**To link to this article**: DOI: 10.1016/j.sigpro.2016.05.027
URL: http://dx.doi.org/10.1016/j.sigpro.2016.05.027

**To cite this version**: Besson, Olivier and Abramovich, Yuri and Johnson, Ben Direction of arrival estimation in a mixture of K-distributed and Gaussian noise. (2016) Signal Processing, vol. 128. pp. 512-520. ISSN 0165-1684

# Direction-of-arrival estimation in a mixture of $K$-distributed and Gaussian noise ☆

Olivier Besson [a,*], Yuri Abramovich [b], Ben Johnson [c]

[a] ISAE-Supaéro, Department Electronics Optronics Signal, 10 Avenue Edouard Belin, 31055 Toulouse, France
[b] W R Systems, Ltd., 11351 Random Hills Road, Suite 400, Fairfax, VA 22030, USA
[c] University of South Australia - ITR, Mawson Lakes SA 5085, Australia

A B S T R A C T

We address the problem of estimating the directions-of-arrival (DoAs) of multiple signals received in the presence of a combination of a strong compound-Gaussian external noise and weak internal white Gaussian noise. Since the exact distribution of the mixture is not known, we get an insight into optimum procedure via a related model where we consider the texture of the compound-Gaussian component as an unknown and deterministic quantity to be estimated together with DoAs or a basis of the signal subspace. Alternate maximization of the likelihood function is conducted and it is shown that it operates a separation between the snapshots with small/large texture values with respect to the additive noise power. The modified Cramér–Rao bound is derived and a prediction of the actual mean-square error is presented, based on separation between external/internal-noise dominated samples. Numerical simulations indicate that the suggested iterative DoA estimation technique comes close to the introduced bound and outperform a number of existing routines.

*Keywords:*
Direction-of-arrival estimation
Noise mixture
Maximum likelihood
Lower bounds

## 1. Problem statement

Localization of sources of interest using an array of spatially distributed sensors is a primordial task in many applications, including sonar and radar [1]. Historically, this problem has been tackled under the assumption of additive white Gaussian noise while the waveforms of interest were considered either drawn from a Gaussian distribution (unconditional model) or deterministic unknowns (conditional model). Whatever the case, the reference approach is the maximum likelihood (ML) estimator, which is known to possess optimal asymptotic (either in the number of snapshots or signal to noise ratio depending on the underlying assumptions) properties [2–5], including achieving the Cramér–Rao bound (CRB). The main drawback of the ML approach lies in its computational complexity as one needs to carry out a $P$-dimensional search, where $P$ stands for the number of signals impinging on the array.

In an attempt to decrease computational cost, the low-rank structure of the array output in the absence of noise was strongly employed in the so-called subspace methods, such as MUSIC [6], MODE [7] or ESPRIT [8]. Most of these methods are based on the eigenvalue decomposition of the sample covariance matrix and a partitioning of the whole space into a signal subspace (spanned by the $P$ eigenvectors associated to the $P$ largest eigenvalues) and its orthogonal complement, often referred to as the noise subspace. As for MUSIC, the directions-of-arrival are then estimated by minimizing the norm of the projection of the array steering vector onto the noise subspace. Hence, the method requires finding $P$ maxima of a scalar function, which is much less computationally involved than the ML estimator. Despite this significant simplification, in [3,4], it was proven that under a number of non-restrictive conditions on independent and identically distributed (i.i.d.) Gaussian samples, MUSIC DoA estimates are asymptotically efficient. The latter means that, for a fixed array size $M$ and $T \to \infty$ i.i.d. training samples, MUSIC DoA estimation accuracy tends to the CRB. Yet, in [9] it was demonstrated that while MUSIC produces consistent and asymptotically efficient DoA estimates in the classical asymptotic assumption where $M$ is held constant and $T \to \infty$, it is not even consistent under the so-called Kolmogorov's assumption whereby $M \to \infty, T \to \infty, M/T \to c$, with $c$ a constant. Most pronounced manifestation of this inconsistency is MUSIC behavior in the so-called threshold area where, due to small sample support $T$ and/or signal to noise ratio (SNR), the mean-square error of MUSIC (and various other subspace-based

☆ The work of O. Besson is supported by the Délégation Générale de l'Armement (DGA), Mission pour la Recherche et l'Innovation Scientifique (MRIS) under grant no. 2015.60.0090.00.470.75.01.

* Corresponding author.
*E-mail addresses:* olivier.besson@isae-supaero.fr (O. Besson), yabramovich@wrsystems.com (Y. Abramovich), ben.a.johnson@lmco.com (B. Johnson).

estimators) begins to depart significantly from the CRB. This departure is caused by the onset of severely erroneous DoA estimates (outliers) which occur under threshold conditions ($T$,SNR) where the genuine ML estimator still provides CRB-consistent estimation accuracy [10]. Considerable attention has focused on trying to explain this phenomenon [11,12,10,13] as well as to counteract it by improving subspace-based estimators threshold behavior. In [9,14,15], an important advance was provided by Mestre with the derivation of G-MUSIC, a variation of MUSIC based on random matrix theory which provides consistent in Kolmogorov's sense estimate of the pseudo-spectrum function. In [10] it was shown that G-MUSIC somewhat reduces the gap between MUSIC-specific and ML-intrinsic threshold conditions, without however completely closing this gap.

The above references mostly deal with an additive noise which is assumed to be Gaussian distributed, and most often with known (up to a scaling factor) covariance matrix, considered to be the identity matrix in most studies. Non-Gaussian noise case has been considered e.g., in [16–18]. For particular noise models, the traditional asymptotic properties of ML estimation and specific estimation routines have been considered in these studies. Yet, in the recent paper [19] we demonstrated that for a specific non-Gaussian noise, namely a $K$-distributed noise, the traditional CRB may not exist. More precisely, when the noise vector is modeled as the product of a Gaussian vector by a positive random variable (the texture) distributed according to a Gamma distribution with shape parameter $\nu$ and scale parameter $\beta$, then the CRB is non-existent for $\nu < 1$. Moreover, the ML estimator exhibits a behavior significantly different from the Gaussian case, since its convergence rate was shown to be $T^{-1/\nu}$ for $\nu < 1$. Interestingly, DoA estimation in compound-Gaussian and more generally elliptically distributed noise models [20] was recently addressed. A first improvement, referred to as R-MUSIC, consists in using eigenvalue decomposition of Maronna-type (robust) covariance matrix estimates and subsequently applying conventional MUSIC [21,20,22]. In the same way as G-MUSIC was obtained from MUSIC, random matrix theory methodology is applied to Maronna covariance matrix estimates in Refs. [23,24] in order to get a consistent (in Kolmogorov's sense) robust MUSIC pseudo-spectrum estimate, yielding the RG-MUSIC estimator. Some improvement compared to R-MUSIC is highlighted in [23,24], yet the actual merits of RG-MUSIC is hard to evaluate without the ML benchmark known.

Furthermore, while ML DoA estimation properties in non-regular case observed in the presence of $K$-distributed noise is of theoretical interest, practical scenarios are more complicated. Indeed, in radar applications particularly, one has to take into account two types of noise, namely external noise and internal (thermal) noise. The two of them are generated by different physical mechanisms and hence have different distributions. While the white Gaussian assumption is natural for thermal noise, external noise is most often non-Gaussian. For instance, in HF DF systems, external noise is dominated by atmospheric activity (thunderstorms) and is white non-Gaussian, being typically 30 dB above the internal noise [25]. When external noise corresponds to clutter, a compound-Gaussian model or an elliptical distribution is deemed relevant amongst the radar community [26,27]. While inter-scan observation justifies independence of texture values, the speckle component is usually assumed to be correlated, even though pulse-to-pulse frequency agility can lead to independent speckle snapshots. Usually, the external to thermal noise ratio is high and hence thermal noise is generally neglected, on the basis that external noise is the main source of disturbance. One of the outcomes of [19] was that, with heavy-tailed distributions (small $\nu$), the DoA estimation performance is mostly dictated by the snapshot corresponding to the minimal texture value. Indeed, if one sets the power of the texture to 1 ($\beta = \nu^{-1}$), the average value of the minimal over $T$=8 texture values is $-25$ dB for $\nu$=0.2 and $-38$ dB for $\nu$=0.1. Moreover, we demonstrated in [19] that this specific sample with minimal texture sets the limit to DoA estimation accuracy. As a corollary, even if the total power of $K$-distributed (external) noise significantly exceeds the power of internal white Gaussian noise, under large enough support of $T$ training samples, a certain number of them may have texture values well below the internal noise power. Consequently, thermal noise cannot be neglected and should be taken into account. Finally, the above findings suggest that all snapshots are not equal, since some of them will correspond to very small values of the texture while others will be buried in strong noise (large texture values). Therefore, not all of them should be treated equally in an estimation procedure. This is the problem we tackle in the present paper: DoA estimation in a mixture of $K$-distributed noise and Gaussian noise.

First, we try to get an insight into the nature of ML DoA estimation. Since the closed-form multivariate probability density function (p.d.f.) for the noise mixture does not exist, we introduce a related model where we treat Gamma distributed texture values as unknown deterministic parameters to be estimated with the DoAs. This approach provides an insight into the nature of close to ML optimum processing and leads to a particular iterative DoA estimation routine.

## 2. Data model and DoA estimation

We begin this section with the assumed data model. We consider a uniform linear array (ULA) of $M$ elements spaced a half wavelength apart and the received signal at time $t = 1, \ldots, T$ can be written as

$$\boldsymbol{x}_t = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{s}_t + \sqrt{\tau_t}\boldsymbol{n}_t + \sigma_w \boldsymbol{w}_t \tag{1}$$

where

- $\boldsymbol{\theta} = \left[\theta_1 \cdots \theta_P\right]^T$ is the vector of the DoA and $\boldsymbol{A}(\boldsymbol{\theta}) = \left[\boldsymbol{a}(\theta_1) \cdots \boldsymbol{a}(\theta_P)\right]$ is the array manifold matrix: for the considered ULA $\boldsymbol{a}(\theta) = \left[1 \; e^{i\pi \sin\theta} \ldots \; e^{i(M-1)\pi \sin\theta}\right]^T$. Note that the specific properties of the ULA manifold are not exploited in any of the methods derived below, only knowledge of the form of the steering vector is required, and hence the methods can be extended to arbitrary geometry.
- $\boldsymbol{s}_t$ stands for the emitted waveforms and we choose to consider them as deterministic unknowns (conditional model).
- $\boldsymbol{n}_t$ and $\boldsymbol{w}_t$ are i.i.d. vectors drawn from $\boldsymbol{n}_t \sim \text{CN}(\boldsymbol{0}, \boldsymbol{I})$ and $\boldsymbol{w}_t \sim \text{CN}(\boldsymbol{0}, \boldsymbol{I})$. $\sigma_w^2$ stands for the thermal noise power and is assumed to be known here. We are mostly interested in situations where $\sigma_w^2 \ll 1$.
- $\tau_t$ is a positive random variable such that $\text{E}\{\tau_t\} = 1$, so that $\sigma_w^{-2}$ represents the ratio of the external compound-Gaussian noise power to the internal white Gaussian noise power. We will refer to it as nGGNR.

The immediate problem one is faced with when considering (1) is that, if one assumes a given prior distribution (for instance a Gamma distribution) for $\tau_t$, then the distribution of $\boldsymbol{x}_t$ cannot be obtained analytically. In such a case, the ML algorithm cannot be derived, nor the associated CRB. Consequently, any practical DoA estimation technique, such as R-MUSIC or RG-MUSIC cannot be evaluated neither by comparison to a bound nor with the results of direct exhaustive $P$-dimensional search for the global maximum of the likelihood function. In this paper, we choose the following approach to address this problem. First, we consider a "clairvoyant" model where the random textures $\tau_t$ are considered to be known a priori. For this non-homogeneous but still purely Gaussian model, the LF function can be derived and, subsequently, the ML estimator and the corresponding CRB. Clearly, if any practical DoA estimation technique approaches this clairvoyant estimator, the proximity of such a technique to the ML-optimal performance would be established. Secondly, and similarly to what has been done many times, we consider $\tau_t$ as deterministic unknowns to be estimated.

### 2.1. Clairvoyant estimation (known $\boldsymbol{\tau}$)

As said previously, we first consider that the textures $\tau_t$ are known, which will provide us with a benchmark to which the adaptive estimation scheme of the next section can be compared. Under the stated assumptions, the p.d.f. of the data matrix $\boldsymbol{X} = \left[\boldsymbol{x}_1 \cdots \boldsymbol{x}_T\right]$ is given by

$$p(\boldsymbol{X}|\boldsymbol{A}(\boldsymbol{\theta}), \boldsymbol{\tau}, \boldsymbol{S}) \propto \prod_{t=1}^{T} \left(\tau_t + \sigma_w^2\right)^{-M} \exp\left\{-\left(\tau_t + \sigma_w^2\right)^{-1}\left[\boldsymbol{x}_t - \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{s}_t\right]^H\left[\boldsymbol{x}_t - \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{s}_t\right]\right\} \tag{2}$$

where $\propto$ means proportional to, $\boldsymbol{\tau} = \left[\tau_1 \cdots \tau_T\right]^T$ and $\boldsymbol{S} = \left[\boldsymbol{s}_1 \cdots \boldsymbol{s}_T\right]$. Using the fact that

$$\left[\boldsymbol{x}_t - \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{s}_t\right]^H\left[\boldsymbol{x}_t - \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{s}_t\right] = \boldsymbol{x}_t^H\boldsymbol{x}_t - \boldsymbol{x}_t^H\boldsymbol{A}(\boldsymbol{\theta})\left(\boldsymbol{A}(\boldsymbol{\theta})^H\boldsymbol{A}(\boldsymbol{\theta})\right)^{-1}\boldsymbol{A}^H\boldsymbol{x}_t + \left[\boldsymbol{s}_t - \left(\boldsymbol{A}(\boldsymbol{\theta})^H\boldsymbol{A}(\boldsymbol{\theta})\right)^{-1}\boldsymbol{A}(\boldsymbol{\theta})^H\boldsymbol{x}_t\right]^H\left(\boldsymbol{A}(\boldsymbol{\theta})^H\boldsymbol{A}\right)\left[\boldsymbol{s}_t - \left(\boldsymbol{A}(\boldsymbol{\theta})^H\boldsymbol{A}(\boldsymbol{\theta})\right)^{-1}\boldsymbol{A}(\boldsymbol{\theta})^H\boldsymbol{x}_t\right] \tag{3}$$

it follows that the maximum of (2) with respect to (w.r.t.) $\boldsymbol{s}_t$ is obtained when

$$\boldsymbol{s}_t = \left(\boldsymbol{A}(\boldsymbol{\theta})^H\boldsymbol{A}(\boldsymbol{\theta})\right)^{-1}\boldsymbol{A}(\boldsymbol{\theta})^H\boldsymbol{x}_t \tag{4}$$

and hence

$$\max_{\boldsymbol{S}} p(\boldsymbol{X}|\boldsymbol{A}(\boldsymbol{\theta}), \boldsymbol{\tau}, \boldsymbol{S}) \propto \prod_{t=1}^{T} \left(\tau_t + \sigma_w^2\right)^{-M} \exp\left\{-\left(\tau_t + \sigma_w^2\right)^{-1}\boldsymbol{x}_t^H\boldsymbol{P}_{\boldsymbol{A}(\boldsymbol{\theta})}^{\perp}\boldsymbol{x}_t\right\} \tag{5}$$

where $\boldsymbol{P}_{\boldsymbol{A}(\boldsymbol{\theta})}^{\perp} = \boldsymbol{I} - \boldsymbol{A}(\boldsymbol{\theta})\left(\boldsymbol{A}(\boldsymbol{\theta})^H\boldsymbol{A}(\boldsymbol{\theta})\right)^{-1}\boldsymbol{A}(\boldsymbol{\theta})^H$.

The clairvoyant ML estimator of $\boldsymbol{\theta}$ is thus obtained as

$$\boldsymbol{\theta}_{\mathrm{ML}|\boldsymbol{\tau}} = \arg\min_{\boldsymbol{\theta}} \sum_{t=1}^{T} \frac{\boldsymbol{x}_t^H\boldsymbol{P}_{\boldsymbol{A}(\boldsymbol{\theta})}^{\perp}\boldsymbol{x}_t}{\tau_t + \sigma_w^2}. \tag{6}$$

The previous estimator, which uses knowledge of $\tau_t$, will constitute a reference to which any adaptive estimator can be compared. However, implementing (6) requires solving a $P$-dimensional optimization problem. In order to simplify it, let us proceed in two steps: first estimate the unstructured steering matrix $\boldsymbol{A}$ for a known $\boldsymbol{\tau}$, then use a MUSIC-like procedure for DoA estimation. Indeed, from (5) and ignoring the dependence of the steering matrix towards $\boldsymbol{\theta}$, we have

$$\max_{\boldsymbol{S}} p(\boldsymbol{X}|\boldsymbol{A}, \boldsymbol{\tau}, \boldsymbol{S}) \propto \prod_{t=1}^{T} \left(\tau_t + \sigma_w^2\right)^{-M} \exp\left\{-\left(\tau_t + \sigma_w^2\right)^{-1}\boldsymbol{x}_t^H\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{x}_t\right\} \tag{7}$$

from which it readily follows that

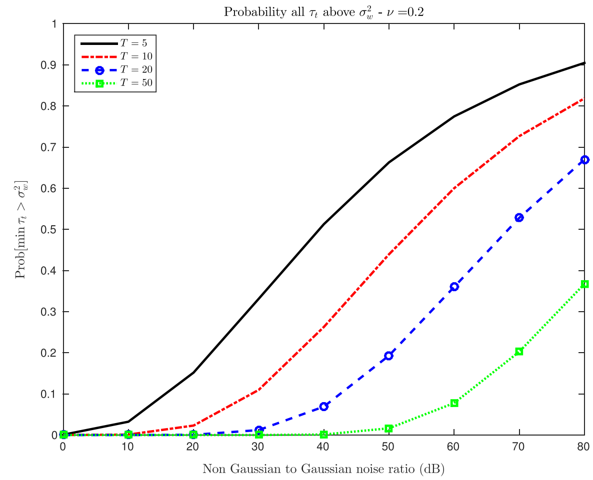



**Fig. 1.** $P(\sigma_w^2, T)$ for various values of nGGNR and $T$. $\nu$=0.2 and $\beta = 1/\nu$.

$$\begin{aligned}\boldsymbol{A}_{\mathrm{ML}|\boldsymbol{\tau}} &= \arg\max_{\boldsymbol{A}}\left[\max_{\boldsymbol{S}} p(\boldsymbol{X}|\boldsymbol{A}, \boldsymbol{\tau}, \boldsymbol{S})\right] \\ &= \arg\max_{\boldsymbol{A}} \sum_{t=1}^{T} \frac{\boldsymbol{x}_t^H\boldsymbol{P}_{\boldsymbol{A}}\boldsymbol{x}_t}{\tau_t + \sigma_w^2} \\ &= \mathcal{P}_P\left(\sum_{t=1}^{T} \frac{\boldsymbol{x}_t\boldsymbol{x}_t^H}{\tau_t + \sigma_w^2}\right)\end{aligned} \tag{8}$$

where $\mathcal{P}_P(.)$ stands for the $P$ principal subspace of the matrix between parentheses. From $\boldsymbol{A}_{\mathrm{ML}|\boldsymbol{\tau}}$, the DoA can be estimated e.g., as

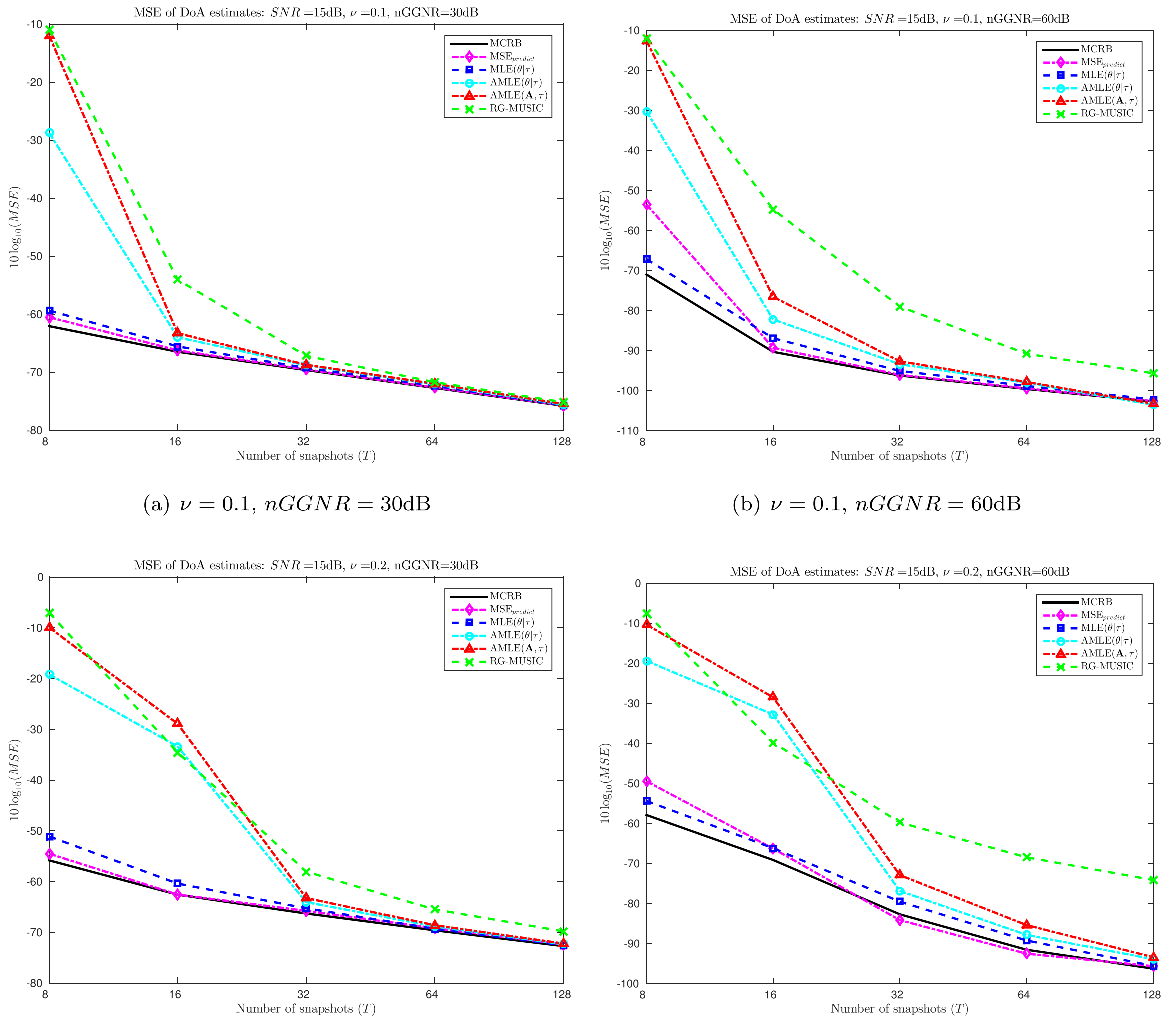
$$\boldsymbol{\theta}_{\mathrm{AML}|\boldsymbol{\tau}} = \mathrm{MUSIC}\left[\boldsymbol{A}_{\mathrm{ML}|\boldsymbol{\tau}}\right] \tag{9}$$

where MUSIC$[.]$ stands for the conventional MUSIC algorithm (either in spectral or root form) applied to a $M \times P$ matrix whose columns form a basis for the signal subspace. The comparison between $\boldsymbol{\theta}_{\mathrm{ML}|\boldsymbol{\tau}}$ and $\boldsymbol{\theta}_{\mathrm{AML}|\boldsymbol{\tau}}$ will provide insight into the loss associated with replacing the search for the global maximum of the $P$-dimensional likelihood function by a MUSIC-like procedure.

Before closing this section, it is noteworthy that the above clairvoyant approach calls for a lower bound for DoA estimation. The true p.d.f. of the observations can in principle be obtained by integrating (2) with respect to the p.d.f. of $\boldsymbol{\tau}$, but it does not seem feasible to obtain a closed-form expression so that the true Cramér–Rao bound for the problem at hand appears to be intractable. In this case, since we have a mixture of deterministic ($\boldsymbol{\theta}, \boldsymbol{S}$) and random ($\tau$) parameters, it is customary to resort to hybrid bounds, such as the Miller–Chang bound [28] or the modified CRB [29]. Let $CRB_G(T, \sigma^2\boldsymbol{I})$ denote the conditional CRB for DoA estimation obtained with $T$ samples and the white Gaussian noise with covariance matrix $\sigma^2\boldsymbol{I}$: note that, for the sake of simplicity, we omit the dependence of this bound to the values of the DoA and signal waveforms. This bound is obtained as the inverse of the Fisher information matrix (FIM), which is given by [1]

$$FIM_G(T, \sigma^2\boldsymbol{I}) = \frac{2T}{\sigma^2}\,\mathrm{Re}\left\{\boldsymbol{H} \odot \hat{\boldsymbol{R}}_s^T\right\} \tag{10}$$

where $\hat{\boldsymbol{R}}_s$ is the sample covariance matrix of the signal waveforms and $\boldsymbol{H}$ depends on the array manifold. For any estimate $\hat{\boldsymbol{\theta}}$ based on $\boldsymbol{X}$ only (i.e., without the knowledge of $\tau_t$), one can write




(a) $\nu = 0.1$, $nGGNR = 30$dB

(b) $\nu = 0.1$, $nGGNR = 60$dB

(c) $\nu = 0.2$, $nGGNR = 30$dB

(d) $\nu = 0.2$, $nGGNR = 60$dB

**Fig. 2.** Mean square error of estimators versus $T$. $SNR$ = 15 dB.

$$\mathrm{E}\left\{\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)^T\right\} = \int\left[\int\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)^T p(\boldsymbol{X}|\boldsymbol{\tau})d\boldsymbol{X}\right]p(\boldsymbol{\tau})d\boldsymbol{\tau}$$

$$\geq \int\frac{1}{2}\left[\mathrm{Re}\{\boldsymbol{H}\odot\hat{\boldsymbol{R}}_s^T\}\right]^{-1}\left(\sum_{t=1}^{T}\frac{1}{\tau_t+\sigma_w^2}\right)^{-1}p(\boldsymbol{\tau})d\boldsymbol{\tau}$$

$$= CRB_G(1,\boldsymbol{I})\int\left(\sum_{t=1}^{T}\frac{1}{\tau_t+\sigma_w^2}\right)^{-1}p(\boldsymbol{\tau})d\boldsymbol{\tau}. \tag{11}$$

Note that (11) is the averaged-over-$\boldsymbol{\tau}$ CRB conditioned on $\boldsymbol{\tau}$, i.e., $\mathrm{E}\{CRB(\boldsymbol{\theta}|\boldsymbol{\tau})\}$ and is referred to in the literature as the Miller–Chang bound (MCB) [28]. Obviously, it constitutes a lower bound for any adaptive estimator which does not have knowledge of $\boldsymbol{\tau}$. It appears that a closed-form expression for the integral in (11) cannot be obtained, mainly because one has to handle the inverse of a sum. However, by using Jensen's inequality, one has

$$\mathrm{E}\{CRB(\boldsymbol{\theta}|\boldsymbol{\tau})\} \geq CRB_G(1,\boldsymbol{I})\left[\int\left(\sum_{t=1}^{T}\frac{1}{\tau_t+\sigma_w^2}\right)p(\boldsymbol{\tau})d\boldsymbol{\tau}\right]^{-1} \tag{12}$$

$$= T^{-1}CRB_G(1,\boldsymbol{I})\left[\int_0^\infty \frac{\beta^{-\nu}}{\Gamma(\nu)}\frac{\tau^{\nu-1}}{\tau+\sigma_w^2}e^{-\beta^{-1}\tau}d\tau\right]^{-1}$$

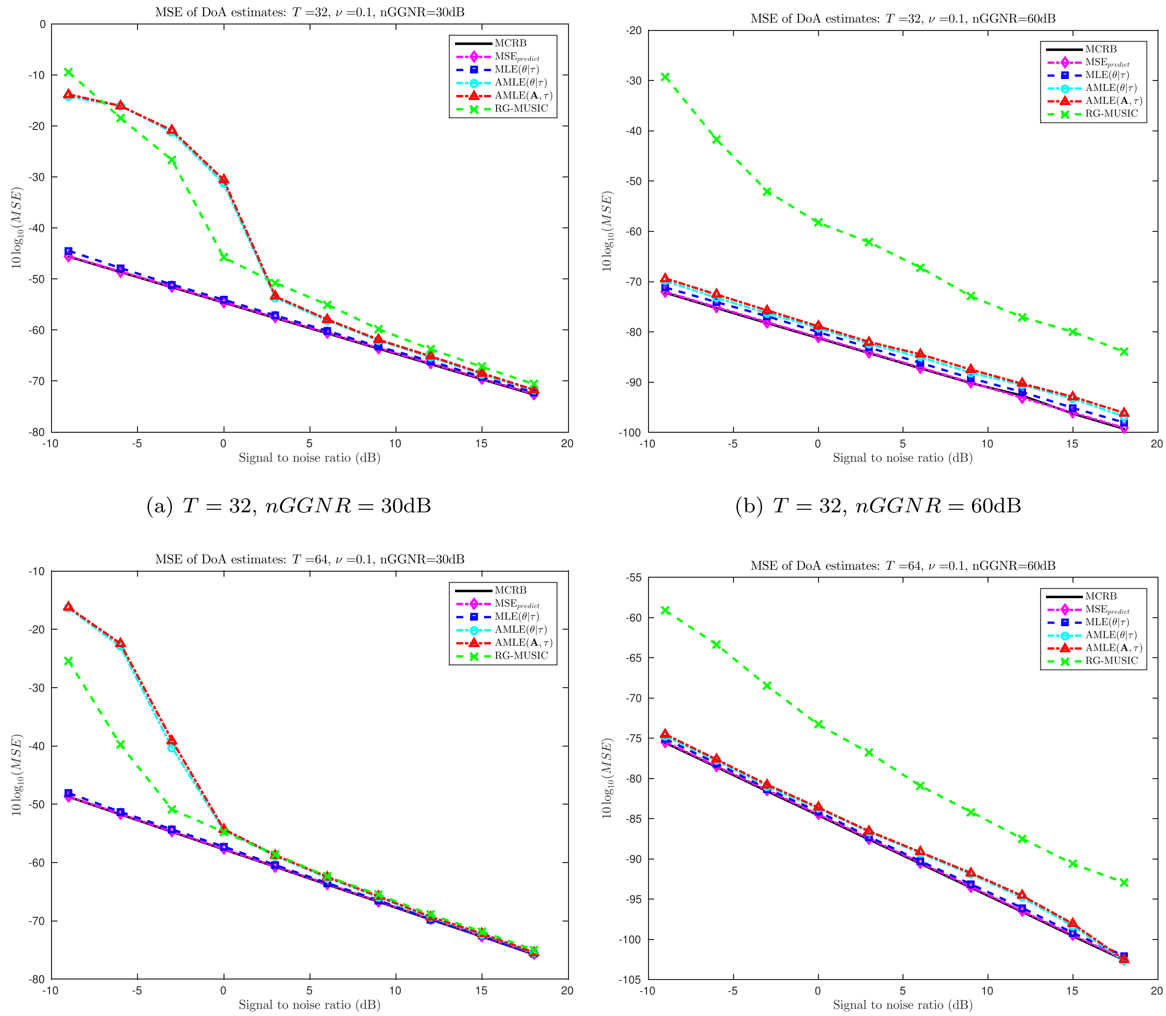
$$= T^{-1}CRB_G(1,\boldsymbol{I})\left[\beta^{-\nu}\sigma_w^{2(\nu-1)}e^{\beta^{-1}\sigma_w^2}\Gamma(1-\nu,\beta^{-1}\sigma_w^2)\right]^{-1}$$

$$= CRB_G(T,\sigma_w^2\boldsymbol{I})\times\frac{(\beta^{-1}\sigma_w^2)^{-\nu}e^{-\beta^{-1}\sigma_w^2}}{\Gamma(1-\nu,\beta^{-1}\sigma_w^2)} \tag{13}$$

where, in the second line we assumed that $\tau_t$ follows a Gamma distribution with shape parameter $\nu$ and scale parameter $\beta$ and, to obtain the penultimate equation, we made use of [30, 3.383.10] with $\Gamma(a,x) = \int_x^\infty t^{a-1}e^{-t}dt$. The expression in (12) corresponds to the inverse of the average FIM i.e., $\mathrm{E}\{FIM(\boldsymbol{\theta}|\boldsymbol{\tau})\}^{-1}$, and is the modified CRB (MCRB) [29]. Eq. (13) is the MCRB for the specific case of Gamma distributed textures.

### 2.2. Adaptive estimation (unknown $\boldsymbol{\tau}$)

Since, in practical situations, $\boldsymbol{\tau}$ is unknown, we now consider this case and proceed to joint estimation of $\boldsymbol{\tau}$ and the DoA. Prior to that, let us make the following observations. Treating $\boldsymbol{\theta}$ as unknowns, as in




(a) $T=32,\ nGGNR=30$dB (b) $T=32,\ nGGNR=60$dB

(c) $T=64,\ nGGNR=30$dB (d) $T=64,\ nGGNR=60$dB

**Fig. 3.** Mean square error of estimators versus *SNR*. $\nu=0.1$.

the conventional ML estimator, results in a $P$-dimensional optimization problem. MUSIC and its variants start from an estimated covariance matrix on which eigenvalue decomposition is applied. Then, a 1-dimensional search of $P$ maxima is performed, which is computationally simpler than $P$-dimensional search of one maximum. Hence, in order to come up with a practical estimation scheme, we choose to set as unknown the array steering matrix $\boldsymbol{A}$ which provides us with a basis for the signal subspace (actually this is also what MUSIC does); then, DoA estimation will be carried out by minimizing the usual MUSIC-like criterion $\boldsymbol{a}^H(\theta)\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{a}(\theta)$ or resorting to the root version of MUSIC to exploit the ULA structure, if desired. In summary, we will derive approximate ML estimates of $\boldsymbol{s}_t$, $\tau_t$ and (unstructured) $\boldsymbol{A}$ from observation of (1).

Let us first proceed to the maximization w.r.t. $\boldsymbol{\tau}$. For a given $\boldsymbol{A}$, we need to maximize (7) with respect to $\tau_t$, which amounts to maximizing $\left(\tau_t+\sigma_w^2\right)^{-M}\exp\left\{-\left(\tau_t+\sigma_w^2\right)^{-1}\boldsymbol{x}_t^H\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{x}_t\right\}$. Let $f(x)=x^{-M}\exp\left\{-\alpha x^{-1}\right\}$. It is readily seen that this function is increasing on $]0, M^{-1}\alpha]$, achieves its maximum at $x_\star=M^{-1}\alpha$ and then decreases monotonically. It follows that the maximum of (7), for a given $\boldsymbol{A}$, is attained for

$$\tau_t+\sigma_w^2=\begin{cases}\dfrac{\boldsymbol{x}_t^H\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{x}_t}{M} & \boldsymbol{x}_t^H\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{x}_t\geq M\sigma_w^2\\ \sigma_w^2 & \boldsymbol{x}_t^H\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{x}_t\leq M\sigma_w^2\end{cases}\tag{14}$$

and is given by

$$\max_{\boldsymbol{S},\boldsymbol{\tau}} p(\boldsymbol{X}|\boldsymbol{A},\boldsymbol{\tau},\boldsymbol{S})\propto\prod_{\boldsymbol{x}_t^H\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{x}_t\geq M\sigma_w^2}\left(\boldsymbol{x}_t^H\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{x}_t\right)^{-M}\times\prod_{\boldsymbol{x}_t^H\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{x}_t\leq M\sigma_w^2}\exp\left\{-\frac{\boldsymbol{x}_t^H\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{x}_t}{\sigma_w^2}\right\}.\tag{15}$$

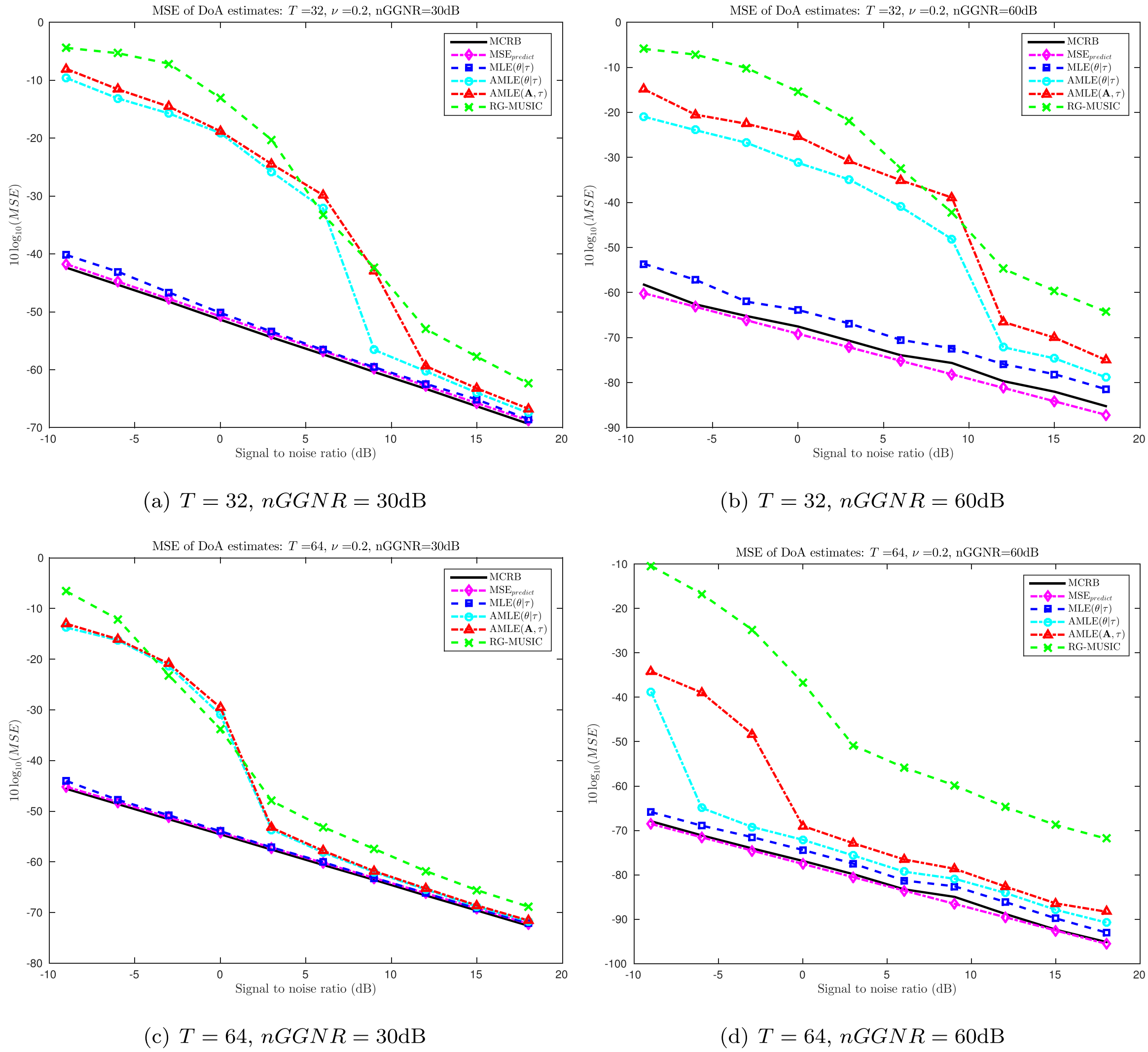
Let us make some comments about the previous result. First, let us note that, at the true steering matrix $\boldsymbol{A}_0$,

$$\boldsymbol{P}_{\boldsymbol{A}_0}^{\perp}\boldsymbol{x}_t=\sqrt{\tau_t}\boldsymbol{P}_{\boldsymbol{A}_0}^{\perp}\boldsymbol{n}_t+\sigma_w\boldsymbol{P}_{\boldsymbol{A}_0}^{\perp}\boldsymbol{w}_t\tag{16}$$

which implies that

$$\boldsymbol{x}_t^H\boldsymbol{P}_{\boldsymbol{A}_0}^{\perp}\boldsymbol{x}_t\simeq\tau_t\boldsymbol{n}_t^H\boldsymbol{P}_{\boldsymbol{A}_0}^{\perp}\boldsymbol{n}_t+\sigma_w^2\boldsymbol{w}_t^H\boldsymbol{P}_{\boldsymbol{A}_0}^{\perp}\boldsymbol{w}_t\tag{17}$$

where both $\boldsymbol{n}_t^H\boldsymbol{P}_{\boldsymbol{A}_0}^{\perp}\boldsymbol{n}_t$ and $\boldsymbol{w}_t^H\boldsymbol{P}_{\boldsymbol{A}_0}^{\perp}\boldsymbol{w}_t$ follow a complex chi-square distribution with $M-P$ degrees of freedom. In the previous equation, we neglected the cross-terms which are zero-mean while those retained have mean $M-P$. Therefore, (14) suggests that, for those snapshots with small value of $\tau_t$, or at least values which fall under white noise power, estimation of $\tau_t$ is meaningless, and the estimate of $\tau_t$ is equal to zero. Correspondingly, these samples in (15) are treated as if no compound Gaussian noise component is present. On




(a) $T=32$, $nGGNR=30$dB

(b) $T=32$, $nGGNR=60$dB

(c) $T=64$, $nGGNR=30$dB

(d) $T=64$, $nGGNR=60$dB

**Fig. 4.** Mean square error of estimators versus *SNR*. $\nu=0.2$.

the contrary, training samples with $\tau_t \gg \sigma_w^2$ are treated as if no white Gaussian noise is present. Hence, there is a natural separation between snapshots corresponding to small $\tau_t$ and snapshots corresponding to large $\tau_t$. Indeed, taking the logarithm of (15) yields

$$\begin{aligned}\log \max_{\boldsymbol{S},\boldsymbol{\tau}} p(\boldsymbol{X}|\boldsymbol{A},\boldsymbol{\tau},\boldsymbol{S}) &= \text{const.} - M \sum_{\boldsymbol{x}_t^H \boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{x}_t \geq M\sigma_w^2} \log(\boldsymbol{x}_t^H \boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{x}_t) \\ &- \sum_{\boldsymbol{x}_t^H \boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{x}_t \leq M\sigma_w^2} \frac{\boldsymbol{x}_t^H \boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{x}_t}{\sigma_w^2}.\end{aligned} \tag{18}$$

In (18), the first term coincides with the conditional ML function derived in the *K*-distributed noise only [19], while the second term coincides with the conditional ML function for internal Gaussian noise only. Accordingly, the derived representation of the log likelihood function in (18) relies upon a natural separation between the snapshots with small $\tau_t < \sigma_w^2$ and large $\tau_t > \sigma_w^2$. Naturally, since $\boldsymbol{A}$ is unknown, this separation is not feasible and one must achieve a similar separation based upon a running estimate of $\boldsymbol{A}$ and $\boldsymbol{\tau}$.

Coming back to (15), it appears that the maximization of the function there w.r.t. $\boldsymbol{A}$ is infeasible. In order to come up with a practical method, we use the fact that if $\boldsymbol{A}$ is known, the maximum of $\max_{\boldsymbol{S}} p(\boldsymbol{X}|\boldsymbol{A},\boldsymbol{\tau},\boldsymbol{S})$ w.r.t. $\boldsymbol{\tau}$ is given by (14). On the other hand, if $\boldsymbol{\tau}$ is known, the ML estimate of $\boldsymbol{A}$ is $\boldsymbol{A}_{\text{ML}|\boldsymbol{\tau}}$ as per (8). This suggests the approximate maximum likelihood estimator described by Algorithm 1, which is referred to as AMLE($\boldsymbol{A}$, $\tau$) in the sequel. The algorithm alternatively estimates $\boldsymbol{\tau}$ and a basis $\boldsymbol{A}$ for the signal subspace. The iterations are run a fixed number of times or until the distance between the subspaces spanned by $\boldsymbol{A}^{(n-1)}$ and $\boldsymbol{A}^{(n)}$ is below some threshold. Note that the concentrated likelihood function in (7) is increasing at each iteration. Indeed, if we let $g\left(\boldsymbol{P}_{\boldsymbol{A}}^{\perp},\boldsymbol{\tau}\right)$ denote the right-hand side of (7), we necessarily have

$$g\left(\boldsymbol{P}_{\boldsymbol{A}^{(n-1)}}^{\perp},\boldsymbol{\tau}^{(n-1)}\right) \leq g\left(\boldsymbol{P}_{\boldsymbol{A}^{(n-1)}}^{\perp},\boldsymbol{\tau}^{(n)}\right) \leq g\left(\boldsymbol{P}_{\boldsymbol{A}^{(n)}}^{\perp},\boldsymbol{\tau}^{(n)}\right). \tag{19}$$

This guarantees convergence to, at least, a local maximum. Yet (7) is not the true (concentrated) likelihood function since the latter is parametrized by $\boldsymbol{\theta}$ and is given by (5). We further discuss convergence below.

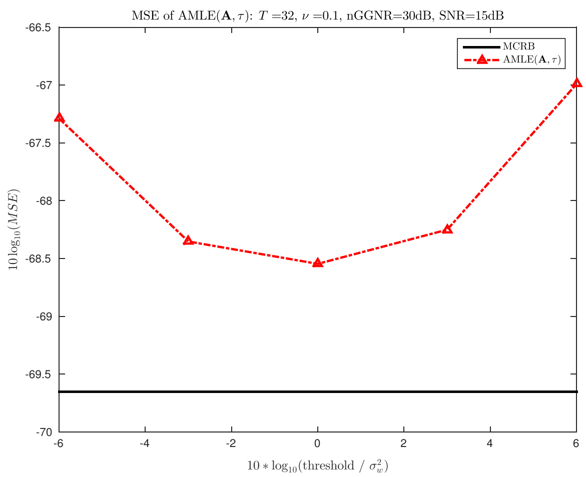
**Algorithm 1.** Approximate maximum likelihood estimation of $\boldsymbol{A}$ and $\boldsymbol{\tau}$.

**Input:** $\boldsymbol{X}$, initial estimate

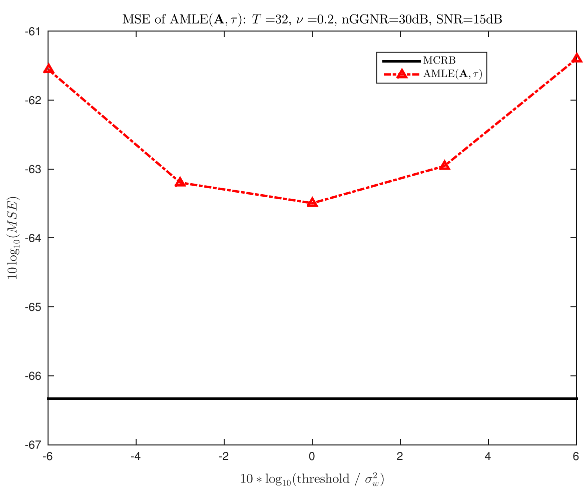
$$\boldsymbol{A}^{(0)} = \begin{bmatrix} \mathbf{0} \\ \boldsymbol{I}_P \end{bmatrix}$$

1: **for** $n = 1, \ldots$ **do**

2: Estimate $\tau_t^{(n)} = \max\left(M^{-1}\boldsymbol{x}_t^H \boldsymbol{P}_{\boldsymbol{A}^{(n-1)}}^{\perp}\boldsymbol{x}_t, \sigma_w^2\right) - \sigma_w^2$




(a) $\nu = 0.1$




(b) $\nu = 0.2$

**Fig. 5.** Mean square error of AMLE($\boldsymbol{A}, \boldsymbol{\tau}$) versus threshold. $T$=32, *nGGNR* = 30 dB and *SNR* = 15 dB.

Estimate $\boldsymbol{A}^{(n)} = \mathcal{P}_P\left(\sum_{t=1}^{T}\left(\tau_t^{(n)} + \sigma_w^2\right)^{-1}\boldsymbol{x}_t\boldsymbol{x}_t^H\right)$

**end for**

Let $\boldsymbol{A}_{\text{AML}}$ denote the value of $\boldsymbol{A}^{(n)}$ at the end of the iterations. Then, the DoAs are simply estimated as MUSIC$\left[\boldsymbol{A}_{\text{AML}}\right]$. Some observations about the above method are in order:

- As already said before, the algorithms, which use a threshold on the estimated value of $\tau_t$, result in a non-linear weighting of the various snapshots, depending on the amount of non-Gaussian noise power present.
- Convergence of Algorithm 1 is clearly an important issue and, admittedly, we are not in a position to provide a formal proof of global convergence, at most of local convergence. We can only surmise such a property from results obtained with similar algorithms. Indeed, the algorithm presented is reminiscent of iterative covariance estimation schemes, such as those found in the framework of elliptical distributions or for RG-MUSIC but the difference here is that one estimates only the signal subspace along the iterations, from an estimate of the covariance matrix. The difference is also in the fact that $\boldsymbol{x}_t^H \boldsymbol{P}_{\boldsymbol{A}}^{\perp} \boldsymbol{x}_t$ is used instead of the usual $\boldsymbol{x}_t^H \boldsymbol{R}^{-1} \boldsymbol{x}_t$, and that some thresholding is introduced. In fact, AMLE ($\boldsymbol{A}, \tau$) also bears resemblance with the recently proposed regularized iterative reweighted least-squares algorithm of [31] which is introduced for the so-called robust subspace recovery problem. This algorithm is closely related to $M$-estimators and Tyler's estimator. More precisely, [31] assumes that, within the provided data set, some of these points are sampled in a fixed subspace (inliers) and the rest of them (outliers) are spread in the whole ambient space. The aim of [31] is to recover the underlying subspace, and is based on an iterative algorithm. The scheme differs from ours in that the weights in IRLS suggested in [31] are calculated differently and a non-negative definite matrix is estimated at each iteration, the subspace being estimated after convergence. In our framework, inliers might be viewed as the samples corresponding to very small values of $\tau_t$ while outliers correspond to large values of the textures. Interestingly enough, the robustness of $M$-estimators to the presence of outliers has recently been shown in [32] and can be attributed to their ability to properly weight the samples. Finally, we mention that our AMLE($\boldsymbol{A}, \boldsymbol{\tau}$) coincides with the fast median subspace algorithm of [33] for a specific choice of the latter algorithm parameters, and some convergence properties of this algorithm have also been demonstrated in [33]. Accordingly, it is not that surprising that our method shares some common features with iterative covariance matrix estimation schemes for elliptical distributed data.

Before closing this section, we come back to the true ML solution which would consist of maximizing, with respect to $\boldsymbol{\theta}$, the likelihood function in (18) with $\boldsymbol{A}$ substituted for $\boldsymbol{A}(\boldsymbol{\theta})$. As indicated above, the function in (18) operates a selection of training samples: for those such that $\tau_t$ is above $\sigma_w^2$ one maximizes the conditional ML function derived in the $K$-distributed noise only (first term of Eq. (18)) while, for $\tau_t$ below $\sigma_w^2$, the second term coincides with the conditional ML function for internal Gaussian noise only. In other words, one can distinguish a situation (mostly for very small $\sigma_w^2$) where all texture values $\tau_t$ are larger than the internal Gaussian noise level. In this regime, the $K$-distributed noise dominates and the performance of ML estimator will be driven by the snapshot corresponding to minimum $\tau_t$, as proved in [19] where we provided an expression for the MSE of the ML estimator. On the other hand, for moderate nGGNR, some snapshots, say $n_T$ of them, will correspond to $\tau_t < \sigma_w^2$. In this case, we might expect the performance of the ML estimator to be driven by these $n_T$ snapshots in internal Gaussian noise only. Of course, these comments are qualitative but they can provide a prediction of the actual value of the mean-square error (MSE) of the ML estimator. More precisely, let

$$P(\sigma_w^2, T) = \Pr\left[\min \tau_t > \sigma_w^2\right] = \left(1 - \Pr\left[\tau_t < \sigma_w^2\right]\right)^T$$

$$= \left[1 - \gamma(\nu, \beta^{-1}\sigma_w^2)\right]^T \tag{20}$$

where $\gamma(a, x) = \frac{1}{\Gamma(a)}\int_0^x t^{a-1}e^{-t}dt$. With probability $P(\sigma_w^2, T)$, all textures will be above white the Gaussian noise level. As illustrated in Fig. 1, this probability is not negligible at least for very large values of the non-Gaussian to Gaussian noise ratio $\mathrm{E}\left\{\tau_t\right\}/\sigma_w^2$.

Hence, with probability $P(\sigma_w^2, T)$, the MSE will be more or less that obtained in $K$-distributed noise only. In [19] we derived an upper bound of this MSE in $K$ distributed noise only, which was seen to predict quite well the actual MSE, at least for $T$ large enough. An important observation made in [19] is that this MSE is of the order $T^{-1/\nu}$ for small $\nu$. On the other hand, with probability $1 - P(\sigma_w^2, T)$, $n_T$ snapshots will correspond to texture $\tau_t$ below $\sigma_w^2$, and the MSE is roughly $CRB_G(n_T, \sigma_w^2\boldsymbol{I})$. Summing up these observations, we propose the following “prediction” of the MSE of the

ML estimator:

$$MSE \simeq P(\sigma_w^2, T)C(\nu, \beta)T^{-1/\nu}CRB_G(1, \boldsymbol{I}) + [1 - P(\sigma_w^2, T)] \sum_{t=1}^{T} \Pr[n_T = t]\frac{1}{t}CRB_G(1, \sigma_w^2\boldsymbol{I}). \quad (21)$$

The first term in (21) is an upper bound for the MSE of the ML estimator in $K$-distributed noise only (and $\nu < 1$) while the second term accounts for randomness of $n_T$ which follow a binomial distribution with parameters $T$ and $\Pr\left[\tau_t < \sigma_w^2\right]$. In the next section, we will assess numerically (21) and compare it with the averaged CRB (11).

Eq. (21) suggests that for $\nu < 1$, there are two regimes specified by $(\sigma_w^2, T)$. For very small $\sigma_w^2$ and $T$, where $P(\sigma_w^2, T) \to 1$, DoA estimation accuracy is dominated by external noise with fast convergence rate $T^{-1/\nu}$ in this regime. As discussed in the introduction, for $T$=8 and $\nu$=0.1, $\mathrm{E}\left\{\min \tau_t\right\} = -38$ dB: it is therefore clear that nGGNR must be higher than 40 dB for internal white Gaussian noise (WGN) to be ignored for $T$=8 training samples. Therefore, for any practically reasonable nGGNR (say 50 dB) and small $\nu$ ($\nu \leq 0.2$), one could expect a rather short transition from $K$-noise dominated to internal WGN-dominated ML estimation regime. The latter is specified by WGN power $\sigma_w^2$ and $n_T$ training samples.

## 3. Numerical simulations

In this section, we aim at evaluating the performance of the estimators derived in the previous section. We consider an array with $M$=20 elements and two sources, with the same power, impinging from DoA $\theta_1 = 16°$ and $\theta_2 = 18°$. The waveforms are generated from a Gaussian distribution with covariance matrix $P_s\boldsymbol{I}_2$. The non-Gaussian noise follows a Gamma distribution with shape parameter $\nu$ and scale parameter $\nu^{-1}$ so that $\mathrm{E}\left\{\tau_t\right\} = 1$. Two values of $\nu$ are considered in what follows, namely $\nu$=0.1 and $\nu$=0.2. The total noise power is thus $1 + \sigma_w^2$ and the non-Gaussian to Gaussian noise ratio is defined as $nGGNR = 10\log_{10}\sigma_w^{-2}$. The signal to noise ratio is defined as $SNR = 10\log_{10}(P_s(\nu\beta + \sigma_w^2)^{-1})$.

The criterion used to assess performance is the sum of the mean-square errors of DoA estimates, i.e., $\sum_{i=1}^{2} \mathrm{E}\left\{(\hat{\theta}_i - \theta_i)^2\right\}$. 2000 Monte-Carlo simulations are run to evaluate the MSE. We compare the MLE$(\boldsymbol{\theta}|\boldsymbol{\tau})$, the AMLE$(\boldsymbol{\theta}|\boldsymbol{\tau})$, the AMLE$(\boldsymbol{A}, \boldsymbol{\tau})$ and the RG-MUSIC estimators. The MLE$(\boldsymbol{\theta}|\boldsymbol{\tau})$ was initialized at the true DoA and search for the global maximum of the likelihood function was restricted to the main beam width, so as to obtain a local behavior. The AMLE$(\boldsymbol{A}, \boldsymbol{\tau})$ was initialized with

$$\boldsymbol{A}^{(0)} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{I}_P \end{bmatrix}.$$

RG-MUSIC [24] is obtained from the eigenvalue decomposition of $\hat{\boldsymbol{R}} = \lim_{k\to\infty}\hat{\boldsymbol{R}}_k$ with

$$\hat{\boldsymbol{R}}_{k+1} = \frac{1}{T}\sum_{t=1}^{T} u\left(\frac{1}{M}\boldsymbol{x}_t^H\hat{\boldsymbol{R}}_k^{-1}\boldsymbol{x}_t\right)\boldsymbol{x}_t\boldsymbol{x}_t^H.$$

In the simulations $u(x) = (1 + \alpha)/(\alpha + x)$ with $\alpha$=0.2 [24]. For AMLE$(\boldsymbol{A}, \boldsymbol{\tau})$ as well as for RG-MUSIC, the number of iterations was set to 50 and the spectral form of MUSIC was used. Finally, we display the MCRB $\mathrm{E}\left\{CRB(\boldsymbol{\theta}|\boldsymbol{\tau})\right\}$ as well as the predicted MSE in (21).

We first study in Fig. 2 the transition between $K$-dominated and Gaussian dominated regimes in the asymptotic case, i.e., where SNR is high enough to prevent any threshold behavior. From this figure, one can observe the steep decrease of MSE for small values of $T$, and then a conventional behavior where the MSE varies as $T^{-1}$. This is especially pronounced when $\nu$=0.1 and $nGGNR$ = 60 dB, as could be expected from analysis of $P(\sigma_w^2, T)$. In this respect, the predicted MSE of Eq. (21) provides an accurate approximation of the actual MSE: it fits very well the MSE of MLE$(\boldsymbol{\theta}|\boldsymbol{\tau})$ and $\mathrm{E}\left\{CRB(\boldsymbol{\theta}|\boldsymbol{\tau})\right\}$, and the latter are nearly equal. This figure also reveals that AMLE$(\boldsymbol{\theta}|\boldsymbol{\tau})$ comes very close to MLE$(\boldsymbol{\theta}|\boldsymbol{\tau})$, at least for large enough $T$: this parallels the classical result that MUSIC-based estimates do as well as MLE asymptotically. However, the value of $T$ for which AMLE$(\boldsymbol{\theta}|\boldsymbol{\tau})$ is equivalent to MLE$(\boldsymbol{\theta}|\boldsymbol{\tau})$ decreases as $\nu$ decreases. Moreover, the AMLE$(\boldsymbol{A}, \boldsymbol{\tau})$ is seen to come very close to AMLE$(\boldsymbol{\theta}|\boldsymbol{\tau})$ despite the fact that it does not know $\tau$. Finally, we notice that RG-MUSIC is asymptotically as good as AMLE$(\boldsymbol{A}, \boldsymbol{\tau})$ for $nGGNR$ = 30 dB. However, for $nGGNR$ = 60 dB, even for large number of snapshots, there is a considerable difference between RG-MUSIC and all other algorithms, and this difference is more important when $\nu$=0.2 than when $\nu$=0.1.

Next, we investigate the threshold behavior of all algorithms in Figs. 3–4 where we plot MSE versus SNR, for different values of $T$. The curves are seen to be very different from $nGGNR$ = 30 dB to $nGGNR$ = 60 dB. In the latter case, all methods except RG-MUSIC achieve the bound even for low SNR, while for $nGGNR$ = 30 dB, one recovers the usual transition between a MSE far from the bound to a MSE close to the bound. It is observed that, for $nGGNR$ = 30 dB, RG-MUSIC has a better MSE than AMLE$(\boldsymbol{A}, \boldsymbol{\tau})$ for low $SNR$, while the two are equivalent for large $SNR$. For $nGGNR$ = 60 dB there is a significant improvement of AMLE$(\boldsymbol{A}, \boldsymbol{\tau})$ compared to RG-MUSIC. Another very interesting property is that AMLE$(\boldsymbol{A}, \boldsymbol{\tau})$ remains very close to AMLE$(\boldsymbol{\theta}|\boldsymbol{\tau})$, even in the threshold area, which means that AMLE$(\boldsymbol{A}, \boldsymbol{\tau})$ is an interesting solution.

Finally, we investigate *robustness of* AMLE$(\boldsymbol{A}, \boldsymbol{\tau})$ *to a non-perfect knowledge of white noise power*. Indeed, AMLE$(\boldsymbol{A}, \boldsymbol{\tau})$ requires knowledge of $\sigma_w^2$ and uses it mainly as a threshold to set to zero those estimated $\tau_t$ that fall below the threshold, see Algorithm 1. In order to figure out if a precise knowledge of $\sigma_w^2$ is required, we implemented AMLE$(\boldsymbol{A}, \boldsymbol{\tau})$ using a possibly wrong guess of $\sigma_w^2$: more precisely, the first step of Algorithm 1 is replaced by $\tau_t^{(n)} = \max\left(M^{-1}\boldsymbol{x}_t^H\boldsymbol{P}_{\boldsymbol{A}^{(n-1)}}^{\perp}\boldsymbol{x}_t, \mathrm{T}\right) - \mathrm{T}$ where the threshold $\mathrm{T} \neq \sigma_w^2$. In Fig. 5, we plot the MSE obtained when T varies around $\sigma_w^2$. Clearly, from this figure it can be seen that AMLE$(\boldsymbol{A}, \boldsymbol{\tau})$ is rather robust up to a 6 dB difference between T and $\sigma_w^2$. Usually, the internal noise level is known with a better accuracy and hence AMLE$(\boldsymbol{A}, \boldsymbol{\tau})$ is applicable even if one does not known perfectly $\sigma_w^2$.

## 4. Conclusions

In this paper, we addressed the direction finding problem when the additive noise consists of a mixture of $K$-distributed noise and Gaussian noise. The modified Cramér–Rao bound was derived. Assuming that the texture values $\tau$are known, we presented the maximum likelihood estimator. When $\tau$ is unknown an iterative procedure for joint estimation of $\tau$and a basis of the signal subspace was proposed. This AMLE$(\boldsymbol{A}, \boldsymbol{\tau})$ method was shown to provide quasi optimal performance in the asymptotic regime and a very good threshold behavior, especially for large ratio between the $K$-distributed noise power and the Gaussian noise power. It was also shown to achieve a better performance than RG-MUSIC, at least for high $K$-distributed noise power to Gaussian noise power ratio. Finally, a formula for predicting MSE, based upon differentiation between a $K$-dominated and a Gaussian-dominated regime, was presented which was shown to accurately fit the actual MSE values.

## References


[1] H.L. Van Trees, Optimum Array Processing, John Wiley, New York, 2002.
[2] J.F. Böhme, Estimation of spectral parameters of correlated signals in wavefields, Signal Process. 11 (4) (1986) 329–337.
[3] P. Stoica, A. Nehorai, MUSIC, maximum likelihood and Cramér–Rao bound, IEEE Trans. Acoust. Speech Signal Process. 37 (5) (1989) 720–741.
[4] P. Stoica, A. Nehorai, Performance study of conditional and unconditional Direction of Arrival Estimation, IEEE Trans. Acoust. Speech Signal Process. 38 (10) (1990) 1793–1795.
[5] P. Stoica, A. Nehorai, MUSIC, maximum likelihood and Cramér-Rao bound: further results and comparisons, IEEE Trans. Acoust. Speech Signal Process. 38 (12) (1990) 2140–2150.
[6] R.O. Schmidt, Multiple emitter location and signal parameter estimation, IEEE Trans. Antennas Propag. 34 (3) (1986) 276–280.
[7] P. Stoica, K. Sharman, Novel eigenanalysis method for direction estimation, Proc. IEE-Radar Sonar Nav. 137 (1) (1990) 19–26.
[8] R. Roy, T. Kailath, ESPRIT—estimation of signal parameters via rotational invariance techniques, IEEE Trans. Acoust. Speech Signal Process. 37 (7) (1989) 984–995.
[9] X. Mestre, M.A. Lagunas, Modified subspace algorithms for DoA estimation with large arrays, IEEE Trans. Signal Process. 56 (2) (2008) 598–614.
[10] B. Johnson, Y. Abramovich, X. Mestre, MUSIC, G-MUSIC and maximum likelihood performance breakdown, IEEE Trans. Signal Process. 56 (8) (2008) 3944–3958.
[11] J.K. Thomas, L.L. Scharf, D.W. Tufts, The probability of a subspace swap in the SVD, IEEE Trans. Signal Process. 43 (3) (1995) 730–736.
[12] B. Johnson, Y. Abramovich, X. Mestre, The role of subspace swap in MUSIC performance breakdown, in: Proceedings ICASSP, 2008, pp. 2473 –2476.
[13] R.R. Nadakuditi, F. Benaych-Georges, The breakdown point of signal subspace estimation, in: Proceedings SAM, Israel, 2010, pp. 177–180.
[14] X. Mestre, Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates, IEEE Trans. Inf. Theory 54 (11) (2008) 5113–5129.
[15] X. Mestre, On the asymptotic behavior of the sample estimates of eigenvalues and eigenvectors of covariance matrices, IEEE Trans. Signal Process. 56 (11) (2008) 5353–5368.
[16] P. Tsakalides, C.L. Nikias, Maximum likelihood localization of sources in noise modeled as a stable process, IEEE Trans. Signal Process. 43 (11) (1995) 2700–2713.
[17] Y. Yardimci, A. Cetin, J.A. Cadzow, Robust direction-of-arrival estimation in non-Gaussian noise, IEEE Trans. Signal Process. 46 (5) (1998) 1443–1451.
[18] R. Kozick, B. Sadler, Maximum-likelihood array processing in non-Gaussian noise with Gaussian mixtures, IEEE Trans. Signal Process. 48 (12) (2000) 3520–3535.
[19] Y. Abramovich, O. Besson, B. Johnson, Bounds for maximum likelihood regular and non-regular DoA estimation in *K*-distributed noise, IEEE Trans. Signal Process. 63 (21) (2015) 5746–5757.
[20] E. Ollila, D. Tyler, V. Koivunen, H. Poor, Complex elliptically symmetric distributions: survey, new results and applications, IEEE Trans. Signal Process. 60 (11) (2012) 5597–5625.
[21] M. Mahot, F. Pascal, P. Forster, J.-P. Ovarlez, Robust covariance matrix estimates with attractive asymptotic properties, in: Proceedings of the 4th IEEE CAMSAP Workshop, San Juan, 2011, pp. 305–308.
[22] M. Mahot, F. Pascal, P. Forster, J.-P. Ovarlez, Asymptotic properties of robust complex covariance matrix estimates, IEEE Trans. Signal Process. 61 (13) (2013) 3348–3356.
[23] R. Couillet, F. Pascal, J.W. Silverstein, Robust estimates of covariance matrices in the large dimensional regime, IEEE Trans. Inf. Theory 60 (2014) 7269–7278.
[24] R. Couillet, Robust spiked random matrices and a robust G-MUSIC estimator, J. Multivar. Anal. 140 (2015) 139–161.
[25] M.A. Wadsworth, S.A. Dianat, Representation of HF atmospheric noise as a spherically invariant random process, in: Proceedings AeroSense'99, International Society for Optics and Photonics, 1999, pp. 181–192.
[26] E. Conte, M. Longo, Characterisation of radar clutter as a spherically invariant process, IEE Proc. Radar, Sonar Nav. 134 (2) (1987) 191–197.
[27] C.J. Baker, Coherent *K*-distributed sea clutter, IEE Proc. Radar, Sonar Nav. 138 (2) (1991) 89–92.
[28] R.W. Miller, C.B. Chang, A modified Cramér–Rao bound and its applications, IEEE Trans. Inf. Theory 24 (3) (1978) 398–400.
[29] F. Gini, R. Reggiannini, On the use of Cramér–Rao-like bounds in the presence of random nuisance parameters, IEEE Trans. Commun. 48 (12) (2000) 2120–2126.
[30] I.S. Gradshteyn, I.M. Ryzhik, in: Alan Jeffrey (Ed.), Table of Integrals, Series and Products, 5th ed., Academic Press, San Diego, 1994.
[31] T. Zhang, G. Lerman, A. Novel, M-estimator for robust PCA, J. Mach. Learn. Res. 15 (2014) 749–808.
[32] D. Morales-Gonzalez, R. Couillet, M. McKay, Large dimensional analysis of robust M-estimators of covariance with outliers, IEEE Trans. Signal Process. 63 (21) (2015) 5784–5797.
[33] G. Lerman, T. Maunu, Fast algorithm for robust subspace recovery URL arxiv.org/abs/1406.6145.