Knowledge-Aided Covariance Matrix Estimation and Adaptive Detection in Compound-Gaussian Noise

Francesco Bandiera, Olivier Besson, and Giuseppe Ricci

***Abstract*—We address the problem of adaptive detection of a signal of interest embedded in colored noise modeled in terms of a compound-Gaussian process. The covariance matrices of the primary and the secondary data share a common structure while having different power levels. A Bayesian approach is proposed here, where both the power levels and the structure are assumed to be random, with some appropriate distributions. Within this framework we propose MMSE and MAP estimators of the covariance structure and their application to adaptive detection using the NMF test statistic and an optimized GLRT herein derived. Some results, also in comparison with existing algorithms, are presented to illustrate the performances of the proposed detectors. The relevant result is that the solutions presented herein allows to improve the performance over conventional ones, especially in presence of a small number of training data.**

***Index Terms*—Adaptive detection, compound-Gaussian clutter, covariance matrix estimation, heterogeneous environments, MAP estimation, MMSE estimation.**

I. Introduction

Detection of a signal of interest in a background of noise is a fundamental task in many applications, including radar, communications or sonar. This is especially the case for radar systems whose core task is to detect a target in presence of clutter, thermal noise and possibly jamming, see [1] for a very good list of publications on this topic. Typically, the presence of a target, with given space and/or time signature $\boldsymbol{v} \in \mathbb{C}^{N\times 1}$ ($\mathbb{C}$ being the complex field), is sought in a (range/doppler) cell under test (CUT), given an observation vector $\boldsymbol{z} \in \mathbb{C}^{N\times 1}$, the so-called primary data, that corresponds to the output of an array of sensors. To be quantitative, the classical problem of detection is often formulated in terms of the following binary hypothesis test

$$\begin{cases} H_0: & \boldsymbol{z} = \boldsymbol{n}, \\ H_1: & \boldsymbol{z} = \alpha \boldsymbol{v} + \boldsymbol{n}, \end{cases} \tag{1}$$

where $\alpha \in \mathbb{C}$ is the unknown amplitude of the target (it takes into account both target and channel effects) and $\boldsymbol{n} \in \mathbb{C}^{N\times 1}$ is the noise component.

As to the noise component, experimental data [2] as well as physical and theoretical arguments [3], have demonstrated that the Gaussian assumption is not always valid; in fact, the clutter can generally be modeled as a compound-Gaussian process that, when observed on sufficiently short time intervals, degenerates into a spherically invariant random process (SIRP) [4]. It follows that $\boldsymbol{n} = \sqrt{\tau}\boldsymbol{g}$, with $\boldsymbol{g}$ a complex normal random vector with covariance matrix $\boldsymbol{R}$ (the so-called speckle component) and τ a positive random variable (rv) independent of $\boldsymbol{g}$ (the so-called texture component).

Optimized detection structures, i.e., solutions designed for a preassigned distribution of τ, have been proposed and assessed, see for example [5]. More importantly, it has been shown in [6] that the normalized matched filter (NMF) is asymptotically optimum[1] when the matrix $\boldsymbol{R}$ is known and τ is an arbitrary (positive) random variable. The NMF is equivalent to the following decision rule:

$$\frac{|\boldsymbol{v}^\dagger \boldsymbol{R}^{-1}\boldsymbol{z}|^2}{(\boldsymbol{v}^\dagger \boldsymbol{R}^{-1}\boldsymbol{v})(\boldsymbol{z}^\dagger \boldsymbol{R}^{-1}\boldsymbol{z})} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \tag{2}$$

where γ is the threshold value that ensures a preassigned probability of false alarm (P_{fa}) and † denotes conjugate transpose. Since the matrix $\boldsymbol{R}$ is usually unknown, the NMF cannot be directly implemented using (2). In general, estimation of $\boldsymbol{R}$ is not feasible based upon the sample $\boldsymbol{z}$ only, unless some *a priori* knowledge or some structure on $\boldsymbol{R}$ is assumed. In order to estimate $\boldsymbol{R}$, a set of noise-only training samples $\boldsymbol{z}_k \in \mathbb{C}^{N\times 1}$, $k = 1, \ldots, K$, is commonly used. These training samples, also referred to as secondary data, are usually obtained from range cells adjacent to the CUT. In fact, clutter returns can be modeled as $\boldsymbol{z}_k = \sqrt{\tau_k}\boldsymbol{g}_k$, $k = 1, \ldots, K$, where the $\boldsymbol{g}_k$'s are complex normal vectors with covariance matrix $\boldsymbol{R}$ and the τ_k's are positive rv's. It is customary to assume that $\boldsymbol{g}, \boldsymbol{g}_1, \ldots, \boldsymbol{g}_K$ are independent random vectors while $\tau, \tau_1, \ldots, \tau_K$ are rv's drawn from a possibly correlated wide sense stationary random process. In this scenario, the problem of estimating the covariance matrix is generally intractable. A normalized sample covariance matrix estimator has been advocated in [7] and [8] that guarantees the constant false alarm rate (CFAR) property with respect to the texture statistics. Also, considering the τ_k's as unknown deterministic quantities, it was shown that the maximum likelihood estimate of $\boldsymbol{R}$ obeys an implicit equation, that can be "solved" through an iterative procedure [9]–[12].

Although the homogeneous assumption is an idealized situation [13], it is possible to select training samples that are most homogeneous with the CUT and use only the retained ones to estimate the noise covariance matrix. The reader is referred to [14] and references therein for examples of applications of this rationale.

More recently, the so-called knowledge-aided space-time adaptive processing (KA-STAP) has been recognized as one of the potentially most efficient way to handle heterogeneities [15]. KA-STAP improves the performance of adaptive detection schemes using additional (*a priori*) information, such as digital elevation and terrain data, synthetic aperture radar imagery, etc. The reader is referred to [16] and [17] and references therein, just to give some examples. Alternatively, the Bayesian approach can be advocated. The basic idea of the Bayesian modeling is to assume that the quantities τ, $\boldsymbol{R}$, and τ_k, $k = 1, \ldots, K$, are random with some preassigned *a priori* distribution. Examples of this modeling can be found in [18]–[21] where secondary data samples are homogeneous and [22] that extends the results of [21] to the heterogeneous case.

In this correspondence, we consider a "knowledge-aided scheme" for solving problem (1) and estimating the covariance matrix $\boldsymbol{R}$. Specifically, we assume that τ and $\boldsymbol{R}$ are random quantities and that[2] $\boldsymbol{n}|\tau, \boldsymbol{R} \sim \mathcal{CN}_N(\boldsymbol{0}, \tau\boldsymbol{R})$. $\boldsymbol{R}$ has known mean, $\overline{\boldsymbol{R}}$ say, that can be obtained, as an example, from the general clutter covariance matrix model of [23]. Additionally, the secondary samples $\boldsymbol{z}_k$'s

Manuscript received October 27, 2009; accepted June 02, 2010. Date of publication June 14, 2010; date of current version September 15, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Namrata Vaswani. The work of O. Besson was supported by the Mission pour la Recherche et l'Innovation Scientifique (MRIS) of the DGA under Contract 2009.60.033.00.470.75.01.

F. Bandiera and G. Ricci are with the Dipartimento di Ingegneria dell'Innovazione, Università del Salento, 73100 Lecce, Italy (e-mail: francesco.bandiera@unisalento.it; giuseppe.ricci@unisalento.it).

O. Besson is with the Départment Electronique Optronique Signal, Université de Toulouse, ISAE, Toulouse 31055, France (e-mail: olivier.besson@isae.fr).

Digital Object Identifier 10.1109/TSP.2010.2052922

[1]Optimality in this case means that the NMF tends to coincide to the generalized likelihood ratio test (GLRT) as the number of integrated pulses diverges.

[2]We will use throughout the correspondence the following notation $\boldsymbol{x} \sim \mathcal{CN}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ to denote that the vector $\boldsymbol{x} \in \mathbb{C}^{m\times 1}$ is ruled by the complex normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

change their distribution on a cell-to-cell basis according to $\boldsymbol{z}_k|\tau_k$, $\boldsymbol{R} \sim \mathcal{CN}_N(\boldsymbol{0}, \tau_k\boldsymbol{R})$, τ_k's being rv's. Within this framework, we consider optimal, i.e., MMSE and maximum *a posteriori* probability (MAP) estimation of both $\boldsymbol{\tau} = [\tau_1 \cdots \tau_K]$ and $\boldsymbol{R}$. Subsequently, we propose ad hoc detectors obtained replacing the true $\boldsymbol{R}$ with its estimates. To this end, we consider two test statistics: the NMF (2) and the GLRT for random τ and known $\boldsymbol{R}$.

The reminder of the correspondence is organized as follows: next section is devoted to the description of the data model while the estimation is the object of Section III. In Section IV we attack the detection problem and a performance assessment is presented in Section V. Finally, some concluding remarks are given in Section VI.

II. Data Model

In this section, we provide the assumptions about our model. As stated previously, we assume that the K vectors $\boldsymbol{z}_k \in \mathbb{C}^{N\times 1}$, conditioned on τ_k and $\boldsymbol{R}$, are independent, zero-mean complex normal random vectors with covariance matrix $\tau_k\boldsymbol{R}$ (in symbols $\boldsymbol{z}_k|\tau_k, \boldsymbol{R} \sim \mathcal{CN}_N(\boldsymbol{0}, \tau_k\boldsymbol{R})$), i.e.,

$$f(\boldsymbol{z}_k|\tau_k, \boldsymbol{R}) = \frac{1}{\pi^N \tau_k^N \det(\boldsymbol{R})} \exp\left\{-\frac{\boldsymbol{z}_k^\dagger \boldsymbol{R}^{-1}\boldsymbol{z}_k}{\tau_k}\right\} \tag{3}$$

where $\det(\cdot)$ stands for the determinant of the matrix argument. Now, since $\boldsymbol{\tau}$ and $\boldsymbol{R}$ are random quantities, one needs to assign prior distributions for them. On one hand, if one wishes to make the least possible assumptions on the random variables, a non-informative, e.g., Jeffreys, prior can be the solution [24]. However, this approach more or less corresponds to maximum likelihood estimation as almost no prior information about the unknown variables is available. When prior information is to be included in the model, a tradeoff must be made between plausibility, relevance of the prior distributions and mathematical tractability [25]. In our case, we propose to choose for $\boldsymbol{\tau}$ and $\boldsymbol{R}$ conjugate priors [24, p. 41]; moreover, we assume that $\boldsymbol{\tau}$ and $\boldsymbol{R}$ are independent. More precisely, we assume that $\boldsymbol{R}$ is drawn from a complex inverse Wishart distribution, with mean $\overline{\boldsymbol{R}}$ (an Hermitian and positive definite matrix) and ν $(> N+1)$ degrees of freedom [21], i.e.,

$$f(\boldsymbol{R}) = \frac{\det\left[(\nu-N)\overline{\boldsymbol{R}}\right]^\nu}{\widetilde{\Gamma}_N(\nu)\det(\boldsymbol{R})^{\nu+N}} \mathrm{etr}\left\{-(\nu-N)\boldsymbol{R}^{-1}\overline{\boldsymbol{R}}\right\} \tag{4}$$

where $\mathrm{etr}(\cdot)$ stands for the exponential of the trace of the matrix argument and $\widetilde{\Gamma}_N(\nu)$ is given by

$$\widetilde{\Gamma}_N(\nu) = \pi^{N(N-1)/2} \prod_{n=1}^{N} \Gamma(\nu-n+1) \tag{5}$$

with $\Gamma(x)$ being, in turn, the Eulerian Gamma function. We denote it as $\boldsymbol{R} \sim \mathcal{CW}^{-1}((\nu-N)\overline{\boldsymbol{R}}, \nu)$. $\overline{\boldsymbol{R}}$ is the expected value of $\boldsymbol{R}$ while ν sets the "distance" between $\overline{\boldsymbol{R}}$ and $\boldsymbol{R}$; in fact, as ν increases $\boldsymbol{R}$ is closer to $\overline{\boldsymbol{R}}$ (in the sense that the variance of $\boldsymbol{R}$ decreases). In order to have a conjugate prior for τ_k, we assume that the τ_k's are independent and distributed according to an inverse Gamma distribution with parameters q_k and β_k [24], i.e.,

$$f(\tau_k) = \frac{\beta_k^{q_k}}{\Gamma(q_k)\tau_k^{q_k+1}} \exp\left\{-\frac{\beta_k}{\tau_k}\right\}, \quad \tau_k \geq 0, q_k > 2, \beta_k > 0 \tag{6}$$

which we denote as $\tau_k \sim \mathcal{IG}(q_k, \beta_k)$. For what concerns, instead, the primary data $\boldsymbol{z}$, we assume that, under the H_i hypothesis, $i = 0, 1$, $\boldsymbol{z}|i\alpha, \tau, \boldsymbol{R} \sim \mathcal{CN}_N(i\alpha\boldsymbol{v}, \tau\boldsymbol{R})$. In addition, the *a priori* distribution of τ is $\tau \sim \mathcal{IG}(q, \beta)$ with τ independent of both $\boldsymbol{R}$ and $\boldsymbol{\tau}$.

Equations (3), (4), and (6) form the model which will be used in the subsequent sections for estimation and detection purposes.

III. Estimation

In this section, we will derive different estimators based upon MMSE and MAP criteria and possible modifications.

A. Posterior Distributions

In order to pursue our goals we need to find the posterior distribution $f(\boldsymbol{R}|\boldsymbol{Z})$, where $\boldsymbol{Z} = [\boldsymbol{z}_1 \cdots \boldsymbol{z}_K] \in \mathbb{C}^{N\times K}$ is the secondary data matrix. Ignoring irrelevant constants, the joint posterior distribution of $\boldsymbol{\tau}$ and $\boldsymbol{R}$ can be written as follows:

$$\begin{aligned} f(\boldsymbol{\tau}, \boldsymbol{R}|\boldsymbol{Z}) &\propto f(\boldsymbol{Z}|\boldsymbol{\tau}, \boldsymbol{R}) f(\boldsymbol{\tau}) f(\boldsymbol{R}) \\ &\propto \frac{1}{\det(\boldsymbol{R})^{\nu+K+N}} \mathrm{etr}\left\{-(\nu-N)\boldsymbol{R}^{-1}\overline{\boldsymbol{R}}\right\} \\ &\quad \times \prod_{k=1}^{K} \frac{1}{\tau_k^{q_k+N+1}} \exp\left\{-\frac{\beta_k + \boldsymbol{z}_k^\dagger \boldsymbol{R}^{-1}\boldsymbol{z}_k}{\tau_k}\right\} \end{aligned} \tag{7}$$

where $\propto$ means "proportional to." Integration over $\boldsymbol{\tau}$ provides

$$\begin{aligned} f(\boldsymbol{R}|\boldsymbol{Z}) &= \int f(\boldsymbol{\tau}, \boldsymbol{R}|\boldsymbol{Z}) d\boldsymbol{\tau} \\ &\propto \frac{1}{\det(\boldsymbol{R})^{\nu+K+N}} \mathrm{etr}\left\{-(\nu-N)\boldsymbol{R}^{-1}\overline{\boldsymbol{R}}\right\} \\ &\quad \times \prod_{k=1}^{K} \int \frac{1}{\tau_k^{q_k+N+1}} \exp\left\{-\frac{\beta_k + \boldsymbol{z}_k^\dagger \boldsymbol{R}^{-1}\boldsymbol{z}_k}{\tau_k}\right\} d\tau_k \\ &\propto \frac{1}{\det(\boldsymbol{R})^{\nu+K+N}} \mathrm{etr}\left\{-(\nu-N)\boldsymbol{R}^{-1}\overline{\boldsymbol{R}}\right\} \\ &\quad \times \prod_{k=1}^{K} \left[\beta_k + \boldsymbol{z}_k^\dagger \boldsymbol{R}^{-1}\boldsymbol{z}_k\right]^{-(q_k+N)} \end{aligned} \tag{8}$$

where we have used the fact that (6) is a density and hence it integrates to one.

B. MMSE Estimation

The MMSE estimate of $\boldsymbol{R}$ is given by the posterior mean

$$\widehat{\boldsymbol{R}}_{\mathrm{MMSE}} = E[\boldsymbol{R}|\boldsymbol{Z}] = \int \boldsymbol{R} f(\boldsymbol{R}|\boldsymbol{Z}) d\boldsymbol{R}. \tag{9}$$

Unfortunately, previous integral cannot be obtained in closed form. Furthermore, distribution (8) is not a classical one; as a consequence, it is not possible to investigate generating samples drawn from it and averaging them to approximate the MMSE estimator. In contrast, as we shall see next, the conditional posterior distributions $f(\boldsymbol{R}|\boldsymbol{\tau}, \boldsymbol{Z})$ and $f(\boldsymbol{\tau}|\boldsymbol{R}, \boldsymbol{Z})$ are not only easy to obtain but also belong to familiar classes of distributions. As a matter of fact, from (7) it is easy to see that

$$f(\boldsymbol{\tau}|\boldsymbol{R}, \boldsymbol{Z}) \propto \prod_{k=1}^{K} \frac{1}{\tau_k^{q_k+N+1}} \exp\left\{-\frac{\beta_k + \boldsymbol{z}_k^\dagger \boldsymbol{R}^{-1}\boldsymbol{z}_k}{\tau_k}\right\} \tag{10}$$

and hence

$$\tau_k|\boldsymbol{R}, \boldsymbol{Z} \sim \mathcal{IG}\left(q_k + N, \beta_k + \boldsymbol{z}_k^\dagger \boldsymbol{R}^{-1}\boldsymbol{z}_k\right). \tag{11}$$

TABLE I
GIBBS SAMPLER TO APPROXIMATE THE MMSE ESTIMATOR OF $\boldsymbol{R}$

Input: $\boldsymbol{Z}$
1) generate an initial value $\boldsymbol{R}^{(0)}$;
2) **for** $n = 1, \ldots, N_{bi} + N_r$,
3) generate $\boldsymbol{\tau}^{(n)}$ according to $f(\boldsymbol{\tau}|\boldsymbol{R}^{(n-1)}, \boldsymbol{Z})$;
4) generate $\boldsymbol{R}^{(n)}$ according to $f(\boldsymbol{R}|\boldsymbol{\tau}^{(n)}, \boldsymbol{Z})$;
5) **end for**
Output: sequence of random variables $\boldsymbol{\tau}^{(n)}$ and $\boldsymbol{R}^{(n)}$.

Using (7) again, we have also that

$$f(\boldsymbol{R}|\boldsymbol{\tau}, \boldsymbol{Z}) \propto \frac{1}{\det(\boldsymbol{R})^{\nu+K+N}} \times \mathrm{etr}\left\{-\boldsymbol{R}^{-1}\left((\nu-N)\overline{\boldsymbol{R}} + \sum_{k=1}^{K}\frac{\boldsymbol{z}_k\boldsymbol{z}_k^\dagger}{\tau_k}\right)\right\} \quad (12)$$

which implies that

$$\boldsymbol{R}|\boldsymbol{\tau}, \boldsymbol{Z} \sim \mathcal{CW}^{-1}\left((\nu-N)\overline{\boldsymbol{R}} + \sum_{k=1}^{K}\frac{\boldsymbol{z}_k\boldsymbol{z}_k^\dagger}{\tau_k}, \nu+K\right). \quad (13)$$

Consequently, it is quite standard to generate samples drawn from $f(\boldsymbol{R}|\boldsymbol{\tau}, \boldsymbol{Z})$ and $f(\boldsymbol{\tau}|\boldsymbol{R}, \boldsymbol{Z})$. This suggests the use of a Gibbs-sampler [20], [24], whose procedure is reported in Table I, where N_{bi} stands for the number of burn-in iterations and N_r is the number of samples which are effectively averaged. Some statistically sound criteria, such as the potential scale reduction factor [24], are available to select the values of N_{bi} and N_r that ensure convergence of the Gibbs sampler. The latter is known to generate random variables which are asymptotically distributed according to the posterior distributions $f(\boldsymbol{R}|\boldsymbol{Z})$ and $f(\boldsymbol{\tau}|\boldsymbol{Z})$, and therefore a natural way to approximate the MMSE estimator is to average the N_r last values generated by the sampler, i.e.,

$$\widehat{\boldsymbol{R}}_{\mathrm{MMSE}} \approx \frac{1}{N_r}\sum_{n=N_{bi}+1}^{N_{bi}+N_r}\boldsymbol{R}^{(n)}. \quad (14)$$

C. MAP Estimation

As an alternative to MMSE estimation, one can consider the MAP estimate of $\boldsymbol{R}$. The marginal MAP estimator of $\boldsymbol{R}$ is obtained as follows:

$$\widehat{\boldsymbol{R}}_{\mathrm{M-MAP}} = \arg\max_{\boldsymbol{R}} f(\boldsymbol{R}|\boldsymbol{Z}) \quad (15)$$

where the subscript M-MAP stands for marginal MAP. As an alternative, one can obtain the joint MAP estimate of both $\boldsymbol{\tau}$ and $\boldsymbol{R}$ by solving the following problem:

$$(\widehat{\boldsymbol{\tau}}_{\mathrm{J-MAP}}, \widehat{\boldsymbol{R}}_{\mathrm{J-MAP}}) = \arg\max_{\boldsymbol{\tau}, \boldsymbol{R}} f(\boldsymbol{\tau}, \boldsymbol{R}|\boldsymbol{Z}) \quad (16)$$

where the subscript J-MAP stands for joint MAP.

The M-MAP estimator of $\boldsymbol{R}$ can be obtained by setting to zero the derivative[3] of $\ln f(\boldsymbol{R}|\boldsymbol{Z})$. It can be shown that this is equivalent to solve the following matrix equation:

$$\frac{\partial \ln f(\boldsymbol{R}|\boldsymbol{Z})}{\partial \boldsymbol{R}} = -(\nu+K+N)\boldsymbol{R}^{-1} + (\nu-N)\boldsymbol{R}^{-1}\overline{\boldsymbol{R}}\boldsymbol{R}^{-1} + \sum_{k=1}^{K}(q_k+N)\frac{\boldsymbol{R}^{-1}\boldsymbol{z}_k\boldsymbol{z}_k^\dagger\boldsymbol{R}^{-1}}{\beta_k + \boldsymbol{z}_k^\dagger\boldsymbol{R}^{-1}\boldsymbol{z}_k} = 0. \quad (17)$$

It follows that the M-MAP estimator of $\boldsymbol{R}$ can be obtained solving

$$(\nu+K+N)\boldsymbol{R} = (\nu-N)\overline{\boldsymbol{R}} + \sum_{k=1}^{K}\frac{(q_k+N)\boldsymbol{z}_k\boldsymbol{z}_k^\dagger}{\beta_k + \boldsymbol{z}_k^\dagger\boldsymbol{R}^{-1}\boldsymbol{z}_k}. \quad (18)$$

On the other hand, the J-MAP estimates can be obtained through the following procedure: first, for a given $\boldsymbol{R}$, $f(\boldsymbol{\tau}, \boldsymbol{R}|\boldsymbol{Z})$ is maximized analytically, i.e., a closed-form expression for the value of $\boldsymbol{\tau}$ that maximizes $f(\boldsymbol{\tau}, \boldsymbol{R}|\boldsymbol{Z})$ is obtained. Plugging this value in $f(\boldsymbol{\tau}, \boldsymbol{R}|\boldsymbol{Z})$, one is left with a maximization problem with respect to $\boldsymbol{R}$ only, that can be still accomplished by setting to zero the derivative of the resulting function. To be analytical, observe that the distribution in (7) can be easily maximized separately with respect to each of the τ_k and the corresponding maximizer is given by

$$\widehat{\tau}_k = \frac{\beta_k + \boldsymbol{z}_k^\dagger\boldsymbol{R}^{-1}\boldsymbol{z}_k}{q_k+N+1}, \quad k = 1, \ldots, K. \quad (19)$$

Substituting the $\widehat{\tau}_k$'s into $f(\boldsymbol{\tau}, \boldsymbol{R}|\boldsymbol{Z})$ we get

$$f(\widehat{\boldsymbol{\tau}}, \boldsymbol{R}|\boldsymbol{Z}) \propto \frac{1}{\det(\boldsymbol{R})^{\nu+K+N}}\mathrm{etr}\left\{-(\nu-N)\boldsymbol{R}^{-1}\overline{\boldsymbol{R}}\right\} \times \prod_{k=1}^{K}\left[\frac{q_k+N+1}{\beta_k + \boldsymbol{z}_k^\dagger\boldsymbol{R}^{-1}\boldsymbol{z}_k}\right]^{q_k+N+1} e^{-(q_k+N+1)}. \quad (20)$$

Setting to zero the derivative of $\ln f(\widehat{\boldsymbol{\tau}}, \boldsymbol{R}|\boldsymbol{Z})$ with respect to $\boldsymbol{R}$ is equivalent to the following equation:

$$\frac{\partial \ln f(\widehat{\boldsymbol{\tau}}, \boldsymbol{R}|\boldsymbol{Z})}{\partial \boldsymbol{R}} = -(\nu+K+N)\boldsymbol{R}^{-1} + (\nu-N)\boldsymbol{R}^{-1}\overline{\boldsymbol{R}}\boldsymbol{R}^{-1} + \sum_{k=1}^{K}(q_k+N+1)\frac{\boldsymbol{R}^{-1}\boldsymbol{z}_k\boldsymbol{z}_k^\dagger\boldsymbol{R}^{-1}}{\beta_k + \boldsymbol{z}_k^\dagger\boldsymbol{R}^{-1}\boldsymbol{z}_k} = 0 \quad (21)$$

which, in turn, produces the following implicit equation in $\boldsymbol{R}$:

$$(\nu+K+N)\boldsymbol{R} = (\nu-N)\overline{\boldsymbol{R}} + \sum_{k=1}^{K}\frac{(q_k+N+1)\boldsymbol{z}_k\boldsymbol{z}_k^\dagger}{\beta_k + \boldsymbol{z}_k^\dagger\boldsymbol{R}^{-1}\boldsymbol{z}_k}. \quad (22)$$

Some remarks are now in order.

- It is instructive to note that (18) and (22) are very similar to that obtained via maximum likelihood estimation assuming that the τ_k's are unknown deterministic [9]. However, in the present scheme, we have introduced the *a priori* knowledge $\overline{\boldsymbol{R}}$ that counterbalances the influence of the snapshots. Indeed, the estimator is somehow a weighted combination of $\overline{\boldsymbol{R}}$ and the (properly compensated) sample covariance matrices of the snapshots. It can also be viewed as a kind of colored loading. Note also that introducing diagonal

[3]We make use of the following definition for the derivative of a real function $g(\boldsymbol{X})$ with respect to a complex variable:

$$\frac{\partial g(\boldsymbol{X})}{\partial x_{h,l}} = \frac{1}{2}\left[\frac{\partial g(\boldsymbol{X})}{\partial x_{h,l}^R} + j\frac{\partial g(\boldsymbol{X})}{\partial x_{h,l}^I}\right]$$

where $x_{h,l} = x_{h,l}^R + j\,x_{h,l}^I$ is the (h,l)th entry of the Hermitian matrix $\boldsymbol{X}$, $h < l$; see [26] for further details.

loading in the iterative scheme of [9] has been proposed in [27]. Herein, this loading technique emerges naturally as the MAP estimator in a Bayesian setting.

- In order to solve (18) or (22), one can advocate an iterative procedure. In practice, since we have that $\boldsymbol{R} = g(\boldsymbol{R}, \boldsymbol{Z})$, one starts with an initial value $\boldsymbol{R}^{(0)}$ and computes $\boldsymbol{R}^{(n+1)} = g(\boldsymbol{R}^{(n)}, \boldsymbol{Z})$ until a preassigned convergence criterion is met, see also [9]–[12] for discussions about convergence of such procedures.

IV. Detection

In this section, we propose ad hoc detection strategies for solving test (1). More in detail, we first assume that $\boldsymbol{R}$ is known and derive the GLRT (in the hypothesis that α is an unknown deterministic constant and that $\boldsymbol{v}$ is the known target signature). Subsequently, we make this detector fully adaptive replacing $\boldsymbol{R}$ with the estimates introduced in Section III. In addition, since the NMF is the asymptotic (with respect to N) approximation of the GLRT for known $\boldsymbol{R}$, we also consider adaptive implementations of the NMF.

Assuming that $\boldsymbol{R}$ is known the GLRT becomes

$$\frac{\max_{\alpha} \int f(\boldsymbol{z}|\alpha, \tau, \boldsymbol{R}) f(\tau) d\tau}{\int f(\boldsymbol{z}|0, \tau, \boldsymbol{R}) f(\tau) d\tau} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \tag{23}$$

where $f(\boldsymbol{z}|i\alpha, \tau, \boldsymbol{R})$ is the probability density function (pdf) of $\boldsymbol{z}|i\alpha, \tau, \boldsymbol{R}$ under the H_i hypothesis, $i = 0, 1$. Based upon the model in force for $\boldsymbol{z}$, α, τ, and $\boldsymbol{R}$ (see Section II), it is promptly verified that, for $i = 0, 1$,

$$f(\boldsymbol{z}|i\alpha, \tau, \boldsymbol{R}) = \frac{1}{\pi^N \tau^N \det(\boldsymbol{R})} \times \exp\left\{ -\frac{(\boldsymbol{z} - i\alpha\boldsymbol{v})^\dagger \boldsymbol{R}^{-1} (\boldsymbol{z} - i\alpha\boldsymbol{v})}{\tau} \right\}. \tag{24}$$

Ignoring irrelevant constants that will simplify computing the ratio (23), integration over τ can be easily accomplished and we obtain

$$\begin{aligned} f(\boldsymbol{z}|i\alpha, \boldsymbol{R}) &\propto \int \frac{1}{\tau^{q+N+1}} \\ &\quad \times \exp\left\{ -\frac{\beta + (\boldsymbol{z} - i\alpha\boldsymbol{v})^\dagger \boldsymbol{R}^{-1} (\boldsymbol{z} - i\alpha\boldsymbol{v})}{\tau} \right\} d\tau \\ &\propto \left[\beta + (\boldsymbol{z} - i\alpha\boldsymbol{v})^\dagger \boldsymbol{R}^{-1} (\boldsymbol{z} - i\alpha\boldsymbol{v}) \right]^{-(q+N)}. \end{aligned} \tag{25}$$

The GLRT at this intermediate stage can thus be written as

$$\frac{\beta + \boldsymbol{z}^\dagger \boldsymbol{R}^{-1} \boldsymbol{z}}{\beta + \min_\alpha \left\{ (\boldsymbol{z} - \alpha\boldsymbol{v})^\dagger \boldsymbol{R}^{-1} (\boldsymbol{z} - \alpha\boldsymbol{v}) \right\}} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma. \tag{26}$$

The optimization problem over α is well known and it provides a rule equivalent to

$$\frac{\left| \boldsymbol{z}^\dagger \boldsymbol{R}^{-1} \boldsymbol{v} \right|^2}{(\beta + \boldsymbol{z}^\dagger \boldsymbol{R}^{-1} \boldsymbol{z}) \boldsymbol{v}^\dagger \boldsymbol{R}^{-1} \boldsymbol{v}} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma. \tag{27}$$

Assuming that a set of training data $\boldsymbol{z}_k$, $k = 1, \ldots, K$, distributed according to the model described in Section II is available, we can implement adaptive detectors using both the NMF (2) and the GLRT (27) statistics, with $\boldsymbol{R}$ replaced by the estimates of Section III. We thus define the following detectors: the MMSE adaptive NMF (MMSE-ANMF) and the MAP adaptive NMF (MAP-ANMF) as

$$\frac{\left| \boldsymbol{v}^\dagger \widehat{\boldsymbol{R}}^{-1} \boldsymbol{z} \right|^2}{(\boldsymbol{v}^\dagger \widehat{\boldsymbol{R}}^{-1} \boldsymbol{v})(\boldsymbol{z}^\dagger \widehat{\boldsymbol{R}}^{-1} \boldsymbol{z})} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \tag{28}$$

and the MMSE adaptive GLRT (MMSE-GLRT) and the MAP adaptive GLRT (MAP-GLRT) as

$$\frac{\left| \boldsymbol{v}^\dagger \widehat{\boldsymbol{R}}^{-1} \boldsymbol{z} \right|^2}{(\beta + \boldsymbol{z}^\dagger \widehat{\boldsymbol{R}}^{-1} \boldsymbol{z})(\boldsymbol{v}^\dagger \widehat{\boldsymbol{R}}^{-1} \boldsymbol{v})} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \tag{29}$$

where $\widehat{\boldsymbol{R}}$ represents either the MAP or the MMSE estimators according to the acronyms. The above detectors incorporate a rough knowledge of $\boldsymbol{R}$, but would require knowledge of $\nu, q, q_1, \ldots, q_K, \beta, \beta_1, \ldots, \beta_K$. However, in realistic scenarios such quantities are not known. A viable approach is to assume that $E[\tau] = E[\tau_1] = \cdots = E[\tau_K] = 1$ and $q = q_1 = \cdots = q_K$; this, in turn, implies that $\beta = \beta_1 = \cdots = \beta_K = q - 1$. As a consequence, the variance of τ_k (τ) decreases to zero as q_k (q) increases to infinity.[4] We can thus use the two design parameters q and ν to tune the detectors according to the expected level of heterogeneities of the scenario.

V. Performance Assessment

In this section, we use standard Monte Carlo counting techniques to evaluate the performance of the proposed algorithms. Towards this end, we evaluate the probability of detection (P_d), i.e., the probability to decide for H_1 when it is actually in force, given P_{fa}. Data are generated according to the model described is Section II. We set $N = 8$, $q = q_1 = \ldots = q_K, \beta = \beta_1 = \ldots, \beta_K = q-1$, and consider two different scenarios $q = 4$ and $\nu = 10$ (weak *a priori* knowledge) and $q = 10$ and $\nu = 16$ (more precise *a priori* knowledge). As to $\overline{\boldsymbol{R}}$ we assume an exponentially correlated covariance matrix with one-lag correlation coefficient $\rho = 0.99$, i.e., the (i, j)th element of $\overline{\boldsymbol{R}}$ is given by $\rho^{|i-j|}$. According to [21], we also set $N_{bi} = 20$ and $N_r = 100$. Finally, we set to 3 the number of iterations used by the MAP estimators and we use the sample covariance matrix (SCM) $\mathrm{SCM} = (1/K) \sum_{k=1}^{K} \boldsymbol{z}_k \boldsymbol{z}_k^\dagger$ as the starting point for the iterations.

In Figs. 1–4 we plot the P_d versus the signal-to-noise ratio (SNR), defined as

$$\mathrm{SNR} = |\alpha|^2 E\left[\boldsymbol{v}^\dagger (\tau \boldsymbol{R})^{-1} \boldsymbol{v} \right] = |\alpha|^2 \frac{q}{q-1} \frac{\nu}{\nu - N} \boldsymbol{v}^\dagger \overline{\boldsymbol{R}}^{-1} \boldsymbol{v}. \tag{30}$$

The P_{fa} is set to 10^{-3} and corresponding thresholds are evaluated over $200/P_{fa}$ independent runs, while the P_d's are computed on 10^3 independent runs. Each figure contains eight curves: four for the NMF-based detectors (dashed lines) and four for the GLRT-based detectors (solid lines). More precisely, we report the performances of NMF, GLRT (i.e., the detectors that assume exact knowledge of $\boldsymbol{R}$), MMSE-

[4]Remember also that ν sets the distance between $\boldsymbol{R}$ and $\overline{\boldsymbol{R}}$.

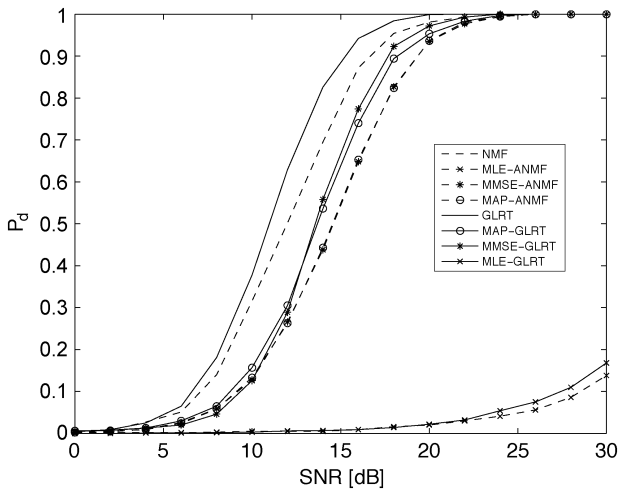

Fig. 1. P_d versus SNR, $N = 8$, $K = 8$, $q = 4$, $\nu = 10$, and $P_{fa} = 10^{-3}$.

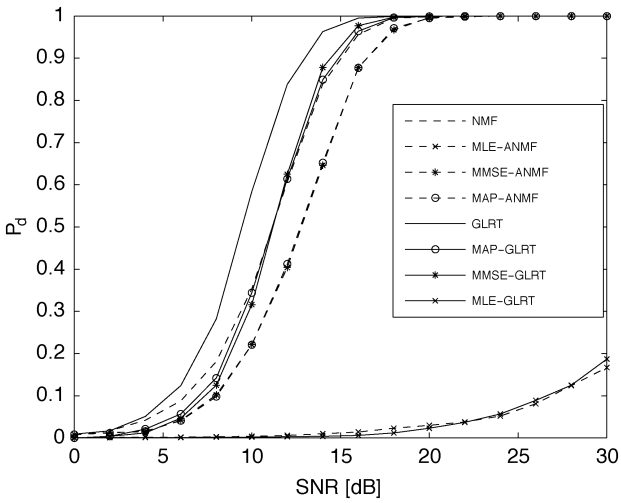

Fig. 2. P_d versus SNR, $N = 8$, $K = 8$, $q = 10$, $\nu = 16$, and $P_{fa} = 10^{-3}$.

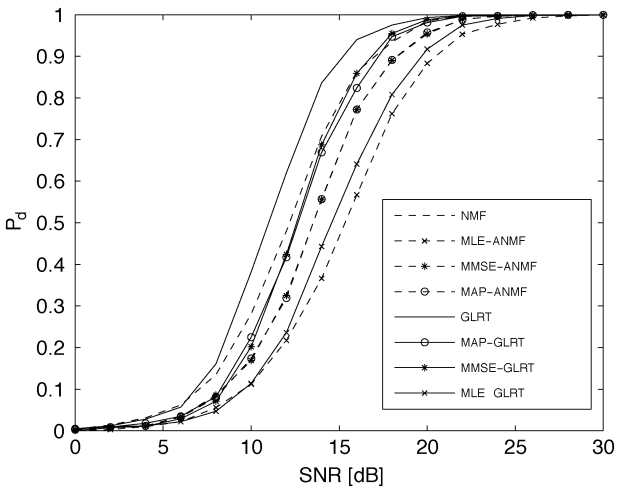

Fig. 3. P_d versus SNR, $N = 8$, $K = 16$, $q = 4$, $\nu = 10$, and $P_{fa} = 10^{-3}$.

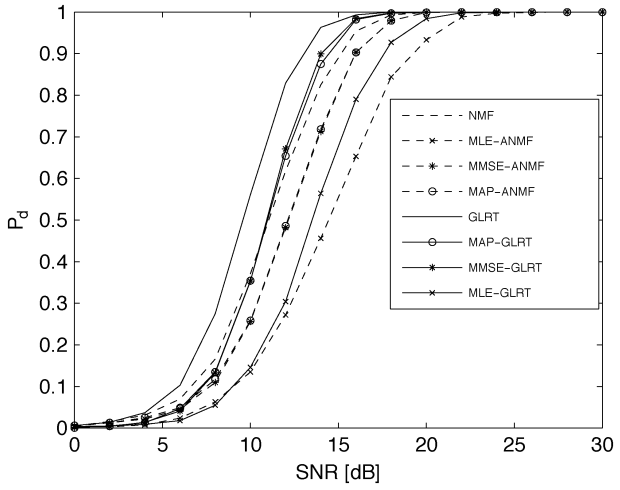

Fig. 4. P_d versus SNR, $N = 8$, $K = 16$, $q = 10$, $\nu = 16$, and $P_{fa} = 10^{-3}$.

ANMF, MMSE-GLRT, MAP-ANMF, and MAP-GLRT,[5] and, for comparison purposes, performances of MLE-ANMF and MLE-GLRT, i.e., adaptive implementations of NMF and GLRT using the recursive estimate proposed in [9] (initialized with the SCM and performing three iterations). Inspection of figures highlights the following.

- GLRT-based detectors perform better than NMF-based detectors, as it could be expected since the GLRT is optimized for the underlying model while the NMF is only asymptotically optimum.
- MMSE and MAP estimators ensure basically the same performance; this indicates that the MAP estimator should be preferred to the MMSE since the former is less computationally intensive.
- MLE recursive estimator is not a viable one when the sample support is small ($K = 8$, Figs. 1 and 2); for $K = 16$ (Figs. 3 and 4), instead, the loss of the MLE-based detectors is reduced.

We have also run other simulations (not reported here) in order to study the sensitivity of the proposed detectors with respect to a wrong choice of parameters ν and q. Results have shown that the relative hierarchy between the adaptive detectors is not affected by this kind of mismatch. Moreover, the detectors seem to be more sensitive to ν rather than to q. Summarizing, the performance assessment has shown how the *a priori* knowledge introduced in the proposed algorithms can help to outperform existing solutions. Obviously, the gain is much higher in the case of a small number of training data.

[5]Simulation studies not reported here show that M-MAP and J-MAP have the same performance in terms of P_d versus SNR; for this reason we only report that of the M-MAP.

VI. Conclusion

We have addressed the problem of adaptive detection of a signal of interest corrupted by correlated noise modeled in terms of a compound-Gaussian process. In order to embed into the design procedure some *a priori* knowledge about the noise, a Bayesian approach has been proposed. We have assumed that the covariance structure $\boldsymbol{R}$ is a random matrix drawn from an inverse complex Wishart distribution with known mean. As to the power levels, we have considered independent rv's distributed according to the inverse Gamma distribution with known parameters. We have proposed different algorithms for the estimation of $\boldsymbol{R}$; the estimators are based upon either the MMSE criterion or the MAP criterion. In order to obtain fully adaptive detectors, such estimators have been subsequently used inside an optimized GLRT and the NMF. The performance assessment has highlighted the effectiveness of the proposed algorithms also in comparison to existing ones. Remarkably, due to embedded *a priori* knowledge about the environment, it turns out that the proposed solutions are particularly suitable in case of small estimation samples.

REFERENCES

[1] F. Gini, A. Farina, and M. Greco, "Selected list of references on radar signal processing," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 37, no. 1, pp. 329–359, Jan. 2001.

[2] K. D. Ward, C. J. Baker, and S. Watts, "Maritime surveillance radar. Part 1: Radar scattering from the ocean surface," *Proc. Inst. Electr. Eng. F*, vol. 137, no. 2, pp. 51–62, 1990.

[3] M. Di Bisceglie and C. Galdi, "Random walk based characterization of radar backscatter from the sea surface," *Proc. Inst. Elect. Eng.—Radar, Sonar Navig.*, vol. 145, no. 4, pp. 216–225, 1998.

[4] E. Conte and M. Longo, "Characterisation of radar clutter as a spherically invariant process," *Proc. Inst. Elect. Eng. F*, vol. 134, no. 2, pp. 191–197, Apr. 1987.

[5] E. Conte, M. Longo, M. Lops, and S. L. Ullo, "Radar detection of signals with unknown parameters in K-distributed clutter," *Proc. Inst. Elect. Eng. F*, vol. 138, no. 2, pp. 131–138, 1991.

[6] 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.

[7] E. Conte, M. Lops, and G. Ricci, "Adaptive matched filter detection in spherically invariant noise," *IEEE Signal Process. Lett.*, vol. 3, no. 8, pp. 248–250, Aug. 1996.

[8] E. Conte, M. Lops, and G. Ricci, "Adaptive detection schemes in compound-Gaussian clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 34, no. 4, pp. 1058–1069, Oct. 1998.

[9] E. Conte, A. De Maio, and G. Ricci, "Recursive estimation of the covariance matrix of a compound Gaussian process and its application to adaptive CFAR detection," *IEEE Trans. Signal Process.*, vol. 50, no. 8, pp. 1908–1915, Aug. 2002.

[10] M. S. Greco and F. Gini, "Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter," *Signal Process.*, vol. 82, no. 12, pp. 1847–1859, Dec. 2002.

[11] E. Conte and A. De Maio, "Mitigation techniques for non-Gaussian sea clutter," *IEEE J. Ocean. Eng.*, vol. 29, no. 2, pp. 284–302, Apr. 2004.

[12] F. Pascal, Y. Chitour, J.-P. Ovarlez, P. Forster, and P. Larzabal, "Covariance structure maximum-likelihood estimates in compound Gaussian noise: Existence and algorithm analysis," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 34–48, Jan. 2008.

[13] W. L. Melvin, "Space-time adaptive radar performance in heterogeneous clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 2, pp. 621–633, Apr. 2000.

[14] M. Rangaswamy, "Statistical analysis of the nonhomogeneity detector for non-Gaussian interference backgrounds," *IEEE Trans. Signal Process.*, vol. 53, no. 6, pp. 2101–2111, Jun. 2005.

[15] *IEEE Signal Process. Mag. (Special Issue on Knowledge-Based Systems for Adaptive Radar: Detection, Tracking and Classification)*, vol. 23, no. 1, Jan. 2006.

[16] A. De Maio, A. Farina, and G. Foglia, "Design and experimental validation of knowledge-based constant false alarm rate detectors," *Proc. Inst. Electr. Eng.—Radar, Sonar, Navig.*, vol. 1, no. 4, pp. 308–316, Aug. 2007.

[17] A. De Maio, S. De Nicola, L. Landi, and A. Farina, "Knowledge-aided covariance matrix estimation: A MAXDET approach," *Proc. Inst. Elect. Eng.—Radar, Sonar Navig.*, vol. 3, no. 4, pp. 341–356, Aug. 2009.

[18] A. De Maio and A. Farina, "Adaptive radar detection: A Bayesian approach," in *Proc Int. Radar Symp. (IRS) 2006*, May 24–26, 2006, pp. 1–4.

[19] A. De Maio, A. Farina, and G. Foglia, "Adaptive radar detection: A Bayesian approach," in *Proc. IEEE Radar Conf. 2007*, Apr. 17–20, 2007, pp. 624–629.

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

[21] 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.

[22] O. Besson, S. Bidon, and J.-Y. Tourneret, "Covariance matrix estimation with heterogeneous samples," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 909–920, Mar. 2008.

[23] J. Ward, "Space-time adaptive processing for airborne radar," Lincoln Lab., Massachusetts Inst. of Technology, Lexington, MA, Tech. Rep. 1015, Dec. 1994.

[24] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, 2nd ed. New York: Chapman & Hall, 2004.

[25] L. Svensson, "Bayesian inference with unknown noise covariance," Ph.D. dissertation, Chalmers Univ. Technology, Göteborg, Nov. 2004.

[26] S. M. Kay, *Fundamentals of Statistical Signal Processing. Vol. 1: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[27] Y. I. Abramovich and N. K. Spencer, "Diagonally loaded normalised sample matrix inversion (LNSMI) for outlier-resistant adaptive filtering," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Honolulu, HI, Apr. 2007, pp. 1105–1108.

On the Stable Recovery of the Sparsest Overcomplete Representations in Presence of Noise

Massoud Babaie-Zadeh and Christian Jutten, *Fellow, IEEE*

***Abstract*—Let $\mathbf{x}$ be a signal to be sparsely decomposed over a redundant dictionary $\mathbf{A}$, i.e., a sparse coefficient vector $\mathbf{s}$ has to be found such that $\mathbf{x} = \mathbf{A}\mathbf{s}$. It is known that this problem is inherently unstable against noise, and to overcome this instability, Donoho, Elad and Temlyakov ["Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006] have proposed to use an "approximate" decomposition, that is, a decomposition satisfying $\|\mathbf{x} - \mathbf{A}\mathbf{s}\|_2 \leq \delta$ rather than satisfying the exact equality $\mathbf{x} = \mathbf{A}\mathbf{s}$. Then, they have shown that if there is a decomposition with $\|\mathbf{s}\|_0 < (1 + M^{-1})/2$, where M denotes the *coherence* of the dictionary, this decomposition would be stable against noise. On the other hand, it is known that a sparse decomposition with $\|\mathbf{s}\|_0 < (1/2)\mathrm{spark}(\mathbf{A})$ is unique. In other words, although a decomposition with $\|\mathbf{s}\|_0 < (1/2)\mathrm{spark}(\mathbf{A})$ is unique, its stability against noise has been proved only for highly more restrictive decompositions satisfying $\|\mathbf{s}\|_0 < (1 + M^{-1})/2$, because usually $(1 + M^{-1})/2 \ll (1/2)\mathrm{spark}(\mathbf{A})$. This limitation maybe had not been very important before, because $\|\mathbf{s}\|_0 < (1 + M^{-1})/2$ is also the bound which guaranties that the sparse decomposition can be found via minimizing the ℓ^1 norm, a classic approach for sparse decomposition. However, with the availability of new algorithms for sparse decomposition, namely SL0 and robust-SL0, it would be important to know whether or not unique sparse decompositions with $(1 + M^{-1})/2 \leq \|\mathbf{s}\|_0 < (1/2)\mathrm{spark}(\mathbf{A})$ are stable. In this correspondence, we show that such decompositions are indeed stable. In other words, we extend the stability bound from $\|\mathbf{s}\|_0 < (1 + M^{-1})/2$ to the whole uniqueness range $\|\mathbf{s}\|_0 < (1/2)\mathrm{spark}(\mathbf{A})$. In summary, we show that *all unique sparse decompositions are stably recoverable*. Moreover, we see that sparser decompositions are "more stable."**

***Index Terms*—Compressed sensing, overcomplete dictionaries, sparse component analysis (SCA), sparse recovery, sparse signal decomposition.**

I. INTRODUCTION

Let $\mathbf{A}$ be an $n \times m$ matrix with $m > n$, and consider the underdetermined system of linear equations (USLE) $\mathbf{A}\mathbf{s} = \mathbf{x}$. Such a linear system has typically infinitely many solutions, but let consider

Manuscript received July 12, 2009; accepted May 04, 2010. Date of publication June 10, 2010; date of current version September 15, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Kostantinos I. Diamantaras. This work has been partially funded by the Iran National Science Foundation (INSF), by the Iran Telecom Research Center (ITRC), and also by the Center for International Research and Collaboration (ISMO) and the French Embassy in Tehran, Iran, in the framework of a GundiShapour collaboration program.

M. Babaie-Zadeh is with the Electrical Engineering Department, Sharif University of Technology, Tehran, Iran (e-mail: mbzadeh@yahoo.com).

C. Jutten is with the GIPSA-Lab, Grenoble, France, and the Institut Universitaire de France, France (e-mail: christian.jutten@gipsa-lab.grenoble-inp.fr).

Digital Object Identifier 10.1109/TSP.2010.2052357