMF takes the following test statistic:

$$T_{\mathrm{B-PAMF}} = \frac{\left| \sum_{n=P}^{N-1} \hat{\tilde{\mathbf{s}}}^H(n) \hat{\mathbf{Q}}_{\mathrm{MAP}}^{-1} \hat{\tilde{\mathbf{x}}}_0(n) \right|^2}{\sum_{n=P}^{N-1} \hat{\tilde{\mathbf{s}}}^H(n) \hat{\mathbf{Q}}_{\mathrm{MAP}}^{-1} \hat{\tilde{\mathbf{s}}}(n)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_{\mathrm{B-PAMF}} \tag{11}$$

where the temporally whitened steering vector $\hat{\tilde{\mathbf{s}}}(n)$ and test signal $\hat{\tilde{\mathbf{x}}}_0(n)$ are obtained in the same way as in the PAMF, and the spatial whitening matrix $\hat{\mathbf{Q}}_{\mathrm{MAP}}$ is obtained by finding the maximum *a posteriori* (MAP) estimate of $\mathbf{Q}$ [39, eq. (24)]. As a result, the B-PAMF also performs the successively temporal and spatial whitening process against the disturbance, like the PAMF and P-Rao test, but it enables a knowledge-aided solution through the MAP estimate of $\mathbf{Q}$ for the spatial whitening.

In summary, the existing parametric detectors except the B-PAMF are unable to utilize any prior knowledge of the spatial covariance matrix. In the following, we present two

knowledge-aided parametric detectors, which are based on the stochastic AR model introduced in the previous section and derived from a Bayesian framework. By integrating out the uncertainty of the spatial covariance matrix, both knowledge-aided parametric detectors take simple closed-form expressions which admit clear and intuitive interpretation on how the prior knowledge is incorporated.

IV. Knowledge-Aided PAMF

The first knowledge-aided parametric detector, referred to as the KA-PAMF detector, employs a two-step procedure, similar to the AMF detector [5, Sec. II]. Specifically, we first derive a *partially adaptive* P-GLRT by assuming $\mathbf{A}$ is known; and then replace $\mathbf{A}$ in the resulting test variable by its ML estimate from training signals only.

A. Partially Adaptive P-GLRT With Known $\mathbf{A}$

With a known $\mathbf{A}$, the partially adaptive P-GLRT is obtained by integrating out the random spatial covariance matrix $\mathbf{Q}$ in the likelihood ratio:

$$T = \frac{\max\limits_{\alpha} \int f_1(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_K \,|\, \alpha, \mathbf{Q}) p(\mathbf{Q}) d\mathbf{Q}}{\int f_0(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_K \,|\, \alpha = 0, \mathbf{Q}) p(\mathbf{Q}) d\mathbf{Q}} \tag{12}$$

where the matrix integral is performed with respect to the set of positive definite Hermitian matrices, and

$$f_i(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_K \,|\, \alpha, \mathbf{Q}) = \left[\frac{1}{\pi^J |\mathbf{Q}|} \exp\left\{-\mathrm{tr}\left(\mathbf{Q}^{-1}\boldsymbol{\Sigma}_i(\alpha)\right)\right\}\right]^{(K+1)(N-P)} \tag{13}$$

with

$$\boldsymbol{\Sigma}_i(\alpha) = \frac{1}{(K+1)(N-P)} \sum_{k=0}^{K} \sum_{n=P}^{N-1} \boldsymbol{\varepsilon}_k(n) \boldsymbol{\varepsilon}_k^H(n) \tag{14}$$

and

$$\boldsymbol{\varepsilon}_0(n) = \mathbf{x}_0(n) - \alpha \mathbf{s}(n) + \sum_{p=1}^{P} \mathbf{A}^H(p)[\mathbf{x}_0(n-p) - \alpha \mathbf{s}(n-p)] \tag{15}$$

$$\boldsymbol{\varepsilon}_k(n) = \mathbf{x}_k(n) + \sum_{p=1}^{P} \mathbf{A}^H(p) \mathbf{x}_k(n-p). \tag{16}$$

Note that $\alpha = 0$ if $i = 0$ (i.e., H_0) and $\alpha \neq 0$ if $i = 1$.

It can be shown that the integration in (12) is given by [33]

$$\int f_i(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_K \,|\, \alpha, \mathbf{Q}) p(\mathbf{Q}) d\mathbf{Q} = \frac{|(\mu - J)\bar{\mathbf{Q}}|^{\mu} \tilde{\Gamma}(J, L-J)}{\pi^{J(K+1)(N-P)} \tilde{\Gamma}(J, \mu)} |\boldsymbol{\Xi}_i|^{-L+J} \tag{17}$$

where $L = (K+1)(N-P) + (\mu + J)$ and

$$\begin{aligned} \boldsymbol{\Xi}_i &= (K+1)(N-P)\boldsymbol{\Sigma}_i + (\mu - J)\bar{\mathbf{Q}} \\ &= \sum_{k=0}^{K} \sum_{n=P}^{N-1} \boldsymbol{\varepsilon}_k(n) \boldsymbol{\varepsilon}_k^H(n) + (\mu - J)\bar{\mathbf{Q}} \\ &= \hat{\mathbf{R}}_{xx}(\alpha) + \hat{\mathbf{R}}_{yx}^H(\alpha)\mathbf{A} + \mathbf{A}^H \hat{\mathbf{R}}_{yx}(\alpha) \\ &\quad + \mathbf{A}^H \hat{\mathbf{R}}_{yy}(\alpha)\mathbf{A} + (\mu - J)\bar{\mathbf{Q}}. \end{aligned} \tag{18}$$

The correlation matrices in (18) are

$$\hat{\mathbf{R}}_{xx}(\alpha) = \sum_{n=P}^{N-1} [\mathbf{x}_0(n) - \alpha \mathbf{s}(n)][\mathbf{x}_0(n) - \alpha \mathbf{s}(n)]^H + \sum_{k=1}^{K} \sum_{n=P}^{N-1} \mathbf{x}_k(n) \mathbf{x}_k^H(n) \tag{19}$$

$$\hat{\mathbf{R}}_{yy}(\alpha) = \sum_{n=P}^{N-1} [\mathbf{y}_0(n) - \alpha \mathbf{t}(n)][\mathbf{y}_0(n) - \alpha \mathbf{t}(n)]^H + \sum_{k=1}^{K} \sum_{n=P}^{N-1} \mathbf{y}_k(n) \mathbf{y}_k^H(n) \tag{20}$$

$$\hat{\mathbf{R}}_{yx}(\alpha) = \sum_{n=P}^{N-1} [\mathbf{y}_0(n) - \alpha \mathbf{t}(n)][\mathbf{x}_0(n) - \alpha \mathbf{s}(n)]^H + \sum_{k=1}^{K} \sum_{n=P}^{N-1} \mathbf{y}_k(n) \mathbf{x}_k^H(n) \tag{21}$$

where the $JP \times 1$ vectors $\mathbf{y}_k(n)$ and $\mathbf{t}(n)$ are defined as

$$\begin{aligned} \mathbf{y}_k(n) &= \left[\mathbf{x}_k^T(n-1), \ldots, \mathbf{x}_k^T(n-P)\right]^T \\ \mathbf{t}(n) &= [\mathbf{s}^T(n-1), \ldots, \mathbf{s}^T(n-P)]^T. \end{aligned}$$

From (12) and (17), the partially adaptive P-GLRT takes the following intermediate test statistic

$$T = \frac{\max\limits_{\alpha} |\boldsymbol{\Xi}_1|^{-L+J}}{|\boldsymbol{\Xi}_0|^{-L+J}} = \left(\frac{|\boldsymbol{\Xi}_0|}{\min\limits_{\alpha} |\boldsymbol{\Xi}_1|}\right)^{L-J}. \tag{22}$$

Next, we need to compute the determinant of $\boldsymbol{\Xi}_i$. Let

$$\tilde{\mathbf{x}}_k(n) = \mathbf{x}_k(n) + \mathbf{A}^H \mathbf{y}_k(n) \tag{23}$$

$$\tilde{\mathbf{s}}(n) = \mathbf{s}(n) + \mathbf{A}^H \mathbf{t}(n) \tag{24}$$

$$\tilde{\mathbf{S}} \in \mathbb{C}^{J \times (N-P)} = [\tilde{\mathbf{s}}(P), \ldots, \tilde{\mathbf{s}}(N-1)] \tag{25}$$

$$\tilde{\mathbf{X}}_k \in \mathbb{C}^{J \times (N-P)} = [\tilde{\mathbf{x}}_k(P), \ldots, \tilde{\mathbf{x}}_k(N-1)]. \tag{26}$$

Then, from (18), $|\boldsymbol{\Xi}_i|$ can be obtained as

$$\begin{aligned} |\boldsymbol{\Xi}_i| &= \left| (\tilde{\mathbf{X}}_0 - \alpha \tilde{\mathbf{S}})(\tilde{\mathbf{X}}_0 - \alpha \tilde{\mathbf{S}})^H + \sum_{k=1}^{K} \tilde{\mathbf{X}}_k \tilde{\mathbf{X}}_k^H + (\mu - J)\bar{\mathbf{Q}} \right| \\ &= |(\tilde{\mathbf{X}}_0 \mathbf{P} - \alpha \tilde{\mathbf{S}})(\tilde{\mathbf{X}}_0 \mathbf{P} - \alpha \tilde{\mathbf{S}})^H + \boldsymbol{\Psi}| \\ &= |(\tilde{\mathbf{X}}_0 \mathbf{P} - \alpha \tilde{\mathbf{S}})(\tilde{\mathbf{X}}_0 \mathbf{P} - \alpha \tilde{\mathbf{S}})^H \boldsymbol{\Psi}^{-1} + \mathbf{I}| \cdot |\boldsymbol{\Psi}| \end{aligned} \tag{27}$$

where

$$\boldsymbol{\Psi} = \tilde{\mathbf{X}}_0 \mathbf{P}^{\perp} \tilde{\mathbf{X}}_0^H + \sum_{k=1}^{K} \tilde{\mathbf{X}}_k \tilde{\mathbf{X}}_k^H + (\mu - J)\bar{\mathbf{Q}} \tag{28}$$

$\mathbf{P}^{\perp}$ denotes the projection matrix operator which projects the signal to the orthogonal complement of the range of $\tilde{\mathbf{S}}^H$:

$$\mathbf{P}^{\perp} = \mathbf{I} - \mathbf{P} = \mathbf{I} - \tilde{\mathbf{S}}^H (\tilde{\mathbf{S}}^H)^{\dagger} \tag{29}$$

and $(\cdot)^{\dagger}$ denotes the Moore–Penrose pseudo-inverse of the argument. Given (27), the ML estimate of α which minimizes (27) is given by [40, Lemma 4]

$$\hat{\alpha}_{\mathrm{ML}} = \frac{\mathrm{tr}\,(\tilde{\mathbf{S}}^H \boldsymbol{\Psi}^{-1} \tilde{\mathbf{X}}_0)}{\mathrm{tr}\,(\tilde{\mathbf{S}}^H \boldsymbol{\Psi}^{-1} \tilde{\mathbf{S}})}. \tag{30}$$

With the ML estimate of α, as shown in the Appendix, the partially adaptive P-GLRT takes the following test statistic in a compact form:

$$T = \frac{\left| \sum_{n=P}^{N-1} \tilde{\mathbf{s}}^H(n) \boldsymbol{\Psi}^{-1} \tilde{\mathbf{x}}_0(n) \right|^2}{\sum_{n=P}^{N-1} \tilde{\mathbf{s}}^H(n) \boldsymbol{\Psi}^{-1} \tilde{\mathbf{s}}(n)}, \tag{31}$$

where the temporally whitened vectors $\tilde{\mathbf{x}}_0(n)$ and $\tilde{\mathbf{s}}(n)$ are obtained by (23) and (24), and the spatially whitening matrix $\boldsymbol{\Psi}$ is obtained by (28).

B. KA-PAMF: Fully Adaptive P-GLRT With the ML Estimate of $\mathbf{A}$ *Obtained From Training Signals*

The above partially adaptive P-GLRT assumes knowledge of $\mathbf{A}$. To arrive at a fully adaptive P-GLRT, an ML estimate of $\mathbf{A}$ is obtained by using the training signals. Since the spatial covariance matrix is random, the likelihood function conditioned on $\mathbf{A}$ is obtained by integrating the likelihood function over the distribution of $\mathbf{Q}$ (see **AS2**) as

$$\begin{aligned} & f(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_K \,|\, \mathbf{A}) \\ &\quad = \int f(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_K \,|\, \mathbf{A}, \mathbf{Q}) p(\mathbf{Q})\, d\mathbf{Q} \\ &\quad \propto \left| \boldsymbol{\Sigma}(\mathbf{A}) + (\mu - J)\bar{\mathbf{Q}} \right|^{-(\mu + K(N-P))} \end{aligned} \tag{32}$$

where

$$\boldsymbol{\Sigma}(\mathbf{A}) = \sum_{k=1}^{K} \sum_{n=P}^{N-1} \boldsymbol{\varepsilon}_k(n) \boldsymbol{\varepsilon}_k^H(n). \tag{33}$$

Therefore, the ML estimate of $\mathbf{A}$ is equivalent to minimizing the determinant in (32). Rewrite the matrix as

$$\begin{aligned} & \Sigma(\mathbf{A}) + (\mu - J)\bar{\mathbf{Q}} \\ &\quad = \left(\hat{\mathbf{R}}_{xx} + \hat{\mathbf{R}}_{yx}^H \mathbf{A} + \mathbf{A}^H \hat{\mathbf{R}}_{yx} + \mathbf{A}^H \hat{\mathbf{R}}_{yy} \mathbf{A} \right) \\ &\qquad + (\mu - J)\bar{\mathbf{Q}} \\ &\quad = \left(\mathbf{A}^H + \hat{\mathbf{R}}_{yx}^H \hat{\mathbf{R}}_{yy}^{-1} \right) \hat{\mathbf{R}}_{yy} \left(\mathbf{A}^H + \hat{\mathbf{R}}_{yx}^H \hat{\mathbf{R}}_{yy}^{-1} \right)^H \\ &\qquad + \hat{\mathbf{R}}_{xx} - \hat{\mathbf{R}}_{yx}^H \hat{\mathbf{R}}_{yy}^{-1} \hat{\mathbf{R}}_{yx} + (\mu - J)\bar{\mathbf{Q}} \end{aligned} \tag{34}$$

where $\hat{\mathbf{R}}_{xx}, \hat{\mathbf{R}}_{yy}$ and $\hat{\mathbf{R}}_{yx}$ are defined as the second term (training signals only), respectively, in (19), (20), and (21). Since $\hat{\mathbf{R}}_{yy}$ is nonnegative definite and the remaining term $\hat{\mathbf{R}}_{xx} - \hat{\mathbf{R}}_{yx}^H \hat{\mathbf{R}}_{yy}^{-1} \hat{\mathbf{R}}_{yx} + (\mu - J)\bar{\mathbf{Q}}$ does not depend on $\mathbf{A}$, it follows that

$$\boldsymbol{\Sigma}(\mathbf{A}) + (\mu - J)\bar{\mathbf{Q}} \geq \boldsymbol{\Sigma}(\hat{\mathbf{A}}_{\mathrm{ML}}) + (\mu - J)\bar{\mathbf{Q}} \tag{35}$$

where

$$\hat{\mathbf{A}}_{\mathrm{ML}} = -\hat{\mathbf{R}}_{yy}^{-1} \hat{\mathbf{R}}_{yx} \tag{36}$$

which coincides with the ML estimate of $\mathbf{A}$ used in the conventional PAMF. When the above matrix is minimized, the ML estimate of $\mathbf{A}$ also minimizes any nondecreasing function including the determinant.

Replacing $\mathbf{A}$ in (31) with the ML estimate $\hat{\mathbf{A}}_{\mathrm{ML}}$, the fully adaptive KA-PAMF takes the following test statistic:

$$T_{\mathrm{KA-PAMF}} = \frac{\left| \sum_{n=P}^{N-1} \hat{\tilde{\mathbf{s}}}^H(n) \hat{\boldsymbol{\Psi}}^{-1} \hat{\tilde{\mathbf{x}}}_0(n) \right|^2}{\sum_{n=P}^{N-1} \hat{\tilde{\mathbf{s}}}^H(n) \hat{\boldsymbol{\Psi}}^{-1} \hat{\tilde{\mathbf{s}}}(n)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_{\mathrm{KA-PAMF}} \tag{37}$$

where $\gamma_{\mathrm{KA-PAMF}}$ is a threshold subject to the probability of false alarm, the temporally whitened vectors $\hat{\tilde{\mathbf{x}}}_k(n)$ and $\hat{\tilde{\mathbf{s}}}(n)$ are similarly obtained as in (23) and (24) with $\mathbf{A}$ replaced by the ML estimate $\hat{\mathbf{A}}_{\mathrm{ML}}$, and the spatially whitening matrix $\hat{\boldsymbol{\Psi}}$ is obtained adaptively:

$$\hat{\boldsymbol{\Psi}} = \hat{\tilde{\mathbf{X}}}_0 \hat{\mathbf{P}}^{\perp} \hat{\tilde{\mathbf{X}}}_0^H + \sum_{k=1}^{K} \hat{\tilde{\mathbf{X}}}_k \hat{\tilde{\mathbf{X}}}_k^H + (\mu - J)\bar{\mathbf{Q}} \tag{38}$$

which is similarly defined as (28).

A comparison of (37) and (10) indicates that the KA-PAMF is similar to the PAMF using a successive spatio-temporal whitening process. However, different from the PAMF where the spatial whitening matrix $\hat{\mathbf{Q}}$ is formed from only training signals, $\hat{\boldsymbol{\Psi}}$ in the KA-PAMF includes also contributions from the target-canceled test signal (the first term in (38)) and the prior knowledge $(\mu - J)\bar{\mathbf{Q}}$. The inclusion of the latter allows the KA-PAMF to better handle cases with limited training data. The proposed KA-PAMF is also different from the B-PAMF of (11). In particular, the B-PAMF involves a more complicated loading process with two loading factors [39, eq. (24)]. In contrast, the KA-PAMF uses a more intuitive loading formula with only one loading factor. Furthermore, the inclusion of the target-free test signal component in the KA-PAMF is a new feature, as compared with the B-PAMF.

We now discuss the complexity of the KA-PAMF. On one hand, the proposed detector is in general significantly simpler than the covariance-matrix based STAP detectors [4], [5], [7], which performs a fully spatial and temporal whitening, with a complexity in order of $\mathcal{O}(J^3N^3)$, especially when the dimension JN is large; see [15], [41] for a comparison of the complexities of covariance matrix based and parametric STAP detectors. On the other hand, the complexity of the KA-PAMF is comparable to that of the conventional PAMF and P-Rao test. As discussed above, the difference lies on the spatial whitening process. Specifically, the complexity of the spatial whitening of the PAMF and the P-Rao test is about $\mathcal{O}(J^2KNP^2)$, while the complexity of computing (38) is about $\mathcal{O}(J^2KNP^2) + \mathcal{O}(JN^2)$. The additional complexity of $\mathcal{O}(JN^2)$ for the KA-PAMF mostly comes from the calculation of the spatial whitening matrix $\hat{\boldsymbol{\Psi}}$ and its inherent Moore–Penrose pseudo-inverse in (29).

V. KNOWLEDGE-AIDED PARAMETRIC GLRT

In this section, different from the two-step procedure of the KA-PAMF detector, the KA-PGLRT detector is fully developed within a Bayesian framework. This approach enables the ML estimate of $\mathbf{A}$ to utilize the test signal, in addition to the training signals, and enhances the final detection performance, although at some higher complexity. Specifically, the KA-PGLRT detector takes the following test statistic:

$$T = \frac{\max\limits_{\alpha}\max\limits_{\mathbf{A}} \int f_1(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_K \,|\, \alpha, \mathbf{A}, \mathbf{Q}) p(\mathbf{Q}) d\mathbf{Q}}{\max\limits_{\mathbf{A}} \int f_0(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_K \,|\, \mathbf{A}, \mathbf{Q}) p(\mathbf{Q}) d\mathbf{Q}}. \tag{39}$$

The integration over the prior distribution of $\mathbf{Q}$ can be similarly performed as (17). As a result, (39) can be rewritten as

$$T = \frac{\max\limits_{\alpha}\max\limits_{A} |\mathbf{\Xi}_1|^{-L+J}}{\max\limits_{A} |\mathbf{\Xi}_0|^{-L+J}} = \left(\frac{\min\limits_{A} |\mathbf{\Xi}_0|}{\min\limits_{\alpha}\min\limits_{A} |\mathbf{\Xi}_1|} \right)^{L-J} \tag{40}$$

where $\mathbf{\Xi}_i, i = 0, 1$, is defined in (18).

Following (40), since $\mathbf{A}$ is unknown, we need first to maximize the determinant of $\mathbf{\Xi}_i$ with respect to $\mathbf{A}$, and then find the ML estimate of α under H_1. The maximization of $|\mathbf{\Xi}_i|$ with respect to $\mathbf{A}$ yields [20]

$$\hat{\mathbf{A}}_{\mathrm{ML}} = -\hat{\mathbf{R}}_{yy}^{-1}(\alpha)\hat{\mathbf{R}}_{yx}(\alpha), \tag{41}$$

where $\hat{\mathbf{R}}_{yy}(\alpha)$ and $\hat{\mathbf{R}}_{yx}(\alpha)$ are defined in (20) and (21), respectively. By substituting the ML estimate of $\mathbf{A}$ back into (40), the KA-PGLRT test statistics is equivalent to

$$T = \frac{\left| \hat{\mathbf{R}}_{xx}(0) - \hat{\mathbf{R}}_{yx}^{H}(0)\hat{\mathbf{R}}_{yy}^{-1}(0)\hat{\mathbf{R}}_{yx}(0) + (\mu - J)\bar{\mathbf{Q}} \right|}{\min\limits_{\alpha} \left| \hat{\mathbf{R}}_{xx}(\alpha) - \hat{\mathbf{R}}_{yx}^{H}(\alpha)\hat{\mathbf{R}}_{yy}^{-1}(\alpha)\hat{\mathbf{R}}_{yx}(\alpha) + (\mu - J)\bar{\mathbf{Q}} \right|}. \tag{42}$$

The next step is to find the ML estimate of α, which is the solution to the following minimization problem:

$$\hat{\alpha}_{\mathrm{ML}} = \arg\min_{\alpha} \left| \hat{\mathbf{R}}_{xx}(\alpha) - \hat{\mathbf{R}}_{yx}^{H}(\alpha)\hat{\mathbf{R}}_{yy}^{-1}(\alpha)\hat{\mathbf{R}}_{yx}(\alpha) + (\mu - J)\bar{\mathbf{Q}} \right|. \tag{43}$$

The cost function in the above equation is highly nonlinear with respect to α and the exact ML estimate of α turns to be intractable. Since a computationally exhaustive search over the two-dimensional parameter space (i.e., the real and imaginary parts of α) is generally impractical, we resort to an asymptotic ML estimate which leads to a simple closed-form solution of α.

The asymptotic ML estimate of α relies on the recognition that [25]

$$[\hat{\mathbf{R}}_{xx}(\alpha) + (\mu - J)\bar{\mathbf{Q}}] - \hat{\mathbf{R}}_{yx}^{H}(\alpha)\hat{\mathbf{R}}_{yy}^{-1}(\alpha)\hat{\mathbf{R}}_{yx}(\alpha) \tag{44}$$

in (43) is the Schur complement of $\hat{\mathbf{R}}_{yy}(\alpha)$ which is a part of a block matrix $\hat{\mathbf{R}}(\alpha)$ defined as

$$\hat{\mathbf{R}}(\alpha) = \begin{bmatrix} \hat{\mathbf{R}}_{yy}(\alpha) & \hat{\mathbf{R}}_{yx}(\alpha) \\ \hat{\mathbf{R}}_{yx}^{H}(\alpha) & \hat{\mathbf{R}}_{xx}(\alpha) + (\mu - J)\bar{\mathbf{Q}} \end{bmatrix}. \tag{45}$$

A similar technique was used in [25, eq. (13)] for the case without the prior knowledge $(\mu - J)\bar{\mathbf{Q}}$. According to the property of the Schur complement, the cost function in (43) is equivalent to

$$\ln|\hat{\mathbf{R}}(\alpha)| - \ln|\hat{\mathbf{R}}_{yy}(\alpha)|. \tag{46}$$

To explicitly express the above function as a function of α, we decompose $\hat{\mathbf{R}}(\alpha)$ and $\hat{\mathbf{R}}_{yy}(\alpha)$ into an α-independent component and an α-dependent one as follows:

$$\hat{\mathbf{R}}(\alpha) = (\mathbf{X}_0 - \alpha\mathbf{S})(\mathbf{X}_0 - \alpha\mathbf{S})^H + \sum_{k=1}^{K} \mathbf{X}_k\mathbf{X}_k^H + \tilde{\mathbf{Q}} \tag{47}$$

$$\hat{\mathbf{R}}_{yy}(\alpha) = (\mathbf{Y}_0 - \alpha\mathbf{T})(\mathbf{Y}_0 - \alpha\mathbf{T})^H + \sum_{k=1}^{K} \mathbf{Y}_k\mathbf{Y}_k^H \tag{48}$$

where

$$\begin{aligned} \mathbf{S} \in \mathbb{C}^{J(P+1)\times(N-P)} &= [\mathbf{s}_{P+1}(P), \ldots, \mathbf{s}_{P+1}(N-1)], \\ \mathbf{X}_k &= [\mathbf{x}_{k,P+1}(P), \ldots, \mathbf{x}_{k,P+1}(N-1)] \\ \mathbf{T} \in \mathbb{C}^{JP\times(N-P)} &= [\mathbf{t}(P), \ldots, \mathbf{t}(N-1)], \\ \mathbf{Y}_k &= [\mathbf{y}_k(P), \ldots, \mathbf{y}_k(N-1)] \end{aligned}$$

and

$$\tilde{\mathbf{Q}} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (\mu - J)\bar{\mathbf{Q}} \end{bmatrix}, \tag{49}$$

with the following definitions:

$$\mathbf{s}_{P+1}(n) = [\mathbf{t}^T(n), \mathbf{s}^T(n)]^T \tag{50}$$

$$\mathbf{x}_{k,P+1}(n) = \left[\mathbf{y}_k^T(n), \mathbf{x}_k^T(n)\right]^T. \tag{51}$$

Define the following orthogonal projection matrices: $\mathbf{P}_S = \mathbf{S}^H(\mathbf{S}^H)^{\dagger}, \mathbf{P}_T = \mathbf{T}^H(\mathbf{T}^H)^{\dagger}, \mathbf{P}_S^{\perp} = \mathbf{I} - \mathbf{P}_S$, and $\mathbf{P}_T^{\perp} = \mathbf{I} - \mathbf{P}_T$. The determinants of $\hat{\mathbf{R}}(\alpha)$ and $\hat{\mathbf{R}}_{yy}(\alpha)$ can be computed as follows:

$$\begin{aligned} |\hat{\mathbf{R}}(\alpha)| &= \left| (\mathbf{X}_0 - \alpha\mathbf{S})(\mathbf{X}_0 - \alpha\mathbf{S})^H + \sum_{k=1}^{K} \mathbf{X}_k\mathbf{X}_k^H + \tilde{\mathbf{Q}} \right| \\ &= \left| (\mathbf{X}_0 - \alpha\mathbf{S}) \left(\mathbf{P}_S + \mathbf{P}_S^{\perp}\right) (\mathbf{X}_0 - \alpha\mathbf{S})^H \right. \\ &\quad \left. + \sum_{k=1}^{K} \mathbf{X}_k\mathbf{X}_k^H + \tilde{\mathbf{Q}} \right| \\ &\triangleq |(\mathbf{X}_0\mathbf{P}_S - \alpha\mathbf{S})(\mathbf{X}_0\mathbf{P}_S - \alpha\mathbf{S})^H + \hat{\mathbf{R}}_X| \\ &= \left| (\mathbf{X}_0\mathbf{P}_S - \alpha\mathbf{S})(\mathbf{X}_0\mathbf{P}_S - \alpha\mathbf{S})^H\hat{\mathbf{R}}_X^{-1} + \mathbf{I} \right| \cdot |\hat{\mathbf{R}}_X| \end{aligned} \tag{52}$$

and, similarly,

$$\begin{aligned} &|\hat{\mathbf{R}}_{yy}(\alpha)| \\ &\quad = \left| (\mathbf{Y}_0\mathbf{P}_T - \alpha\mathbf{T})(\mathbf{Y}_0\mathbf{P}_T - \alpha\mathbf{T})^H\hat{\mathbf{R}}_Y^{-1} + \mathbf{I} \right| \cdot |\hat{\mathbf{R}}_Y| \end{aligned} \tag{53}$$

where

$$\hat{\mathbf{R}}_X = \mathbf{X}_0\mathbf{P}_S^{\perp}\mathbf{X}_0^H + \sum_{k=1}^{K}\mathbf{X}_k\mathbf{X}_k^H + \tilde{\mathbf{Q}} \tag{54}$$

$$\hat{\mathbf{R}}_Y = \mathbf{Y}_0\mathbf{P}_T^{\perp}\mathbf{Y}_0^H + \sum_{k=1}^{K}\mathbf{Y}_k\mathbf{Y}_k^H. \tag{55}$$

It is noted that $\hat{\mathbf{R}}_X$ of (54) is a colored-loading version of the original $\hat{\mathbf{R}}_X$ in the simplified P-GLRT, i.e., [25, eq. (24)], whereas $\hat{\mathbf{R}}_Y$ of (55) is the same as the one of the simplified P-GLRT, i.e., [25, eq. (25)]. Following a similar procedure as in [25, App. I and II], the ML estimate of α can be obtained asymptotically as

$$\hat{\alpha}_{\mathrm{ML}} = \frac{\mathrm{tr}\left\{\mathbf{S}^H\hat{\mathbf{R}}_X^{-1}\mathbf{X}_0\right\} - \mathrm{tr}\left\{\mathbf{T}^H\hat{\mathbf{R}}_Y^{-1}\mathbf{Y}_0\right\}}{\mathrm{tr}\left\{\mathbf{S}^H\hat{\mathbf{R}}_X^{-1}\mathbf{S}\right\} - \mathrm{tr}\left\{\mathbf{T}^H\hat{\mathbf{R}}_Y^{-1}\mathbf{T}\right\}}. \tag{56}$$

Taking the above ML estimate of α back into the likelihood ratio of (42) followed by simplifications (similar to Appendix III of [25]), the KA-PGLRT takes the final test statistic as

$$T_{\mathrm{KA-PGLR}} = \frac{\left|\sum_{n=P}^{N-1}\mathbf{s}_{P+1}^H(n)\hat{\mathbf{R}}_X^{-1}\mathbf{x}_{0,P+1}(n) - \sum_{n=P}^{N-1}\mathbf{t}^H(n)\hat{\mathbf{R}}_Y^{-1}\mathbf{y}_0(n)\right|^2}{\sum_{n=P}^{N-1}\mathbf{s}_{P+1}^H(n)\hat{\mathbf{R}}_X^{-1}\mathbf{s}_{P+1}(n) - \sum_{n=P}^{N-1}\mathbf{t}^H(n)\hat{\mathbf{R}}_Y^{-1}\mathbf{t}(n)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_{\mathrm{KA-PGLR}} \tag{57}$$

where $\gamma_{\mathrm{KA-PGLR}}$ is a threshold subject to a probability of false alarm.

It is seen from (57) that the KA-PGLRT detector extends the simplified P-GLRT of [25, eq. (28)] through a knowledge-aided colored-loading step via $\hat{\mathbf{R}}_X$ of (54) to incorporate the prior $\bar{\mathbf{Q}}$. It can further be shown, similarly as in [25, App. IV], that the KA-PGLRT shares the same interpretation of the simplified P-GLRT as a spatio-and-subtemporal whitening across $J(P+1)$ dimensions. The essential difference is that the KA-PGLRT is able to use the prior knowledge through a linear combination of the conventional estimate and the prior knowledge. This is helpful for the KA-PGLRT to improve the detection performance when the number of pulses is comparable to the number of elements, as verified in Section VI. The performance improvement of the KA-PGLRT is achieved with slightly additional complexity, due to the construction of $\tilde{\mathbf{Q}}$ in (49) and the addition of $\tilde{\mathbf{Q}}$ in (54). Overall, the computational complexity of the KA-PGLRT is similar to that of the conventional P-GLRT.

The spatio-and-subtemporal whitening of the KA-PGLRT is in between a fully adaptive STAP detector such as the AMF detector [5] that employs a joint spatio-temporal whitening across all JN dimensions and the conventional PAMF [15] or the P-Rao test [19] that utilizes successively temporal whitening followed by spatial whitening. In terms of the complexity of the whitening process, the KA-PGLRT is about $\mathcal{O}(J^2KN(P+1)^2) + \mathcal{O}(J(P+1)N^2)$, which is more than $\mathcal{O}(J^2KNP^2) + \mathcal{O}(JN^2)$ of the KA-PAMF, mainly due to the larger dimension $J(P+1)\times(N-P)$ of $\mathbf{S}$ in (54) compared with the dimension $J\times(N-P)$ of $\tilde{\mathbf{S}}$ in (25).

VI. Numerical Results

Simulation results are provided to verify the proposed knowledge-aided parametric detectors. We compare the proposed detectors with the following for several different scenarios:

- the **PMF** of (9) that uses $\bar{\mathbf{Q}}$ as a nonadaptive estimate of $\mathbf{Q}$;
- the standard **PAMF** of (10) that employs a spatial covariance matrix estimate $\hat{\mathbf{Q}}$ obtained from training signals [15];
- the Bayesian PAMF (**B-PAMF**) of (11) [39];
- the **simplified P-GLRT** [25].

Two types of testing data are employed. In the first, the disturbance signal $\mathbf{d}_k$ is generated as a multi-channel AR(2) process with AR coefficients $\mathbf{A}$ and a spatial covariance matrix $\mathbf{Q}$. In the second, we consider the widely used KASSPER dataset, where the disturbance signal takes into account many real-world effects of the clutter and, in general, is not an AR signal (additional details about the KASSPER dataset can be found in Section VI-D) [42]. In both cases, the signal vector $\mathbf{s}$ corresponds to a uniform linear array with randomly selected normalized spatial and Doppler frequencies. The SINR in this section is defined as

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

where $\bar{\mathbf{R}}$ is the spatial-temporal covariance matrix corresponding to $\mathbf{A}$ and the *a priori* matrix $\bar{\mathbf{Q}}$. The above SINR is interpreted as the average output SINR and provides a reference to the power level of the signal relative to the disturbance. Once $\bar{\mathbf{Q}}$ is selected, the spatial covariance matrix $\mathbf{Q}$ is generated in each Monte Carlo run from an inverse Wishart distribution with mean $\bar{\mathbf{Q}}$.

A. Limited-Training Scenarios

The training-deficient cases are particularly challenging in practice. As shown in [15], [19], [20], and [25], the parametric approach provides an effective way to improve the detection performance when the amount of training signals is limited. Specifically, when only $K = 2$ training signals are available, three scenarios are considered: 1) $N = 4J$, a preferable case for the conventional parametric detectors (i.e., the PAMF and P-GLRT), which benefit from having much more pulses (temporal observations) than spatial channels; 2) $N = 2J$, an intermediate case; and 3) the case of $N = J + 2$, an unfavorable case for the conventional parametric detectors.

In the first scenario, the number of pulses is $N = 16$ and the number of channels is $J = 4$. It has been shown that, if $N \gg J$, the conventional parametric detectors such as the P-Rao test and P-GLRT can cope with very limited or even no range training signals [20]. Therefore, it is extremely competitive to compare the knowledge-aided parametric detectors such as the KA-PAMF and the KA-PGLRT with such parametric detectors in this case. Fig. 1 shows the probability of detection versus SINR when $\mu = 12$ and $P_f = 0.01$. It is seen

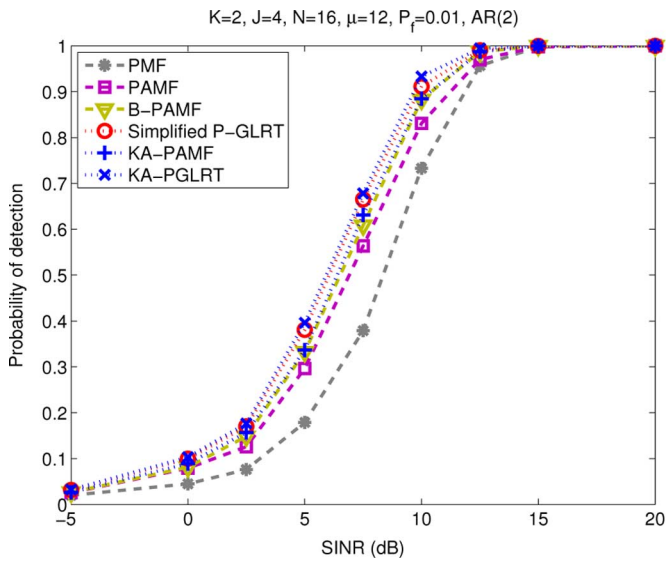


Fig. 1. Training-limited case with adequate temporal observations ($N \gg J$): Probability of detection versus SINR when $J = 4$, $N = 16$, $P = 2$, $K = 2$, $\mu = 12$ and $P_{\mathrm{f}} = 0.01$.

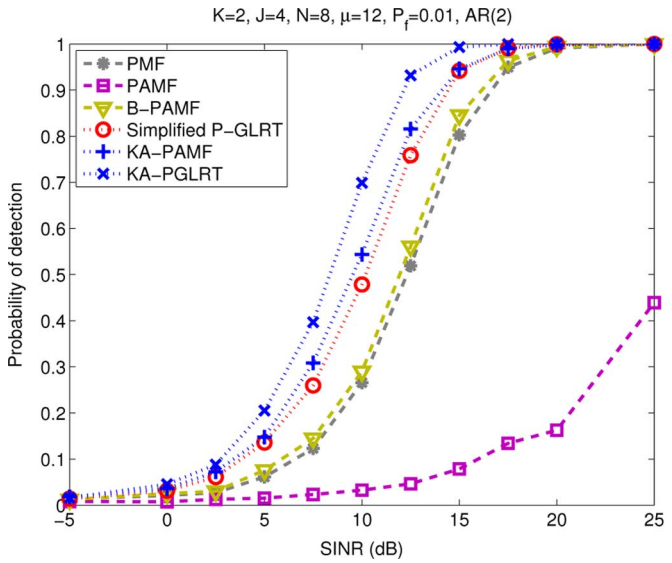


Fig. 2. Training-limited case with moderate temporal observation ($N = 2J$): Probability of detection versus SINR when $J = 4$, $N = 8$, $P = 2$, $K = 2$, $\mu = 12$ and $P_{\mathrm{f}} = 0.01$.

that, except the nonadaptive PMF, all adaptive parametric detectors achieve a close detection performance, which indicates that $K = 2$ range training signals and $N = 16$ pulses are adequate in the current case for the estimation of the unknown parameters $\mathbf{A}$ and $\mathbf{Q}$. It is also noted that the proposed KA-PGLRT exhibits the best performance among all parametric detectors. The two-step KA-PAMF, which performs successive temporal and spatial whitenings, is slightly worse than the simplified P-GLRT, but outperforms the PAMF and the B-PAMF.

In the second scenario, we reduce the number of pulses to $N = 8$, a less favorable case for the conventional parametric detectors. With the same amount of range training signals $K = 2$ as in the first scenario, the conventional parametric detectors are expected to experience some degradation as the number of pulses N is reduced. The results of the simulation are shown in Fig. 2. It is evident from this figure that, in this more challenging case, the proposed KA-PAMF and KA-PGLRT detectors provide the best results among all considered parametric detectors. Moreover, the KA-PGLRT achieves a performance that is better than that of the two-step KA-PAMF. With only $K = 2$ training signals and $N = 8$ pulses, the standard PAMF is unable to reliably estimate $\mathbf{Q}$ and $\mathbf{A}$ from the training signals and, hence, gives much worse results. In contrast, the PMF provides a better performance by using $\bar{\mathbf{Q}}$ as a spatially (nonadaptive) whitening matrix. The B-PAMF, using both prior knowledge and training signals, attains a slight performance gain over the PMF. Interestingly, the simplified P-GLRT which estimates the unknown parameters from both test and training signals achieves a performance similar to that of the KA-PAMF; however with a 0.8-dB performance loss.

The third scenario is the most challenging case with $K = 2$ training signals, $J = 8$ channels and $N = 10$ temporal observations, which is considered as an unfavorable case since $J \approx N$ (see [15, Sec. VII-B.2]). Fig. 3 shows the simulation result. It is seen from this figure that the two knowledge-aided KA-PAMF and KA-PGLRT detectors which combine knowledge learned

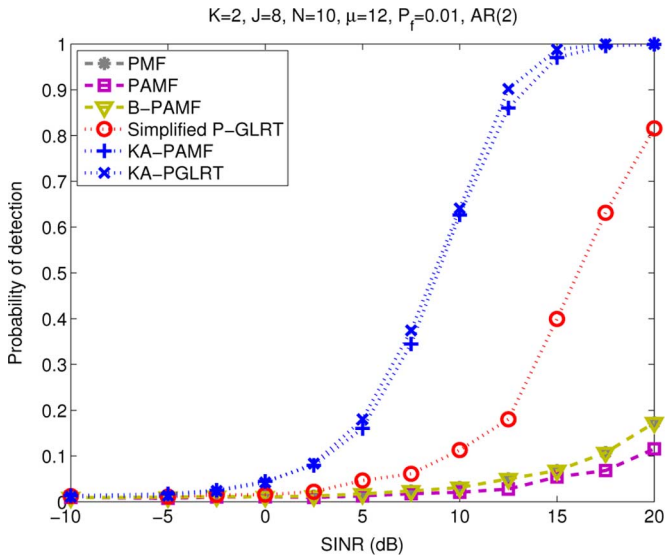


Fig. 3. Training-limited case with limited temporal observation ($N \approx J$): Probability of detection versus SINR when $J = 8$, $N = 10$, $P = 2$, $K = 2$, $\mu = 12$ and $P_{\mathrm{f}} = 0.01$.

from the test signal, training signals and prior information are able to significantly outperform the other parametric detectors. In particular, the performance gains of the KA-PAMF and the KA-PGLRT over the simplified P-GLRT are, respectively, 6 and 6.6 dB. The performance gain is even larger by comparing the knowledge-aided parametric detectors with the other parametric detectors.

Finally, we consider the performance of detection as a function of the number of pulses N, when $J = 4, K = 2, P = 2, \mu = 12$ and SINR $= 10$ dB. The number of pulses increases from $N = 6$ to $N = 32$, which corresponds to a transition from a low value of N/J to a high N/J. The purpose of this simulation is to evaluate the convergence performance of the simulated parametric detectors. As shown in Fig. 4, it is observed

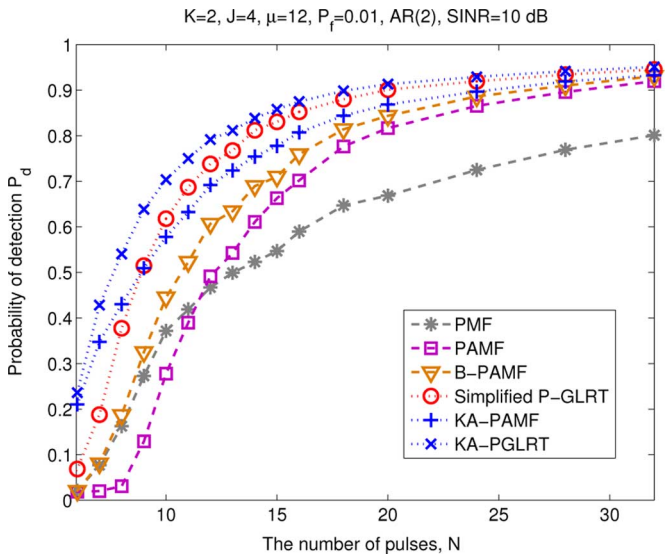


Fig. 4. Performance convergence: Probability of detection versus the number of pulses N when $J = 4$, $P = 2$, $K = 2$, $\mu = 12$, SINR $= 10$ dB and $P_{\rm f} = 0.01$.

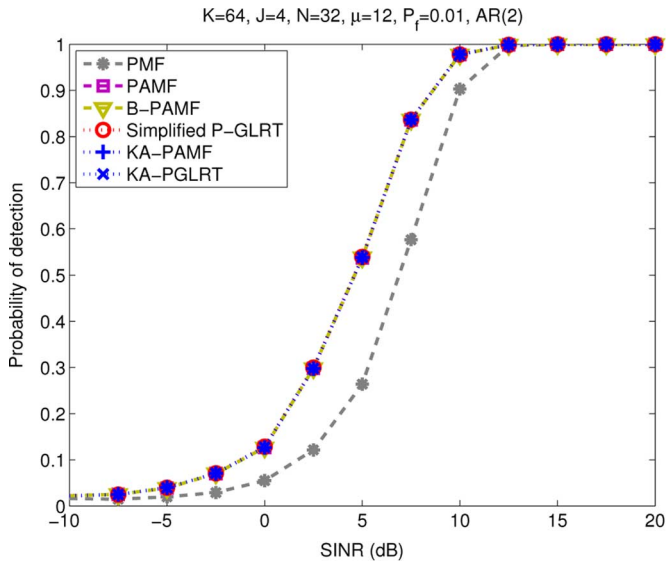


Fig. 5. Asymptotic case: Probability of detection versus SINR when $J = 4$, $N = 32$, $P = 2$, $K = 64$, $\mu = 12$ and $P_{\rm f} = 0.01$.

that the KA-PGLRT provides the best detection performance among all considered detectors, and the KA-PAMF outperforms the competitive P-GLRT when the number of pulses is small (i.e., $N < 10$) but becomes slightly inferior when N is larger than 10. It is also observed that the KA-PAMF is better than its counterparts, e.g., the B-PAMF and the PAMF, for almost all N. When $N = 32$ or larger, all adaptive parametric detectors converge and yield similar detection probability, while the nonadaptive PMF becomes the worst since it does not explore the useful information from the training signals.

B. Sufficient-Training Scenario

In this section, an asymptotic scenario with a large number of training signals is considered to show the convergence of the parametric adaptive detectors. Specifically, the simulation parameters are $J = 4$, $N = 32$, $P = 2$, $K = 64$, $\mu = 12$ and $P_f = 0.01$. The result is shown in Fig. 5. It is seen that, all parametric adaptive detectors have the same performance, while the nonadaptive PMF has a performance loss of about 3 dB. The results reveal that, with sufficient training signals, the parametric adaptive detectors can fully learn the knowledge about the unknown disturbance covariance matrix from the training signals. For the two proposed knowledge-aided parametric detectors, they put less weight on the prior knowledge, and rely more on training signals. In contrast, the nonadaptive PMF uses only the prior knowledge $\bar{\mathbf{Q}}$ to perform spatial whitening and suffers from ignoring the useful information from the large number of training signals.

We now examine the constant false alarm rate (CFAR) property of the proposed detectors. Recall that the disturbance covariance matrix and its spectral property are characterized by the AR coefficient matrices $\mathbf{A}$ and the spatial covariance matrix $\mathbf{Q}$. In the following, we vary these two parameters $\mathbf{A}$ and $\mathbf{Q}$, and then evaluate the detection performance subject to a constant probability of false alarm. Note that a similar procedure was used in [15, Table I] for the same purpose. Specifically, three cases are considered: 1) **Case I**: $\mathbf{A}$ is fixed while $\mathbf{Q}$ varies; 2) **Case II**: $\mathbf{A}$ varies while $\mathbf{Q}$ is fixed; and 3) **Case III**: Both $\mathbf{A}$ and $\mathbf{Q}$ vary. In each case, we compute the detection probability via the Monte Carlo simulation for five different sets of disturbance parameters, and then calculate the average detection probability and its standard deviation.

The results are shown in Table I where the system parameters are chosen as $J = 4$, $N = 16$, $K = 32$, $\mu = 12$, $P_f = 0.01$, and SINR $= 7.5$ dB. It is noted that this corresponds to a relatively large-sample (i.e., asymptotic) case. From Table I, it is seen that the standard deviation on the detection probability is relatively small compared to the detection probability, and the detection probability is similar across the three considered cases for each detector. Overall, the results confirm that the proposed parametric detectors are insensitive to the disturbance spectral variation, which implies that they approach the CFAR property in the asymptotic region. Nevertheless, it is noted that the CFAR property of the parametric detectors may vanish in the case of $N \approx J$ and when the disturbance does not follow the AR process, similarly to the PAMF as noted in [15].

C. Accuracy of the Prior Knowledge

The impact of the accuracy of the prior knowledge on the detection performance is examined in this section. We consider scenarios similar to that of Fig. 2 but with different values of μ, a parameter that controls the accuracy of the prior knowledge. Specifically, we choose $\mu = 5$, a case of less reliable prior knowledge, and $\mu = 36$, a case of more accurate prior knowledge. The results of the simulation are shown in Figs. 6 and 7, respectively. In both cases, it is seen that the proposed KA-PAMF and KA-PGLRT are able to provide better detection performance than the other detectors, while the standard PAMF gives the worst performance due to lack of training. A comparison between Figs. 6 and 7 shows that 1) the KA-PAMF and the KA-PGLRT offer a larger performance gain over the competitive simplified P-GLRT when the prior knowledge is more reliable; 2) the nonadaptive PMF performs better in the case of

TABLE I
PROBABILITY OF DETECTION (AND ITS STANDARD DEVIATION SHOWN IN THE PARENTHESIS) IN DIFFERENT SIMULATION SCENARIOS WHEN $J = 4$, $N = 16$, $P = 2$, $K = 32$, AND SINR $= 7.5$ dB

Scenarios	PMF	PAMF	B-PAMF	P-GLRT	KA-PAMF	KA-PGLRT
Case-I	0.5449 (0.0210)	0.8152 (0.0072)	0.8160 (0.0065)	0.8156 (0.0071)	0.8154 (0.0064)	0.8173 (0.0072)
Case-II	0.5384 (0.0156)	0.8088 (0.0101)	0.8101 (0.0101)	0.8085 (0.0095)	0.8102 (0.0102)	0.8107 (0.0106)
Case-III	0.5314 (0.0230)	0.8037 (0.0102)	0.8052 (0.0106)	0.8044 (0.0102)	0.8049 (0.0106)	0.8055 (0.0100)

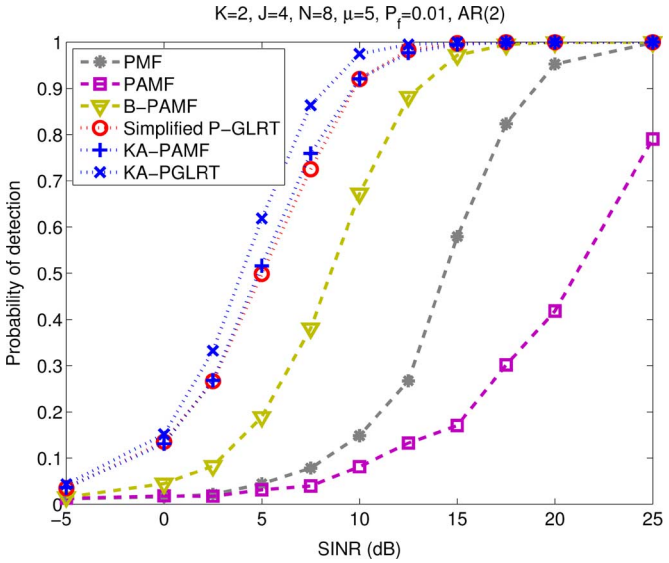


Fig. 6. Less reliable prior knowledge: Probability of detection versus SINR when $J = 4$, $N = 8$, $P = 2$, $K = 2$, $\mu = 5$ and $P_{\rm f} = 0.01$.

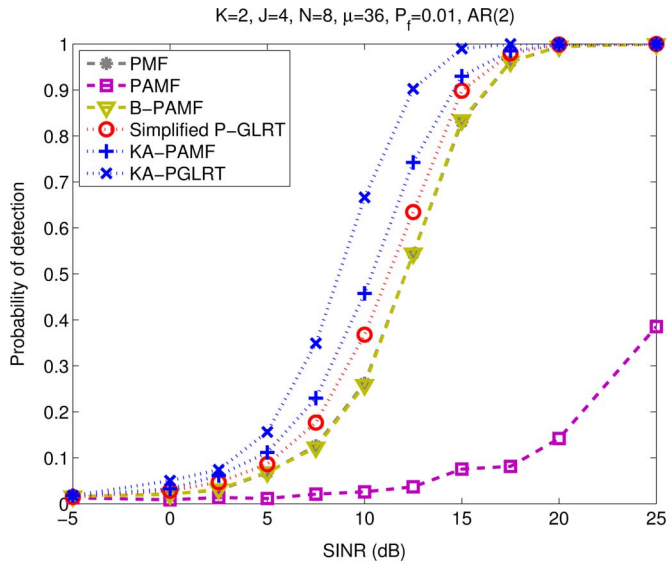


Fig. 7. More accurate prior knowledge: Probability of detection versus SINR when $J = 4$, $N = 8$, $P = 2$, $K = 2$, $\mu = 36$ and $P_{\rm f} = 0.01$.

$\mu = 36$ than in the case of $\mu = 5$ and yields almost the same performance as the B-PAMF when $\mu = 36$ since they both rely heavily on the prior knowledge; and 3) the performance loss between the PAMF and other knowledge-aided detectors increases from the case of $\mu = 5$ to the case of $\mu = 36$. We also noted that both the simplified P-GLRT and the PAMF, which use no prior knowledge, give different detection performance in cases of $\mu = 5$ and $\mu = 36$. This effect is related to the definition of SINR in (58). By using the deterministic $\bar{\mathbf{R}}$ instead of the stochastic $\mathbf{R}$ in the SINR definition, different values of μ lead to different realizations of $\mathbf{R}$ and, therefore, the simplified P-GLRT and the PAMF give different detection performance even with the same SINR.

D. Model Mismatch

In the above examples, the disturbance is generated from a multi-channel AR process. In this section, the detection performance of the proposed knowledge-aided parametric detectors is examined when the disturbance is not exactly an AR process. For this purpose, we use the KASSPER dataset, which contains many challenging real-world effects, including a heterogeneous terrain, array errors, and dense ground targets. The KASSPER dataset uses a $J = 11$ uniform linear array with half-wavelength inter-element spacing and $N = 32$ temporal pulses in each coherent processing interval (see [42] for a detailed description of the KASSPER dataset). Specifically, the disturbance covariance matrix at range bin $r = 200$ is used as the prior knowledge $\bar{\mathbf{R}}$ and, at each trial, the disturbance covariance matrix is generated as $\mathbf{R} \sim \mathcal{CW}^{-1}((\mu - JN)\bar{\mathbf{R}}, \mu)$ with $\mu \geq JN + 1$. In this case, the overall space-time covariance matrix is random, instead of only the spatial covariance matrix is random as assumed in **AS2**. For a feasible Monte Carlo simulation, we downsample the original 352×352 covariance matrix $\mathbf{R}$ into two cases: 1) $J = 4$ and $N = 8$ with $\mu = 33$; and 2) $J = 8$ and $N = 10$ with $\mu = 81$.

The simulation results for both KASSPER cases are shown in Figs. 8 and 9, respectively, where the parametric detectors model the disturbance using a multi-channel AR(2) process, i.e., $P = 2$. For the proposed KA-PAMF and KA-PGLRT detectors and the B-PAMF, we need the prior spatial covariance matrix $\bar{\mathbf{Q}}$ extracted from the available prior space-time covariance matrix $\bar{\mathbf{R}}$. As mentioned in Section II, the prior knowledge $\bar{\mathbf{Q}}$ can be obtained in multiple ways. In this simulation, we choose the last approach. For both cases, we use $\mu = J + 8$ for the KA-PAMF and the KA-PGLRT. As seen from Figs. 8 and 9, the proposed KA-PAMF and KA-PGLRT are still the best detectors even with model mismatch. It is also noted that, with the KASSPER dataset, the two-step KA-PAMF performs slightly better than the one-step KA-PGLRT, which is not unexpected since, due to the presence of model mismatch, the KA-PGLRT is not necessarily better than the KA-PAMF.

VII. CONCLUSION

This paper introduces a new multichannel AR model with a random spatial covariance matrix for STAP application. Based on this model, we present two knowledge-aided parametric detectors, referred to as the KA-PAMF and the KA-PGLRT, re-

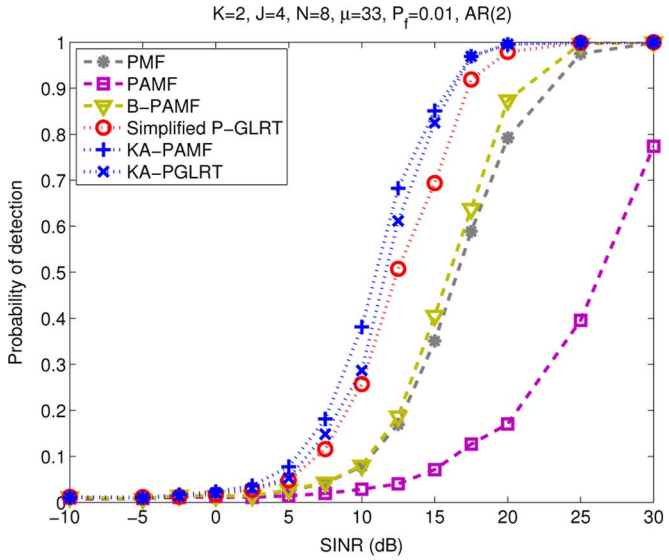


Fig. 8. KASSPER dataset: Probability of detection versus SINR when $J = 4$, $N = 8$, $P = 2$, $K = 2$, $\mu = 33$ and $P_{\rm f} = 0.01$.

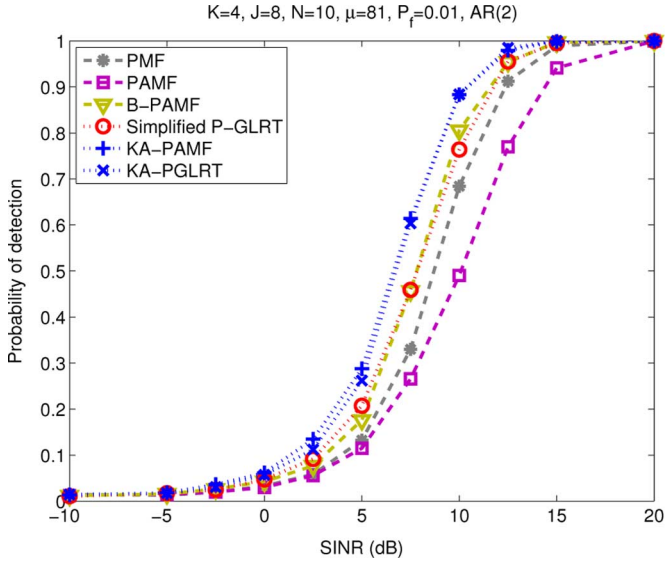


Fig. 9. KASSPER dataset: Probability of detection versus SINR when $J = 8$, $N = 10$, $P = 2$, $K = 4$, $\mu = 81$ and $P_{\rm f} = 0.01$.

spectively, by integrating out the random spatial covariance matrix and maximizing the likelihood function over the AR coefficient matrices. The KA-PAMF utilizes a similar whitening process as the conventional PAMF, but with a distinctively different spatial whitening matrix that is formed from the training data, target-canceled test signal, and prior knowledge. Meanwhile, the KA-PGLRT takes a colored-loading form of the conventional parametric GLRT and enables the utilization of the prior knowledge. The significance (or accuracy) of the prior knowledge is directly reflected in the test statistics of both detectors. Our simulation results show that the proposed knowledge-aided detectors provide enhanced performance when the training signals are limited and when the number of pulse is comparable to the number of antennas. In the asymptotic case with sufficient training signals, the proposed detectors perform similarly as the conventional parametric detectors, which is expected as the significance of the prior knowledge diminishes. Performance examination with the KASSPER dataset also confirms that the knowledge-aided parametric detectors yield better detection performance than the conventional ones.

Appendix I
Derivation of (31)

With the ML estimate of α in (30), the test variable of the partially adaptive P-GLRT in (22) is equivalent to

$$\begin{aligned} T &= \frac{|\boldsymbol{\Xi}_0|_{\alpha=0}}{|\boldsymbol{\Xi}_1(\hat{\alpha}_{\rm ML})|} \\ &= \frac{\left|\tilde{\mathbf{X}}_0\mathbf{P}\tilde{\mathbf{X}}_0^H\boldsymbol{\Psi}^{-1}+\mathbf{I}\right|}{|(\tilde{\mathbf{X}}_0\mathbf{P}-\hat{\alpha}_{\rm ML}\tilde{\mathbf{S}})(\tilde{\mathbf{X}}_0\mathbf{P}-\hat{\alpha}_{\rm ML}\tilde{\mathbf{S}})^H\boldsymbol{\Psi}^{-1}+\mathbf{I}|} \\ &\propto \ln\left|\tilde{\mathbf{X}}_0\mathbf{P}\tilde{\mathbf{X}}_0^H\boldsymbol{\Psi}^{-1}+\mathbf{I}\right| \\ &\quad -\ln|(\tilde{\mathbf{X}}_0\mathbf{P}-\hat{\alpha}_{\rm ML}\tilde{\mathbf{S}})(\tilde{\mathbf{X}}_0\mathbf{P}-\hat{\alpha}_{\rm ML}\tilde{\mathbf{S}})^H\boldsymbol{\Psi}^{-1}+\mathbf{I}|. \end{aligned} \tag{59}$$

According to [40, Lemma 4], the eigenvalues of the matrix $(\tilde{\mathbf{X}}_0\mathbf{P}-\alpha\tilde{\mathbf{S}})(\tilde{\mathbf{X}}_0\mathbf{P}-\alpha\tilde{\mathbf{S}})^H\boldsymbol{\Psi}^{-1}$, denoted as λ_α, satisfy $0 \le \lambda_m(\alpha) \ll 1$. Then, we have

$$\begin{aligned} &\ln|(\tilde{\mathbf{X}}_0\mathbf{P}-\alpha\tilde{\mathbf{S}})(\tilde{\mathbf{X}}_0\mathbf{P}-\alpha\tilde{\mathbf{S}})^H\boldsymbol{\Psi}^{-1}+\mathbf{I}| \\ &\quad = \ln\left[\prod_{m=1}^{M}(1+\lambda_m(\alpha))\right] \\ &\quad \overset{(a)}{\approx} \ln\left[1+\sum_{m=1}^{M}\lambda_m(\alpha)\right] \\ &\quad = \ln[1+\mathrm{tr}\{(\tilde{\mathbf{X}}_0\mathbf{P}-\alpha\tilde{\mathbf{S}})^H\boldsymbol{\Psi}^{-1}(\tilde{\mathbf{X}}_0\mathbf{P}-\alpha\tilde{\mathbf{S}})\}] \\ &\quad \overset{(b)}{\approx} \mathrm{tr}\{(\tilde{\mathbf{X}}_0\mathbf{P}-\alpha\tilde{\mathbf{S}})^H\boldsymbol{\Psi}^{-1}(\tilde{\mathbf{X}}_0\mathbf{P}-\alpha\tilde{\mathbf{S}})\} \end{aligned} \tag{60}$$

where the approximation (a) holds asymptotically in a first-order sense as the number of samples N is large enough and (b) occurs by using the fact that $\mathrm{tr}\{(\tilde{\mathbf{X}}_0\mathbf{P}-\alpha\tilde{\mathbf{S}})^H\boldsymbol{\Psi}^{-1}(\tilde{\mathbf{X}}_0\mathbf{P}-\alpha\tilde{\mathbf{S}})\} \ll 1$ and then invoking the approximation $\ln(1+x) \approx x$, for $x \ll 1$. As a result, the test variable in (59) is simplified as

$$\begin{aligned} T = &\mathrm{tr}\left\{\tilde{\mathbf{X}}_0^H\boldsymbol{\Psi}^{-1}\tilde{\mathbf{X}}_0\mathbf{P}\right\} \\ &-\mathrm{tr}\{(\tilde{\mathbf{X}}_0\mathbf{P}-\hat{\alpha}_{\rm ML}\tilde{\mathbf{S}})^H\boldsymbol{\Psi}^{-1}(\tilde{\mathbf{X}}_0\mathbf{P}-\hat{\alpha}_{\rm ML}\tilde{\mathbf{S}})\}. \end{aligned} \tag{61}$$

Substituting the ML estimate of α (30) back into the above equation yields

$$T = \frac{|\mathrm{tr}\{\tilde{\mathbf{S}}^H\boldsymbol{\Psi}^{-1}\tilde{\mathbf{X}}_0\}|^2}{\mathrm{tr}\{\tilde{\mathbf{S}}^H\boldsymbol{\Psi}^{-1}\tilde{\mathbf{S}}\}} \tag{62}$$

which is the matrix form of (31).

References

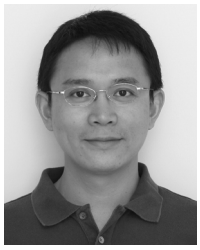
[1] J. Ward, "Space-time adaptive processing for airborne radar," Lincoln Laboratory, MIT, Lexington, MA, Tech. Rep. 1015, Dec. 1994.
[2] R. Klemm, *Principles of Space-Time Adaptive Processing*. London, U.K.: Institute of Electrical Engineers, 2002.
[3] I. S. Reed, J. D. Mallett, and L. E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 10, no. 6, pp. 853–863, 1974.

[4] E. J. Kelly, "An adaptive detection algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 22, no. 1, pp. 115–127, Mar. 1986.
[5] 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.
[6] A. De Maio, "A new derivation of the adaptive matched filter," *IEEE Signal Process. Lett.*, vol. 11, no. 10, pp. 792–793, Oct. 2004.
[7] S. Kraut and L. L. Scharf, "The CFAR adaptive subspace detector is a scale-invariant GLRT," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2538–2541, Sep. 1999.
[8] E. Conte, M. Lops, and G. Ricci, "Asymptotically optimum radar detection in compound Gaussian clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 2, pp. 617–625, Apr. 1995.
[9] 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.
[10] J. S. Goldstein and I. S. Reed, "Reduced-rank adaptive filtering," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 492–496, Feb. 1997.
[11] M. Shen, D. Zhu, and Z. Zhu, "Reduced-rank space-time adaptive processing using a modified projection approximation subspace tracking deflation approach," *IET Radar, Sonar, Navigat.*, vol. 3, no. 1, pp. 93–100, Feb. 2009.
[12] R. Fa, R. C. de Lamare, and L. Wang, "Reduced-rank stap schemes for airborne radar based on switched joint interpolation, decimation and filtering algorithm," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4182–4194, Aug. 2010.
[13] K. Gerlach and M. L. Picciolo, "Robust, reduced rank, loaded reiterative median cascaded canceller," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, no. 1, pp. 15–25, Feb. 2011.
[14] H. Wang and L. Cai, "On adaptive spatial-temporal processing for airborne surveillance radar systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 30, no. 3, pp. 660–670, Jul. 1994.
[15] 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.
[16] J. H. Michels, B. Himed, and M. Rangaswamy, "Performance of STAP tests in Gaussian and compound-Gaussian clutter," *Digit. Signal Process.*, vol. 10, no. 4, pp. 309–324, Oct. 2000.
[17] P. Parker and A. L. Swindlehurst, "Space-time auto-regressive filtering for matched subspace STAP," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 39, no. 2, pp. 510–520, Apr. 2003.
[18] 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.
[19] 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.
[20] K. J. Sohn, H. Li, and B. Himed, "Parametric GLRT for multichannel adaptive signal detection," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5351–5360, Nov. 2007.
[21] S. L. Marple, Jr., P. M. Corbell, and M. Rangaswamy, "Multi-channel fast parametric algorithms and performance for adaptive radar," in *Proc. 41th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 2007, pp. 1835–1838.
[22] S. L. Marple, Jr., P. M. Corbell, and M. Rangaswamy, "Performance tradeoffs for multi-channel parametric adaptive radar algorithms," in *Proc. 2008 Int. Conf. Radar*, Sep. 2008, pp. 154–159.
[23] Y. I. Abramovich, N. K. Spencer, and M. D. E. Turley, "Order estimation and discrimination between stationary and time-varying (TVAR) autoregressive models," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2861–2876, Jun. 2007.
[24] Y. I. Abramovich, N. K. Spencer, and M. D. E. Turley, "Time-varying autoregressive (TVAR) models for multiple radar observations," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1298–1311, Apr. 2007.
[25] P. Wang, H. Li, and B. Himed, "A new parametric GLRT for multichannel adaptive signal detection," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 317–325, Jan. 2010.
[26] M. Rangaswamy, B. Himed, and J. H. Michels, "Performance analysis of the nonhomogeneity detector for STAP applications," in *Proc. 2001 IEEE Radar Conf.*, 2001, pp. 193–197.
[27] G. T. Capraro, A. Farina, H. Griffiths, and M. Wicks, "Knowledge-based radar signal and data processing," *IEEE Signal Process. Mag.*, vol. 23, no. 1, pp. 18–29, Jan. 2006.
[28] J. R. Guerci and E. J. Baranoski, "Knowledge-aided adaptive radar at DARPA: An overview," *IEEE Signal Process. Mag.*, vol. 23, no. 1, pp. 41–50, Jan. 2006.
[29] 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.
[30] W. L. Melvin and J. R. Guerci, "Knowledge-aided signal processing: A new paradigm for radar and other advanced sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 3, pp. 983–996, Jul. 2006.
[31] 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.
[32] X. Zhu, J. Li, P. Stoica, and J. R. Guerci, "Knowledge-aided space-time adaptive processing," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, no. 2, pp. 1325–1336, Apr. 2011.
[33] L. Svensson and M. Lundberg, "On posterior distributions for signals in Gaussian noise with unknown covariance matrix," *IEEE Trans. Signal Process.*, vol. 53, no. 9, pp. 3554–3571, Sep. 2005.
[34] 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.
[35] 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.
[36] 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.
[37] S. Bidon, O. Besson, and J.-Y. Tourneret, "The adaptive coherence estimator is the generalized likelihood ratio test for a class of heterogeneous environments," *IEEE Signal Process. Lett.*, vol. 15, no. 2, pp. 281–284, Feb. 2008.
[38] 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.
[39] 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.
[40] Y. Jiang, P. Stoica, and J. Li, "Array signal processing in the known waveform and steering vector case," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 23–35, Jan. 2004.
[41] K. J. Sohn, H. Li, and B. Himed, "Recursive parametric tests for multichannel adaptive signal detection," *IET Radar, Sonar, Navigat.*, vol. 2, no. 1, pp. 63–70, Feb. 2008.
[42] J. S. Bergin and P. M. Techau, "High-fidelity site-specific radar simulation: KASSPER'02 Workshop Datacube," Information Systems Laboratories, Inc., Vienna, VA, Tech. Rep. ISL-SCRD-TR-02-105, May 2002.

Pu Wang (S'05–M'11) 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, in 2011, all in electrical engineering.

He was an intern at the Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA, during summer 2010. Since May 2011, he has been with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, where he is currently a Research Assistant Professor. His current research interests include statistical signal processing for multi-antenna systems, nonstationary systems, wireless communications, and networks.

Dr. Wang 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 from the Stevens Institute of Technology, the Excellent Master Thesis Award of Sichuan Province in 2007, and the Excellent Master Thesis Award of UESTC in 2006. He was one of the finalists in the Best Student Paper Competition at the 2011 IEEE Radar Conference.



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, 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 during summers 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 is currently a Professor. His current research interests include statistical signal processing, wireless communications, and radars.

Dr. Li is a member of Tau Beta Pi and Phi Kappa Phi. He received 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 both the Sensor Array and Multichannel (SAM) Technical Committee and the Signal Processing Theory and Methods (SPTM) Technical Committee of the IEEE Signal Processing Society. He has served as an Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE SIGNAL PROCESSING LETTERS, and the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and a Guest Editor for the *EURASIP Journal on Applied Signal Processing* Special Issue on Distributed Signal Processing Techniques for Wireless Sensor Networks.

Braham Himed (S'88–M'90–SM'01–F'07) received the B.S. degree in electrical engineering from Ecole Nationale Polytechnique of Algiers in 1984 and the M.S. and Ph.D. degrees, both in electrical engineering, from Syracuse University, Syracuse, NY, in 1987 and 1990, respectively.

He is a Technical Advisor with the Air Force Research Laboratory, Sensors Directorate, RF Technology Branch, Dayton OH, where he is involved with several aspects of airborne and spaceborne phased array radar systems. His research interests include detection, estimation, multichannel adaptive signal processing, time series analyses, array processing, space-time adaptive processing, and waveform diversity.

Dr. Himed is the recipient of the 2001 IEEE Region I Award for his work on bistatic radar systems, algorithm development, and phenomenology. He is a member of the AES Radar Systems Panel.