Adaptive Detection of Structured Signals in Low-Rank Interference

Philip Schniter and Evan Byrne
Dept. of ECE, The Ohio State Univ., Columbus, OH, 43210, USA. (byrne.133@osu.edu, schniter.1@osu.edu)

***Abstract*—In this paper, we consider the problem of detecting the presence (or absence) of an unknown but structured signal from the space-time outputs of an array under strong, non-white interference. Our motivation is the detection of a communication signal in jamming, where often the "training" portion is known but the "data" portion is not. We assume that the measurements are corrupted by additive white Gaussian noise of unknown variance and a few strong interferers, whose number, powers, and array responses are unknown. We also assume the desired signal's array response is unknown. To address the detection problem, we propose two GLRT-based detection schemes that employ a probabilistic signal model and use the EM algorithm for likelihood maximization. Numerical experiments are presented to assess the performance of the proposed schemes.**

I. Introduction

Consider the problem of detecting the presence or absence of a signal $\boldsymbol{s} \in \mathbb{C}^L$ from the measured output $\boldsymbol{Y} \in \mathbb{C}^{M\times L}$ of an M-element antenna array. We are interested in the case where $\boldsymbol{s}$ is known only in probability. A motivating example arises with communications signals, where typically a few "training" samples are known and the remainder (i.e., the "data" samples) are unknown, except for their alphabet.

The signal-detection problem can be formulated as a binary hypothesis test [1] between hypotheses $\mathcal{H}_1$ (signal present) and $\mathcal{H}_0$ (signal absent), i.e.,

$$\mathcal{H}_1 : \boldsymbol{Y} = \boldsymbol{h}\boldsymbol{s}^{\mathsf{H}} + \boldsymbol{B}\boldsymbol{\Phi}^{\mathsf{H}} + \boldsymbol{W} \in \mathbb{C}^{M\times L} \quad \text{(1a)}$$

$$\mathcal{H}_0 : \boldsymbol{Y} = \boldsymbol{B}\boldsymbol{\Phi}^{\mathsf{H}} + \boldsymbol{W} \in \mathbb{C}^{M\times L}. \quad \text{(1b)}$$

In (1), $\boldsymbol{h} \in \mathbb{C}^M$ models the array response, which we assume is completely unknown (as in the case of a dense multipath environment). $\boldsymbol{W}$ models additive white Gaussian noise (AWGN) with unknown variance $\nu > 0$, and $\boldsymbol{B}\boldsymbol{\Phi}^{\mathsf{H}}$ models interference from N interferers, where N is unknown. If the array responses of these N interferers are constant over the measurement epoch and bandwidth, then $\operatorname{rank}(\boldsymbol{B}\boldsymbol{\Phi}^{\mathsf{H}}) = N$. The temporal interference component $\boldsymbol{\Phi}^{\mathsf{H}}$ is assumed white and Gaussian, while the spatial interference component $\boldsymbol{B}$ is deterministic and unknown.

Communications signals often take a form like

$$\boldsymbol{s}^{\mathsf{H}} = \left[\boldsymbol{s}_{\mathsf{t}}^{\mathsf{H}} \; \boldsymbol{s}_{\mathsf{d}}^{\mathsf{H}}\right], \quad (2)$$

where $\boldsymbol{s}_{\mathsf{t}} \in \mathbb{C}^Q$ is a known training sequence, $\boldsymbol{s}_{\mathsf{d}} \in \mathcal{A}^{L-Q}$ is an unknown data sequence, $\mathcal{A} \subset \mathbb{C}$ is a finite alphabet, and $Q \ll L$. Suppose that the measurements are partitioned as $\boldsymbol{Y} = \left[\boldsymbol{Y}_{\mathsf{t}} \, \boldsymbol{Y}_{\mathsf{d}}\right]$, conformal with (2). For signal detection and/or synchronization, the data measurements $\boldsymbol{Y}_{\mathsf{d}}$ are often ignored (see, e.g., [2]). But these data measurements can be very useful, especially when the training symbols (and thus the training measurements $\boldsymbol{Y}_{\mathsf{t}}$) are few. Our goal is to develop detection schemes that use all measurements $\boldsymbol{Y}$ while handling the incomplete knowledge of $\boldsymbol{s}$ in a principled manner.

We propose to probabilistically model the signal by treating $\boldsymbol{s}$ as a random vector with prior pdf $p(\boldsymbol{s})$. Although our method supports arbitrary $p(\boldsymbol{s})$, we sometimes focus (for simplicity) on the case of statistically independent components, i.e.,

$$p(\boldsymbol{s}) = \textstyle\prod_{l=1}^{L} p_l(s_l). \quad (3)$$

For example, with uncoded communication signals partitioned as in (2), we would use (3) with

$$p_l(s_l) = \begin{cases} \delta(s_l - s_{\mathsf{t},l}) & l = 1,\ldots,Q \\ \frac{1}{|\mathcal{A}_l|}\sum_{s\in\mathcal{A}_l}\delta(s_l - s) & l = Q+1,\ldots,L, \end{cases} \quad (4)$$

where $\delta(\cdot)$ the Dirac delta and $s_{\mathsf{t},l}$ the lth training symbol.

We now describe relevant prior work. For the case where the entire signal $\boldsymbol{s} \in \mathbb{C}^L$ is *known*, the detection problem (1) has been studied in detail. For example, in the classical work of Kelly [3], the interference-plus-noise $\boldsymbol{B}\boldsymbol{\Phi}^{\mathsf{H}} + \boldsymbol{W}$ was modeled as temporally white and Gaussian with unknown (and unstructured) spatial covariance $\boldsymbol{\Sigma} > 0$, and the generalized likelihood ratio test (GLRT) [1] was derived. Detector performance can be improved when the interference is known to have low rank. For example, Gerlach and Steiner [4] assumed temporally white Gaussian interference with known noise variance ν and unknown interference rank N and derived the GLRT. More recently, Kang, Monga, and Rangaswamy [5] assumed temporally white Gaussian interference with unknown ν and known N and derived the GLRT. Other structures on $\boldsymbol{\Sigma}$ were considered by Aubry et al. in [6]. In a departure from the above methods, McWhorter [7] proposed to treat the interference components $\boldsymbol{B} \in \mathbb{C}^{M\times N}$ and $\boldsymbol{\Phi} \in \mathbb{C}^{L\times N}$, as well as the noise variance ν, as deterministic unknowns, and then derived the corresponding GLRT. Bandiera et al. [8] proposed yet a different approach, based on a Bayesian perspective.

For adaptive detection of *unknown* but structured signals $\boldsymbol{s}$, Forsythe [9, p.110] described an iterative scheme for signals with deterministic (e.g., finite-alphabet, constant envelope) structure that alternates between maximum-likelihood (ML) signal estimation and least-squares beamforming.

In this work, we propose GLRT-based detectors for (1).

E. Byrne and P. Schniter acknowledge support from NSF grant 1716388 and MIT Lincoln Labs, and insightful discussions with Adam R. Margetts.

Under the probabilistic model $\boldsymbol{s} \sim p(\boldsymbol{s})$, direct evaluation of the GLRT numerator becomes intractable. Thus, we use expectation maximization (EM) [10], and we derive computationally efficient EM procedures for the independent prior (3).

Our first approach makes no attempt to leverage low interference rank, like Kelly [3]. We show that this approach is a variation on Forsythe's iterative scheme [9, p.110] that uses "soft" instead of "hard" symbol estimation. Our second approach is an extension that exploits the possibly low-rank nature of the interference. As in [4]–[6], the interference is modeled as temporally white Gaussian but, different from [4]–[6], the interference rank N and noise variance ν are unknown, and the signal is probabilistic.

This paper is an abbreviated version of [11], which provides the details of many derivations, as well as a third GLRT approach that treats $\boldsymbol{\Phi}$ as deterministic, as in McWhorter [7].

II. Background

In our discussions below, we will use $\boldsymbol{P}_{\boldsymbol{A}} \triangleq \boldsymbol{A}(\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathsf{H}}$ to denote orthogonal projection onto the column space of a given matrix $\boldsymbol{A}$, i.e., and $\boldsymbol{P}_{\boldsymbol{A}}^{\perp} \triangleq \boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A}}$ to denote the orthogonal complement.

The classical work of Kelly [3] tackled the binary hypothesis test (1) by treating the interference-plus-noise $\boldsymbol{B}\boldsymbol{\Phi}^{\mathsf{H}} + \boldsymbol{W}$ as temporally white and Gaussian with unknown spatial covariance $\boldsymbol{\Sigma} > 0$. This reduces (1) to

$$\mathcal{H}_1 : \boldsymbol{Y} = \boldsymbol{h}\boldsymbol{s}^{\mathsf{H}} + \mathcal{CN}(\boldsymbol{0}, \boldsymbol{\Sigma}) \tag{5a}$$

$$\mathcal{H}_0 : \boldsymbol{Y} = \mathcal{CN}(\boldsymbol{0}, \boldsymbol{\Sigma}). \tag{5b}$$

With known $\boldsymbol{s}$, the GLRT [1] takes the form

$$\frac{\max_{\boldsymbol{h},\boldsymbol{\Sigma}>0} p(\boldsymbol{Y}|\mathcal{H}_1; \boldsymbol{h}, \boldsymbol{\Sigma})}{\max_{\boldsymbol{\Sigma}>0} p(\boldsymbol{Y}|\mathcal{H}_0; \boldsymbol{\Sigma})} \gtrless \eta, \tag{6}$$

for some threshold η. Using results from [12], it was shown in [3] that (6) reduces to

$$\frac{\prod_{m=1}^{M} \lambda_{0,m}}{\prod_{m=1}^{M} \lambda_{1,m}} \gtrless \eta, \tag{7}$$

for decreasing ordered (i.e., $\lambda_{i,m} \geq \lambda_{i,m+1}\ \forall m, i$) eigenvalues

$$\{\lambda_{0,m}\}_{m=1}^{M} \triangleq \text{eigenvalues}\big(\tfrac{1}{L}\boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}}\big) \tag{8a}$$

$$\{\lambda_{1,m}\}_{m=1}^{M} \triangleq \text{eigenvalues}\big(\tfrac{1}{L}\boldsymbol{Y}\boldsymbol{P}_{\boldsymbol{s}}^{\perp}\boldsymbol{Y}^{\mathsf{H}}\big). \tag{8b}$$

Kelly's approach was applied to the detection/synchronization of communications signals by Bliss and Parker in [2] after discarding the contribution of the unknown data symbols $\boldsymbol{s}_{\mathsf{d}}$.

The low-rank property of the interference $\boldsymbol{B}\boldsymbol{\Phi}^{\mathsf{H}}$ can be exploited to improve detector performance. For example, Kang, Monga, and Rangaswamy [5] proposed the GLRT

$$\frac{\max_{\boldsymbol{h},\boldsymbol{\Sigma}\in\mathcal{S}_N} p(\boldsymbol{Y}|\mathcal{H}_1; \boldsymbol{h}, \boldsymbol{\Sigma})}{\max_{\boldsymbol{\Sigma}\in\mathcal{S}_N} p(\boldsymbol{Y}|\mathcal{H}_0; \boldsymbol{\Sigma})} \gtrless \eta, \tag{9}$$

where

$$\mathcal{S}_N \triangleq \{\boldsymbol{R} + \nu\boldsymbol{I} : \text{rank}(\boldsymbol{R}) = N, \boldsymbol{R} \geq 0, \nu > 0\}. \tag{10}$$

Using [13], it can be shown that the GLRT (9) simplifies to

$$\frac{\prod_{m=1}^{M} \widehat{\lambda}_{0,m}}{\prod_{m=1}^{M} \widehat{\lambda}_{1,m}} \gtrless \eta, \tag{11}$$

with $\{\widehat{\lambda}_{i,m}\}_{m=1}^{M}$ a smoothed version of $\{\lambda_{i,m}\}_{m=1}^{M}$ from (8):

$$\widehat{\lambda}_{i,m} = \begin{cases} \lambda_{i,m} & m = 1, \ldots, N, \\ \widehat{\nu}_i & m = N+1, \ldots, M. \end{cases} \tag{12}$$

$$\widehat{\nu}_i = \tfrac{1}{M-N} \textstyle\sum_{m=N+1}^{M} \lambda_{i,m}. \tag{13}$$

III. Proposed GLRT

We now consider the hypothesis test (1) with probabilistic $\boldsymbol{s} \sim p(\boldsymbol{s})$. Recalling that $\boldsymbol{B}$ is a deterministic unknown and $\boldsymbol{\Phi}^{\mathsf{H}}$ is white and Gaussian, the interference-plus-noise matrix

$$\boldsymbol{N} \triangleq \boldsymbol{B}\boldsymbol{\Phi}^{\mathsf{H}} + \boldsymbol{W} \tag{14}$$

is temporally white Gaussian with spatial covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{R} + \nu\boldsymbol{I}_M$, where both $\boldsymbol{R} \geq 0$ and $\nu > 0$ are unknown. For now, we will model $\boldsymbol{R}$ using a fixed rank $N \leq M$. The $N = M$ case is reminiscent of Kelly [3], and the $N < M$ case is reminiscent of Kang, Monga, and Rangaswamy [5].

For a fixed rank N, the hypothesis test (1) reduces to

$$\mathcal{H}_1 : \boldsymbol{Y} = \boldsymbol{h}\boldsymbol{s}^{\mathsf{H}} + \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_L \otimes \boldsymbol{\Sigma}) \tag{15a}$$

$$\mathcal{H}_0 : \boldsymbol{Y} = \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_L \otimes \boldsymbol{\Sigma}), \tag{15b}$$

where $\boldsymbol{h}$ and $\boldsymbol{\Sigma} \in \mathcal{S}_N$ (defined in (10)) are unknown and $\boldsymbol{s} \sim p(\boldsymbol{s})$. When $N = M$, note that $\boldsymbol{\Sigma} \in \mathcal{S}_N$ reduces to $\boldsymbol{\Sigma} > 0$. The corresponding GLRT is

$$\frac{\max_{\boldsymbol{h},\boldsymbol{\Sigma}\in\mathcal{S}_N} p(\boldsymbol{Y}|\mathcal{H}_1; \boldsymbol{h}, \boldsymbol{\Sigma})}{\max_{\boldsymbol{\Sigma}\in\mathcal{S}_N} p(\boldsymbol{Y}|\mathcal{H}_0; \boldsymbol{\Sigma})} \gtrless \eta. \tag{16}$$

As a consequence of $\boldsymbol{s} \sim p(\boldsymbol{s})$, the numerator likelihood in (16) differs from that in (9), as detailed in the sequel.

A. GLRT Denominator

For the denominator of (16), equations (15b) and (10) imply

$$p(\boldsymbol{Y}|\mathcal{H}_0; \boldsymbol{\Sigma}) = \left[\frac{\exp(-\operatorname{tr}\{\frac{1}{L}\boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}\})}{\pi^M|\boldsymbol{\Sigma}|}\right]^L. \tag{17}$$

We first find the ML estimate $\widehat{\boldsymbol{\Sigma}}_0$ of $\boldsymbol{\Sigma} \in \mathcal{S}_N$ under $\mathcal{H}_0$. When $N < M$, the results in [13] (see also [5]) imply that

$$\widehat{\boldsymbol{\Sigma}}_0 = \boldsymbol{V}_0\widehat{\boldsymbol{\Lambda}}_0\boldsymbol{V}_0^{\mathsf{H}}, \quad \widehat{\boldsymbol{\Lambda}}_0 = \text{Diag}(\widehat{\lambda}_{0,1}, \ldots, \widehat{\lambda}_{0,M}), \tag{18}$$

where $\{\widehat{\lambda}_{0,m}\}_{m=1}^{M}$ follow the definition in (12) with $i = 0$. That is, $\{\widehat{\lambda}_{0,m}\}_{m=1}^{M}$ is a smoothed version of the eigenvalues $\{\lambda_{0,m}\}$ of the sample covariance matrix $\frac{1}{L}\boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}}$ in decreasing order, where the smoothing averages the $M - N$ smallest eigenvalues to form the noise variance estimate $\widehat{\nu}_0$, as in (13). When $N = M$, the results in [12] (see also [3]) imply that $\widehat{\lambda}_{0,m} = \lambda_{0,m}\ \forall m$. In either case, the columns of $\boldsymbol{V}_0$ are the corresponding eigenvectors of the sample covariance matrix $\frac{1}{L}\boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}}$. It is straightforward to show [11] that

$$\tfrac{1}{L}\ln p(\boldsymbol{Y}|\mathcal{H}_0; \widehat{\boldsymbol{\Sigma}}_0) = -M - \textstyle\sum_{m=1}^{M} \ln\widehat{\lambda}_{0,m} - M\ln\pi. \tag{19}$$

B. GLRT Numerator

For the numerator of (16), $\boldsymbol{s} \sim p(\boldsymbol{s})$ and (15a) imply

$$p(\boldsymbol{Y}|\mathcal{H}_1;\boldsymbol{h},\boldsymbol{\Sigma}) = \int p(\boldsymbol{Y}|\boldsymbol{s},\mathcal{H}_1;\boldsymbol{h},\boldsymbol{\Sigma})\, p(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s} \tag{20}$$

$$= \int \frac{\exp(-\operatorname{tr}\{(\boldsymbol{Y}-\boldsymbol{h}\boldsymbol{s}^{\mathsf{H}})^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{h}\boldsymbol{s}^{\mathsf{H}})\})}{\pi^{ML}|\boldsymbol{\Sigma}|^{L}} p(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s}. \tag{21}$$

Exact maximization of $p(\boldsymbol{Y}|\mathcal{H}_1;\boldsymbol{h},\boldsymbol{\Sigma})$ over $\boldsymbol{h}$ and $\boldsymbol{\Sigma} \in \mathcal{S}_N$ appears to be intractable. We thus propose to approximate the maximization by applying EM [10] with hidden data $\boldsymbol{s}$. This implies that we iterate the following over $t = 0, 1, 2, \ldots$:

$$\big(\widehat{\boldsymbol{h}}^{(t+1)}, \widehat{\boldsymbol{\Sigma}}_1^{(t+1)}\big) = \underset{\boldsymbol{h}\in\mathbb{C}^M,\boldsymbol{\Sigma}\in\mathcal{S}_N}{\arg\max}\ \mathbb{E}\big\{\ln p(\boldsymbol{Y},\boldsymbol{s}|\mathcal{H}_1;\boldsymbol{h},\boldsymbol{\Sigma})\,\big|\,\boldsymbol{Y};\widehat{\boldsymbol{h}}^{(t)},\widehat{\boldsymbol{\Sigma}}_1^{(t)}\big\} \tag{22}$$

The EM algorithm is guaranteed to converge to a local maxima or saddle point of the likelihood (20) [14]. Furthermore, at each iteration t, the EM-approximated log-likelihood increases and lower bounds the true log-likelihood [15].

Because $p(\boldsymbol{s})$ is invariant to $\boldsymbol{h}$ and $\boldsymbol{\Sigma}$, (22) becomes

$$\underset{\boldsymbol{h}\in\mathbb{C}^M,\boldsymbol{\Sigma}_1\in\mathcal{S}_N}{\arg\max}\ \mathbb{E}\big\{\ln p(\boldsymbol{Y}|\boldsymbol{s},\mathcal{H}_1;\boldsymbol{h},\boldsymbol{\Sigma})\,\big|\,\boldsymbol{Y};\widehat{\boldsymbol{h}}^{(t)},\widehat{\boldsymbol{\Sigma}}_1^{(t)}\big\}. \tag{23}$$

It was shown [11] that

$$\widehat{\boldsymbol{h}}^{(t+1)} = \boldsymbol{Y}\widehat{\boldsymbol{s}}^{(t)}/E^{(t)} \tag{24}$$

$$\widehat{\boldsymbol{s}}^{(t)} \triangleq \mathbb{E}\{\boldsymbol{s}|\boldsymbol{Y};\widehat{\boldsymbol{h}}^{(t)},\widehat{\boldsymbol{\Sigma}}_1^{(t)}\} \tag{25}$$

$$E^{(t)} \triangleq \mathbb{E}\{\|\boldsymbol{s}\|^2|\boldsymbol{Y};\widehat{\boldsymbol{h}}^{(t)},\widehat{\boldsymbol{\Sigma}}_1^{(t)}\}. \tag{26}$$

Setting $\boldsymbol{h} = \widehat{\boldsymbol{h}}^{(t+1)}$ in (23) and simplifying, [11] shows that

$$\operatorname{tr}\big\{\boldsymbol{Y}\widetilde{\boldsymbol{P}}^{\perp}_{\widehat{\boldsymbol{s}}^{(t)}}\boldsymbol{Y}^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}\big\} + \ln|\boldsymbol{\Sigma}|^{L} \tag{27}$$

is the cost that must be minimized over $\boldsymbol{\Sigma} \in \mathcal{S}_N$, where

$$\widetilde{\boldsymbol{P}}^{\perp}_{\widehat{\boldsymbol{s}}^{(t)}} \triangleq \boldsymbol{I}_L - \widehat{\boldsymbol{s}}^{(t)}\widehat{\boldsymbol{s}}^{(t)\mathsf{H}}/E^{(t)}. \tag{28}$$

Note that $\widetilde{\boldsymbol{P}}^{\perp}_{\widehat{\boldsymbol{s}}^{(t)}}$ mimics a projection matrix, but is not one in general. Minimizing (27) is equivalent to maximizing

$$\frac{\exp(-\operatorname{tr}\{\boldsymbol{Y}\widetilde{\boldsymbol{P}}^{\perp}_{\widehat{\boldsymbol{s}}^{(t)}}\boldsymbol{Y}^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}\})}{\pi^{ML}|\boldsymbol{\Sigma}|^{L}}. \tag{29}$$

As with (17), when $N < M$, the results in [13] imply

$$\widehat{\boldsymbol{\Sigma}}_1^{(t+1)} = \boldsymbol{V}_1^{(t+1)}\widehat{\boldsymbol{\Lambda}}_1^{(t+1)}\boldsymbol{V}_1^{(t+1)\mathsf{H}}, \tag{30}$$

$$\widehat{\boldsymbol{\Lambda}}_1^{(t+1)} = \operatorname{Diag}(\widehat{\lambda}_{1,1}^{(t+1)},\ldots,\widehat{\lambda}_{1,M}^{(t+1)}) \tag{31}$$

$$\widehat{\lambda}_{1,m}^{(t+1)} = \begin{cases}\lambda_{1,m}^{(t+1)} & m = 1,\ldots,N\\ \widehat{\nu}_1^{(t+1)} & m = N+1,\ldots,M\end{cases} \tag{32}$$

$$\widehat{\nu}_1^{(t+1)} \triangleq \tfrac{1}{M-N}\textstyle\sum_{m=N+1}^{M}\lambda_{1,m}^{(t+1)}, \tag{33}$$

where $\{\lambda_{1,m}^{(t+1)}\}_{m=1}^{M}$ are the eigenvalues of the matrix $\frac{1}{L}\boldsymbol{Y}\widetilde{\boldsymbol{P}}^{\perp}_{\widehat{\boldsymbol{s}}^{(t)}}\boldsymbol{Y}^{\mathsf{H}}$ in decreasing order, and the columns of $\boldsymbol{V}_1^{(t+1)}$ are the corresponding eigenvectors.

Algorithm 1 EM update under white Gaussian interference

Require: Data $\boldsymbol{Y} \in \mathbb{C}^{M\times L}$, signal prior $p(\boldsymbol{s}) = \prod_{l=1}^{L} p_l(s_l)$.

1: Initialize $\widehat{\boldsymbol{s}}$ and $E > 0$ as described in text
2: **repeat**
3: $\quad \widehat{\boldsymbol{h}} \leftarrow \frac{1}{E}\boldsymbol{Y}\widehat{\boldsymbol{s}}$
4: $\quad \widehat{\boldsymbol{\Sigma}}_1 \leftarrow \frac{1}{L}\boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}} - \frac{E}{L}\widehat{\boldsymbol{h}}\widehat{\boldsymbol{h}}^{\mathsf{H}}$
5: $\quad$ Estimate interference rank N via (39).
6: $\quad \boldsymbol{g} \leftarrow \widehat{\boldsymbol{\Sigma}}_1^{-1}\widehat{\boldsymbol{h}}$
7: $\quad \xi \leftarrow \widehat{\boldsymbol{h}}^{\mathsf{H}}\boldsymbol{g}$
8: $\quad \boldsymbol{r} \leftarrow \frac{1}{\xi}\boldsymbol{Y}^{\mathsf{H}}\boldsymbol{g}$ where $\boldsymbol{r} \sim \mathcal{CN}(\boldsymbol{s}, \boldsymbol{I}/\xi)$
9: $\quad \widehat{s}_l \leftarrow \mathbb{E}\{s_l|r_l;\xi\}\ \forall l = 1,\ldots,L$
10: $\quad E \leftarrow \sum_{l=1}^{L}\mathbb{E}\{|s_l|^2|r_l;\xi\}$
11: **until** Terminated

The EM updates of $\widehat{\boldsymbol{s}}^{(t)}$ and $E^{(t)}$ in (25)-(26) depend on $p(\boldsymbol{s})$. For any independent prior (3), we can MMSE-estimate the symbols one at a time from the measurement equation

$$\boldsymbol{y}_l = \widehat{\boldsymbol{h}}^{(t)}s_l^{*} + \mathcal{CN}(\boldsymbol{0},\widehat{\boldsymbol{\Sigma}}_1^{(t)}). \tag{34}$$

Since whitened matched-filter (WMF) outputs form sufficient statistics [1], we can also estimate the symbols from

$$r_l^{(t)} \triangleq \frac{1}{\xi^{(t)}}\boldsymbol{y}_l^{(t)\mathsf{H}}(\widehat{\boldsymbol{\Sigma}}_1^{(t)})^{-1}\widehat{\boldsymbol{h}}^{(t)} = s_l + \mathcal{CN}\Big(0,\frac{1}{\xi^{(t)}}\Big) \tag{35}$$

$$\xi^{(t)} \triangleq \widehat{\boldsymbol{h}}^{(t)\mathsf{H}}(\widehat{\boldsymbol{\Sigma}}_1^{(t)})^{-1}\widehat{\boldsymbol{h}}^{(t)}. \tag{36}$$

For the Gaussian prior $p(s_l) = \mathcal{CN}(s_l;\mu_l,v_l)$ and discrete priors of the form $p(s_l) = \sum_{k=1}^{K_l}\omega_{lk}\delta(s_l - d_{lk})$, with alphabet $\mathcal{A}_l = \{d_{lk}\}_{k=1}^{K_l}$ and prior symbol probabilities $\omega_{lk} \geq 0$, expressions for $\mathbb{E}\{s_l|r_l\}$ and $\mathbb{E}\{|s_l|^2|r_l\}$ are given in [11]. The EM update procedure is summarized in Alg. 1.

Let us denote the final EM-based estimates of $\boldsymbol{s}$, $\boldsymbol{h}$, and $\boldsymbol{\Sigma}$ under $\mathcal{H}_1$ as $\widehat{\boldsymbol{s}}$, $\widehat{\boldsymbol{h}}$, and $\widehat{\boldsymbol{\Sigma}}_1$, respectively. Then

$$\tfrac{1}{L}\ln p(\boldsymbol{Y}|\mathcal{H}_1;\widehat{\boldsymbol{h}},\widehat{\boldsymbol{\Sigma}}_1) = -M - \sum_{m=1}^{M}\ln\widehat{\lambda}_{1,m} - M\ln\pi \tag{37}$$

following steps similar to (19). Recalling (16), the log-domain GLRT is obtained by subtracting (19) from (37), yielding

$$\sum_{m=1}^{M}\ln\frac{\widehat{\lambda}_{0,m}}{\widehat{\lambda}_{1,m}} \gtrless \eta'. \tag{38}$$

It was shown in [11] that Alg. 1 becomes equivalent to Forsythe's iterative scheme from [9, p.110] if $N = M$ and the "soft" estimates $\widehat{\boldsymbol{s}} = \mathbb{E}\{\boldsymbol{s}|\boldsymbol{r};\xi\}$ and $E = \mathbb{E}\{\|\boldsymbol{s}\|^2|\boldsymbol{r};\xi\}$ in lines 9-10 are replaced by the "hard" ML estimates $\widehat{\boldsymbol{s}}_{\mathsf{ML}} = \arg\min_{\boldsymbol{s}\in\mathcal{A}^L}\|\boldsymbol{r}-\boldsymbol{s}\|^2$ and $E_{\mathsf{ML}} = \|\widehat{\boldsymbol{s}}_{\mathsf{ML}}\|^2$. In Sec. IV, we show that soft estimates lead to improved detection performance due to reduced error propagation.

The interference rank $N = \operatorname{rank}(\boldsymbol{R})$ in line 5 can be estimated using the standard model-order selection approach described in, e.g., [16], [17], which specifies

$$\widehat{N} = \underset{N=0,\ldots,N_{\max}}{\arg\max}\ \ln p(\boldsymbol{Y}|\mathcal{H}_1;\widehat{\boldsymbol{\Theta}}_N) - J(D(N)), \tag{39}$$

where $J(\cdot)$ is a penalty function, $\widehat{\boldsymbol{\Theta}}_N$ is the ML parameter estimate under rank hypothesis N, and $D(N)$ is the degrees-of-freedom (DoF) in the parameters $\boldsymbol{\Theta}_N = \{\boldsymbol{h}, \boldsymbol{\Sigma}\}$ for $\boldsymbol{h} \in \mathbb{C}^M$ and $\boldsymbol{\Sigma} \in \mathcal{S}_N$. Thus, $D(N) = (2M-N)N+2M+1$. Common choices of $J(\cdot)$ result from AIC, BIC, and MDL. For our numerical experiments, we used the Generalized Information Criterion, i.e., $J(D) = GD$, with $G = 12.5$.

The initialization of $(\widehat{\boldsymbol{s}}, E)$ in Alg. 1 affects the quality of the final EM estimate. We now propose an initialization for the training/data structure in (2), i.e., $\boldsymbol{Y} = \left[\boldsymbol{Y}_{\mathsf{t}}\, \boldsymbol{Y}_{\mathsf{d}}\right]$ with

$$\boldsymbol{Y}_{\mathsf{t}} = \boldsymbol{h}\boldsymbol{s}_{\mathsf{t}}^{\mathsf{H}} + \boldsymbol{N}_{\mathsf{t}},\ \ \mathrm{vec}(\boldsymbol{N}_{\mathsf{t}}) \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_Q \otimes \boldsymbol{\Sigma}) \quad (40)$$

$$\boldsymbol{Y}_{\mathsf{d}} = \boldsymbol{h}\boldsymbol{s}_{\mathsf{d}}^{\mathsf{H}} + \boldsymbol{N}_{\mathsf{d}},\ \ \mathrm{vec}(\boldsymbol{N}_{\mathsf{d}}) \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_{L-Q} \otimes \boldsymbol{\Sigma}), \quad (41)$$

and $\boldsymbol{s} = [\boldsymbol{s}_{\mathsf{t}}^{\mathsf{H}}, \boldsymbol{s}_{\mathsf{d}}^{\mathsf{H}}]^{\mathsf{H}}$. Essentially, we would like to estimate the random vector $\boldsymbol{s}_{\mathsf{d}} \sim \prod_{l=Q+1}^{L} p_l(s_l)$ from measurements $\boldsymbol{Y}$ under known $\boldsymbol{s}_{\mathsf{t}}$ but unknown $\boldsymbol{s}_{\mathsf{d}}, \boldsymbol{h}, \boldsymbol{\Sigma}, \boldsymbol{N}$.

Recall that the WMF outputs (35) are sufficient statistics [1] for estimating $\boldsymbol{s}_{\mathsf{d}}$. Because $\boldsymbol{\Sigma}$ and $\boldsymbol{h}$ are unknown during initialization, we propose to estimate them from the training data $\boldsymbol{Y}_{\mathsf{t}}$ and use the results to compute approximate-WMF outputs of the form

$$\widehat{r}_l \triangleq \boldsymbol{y}_l^{\mathsf{H}} \widehat{\boldsymbol{\Sigma}}_{\mathsf{t}}^{-1} \widehat{\boldsymbol{h}}_{\mathsf{t}}. \quad (42)$$

With appropriate scaling $\beta \in \mathbb{C}$, we get an unbiased statistic

$$\beta \widehat{r}_l \approx s_l + \mathcal{CN}(0, 1/\widehat{\xi}) \text{ for } l \in \{Q+1, \ldots, L\} \quad (43)$$

that can be converted to MMSE symbol estimates $\widehat{s}_l$ and energy E as described in [11]. For $(\widehat{\boldsymbol{\Sigma}}_{\mathsf{t}}, \widehat{\boldsymbol{h}}_{\mathsf{t}})$ in (42), we use the ML estimate

$$\widehat{\boldsymbol{h}}_{\mathsf{t}} \triangleq \boldsymbol{Y}_{\mathsf{t}} \boldsymbol{s}_{\mathsf{t}} / \|\boldsymbol{s}_{\mathsf{t}}\|^2 \quad (44)$$

and a regularized covariance estimate of the form [18]

$$\widehat{\boldsymbol{\Sigma}}_{\mathsf{t}}^{(\alpha)} = (1-\alpha)\widehat{\boldsymbol{\Sigma}}_{\mathsf{t}} + \alpha c \boldsymbol{I}_M,\ \ \alpha \in (0, 1], \quad (45)$$

with $\widehat{\boldsymbol{\Sigma}}_{\mathsf{t}} \triangleq \frac{1}{Q}\boldsymbol{Y}_{\mathsf{t}} \boldsymbol{P}_{\boldsymbol{s}_{\mathsf{t}}}^{\perp} \boldsymbol{Y}_{\mathsf{t}}^{\mathsf{H}}$ and $c \triangleq \mathrm{tr}(\widehat{\boldsymbol{\Sigma}}_{\mathsf{t}})/M$. We choose α to maximize the precision $\widehat{\xi}$, as estimated by leave-one-out cross-validation (LOOCV) [19] on the training data. Our LOOCV approach is similar to the "SEO" scheme from [20] but targets minimum-variance unbiased estimation rather than MMSE estimation and, more significantly, handles non-white interference. Details are provided in [11].

For a given α, the unbiasing gain $\beta^{(\alpha)}$ (recall (43)) obeys

$$\mathbb{E}\left\{\beta^{(\alpha)} \widehat{r}_l^{(\alpha)} \middle| s_l\right\} = s_l,\quad l \in \{1, \ldots, Q\}, \quad (46)$$

and thus can be estimated as

$$\beta^{(\alpha)} = \frac{1}{\mathbb{E}\left\{\widehat{r}_l^{(\alpha)}/s_l\right\}} \approx \frac{Q}{\sum_{l=1}^{Q} \widehat{r}_l^{(\alpha)}/s_l} \triangleq \widehat{\beta}^{(\alpha)}. \quad (47)$$

After scaling by $\widehat{\beta}^{(\alpha)}$, the error precision $\widehat{\xi}^{(\alpha)}$ is

$$\widehat{\xi}^{(\alpha)} = \frac{1}{\frac{1}{Q}\sum_{l=1}^{Q} \left|\widehat{\beta}^{(\alpha)} \widehat{r}_l^{(\alpha)} - s_l\right|^2}. \quad (48)$$

The value of α can be optimized by maximizing $\widehat{\xi}^{(\alpha)}$ over a grid of possible values.

IV. Numerical Experiments

We now present numerical experiments to evaluate the proposed detectors. Unless otherwise noted, we used $M = 64$ array elements, $L = 1024$ total symbols, $Q = 32$ training symbols, and $N = 5$ interferers. (Note that $Q \ll M$.) The signal $\boldsymbol{s}$ was i.i.d. QPSK with variance 1, the noise $\boldsymbol{W}$ was i.i.d. Gaussian with variance ν, and the interference $\boldsymbol{\Phi}$ was i.i.d. Gaussian with variance σ_i^2/N, giving a total interference power of σ_i^2. For the array, we assumed a uniform planar array (UPA) with half-wavelength element spacing operating in the narrowband regime. For the signal's array response $\boldsymbol{h}$, we assumed that the signal arrived from a random (horizontal,vertical) angle pair drawn uniformly on $[0, 2\pi)^2$. For the nth interferer's array response $\boldsymbol{b}_n$, we used the arrival angle corresponding to the nth largest sidelobe in $\boldsymbol{h}$. Detection performance was quantified using the rate of correct detection when the detector threshold η is set to achieve a false-alarm rate of 10^{-4}. All simulation results represent the average of 10 000 independent draws of $\{\boldsymbol{h}, \boldsymbol{s}, \boldsymbol{B}, \boldsymbol{\Phi}, \boldsymbol{W}\}$.

We considered two existing methods that use only the training data $\boldsymbol{Y}_{\mathsf{t}}$: Kelly's full-rank approach (7), i.e., "kel-tr," and the Kang/Monga/Rangaswamy approach (11) with rank N estimated as in (39), i.e., "kmr-tr." We also tested the proposed EM-based methods, which use the full data $\boldsymbol{Y}$: Alg. 1 with full rank $N = M$, i.e., "kel-em," and Alg. 1 with N estimated as in (39), i.e., "kmr-em." For the EM algorithm, we used a maximum of 50 iterations but terminated at iteration $i > 1$ if $\|\widehat{\boldsymbol{s}}^{(i)} - \widehat{\boldsymbol{s}}^{(i-1)}\|/\|\widehat{\boldsymbol{s}}^{(i)}\| < 0.01$. We also tested Forsythe's iterative method [9, p. 110] by running Alg. 1 with full rank $N = M$ and hard symbol estimates. Finally, we tested a low-rank version of Forsythe's method by running Alg. 1 with hard estimates and N estimated via (39).

A. Performance versus SINR

Figure 1 shows detection-rate at false-alarm-rate=10^{-4} versus $\nu = \sigma_i^2$ for various detectors. There we see that the proposed EM-based, full-data detectors, kel-em and kmr-em, significantly outperformed their training-based counterparts, kel-tr and kmr-tr. We also see that the proposed detectors outperformed their hard-detection counterparts, forsythe and forsythe-lowrank, and we attribute this result to a reduction in error propagation. Finally, we see that the low-rank methods outperformed the corresponding full-rank methods.

B. Performance versus SIR at fixed SNR

Figure 2 shows detection-rate at false-alarm-rate=10^{-4} versus interference power σ_i^2 at the fixed noise power $\nu = Q$. The proposed EM-based, low-rank detector kmr-em gave no errors over 10 000 trials. The proposed EM-based, full-rank detector kel-em outperformed its training-based counterpart kel-tr, but succumbed to error propagation at low σ_i^2. The non-monotonic behavior of the training based schemes kel-tr, kmr-tr, and mcw-tr results from imperfect rank estimation.

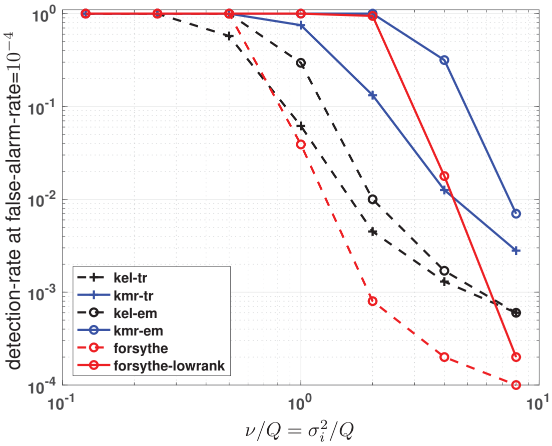

Fig. 1. Detection-rate at false-alarm-rate=10^{-4} versus $\nu = \sigma_i^2$.

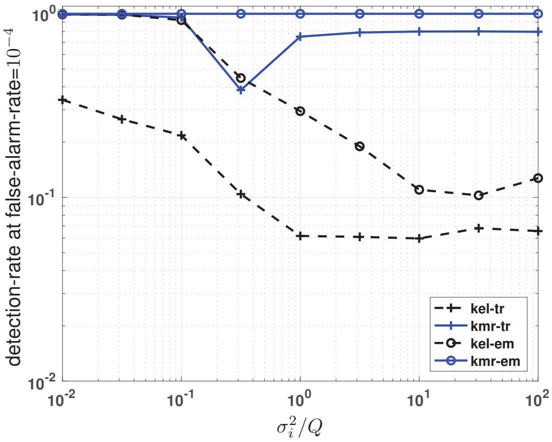

Fig. 2. Detection-rate at false-alarm-rate=10^{-4} versus σ_i^2 under $\nu = Q$.

C. Performance versus training length Q

Figure 3 shows detection-rate at false-alarm-rate=10^{-4} versus training length Q for various detectors under $\nu = \sigma_i^2 = Q$. Here, ν and σ_i^2 grew with Q to prevent the error-rate from vanishing with Q due to spreading gain. The kel-tr trace is clipped on the left because Kelly's approach cannot be applied when $Q < M$. Figure 3 shows that the proposed EM-based, low-rank detectors kmr-em and mcw-em far outperformed the others when Q was between 32 and 128. When $Q = 1024 = L$, there are no data symbols, in which case kmr-em is equivalent to kmr-tr.

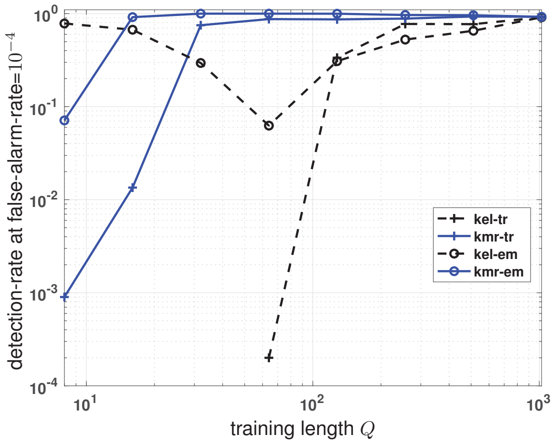

Fig. 3. Detection-rate at false-alarm-rate=10^{-4} versus training length Q under $\nu = \sigma_i^2 = Q$.

References

[1] L. L. Scharf, *Statistical Signal Processing*. Reading, MA: Addison-Wesley, 1991.

[2] D. W. Bliss and P. A. Parker, "Temporal synchronization of MIMO wireless communication in the presence of interference," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1794–1806, 2010.

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

[4] K. Gerlach and M. J. Steiner, "Fast converging adaptive detection of Doppler-shifted, range-distributed targets," *IEEE Trans. Signal Process.*, vol. 48, no. 9, pp. 2686–2690, 2000.

[5] B. Kang, V. Monga, and M. Rangaswamy, "Rank-constrained maximum likelihood estimation of structured covariance matrices," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 1, pp. 501–515, 2014.

[6] A. Aubry, A. De Maio, L. Pallotta, and A. Farina, "Radar detection of distributed targets in homogeneous interference whose inverse covariance structure is defined via unitary invariant functions," *IEEE Trans. Signal Process.*, vol. 61, pp. 4949–4961, Oct. 2013.

[7] L. T. McWhorter, "A high resolution detector in multi-path environments," in *Proc. Workshop Adapt. Sensor Array Process.*, (Lexington, MA), 2004.

[8] F. Bandiera, A. De Maio, A. S. Greco, and G. Ricci, "Adaptive radar detection of distributed targets in homogeneous and partially homogeneous noise plus subspace interference," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1223–1237, 2007.

[9] K. W. Forsythe, "Utilizing waveform features for adaptive beamforming and direction finding with narrowband signals," *Lincoln Lab. J.*, vol. 10, no. 2, pp. 99–126, 1997.

[10] A. Dempster, N. M. Laird, and D. B. Rubin, "Maximum-likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc.*, vol. 39, pp. 1–17, 1977.

[11] E. Bryne and P. Schniter, "Adaptive detection of structured signals in low-rank interference," *arXiv:1808.05650*, 2018.

[12] N. R. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction)," *Ann. Math. Statist.*, vol. 34, no. 1, pp. 152–177, 1963.

[13] T. W. Anderson, "Asymptotic theory for principal component analysis," *Ann. Math. Statist.*, vol. 34, no. 1, pp. 122–148, 1963.

[14] C. F. J. Wu, "On the convergence properties of the EM algorithm," *Ann. Statist.*, vol. 11, no. 1, pp. 95–103, 1983.

[15] R. Neal and G. Hinton, "A view of the EM algorithm that justifies incremental, sparse, and other variants," in *Learning in Graphical Models* (M. I. Jordan, ed.), pp. 355–368, MIT Press, 1998.

[16] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust. Speech & Signal Process.*, vol. 33, no. 2, pp. 387–392, 1986.

[17] P. Stoica and Y. Selén, "Model-order selection: A review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, pp. 36–47, July 2004.

[18] J. P. Hoffbeck and D. A. Landgrebe, "Covariance matrix estimation and classification with limited training data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, pp. 763–767, 1996.

[19] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," *Statist. Surv.*, vol. 4, pp. 40–79, 2010.

[20] J. Tong, P. J. Schreier, Q. Guo, S. Tong, J. Xi, and Y. Yu, "Shrinkage of covariance matrices for linear signal estimation using cross-validation," *IEEE Trans. Signal Process.*, vol. 64, pp. 2965–2975, June 2016.