Signal Processing 160 (2019) 150–163

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

MSE based optimization of the measure-transformed MUSIC algorithm

Nir Halay, Koby Todros*

Ben-Gurion University of the Negev, Israel

ARTICLE INFO

Article history:
Received 2 July 2018
Revised 28 November 2018
Accepted 31 January 2019
Available online 5 February 2019

Keywords:
Array processing
DOA estimation
Probability measure transform
Robust statistics

ABSTRACT

Recently, we developed a robust generalization of the MUltiple SIgnal Classification (MUSIC) algorithm. This generalization, called measure-transformed MUSIC (MT-MUSIC), operates by applying a transform to the probability measure (distribution) of the data. The considered transform is structured by a non-negative data-weighting function, called MT-function, that, when properly chosen, mitigates the effect of non-Gaussian heavy-tailed noise that produces outliers. In this paper, we characterize the asymptotic mean-squared-error (MSE) performance of the MT-MUSIC algorithm. Under some mild regularity conditions, we show that the MT-MUSIC estimator is asymptotically normal and unbiased, and obtain an analytic expression for the asymptotic MSE matrix. We go on to develop a strongly consistent estimator for the asymptotic MSE matrix that is constructed via the *same* sequence of samples being used for implementation of the MT-MUSIC. This paves the way for development of a data-driven procedure for optimal selection of the MT-function parameters that minimizes an empirical estimate of the asymptotic average root MSE (RMSE). The performance advantage of the proposed MSE based optimization of the MT-MUSIC is illustrated in simulation examples.

1. Introduction

The multiple signal classification (MUSIC) algorithm [1,2] is a popular technique for estimating direction-of-arrivals (DOAs) of noisy signals received by an array of sensors. This algorithm has gained popularity due to its appealing implementation simplicity and computational load that is significantly lower as compared to the maximum-likelihood estimator (MLE) [2], which unlike MUSIC, involves multi-dimensional optimization. The operation principle of the MUSIC algorithm is based on finding DOAs with corresponding steering vectors that have minimal projections onto an empirical estimate of the noise-subspace. The spanning vectors of the empirical noise-subspace are obtained from the eigenvectors corresponding to the minimal eigenvalues of the sample covariance matrix (SCM).

The SCM is the maximum-likelihood estimator of the covariance under normally distributed observations [3]. This estimator is not resilient against large deviations from normality, that can occur in the presence of heavy-tailed noise that produces outliers, resulting in poor DOAs estimates. In order to alleviate this limitation, several robust MUSIC generalizations have been proposed in the literature that replace the SCM with robust scatter matrix estimators, for which the empirical noise subspace can be determined from their eigen-decomposition. In [4], two robust MUSIC generalizations were developed that replace the SCM with robust sign and rank covariance estimates. In [5], it was proposed to replace the non-robust SCM with robust M-estimators of scatter [6], such as the maximum-likelihood (under elliptical observations), Huber's [7], and Tyler's [8]. In this context, we note that a consistent robust MUSIC generalization based on M-estimators of scatter was developed in [9,10] for the high-dimensional sample-starved scenario (which is not considered in this manuscript). A more elaborate discussion about the methods proposed in [4,5], and other important robust MUSIC generalizations [11–15], including their advantages and disadvantages, appears in [16, Section 1].

Recently, we developed in [16] a robust MUSIC generalization, called measure-transformed MUSIC (MT-MUSIC) that operates by applying a transform to the probability measure of the data. The considered probability measure transformation, also applied in [17–22], is structured by a nonnegative function, called MT-function, that weights the data points. In MT-MUSIC, the SCM is replaced by an empirical estimate of the measure-transformed (MT) covariance. In [16], the MT-MUSIC was implemented with a spherically contoured Gaussian MT-function parameterized by a width parameter. Under this type of MT-function, it was shown that the empirical MT-covariance is B-robust [23], and, unlike robust scatter matrix estimators, such as the empirical sign-covariance and Tyler's M-estimator of scatter, its influence function [23] decays to zero as the outlier norm approaches infinity, resulting in enhanced robustness to large-norm outliers. Furthermore, under the assumption of compound Gaussian (CG)

* Corresponding author.
E-mail address: todros@bgu.ac.il (K. Todros).

https://doi.org/10.1016/j.sigpro.2019.01.025
0165-1684/

[5] spherically symmetric noise, we have shown that the noise subspace can be determined from the eigen-decomposition of the Gaussian MT-covariance. In [16], selection of the width parameter of the Gaussian MT-function was carried out via suboptimal procedure that controls the transform-domain Fisher-information loss under nominal Gaussian distribution of the data.

This paper provides an important extension of the work [16]. Here, we analyze the asymptotic mean-squared-error (MSE) performance of the MT-MUSIC algorithm. We emphasize that the analysis is not restricted to the Gaussian MT-function considered in [16]. We further note that the analysis is carried out *without* assuming that the probability distribution of the data belongs to a specific family (e.g., Gaussian, elliptical). Under some mild regularity conditions on the MT-function, we show that the MT-MUSIC estimator is asymptotically normal and unbiased, and obtain an analytic expression for the asymptotic MSE matrix. We go on to develop a strongly consistent estimator for the asymptotic MSE matrix that is constructed via the *same* sequence of samples being used for implementation of the MT-MUSIC. This paves the way for development of a data-driven procedure for optimal selection of the MT-function parameters that, unlike the suboptimal procedure considered in [16], minimizes an empirical estimate of the asymptotic average root MSE (RMSE).

The proposed MSE based optimization of the MT-MUSIC algorithm is illustrated in simulation examples that involve Gaussian and heavy-tailed non-Gaussian noise. In these examples, the MT-MUSIC is implemented with a non-Gaussian outlier-suppressing MT-function whose parameter is selected via the optimal MSE based data-driven procedure discussed above. We show that the optimized non-Gaussian MT-MUSIC outperforms the suboptimal Gaussian MT-MUSIC [16] and other state-of-the-art robust MUSIC generalizations.

The paper is organized as follows. In Section 2, we review the MT-MUSIC algorithm [16]. In Section 3, an asymptotic MSE analysis is provided and a data-driven procedure for optimal selection of the measure-transformation parameters is developed. The MSE based optimization of the MT-MUSIC is illustrated in simulation examples in Section 4. In Section 5 we provide concluding remarks. Proofs for the theorem and proposition stated in the manuscript are provided in the Appendix.

2. Measure-transformed MUSIC: review

We begin by introducing the considered sensor array model. Then, the relevant principles of the probability measure transform [16–22] are presented. Finally, the MT-MUSIC algorithm [16] is reviewed.

2.1. Array model

Consider an array of p sensors that receives signals from $q<p$ narrow-band far-field incoherent point sources with distinct azimuthal DOAs $\{\theta_1,\ldots,\theta_q\}\subset\Theta$, where $\Theta\subseteq[-\pi,\pi)$ denotes the parameter space. Under this setting, the array output satisfies the following observation model [2]:

$$\mathbf{x}_n=\mathbf{A}\boldsymbol{s}_n+\mathbf{w}_n,\quad n=1,\ldots,N, \tag{1}$$

where $\{\mathbf{x}_n\in\mathbb{C}^p\}$ is an observation process, $\{\boldsymbol{s}_n\in\mathbb{C}^q\}$ is a first-order stationary latent signal process with zero-mean and non-singular covariance, and $\{\mathbf{w}_n\in\mathbb{C}^p\}$ denotes a first-order stationary zero-mean spatially white noise processes that is independent of $\{\boldsymbol{s}_n\}$. The matrix $\mathbf{A}\triangleq[\mathbf{a}(\theta_1),\ldots,\mathbf{a}(\theta_q)]\in\mathbb{C}^{p\times q}$ is the array steering matrix, where $\mathbf{a}(\vartheta)\in\mathbb{C}^p$ is a steering vector toward direction $\vartheta\in\Theta$. The array is assumed to be unambiguous, i.e., any collection of p steering vectors corresponding to distinct DOAs forms a linearly independent set. Under this assumption the steering matrix $\mathbf{A}$ has a full column rank, and therefore, identification of its column vectors amounts to identification of the DOAs.

2.2. Probability measure transform

We define the measure space $(\mathcal{X},\mathcal{S}_{\mathcal{X}},P_{\mathbf{x}})$, where $\mathcal{X}\subseteq\mathbb{C}^p$ is the observation space of a random vector $\mathbf{x}$, $\mathcal{S}_{\mathcal{X}}$ is a σ-algebra over $\mathcal{X}$ and $P_{\mathbf{x}}$ is a probability measure on $\mathcal{S}_{\mathcal{X}}$.

Definition 1. Given a non-negative function $u:\mathbb{C}^p\to\mathbb{R}_+$ satisfying $0<\mathrm{E}[u(\mathbf{x});P_{\mathbf{x}}]<\infty$, where $\mathrm{E}[u(\mathbf{x});P_{\mathbf{x}}]\triangleq\int_{\mathcal{X}}u(\mathbf{r})dP_{\mathbf{x}}(\mathbf{r})$, a transform on $P_{\mathbf{x}}$ is defined as:

$$Q_{\mathbf{x}}^{(u)}(A)\triangleq\mathbb{T}_u[P_{\mathbf{x}}](A)=\int_A\varphi_u(\mathbf{r})dP_{\mathbf{x}}(\mathbf{r}), \tag{2}$$

where $A\in\mathcal{S}_{\mathcal{X}}$ and $\varphi_u(\mathbf{r})\triangleq u(\mathbf{r})/\mathrm{E}[u(\mathbf{x});P_{\mathbf{x}}]$. The function $u(\cdot)$ is called the MT-function.

By Proposition 1 in [16], $Q_{\mathbf{x}}^{(u)}$ is a probability measure on $\mathcal{S}_{\mathcal{X}}$ that is absolutely continuous w.r.t. $P_{\mathbf{x}}$, with Radon–Nikodym derivative [24] $\frac{dQ_{\mathbf{x}}^{(u)}}{dP_{\mathbf{x}}}(\mathbf{r})=\varphi_u(\mathbf{r})$. Thus, the covariance matrix of $\mathbf{x}$ under $Q_{\mathbf{x}}^{(u)}$, called the MT-covariance, is given by:

$$\boldsymbol{\Sigma}_{\mathbf{x}}^{(u)}\triangleq\mathrm{E}\big[\mathbf{x}\mathbf{x}^H\varphi_u(\mathbf{x});P_{\mathbf{x}}\big]-\boldsymbol{\mu}_{\mathbf{x}}^{(u)}\boldsymbol{\mu}_{\mathbf{x}}^{(u)H}, \tag{3}$$

where $\boldsymbol{\mu}_{\mathbf{x}}^{(u)}\triangleq\mathrm{E}[\mathbf{x}\varphi_u(\mathbf{x});P_{\mathbf{x}}]$ denotes the MT-mean, i.e., the mean vector under $Q_{\mathbf{x}}^{(u)}$. Eq. (3) implies that the MT-covariance is a weighted covariance of $\mathbf{x}$ under $P_{\mathbf{x}}$, with the weighting function $\varphi_u(\cdot)$ defined below (2). Notice that when the MT-function $u(\cdot)$ is non-zero and constant valued, the standard covariance is obtained.

Given a sequence of N i.i.d. samples from $P_{\mathbf{x}}$, the empirical MT-covariance is defined as:

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{(u)}\triangleq\sum_{n=1}^{N}\mathbf{x}_n\mathbf{x}_n^H\hat{\varphi}_u(\mathbf{x}_n)-\hat{\boldsymbol{\mu}}_{\mathbf{x}}^{(u)}\hat{\boldsymbol{\mu}}_{\mathbf{x}}^{(u)H}, \tag{4}$$

where $\hat{\varphi}_u(\mathbf{x}_n)\triangleq u(\mathbf{x}_n)/\sum_{n=1}^{N}u(\mathbf{x}_n)$, and the empirical MT-mean $\hat{\boldsymbol{\mu}}_{\mathbf{x}}^{(u)}\triangleq\sum_{n=1}^{N}\mathbf{x}_n\hat{\varphi}_u(\mathbf{x}_n)$. According to Proposition 2 in [16], if $\mathrm{E}[\|\mathbf{x}\|^2u(\mathbf{x});P_{\mathbf{x}}]<\infty$ then $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{(u)}\xrightarrow{\text{w.p.1}}\boldsymbol{\Sigma}_{\mathbf{x}}^{(u)}$, as $N\to\infty$, where "$\xrightarrow{\text{w.p.1}}$" denotes convergence with probability (w.p.) 1 [25].

Robustness of the empirical MT-covariance (4) to outliers was studied in [16] using its influence function [23], which describes the effect on the estimator introduced by an infinitesimal contamination at some point $\mathbf{r}\in\mathbb{C}^p$. An estimator is said to be B-robust if its influence function is bounded [23]. In Proposition 3 in [16], it was shown that if there exists a finite positive constant M, such that

$$u(\mathbf{r})\le M\ \text{ and }\ u(\mathbf{r})\|\mathbf{r}\|^2\le M,\quad\forall\mathbf{r}\in\mathbb{C}^p \tag{5}$$

then the influence function of $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{(u)}$ is bounded.

2.3. The MT-MUSIC algorithm

The MT-MUSIC [16] is comprised of two steps. First, the MT-function $u(\cdot)$ is chosen such that the following conditions are satisfied:

(A-1) The resulting empirical MT-covariance $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{(u)}$ is B-robust.

(A-2) Let $\lambda_1^{(u)}\ge\ldots\ge\lambda_p^{(u)}$ denote the eigenvalues of the MT-covariance $\boldsymbol{\Sigma}_{\mathbf{x}}^{(u)}$. The $p-q$ smallest eigenvalues of $\boldsymbol{\Sigma}_{\mathbf{x}}^{(u)}$ satisfy

$$\lambda_q^{(u)}>\lambda_{q+1}^{(u)}=\ldots=\lambda_p^{(u)}\triangleq\gamma^{(u)} \tag{6}$$

and their corresponding eigenvectors span the null-space of $\mathbf{A}^H$, also called the noise-subspace.

Second, the DOAs are estimated by finding the q highest maxima of the measure-transformed pseudo-spectrum:

$$\hat{P}^{(u)}(\vartheta) \triangleq \|\hat{\mathbf{G}}^{(u)H}\mathbf{a}(\vartheta)\|^{-2}, \tag{7}$$

where $\hat{\mathbf{G}}^{(u)} \in \mathbb{C}^{p\times p-q}$ denotes the matrix of $p-q$ eigenvectors of $\hat{\mathbf{\Sigma}}_{\mathbf{x}}^{(u)}$ corresponding to the smallest eigenvalues.

We note that in [16], it was shown that condition A-2 is satisfied when the noise obeys a spherical CG distribution and the MT-function belongs to the class of spherically contoured *Gaussian* functions. In this paper, we show in Appendix C that, under the spherical CG noise assumption, condition A-2 is satisfied for *any* strictly-positive spherically contoured function.

3. MSE based optimization of the MT-MUSIC

In this section we characterize the asymptotic MSE performance of the MT-MUSIC algorithm [16]. We then utilize an empirical estimate of the asymptotic MSE to develop a data-driven procedure for optimal selection of the MT-function within some parametric class.

3.1. Asymptotic performance analysis

Throughout the analysis that follows, the deterministic vector $\boldsymbol{\theta} \triangleq [\theta_1, \ldots, \theta_q]$ denotes the true DOAs. The random vector $\hat{\boldsymbol{\theta}}_u \triangleq [\hat{\theta}_1, \ldots, \hat{\theta}_q]$ denotes their estimates obtained by the MT-MUSIC. For simplicity, we shall assume that the MT-mean $\boldsymbol{\mu}_{\mathbf{x}}^{(u)} = \mathbf{0}$ and that the MT-MUSIC is implemented with a sequence of i.i.d. samples from $P_{\mathbf{x}}$. Furthermore, it is assumed that condition A-2, stated above, is satisfied. We note that the first simplifying assumption is satisfied whenever the observations are symmetrically distributed about the origin and the user-defined MT-function $u(\cdot)$ is zero-centered and symmetric. We note that when $\boldsymbol{\mu}_{\mathbf{x}}^{(u)} \neq \mathbf{0}$, the following performance analysis still applies to the case when the MT-MUSIC is implemented with the empirical MT-autocorrelation matrix $\hat{\mathbf{C}}_{\mathbf{x}}^{(u)} \triangleq \hat{\mathbf{\Sigma}}_{\mathbf{x}}^{(u)} + \hat{\boldsymbol{\mu}}_{\mathbf{x}}^{(u)}\hat{\boldsymbol{\mu}}_{\mathbf{x}}^{(u)H}$ instead of the empirical MT-covariance $\hat{\mathbf{\Sigma}}_{\mathbf{x}}^{(u)}$.

The following theorem states sufficient conditions for asymptotic normality and unbiasedness.

Theorem 1. *Assume that the following conditions hold:*

(B-1) *The expectation* $\mathrm{E}[\|\mathbf{x}\|^4 u^2(\mathbf{x}); P_{\mathbf{x}}]$ *is finite.*
(B-2) *The steering vector* $\mathbf{a}(\vartheta)$ *has a bounded Euclidean norm.*
(B-3) $\mathbf{a}(\vartheta)$ *is twice continuously differentiable.*
(B-4) *The true DOAs* $\theta_1, \ldots, \theta_q$ *lie in the interior of* Θ.
(B-5) *The eigenvalues* $\lambda_1^{(u)}, \ldots, \lambda_q^{(u)}$ *associated with the signal subspace are distinct.*

Then,

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_u - \boldsymbol{\theta}) \xrightarrow{\mathrm{D}} \mathcal{N}(\mathbf{0}, \mathbf{R}^{(u)}(\boldsymbol{\theta})) \quad \text{as} \quad N \to \infty, \tag{8}$$

where "$\xrightarrow{\mathrm{D}}$*" denotes converges in distribution [25],*

$$\mathbf{R}^{(u)}(\boldsymbol{\theta}) \triangleq \mathrm{E}[\varphi_u^2(\mathbf{x})\boldsymbol{\psi}_u(\mathbf{x},\boldsymbol{\theta})\boldsymbol{\psi}_u^T(\mathbf{x},\boldsymbol{\theta}); P_{\mathbf{x}}], \tag{9}$$

$$[\boldsymbol{\psi}_u(\mathbf{x},\boldsymbol{\theta})]_i \triangleq \frac{\mathrm{Re}\{\dot{\mathbf{a}}^H(\theta_i)\mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{x}\mathbf{x}^H\mathbf{E}^{(u)}\mathbf{a}(\theta_i)\}}{\dot{\mathbf{a}}^H(\theta_i)\mathbf{P}_{\mathbf{A}}^{\perp}\dot{\mathbf{a}}(\theta_i)}, \tag{10}$$

$$\mathbf{E}^{(u)} \triangleq \mathbf{A}(\mathbf{A}^H(\mathbf{\Sigma}_{\mathbf{x}}^{(u)} - \gamma^{(u)}\mathbf{I}_p)\mathbf{A})^{-1}\mathbf{A}^H, \tag{11}$$

$\varphi_u(\cdot)$ is defined below (2), $\dot{\mathbf{a}}(\vartheta) \triangleq d\mathbf{a}(\vartheta)/d\vartheta$, $[\mathbf{v}]_i$ denotes the ith coordinate of a vector $\mathbf{v}$, $\mathbf{P}_{\mathbf{A}}^{\perp}$ is the projection matrix onto the null-space of $\mathbf{A}^H$, $\gamma^{(u)}$ is defined in (6) and $\mathbf{I}_p$ denotes a $p\times p$ identity matrix. [A proof appears in Appendix A]

Theorem 1 implies that the asymptotic MSE matrix of the MT-MUSIC estimator is given by:

$$\mathbf{C}^{(u)}(\boldsymbol{\theta}) \triangleq N^{-1}\mathbf{R}^{(u)}(\boldsymbol{\theta}). \tag{12}$$

We note that the asymptotic MSE (12) was obtained without assuming that the probability distribution of the data, $P_{\mathbf{x}}$, belongs to a specific family (e.g., Gaussian, elliptical). In particular, when the noise obeys a proper complex normal distribution and the MT-function $u(\cdot)$ is non-zero and constant valued, it can be shown that the asymptotic MSE (12) coincides with the one reported in [26, Eq. (3.11a)], [27, Eq. (55)] and [28, Eq. (69)] for the standard SCM based MUSIC. This result is intuitive since for any non-zero and constant valued MT-function the MT-MUSIC coincides with the standard SCM based MUSIC. In this context, it is worthwhile noting that a general framework for analyzing the asymptotic MSE performance of second-order estimation methods, that are based on the SCM, was developed in [29]. Note that, in this paper, we analyze the asymptotic performance of the MT-MUSIC, that is not based on the SCM, and uses a different covariance, namely, the empirical measure-transformed covariance (4).

In the following Proposition, a strongly consistent estimate of (12) is developed. This quantity will be used in the following subsection for optimal selection of the MT-function.

Proposition 1. *Define the empirical asymptotic MSE:*

$$\hat{\mathbf{C}}^{(u)}(\hat{\boldsymbol{\theta}}_u) \triangleq N^{-1}\hat{\mathbf{R}}^{(u)}(\hat{\boldsymbol{\theta}}_u), \tag{13}$$

where

$$\hat{\mathbf{R}}^{(u)}(\hat{\boldsymbol{\theta}}_u) \triangleq N\sum_{n=1}^{N}\hat{\varphi}_u^2(\mathbf{x}_n)\hat{\boldsymbol{\psi}}_u(\mathbf{x}_n,\hat{\boldsymbol{\theta}}_u)\hat{\boldsymbol{\psi}}_u^T(\mathbf{x}_n,\hat{\boldsymbol{\theta}}_u), \tag{14}$$

$$[\hat{\boldsymbol{\psi}}_u(\mathbf{x},\hat{\boldsymbol{\theta}}_u)]_i \triangleq \frac{\mathrm{Re}\{\dot{\mathbf{a}}^H(\hat{\theta}_i)\mathbf{P}_{\hat{\mathbf{A}}}^{\perp}\mathbf{x}\mathbf{x}^H\hat{\mathbf{E}}^{(u)}\mathbf{a}(\hat{\theta}_i)\}}{\dot{\mathbf{a}}^H(\hat{\theta}_i)\mathbf{P}_{\hat{\mathbf{A}}}^{\perp}\dot{\mathbf{a}}(\hat{\theta}_i)}, \tag{15}$$

$$\hat{\mathbf{E}}^{(u)} \triangleq \hat{\mathbf{A}}(\hat{\mathbf{A}}^H(\hat{\mathbf{\Sigma}}_{\mathbf{x}}^{(u)} - \hat{\gamma}^{(u)}\mathbf{I}_p)\hat{\mathbf{A}})^{-1}\hat{\mathbf{A}}^H, \tag{16}$$

$\hat{\varphi}_u(\cdot)$ is defined below (4), $\hat{\mathbf{A}} \triangleq [\mathbf{a}(\hat{\theta}_1), \ldots, \mathbf{a}(\hat{\theta}_q)]$, and $\hat{\gamma}^{(u)} \triangleq \mathrm{trace}\{\mathbf{P}_{\hat{\mathbf{A}}}^{\perp}\hat{\mathbf{\Sigma}}_{\mathbf{x}}^{(u)}\}/(p-q)$. Furthermore, assume that conditions B-1 - B-3, stated in Theorem 1, are satisfied. Then,

$$N\|\hat{\mathbf{C}}^{(u)}(\hat{\boldsymbol{\theta}}_u) - \mathbf{C}^{(u)}(\boldsymbol{\theta})\| \xrightarrow{\text{w.p.1}} 0 \quad \text{as} \quad N \to \infty. \tag{17}$$

[A proof appears in Appendix B]

3.2. Optimal selection of the MT-function

We propose to choose the MT-function $u(\cdot)$ that minimizes an empirical estimate of the asymptotic average RMSE. This estimate, that is obtained by taking the arithmetic mean over the square roots of the diagonal terms comprising the empirical asymptotic MSE matrix (13), takes the following form:

$$J_u \triangleq \frac{1}{q}\sum_{i=1}^{q}\frac{\sqrt{\sum_{n=1}^{N}\hat{\varphi}_u^2(\mathbf{x}_n)\mathrm{Re}^2\{\dot{\mathbf{a}}^H(\hat{\theta}_i)\mathbf{P}_{\hat{\mathbf{A}}}^{\perp}\mathbf{x}_n\mathbf{x}_n^H\hat{\mathbf{E}}^{(u)}\mathbf{a}(\hat{\theta}_i)\}}}{\dot{\mathbf{a}}^H(\hat{\theta}_i)\mathbf{P}_{\hat{\mathbf{A}}}^{\perp}\dot{\mathbf{a}}(\hat{\theta}_i)}. \tag{18}$$

We emphasize that the objective function (18) is constructed via the *same* sequence of data samples used for implementation of the MT-MUSIC algorithm. Here, we restrict the class of MT-functions to some parametric family $\{u(\mathbf{x};\boldsymbol{\omega}), \boldsymbol{\omega} \in \mathbf{\Omega} \subseteq \mathbb{C}^r\}$ that satisfies conditions A-1, A-2, B-1 and B-5. An example for a non-Gaussian parametric MT-function that induces outlier resilience is given in the following section. The optimal MT-function parameter $\boldsymbol{\omega}_{\mathrm{opt}}$ is obtained through minimization of $J_u(\boldsymbol{\omega})$. This minimization is carried out numerically. When $\boldsymbol{\omega}$ is one-dimensional,

a simple line search can be implemented. Otherwise, the minimization can be carried out via greedy search or gradient descend [30].

Finally, we note that throughout the rest of this paper, the MT-MUSIC algorithm implemented with MT-function whose parameters are selected according to the MSE-based optimization approach, discussed above, will be referred to as "optimized MT-MUSIC".

4. Numerical examples

In this section, we evaluate the performance of the optimized MT-MUSIC (MT-MUSIC$_{\text{opt}}$) as compared to the suboptimal MT-MUSIC (MT-MUSIC$_{\text{sub}}$) [16], the non-robust SCM based MUSIC (SCM-MUSIC) [1,2], the robust MUSIC generalizations based on the empirical sign-covariance (SGN-MUSIC) [4] and Tyler's scatter M-estimator (TYLER-MUSIC) [5,8], to the RG-MUSIC algorithm [9,10], and to the approximate MLE (AMLE) [39, Algorithm 1]. The AMLE provides a conditional maximum-likelihood estimate of the basis vectors spanning the signal subspace. The DOAs are then estimated based on these basis vectors similarly to MUSIC. The MT-MUSIC$_{\text{sub}}$ (implemented with a Gaussian MT-function) and TYLER-MUSIC are implemented exactly as described in the third and fourth paragraphs of Section 6 in [16]. The scalar weight function $u(x)$ in the RG-MUSIC algorithm was set to $u(x)=(1+\alpha)/(\alpha+x)$ with $\alpha=0.2$. The variance parameter of the additional Gaussian noise component in the AMLE was set here to 0. In both RG-MUSIC and AMLE the number of iterations were set to 100.

We begin with description of the general settings under which the simulation studies were conducted. Under these settings, we then derive the extended Miller–Chang bound (EMCB) [36] that is used as a performance benchmark. Afterwards, we specify the parametric non-Gaussian MT-function used for implementation of the optimized MT-MUSIC and discuss its properties. The performance simulation studies are then described. Finally, we analyze the asymptotic computational load (ACL) of the optimized MT-MUSIC as compared to the other examined algorithms.

4.1. General settings

In all simulation examples that follow, we consider an i.i.d. signal process $\{\boldsymbol{s}_n\}$ that is comprised of 4-QAM signals with equal variance $\sigma^2_{\boldsymbol{s}}$. The steering vector $\mathbf{a}(\theta)\triangleq\left[1,e^{-i\pi\sin(\theta)},\ldots,e^{-i\pi(p-1)\sin(\theta)}\right]^T$ represents a uniform linear array (ULA) with half wavelength spacing corresponding to a far-field narrow band signal with $p=16$ elements. Note that under the ULA setting the parameter space $\Theta=[-\pi/2,\pi/2)$. In all compared algorithms $K_\Theta=10^5$ equally spaced samples of Θ were used in order to obtain the empirical pseudo-spectra. We considered a spherically symmetric CG noise that satisfies the following stochastic representation [5]:

$$\mathbf{w}_n=y_n\mathbf{z}_n, \tag{19}$$

where $y_n\triangleq\sigma_{\mathbf{w}}\tau_n$, $\sigma_{\mathbf{w}}\in\mathbb{R}_{++}$ is a scale parameter, $\{\tau_n\in\mathbb{R}_{++}\}$ is an i.i.d. process and $\{\mathbf{z}_n\in\mathbb{C}^p\}$ is a proper-complex i.i.d. Gaussian process with zero-mean and unit covariance. The processes $\{\tau_n\}$ and $\{\mathbf{z}_n\}$ are assumed to be statistically independent. We note that the family of CG-distributions encompasses some well-known distributions, such as the t-distribution and K-distribution that are widely used for modeling heavy-tailed radar clutter [31–35]. Based on the settings above, the generalized signal-to-noise-ratio (GSNR) is defined here as $\text{GSNR}\triangleq 10\log_{10}\sigma^2_{\boldsymbol{s}}/\sigma^2_{\mathbf{w}}$.

4.2. Performance benchmark

Under the stochastic array model (1) and the CG noise model (19), the estimation problem at hand involves a deterministic vector parameter of interest $\boldsymbol{\theta}$ and random nuisance parameters $\mathbf{S}\triangleq[\boldsymbol{s}_n,\ldots,\boldsymbol{s}_n]$ and $\mathbf{y}\triangleq[y_1,\ldots y_N]^T$, whom exact probability distributions are unspecified. Thus, in order to benchmark the performance of the compared algorithms, we use the extended Miller–Chang bound (EMCB) [36], which is a variant of the Cramér–Rao bound (CRB) [37] for the estimation of a deterministic parameter in the presence of random nuisance parameters. This bound is obtained in two steps. First, the conditional CRB for estimation of the deterministic parameter of interest is calculated using the conditional likelihood function of the observations given the nuisance parameters. The EMCB is then obtained by taking the expectation of the conditional CRB w.r.t. the nuisance parameters. Note that unlike the Miller–Chang bound [38] the conditional CRB associated with the EMCB is obtained under the assumption that the nuisance parameters are unknown. Using similar derivations reported in [26, Appendix E], it can be shown that under (1) and the i.i.d. CG noise (19), the conditional CRB for estimating $\boldsymbol{\theta}$, assuming that $\mathbf{S}$ and $\mathbf{y}$ are deterministic unknown is given by:

$$\mathbf{CRB}(\boldsymbol{\theta}|\mathbf{S},\mathbf{y})=\left(\sum_{n=1}^{N}2y_n^{-2}\text{Re}\left\{\boldsymbol{\Lambda}^H_{\boldsymbol{s}_n}\mathbf{B}^H(\boldsymbol{\theta})\mathbf{P}^{\perp}_{\mathbf{A}}\mathbf{B}(\boldsymbol{\theta})\boldsymbol{\Lambda}_{\boldsymbol{s}_n}\right\}\right)^{-1}, \tag{20}$$

where $\boldsymbol{\Lambda}_{\boldsymbol{s}_n}$ is a diagonal matrix whose main diagonal entries are given by the entires of the signal vector $\boldsymbol{s}_n$, i.e., $[\boldsymbol{\Lambda}_{\boldsymbol{s}_n}]_{i,i}=[\boldsymbol{s}_n]_i$, $i=1,\ldots,q$, and $\mathbf{B}(\boldsymbol{\theta})\triangleq[\dot{\mathbf{a}}(\theta_1),\ldots,\dot{\mathbf{a}}(\theta_q)]$. The vector $\dot{\mathbf{a}}(\vartheta)$ and the matrix $\mathbf{P}^{\perp}_{\mathbf{A}}$ are defined below (11). The EMCB is given by:

$$\mathbf{EMCB}(\boldsymbol{\theta})=\text{E}[\mathbf{CRB}(\boldsymbol{\theta}|\mathbf{S},\mathbf{y});P_{\mathbf{S},\mathbf{y}}]. \tag{21}$$

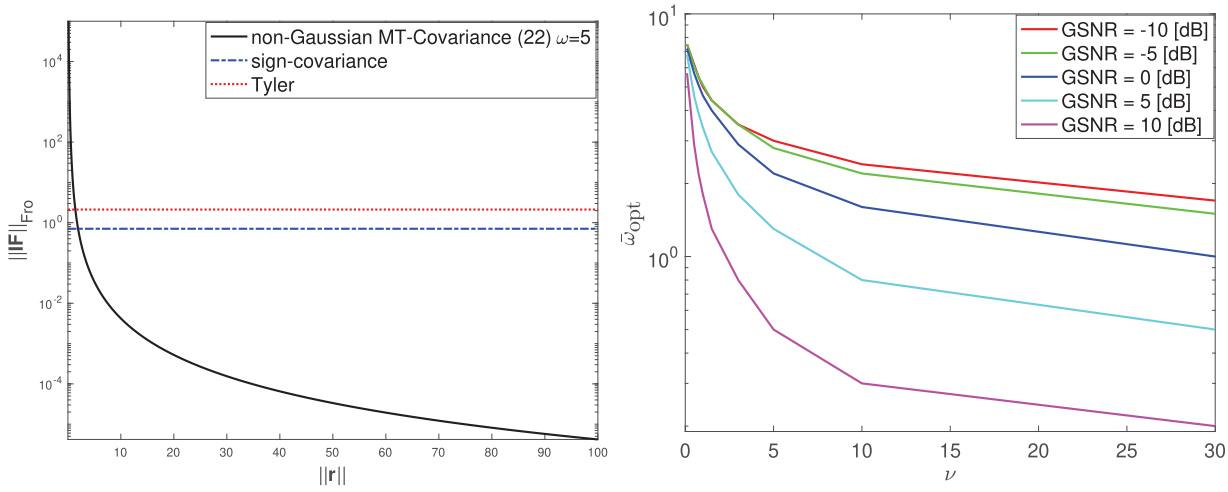
We note that the expectation in (21) was approximated by averaging (20) over 10^4 Monte–Carlo trials.

4.3. Implementation of the optimized MT-MUSIC with a non-Gaussian MT-function

Here, we implemented the optimized MT-MUSIC with the following non-Gaussian parametric MT-function:

$$u(\mathbf{x};\omega)\triangleq\|\mathbf{x}\|^{-\omega},\quad \omega\in\mathbb{R}_+. \tag{22}$$

Notice that when $\omega=0$ the resulting MT-covariance (3) coincides with the standard covariance. Furthermore, when $\omega=2$, one can verify that for centered symmetric distributions, the MT-covariance is a scaled version of the sign-covariance [4]. Also note that for $\omega\geq 2$, the MT-function (22) satisfies the robustness conditions stated in (5) over a sufficiently large subset of $\mathbb{C}^p$ that does not contain the origin. To see this, define the set $\mathcal{B}_\epsilon\triangleq\{\mathbf{r}\in\mathbb{C}^p:\|\mathbf{r}\|>\epsilon\}$, where $\epsilon>0$ is some fixed small positive constant. Clearly, $u(\mathbf{r};\omega)<\epsilon^{-\omega}$ and $u(\mathbf{r};\omega)\|\mathbf{r}\|^2\leq\epsilon^{2-\omega}$ for any $\mathbf{r}\in\mathcal{B}_\epsilon$ and any fixed $\omega\geq 2$. Thus, since $P_{\mathbf{x}}(\mathcal{B}_\epsilon)\approx 1$ for sufficiently small ϵ we conclude that the empirical MT-covariance is B-robust with high probability, i.e., condition A-1 is satisfied with high probability. Furthermore, similarly to the proof of Proposition 4 in [16], it can be shown that, unlike the empirical sign-covariance [4] and Tyler's M-estimator of scatter [8], for any fixed $\omega>2$ the influence function of the empirical MT-covariance [16, Eq. 17] associated with the non-Gaussian MT-function (22) approaches zero as the outlier norm approaches infinity. This property results in enhanced robustness to large-norm outliers. To further illustrate this point, in Fig. 1a, we depict the Frobenius norms of the matrix valued influence functions associated with the empirical MT-covariance under the non-Gaussian MT-function (22) with $\omega=5$, the empirical sign-covariance, and Tyler's M-estimator of scatter versus the norm of an outlier that contaminates a



(a) (b)

Fig. 1. Properties of the non-Gaussian MT-function (22): (a) Frobenius norms of the influence functions associated with the empirical MT-covariance under the non-Gaussian MT-function (22) with $\omega = 5$, the empirical sign covariance, and Tyler's M-estimator of scatter versus the norm of an outlier contaminating a bivariate proper complex standard normal distribution. (b) The optimal power parameter $\bar{\omega}_{\text{opt}}$ of the non-Gaussian MT-function (22) versus the shape parameter ν of a K-distributed noise for several values of GSNR ($q = 5$, $p = 16$, $N = 1000$).

bivariate proper complex standard normal distribution. Observing Fig. 1a, one can notice that, indeed, unlike the empirical sign-covariance and Tyler's M-estimator of scatter, the influence function associated with the empirical MT-covariance approaches zero as the outlier norm approaches infinity, indicating enhanced robustness to large-norm outliers. Finally, notice that the MT-function (22) is spherically contoured and strictly-positive. Hence, as shown in Appendix C, the key condition A-2, required for implementation of the MT-MUSIC, is satisfied under the CG noise (19).

Next, we study the effect of the heaviness of noise tails and the GSNR on the optimal power parameter $\bar{\omega}_{\text{opt}}$ of the MT-function (22) that minimizes the averaged asymptotic RMSE. This quantity is obtained by taking the arithmetic mean over the square roots of the diagonal terms comprising the asymptotic theoretical MSE matrix (12). Note that by (12) it follows that $\bar{\omega}_{\text{opt}}$ is not affected by the sample size N. Here, $q = 5$ statistically independent signals were considered that satisfy the assumptions stated in Section 4.1. The DOAs were set to $\theta_1 = -10^\circ$, $\theta_2 = 0^\circ$, $\theta_3 = 5^\circ$, $\theta_4 = 15^\circ$ and $\theta_5 = 25^\circ$ and $N = 1000$ snapshots were considered. Fig. 1b depicts $\bar{\omega}_{\text{opt}}$ for several GSNR values versus the shape parameter ν of a K-distributed noise that controls the heaviness of the tails. Note that as $\nu \to 0$ the tails of the K-distributed noise become heavier [5], and therefore, defective large-norm outliers become more frequent. On the other hand, as $\nu \to \infty$ the K-distribution approaches a light-tailed Gaussian distribution [5]. Observing Fig. 1b, one sees that $\bar{\omega}_{\text{opt}}$ increases with the decrease in ν. In other words, the optimal MT-function becomes narrower as the tails of the noise distribution become heavier. This result is intuitive since in the presence of heavy-tailed noise that produces large-norm outliers a narrow MT-function with rapid decay toward zero should be used in-order to effectively mitigate their effect on the estimation performance. Furthermore, in Fig. 1b one sees that for any fixed value of the shape parameter ν the optimal power parameter $\bar{\omega}_{\text{opt}}$ increases with the decrease in GSNR, i.e., the optimal MT-function becomes narrower as the noise becomes stronger relative to the signal. This result is intuitive since as the GSNR decreases, defective large-norm outliers become more likely, and therefore, a narrower MT-function with faster decay toward zero should be used to effectively mitigate their effect on the DOAs estimation performance.

4.4. Performance simulation studies

This subsection describes three simulation examples that were conducted to evaluate the performance of the optimized MT-MUSIC. All simulation examples were carried out under the general settings described in Section 4.1. Here, under the CG noise model (19) three types of noise distributions were examined: 1) Gaussian, 2) K-distributed noise, and 3) t-distributed noise.

4.4.1. Simulation example 1

We compared the asymptotic average RMSE (obtained by taking the arithmetic mean over the square roots of the diagonal terms comprising (12)) to its empirical estimate (18) as a function of the power parameter ω of the MT-function (22). Here, $q = 5$ statistically independent signals were considered that satisfy the signal assumptions described in Section 4.1. The DOAs were set to $\theta_1 = -10^\circ$, $\theta_2 = 0^\circ$, $\theta_3 = 5^\circ$, $\theta_4 = 15^\circ$ and $\theta_5 = 25^\circ$. The GSNR and the sample size N were set to -10 [dB] and 1000, respectively. The shape parameter of the K-distributed noise was set to $\nu = 1$. For the t-distributed noise $d = 0.25$ degrees of freedom were considered. Observing Figs. 2 a (Gaussian noise), 3 a (K-distributed noise) and 4 a (t-distributed noise) one sees that due to the consistency of (13), that follows from Proposition 1, the compared quantities are close. This illustrates the reliability of (18) for optimal choice of the MT-function parameter, as discussed in Section 3.2. Also note that for the heavy-tailed K-distributed and t-distributed noise, the asymptotic average RMSE is minimized for a larger value of ω, that corresponds to a narrower MT-function, as compared to the light-tailed Gaussian noise. As discussed in the second paragraph of Section 4.3, this result is intuitive since under the heavy-tailed noise a narrower MT-function with faster decay toward zero should be applied to effectively mitigate the effect of outliers.

4.4.2. Simulation example 2

Here, we compared the empirical average RMSE of the optimized MT-MUSIC to those obtained by the other compared

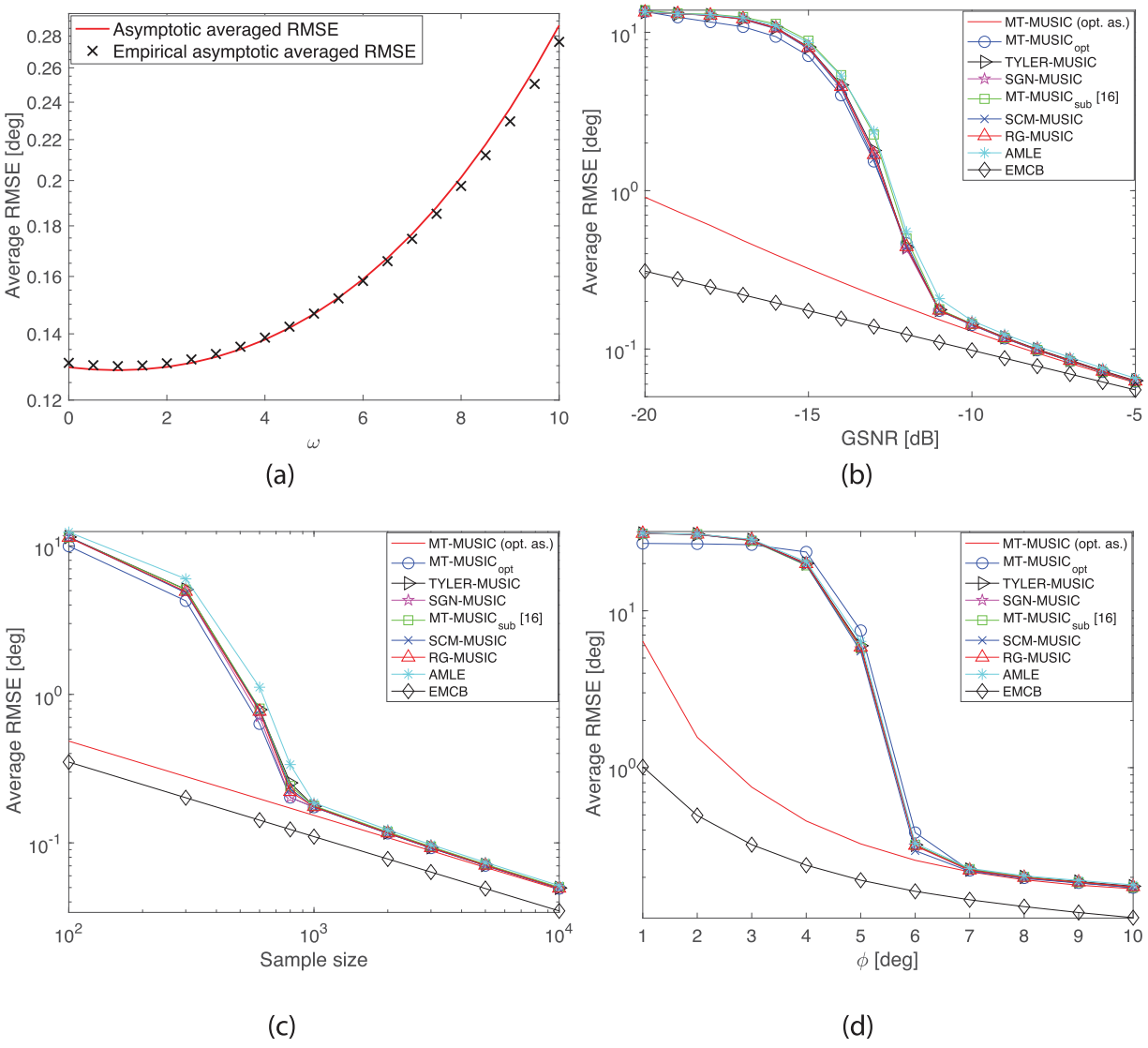


Fig. 2. Localization in Gaussian noise: (a) Asymptotic average RMSE obtained from (12) and its empirical estimate (18) versus the power parameter ω of the non-Gaussian MT-function (22) ($q=5$, $p=16$, GSNR $=-10$ [dB], $N=1000$). Notice that due to the consistency of (13), the compared quantities are close. (b)–(d) Optimum asymptotic (opt. as.) average RMSE of the MT-MUSIC obtained from (12), and the empirical average RMSE of the optimized MT-MUSIC as a function of (b) GSNR ($q=5$, $p=16$, $N=1000$), (c) sample size N ($q=5$, $p=16$, GSNR $=-11$ [dB]) and (d) separation angle ϕ ($q=2$, $p=16$, GSNR $=-11$ [dB], $N=1000$), as compared to the empirical average RMSEs of the MT-MUSIC$_{\text{sub}}$ [16], TYLER-MUSIC, SGN-MUSIC, RG-MUSIC, AMLE, and the standard SCM-MUSIC. The averaged RMSE associated with the EMCB (21) is used here as a performance benchmark.

estimators versus GSNR and samples size N. For the K-distributed and the t-distributed noise the localization performance were also compared versus the shape parameter ν and the degrees of freedom d, respectively, that control the heaviness of the tails. Similarly to the previous example, $q=5$ statistically independent signals were considered (that satisfy the general signal assumptions in Section 4.1) and the DOAs were set to $\theta_1=-10^\circ$, $\theta_2=0^\circ$, $\theta_3=5^\circ$, $\theta_4=15^\circ$ and $\theta_5=25^\circ$. All empirical average RMSE curves were obtained via 10^4 Monte-Carlo trials.

The optimal power parameter ω_{opt} of the MT-function $u(\mathbf{x};\omega)$ (22) was obtained by minimizing the objective function $J_u(\omega)$ (18) over $K_\Omega=17$ equally spaced grid points of the interval $\Omega\triangleq[2,\,10]$. We note that the lower endpoint of the interval Ω was set to 2 since following the discussion below Eq. (22) the empirical MT-covariance is B-robust with high probability for any $\omega\geq 2$. We further note that the parameter K_Ω controls the trade-off between the accuracy in finding the global minimum of $J_u(\omega)$ (over the interval Ω) and the computational load of the optimized MT-MUSIC detailed in Section 4.5. Clearly, as K_Ω increases the sampling of the interval Ω becomes denser and ω_{opt} better approximate the minimum point of $J_u(\omega)$. However, as follows from Table 1, this comes at the expense of increased computational load. On the other hand, decreasing the value of K_Ω results in a decrease in the computa-

Table 1
Asymptotic computational load (flops). Notation: p is the dimension of the observation vectors. N denotes the sample size. K_Θ denotes the number of grid points of Θ. K_Ω denotes the number of grid points of the Ω-AXIS (the parameter space of the non-Gaussian MT-function (22). I denotes number of iterations.

Method	Asymptotic computational load (flops)
MT-MUSIC$_{\text{opt}}$	$O((Np^2+p^3+K_\Theta)K_\Omega)$
MT-MUSIC$_{\text{sub}}$[16]	$O(Np^2I+p^3+K_\Theta)$
SCM-MUSIC	$O(Np^2+p^3+K_\Theta)$
Tyler-MUSIC	$O(Np^2I+p^3+K_\Theta)$
SGN-MUSIC	$O(Np^2+p^3+K_\Theta)$
RG-MUSIC	$O((Np^2+p^3)I+K_\Theta)$
AMLE	$O((Np^2+p^3)I+K_\Theta)$

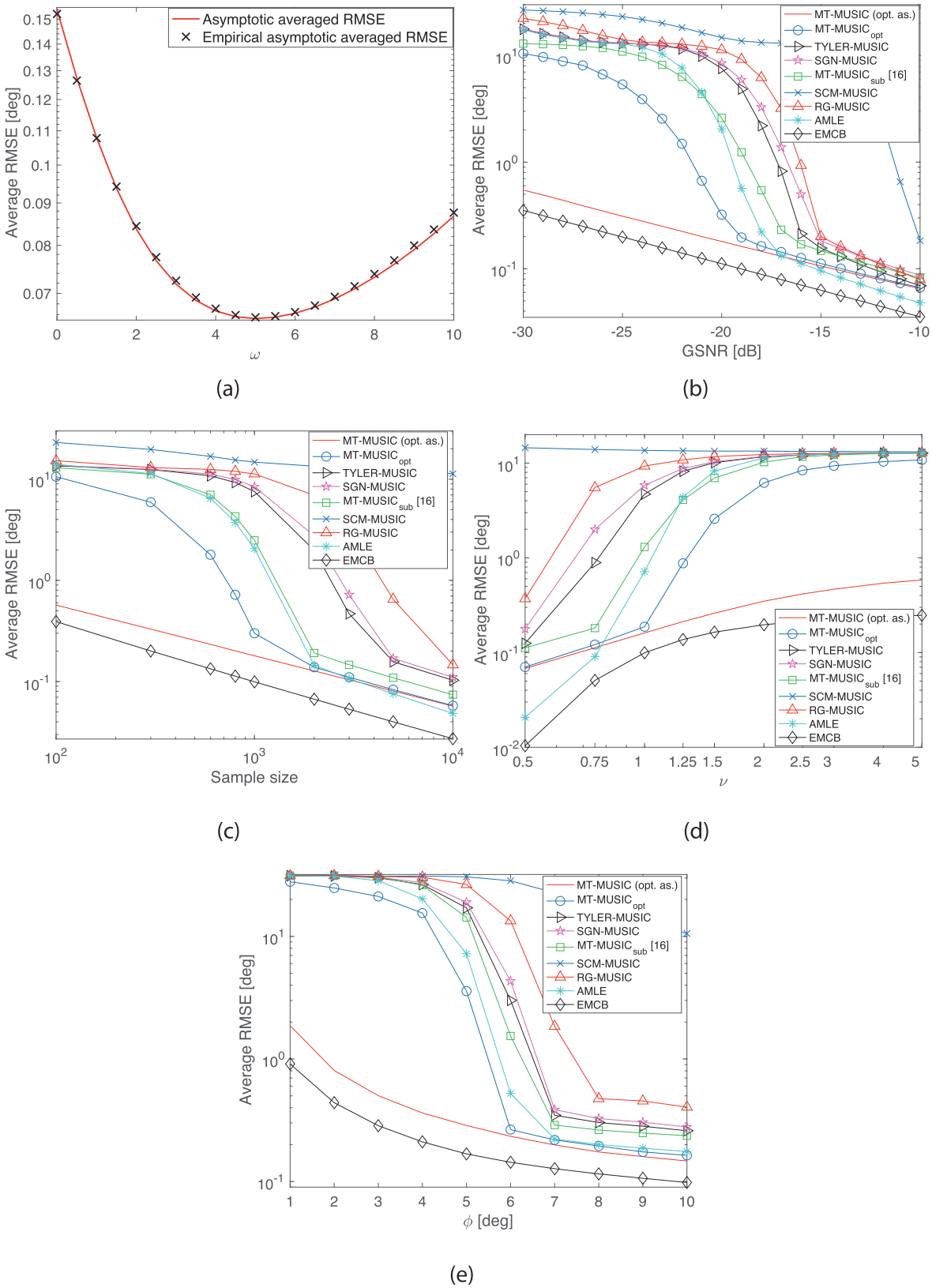


Fig. 3. Localization in K-distributed noise: (a) Asymptotic average RMSE obtained from (12) and its empirical estimate (18) versus the power parameter ω of the non-Gaussian MT-function (22) ($q = 5,\ p = 16,$ GSNR $= -10$ [dB], $N = 1000,\ \nu = 1$). Notice that due to the consistency of (13), the compared quantities are close. (b)-(e) Optimum asymptotic (opt. as.) average RMSE of the MT-MUSIC obtained from (12), and the empirical average RMSE of the optimized MT-MUSIC as a function of (b) GSNR ($q = 5,\ p = 16,\ N = 1000,\ \nu = 1$), (c) sample size N ($q = 5,\ p = 16,$ GSNR $= -20$ [dB], $\nu = 1$), (d) shape parameter ν ($q = 5,\ p = 16,$ GSNR $= -20$ [dB], $N = 1000$), and (e) separation angle ϕ ($q = 2,\ p = 16,$ GSNR $= -20$ [dB], $N = 1000,\ \nu = 1$), as compared to the empirical average RMSEs of the MT-MUSIC$_{\text{sub}}$ [16], TYLER-MUSIC, SGN-MUSIC, RG-MUSIC, AMLE, and the standard SCM-MUSIC. The averaged RMSE associated with the EMCB (21) is used here as a performance benchmark.

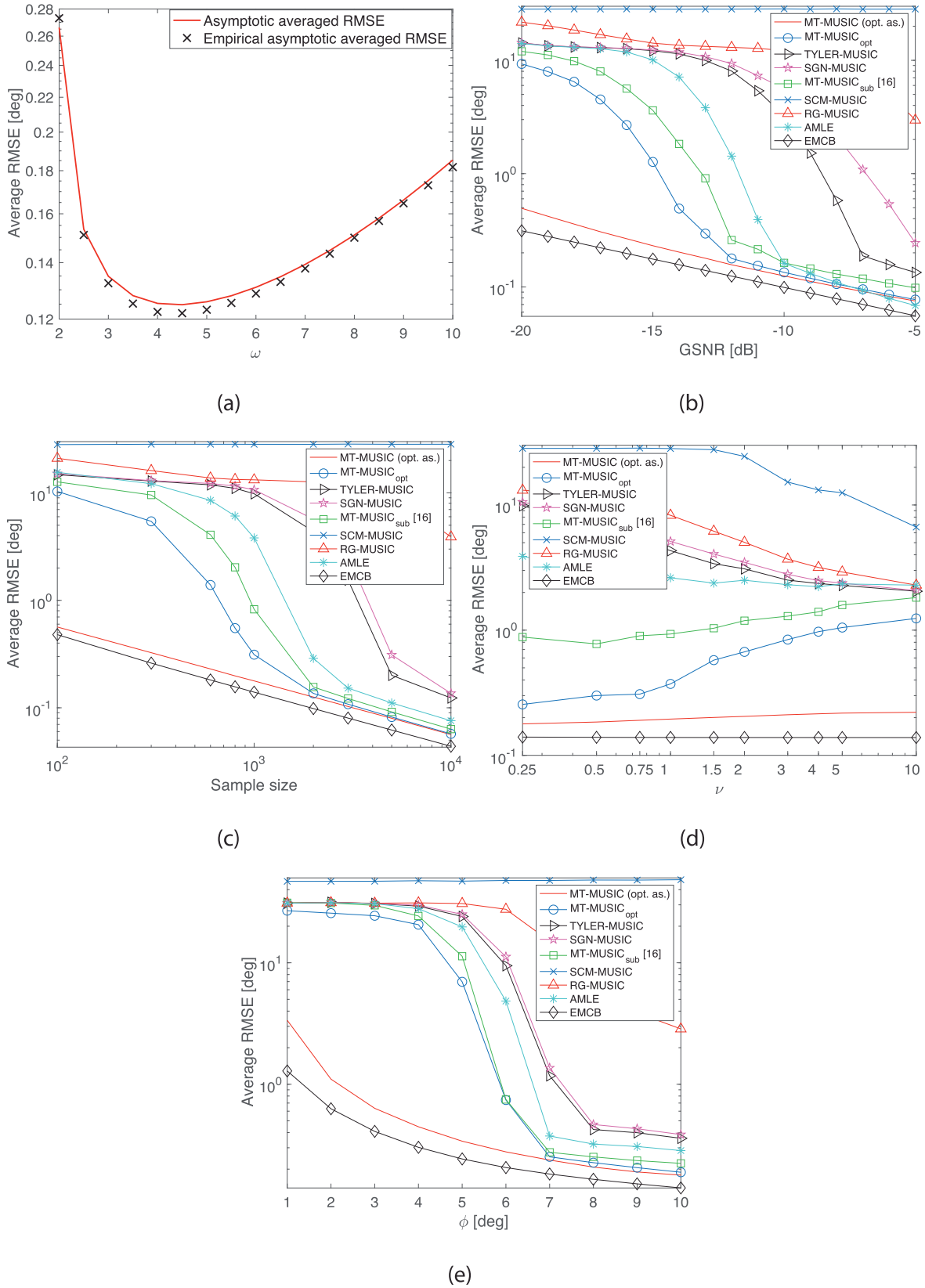


Fig. 4. Localization in t-distributed noise: (a) Asymptotic average RMSE obtained from (12) and its empirical estimate (18) versus the power parameter ω of the non-Gaussian MT-function (22) ($q = 5,\ p = 16,\ \mathrm{GSNR} = -10$ [dB], $N = 1000,\ d = 0.25$). Notice that due to the consistency of (13), the compared quantities are close. (b)–(e) Optimum asymptotic (opt. as.) average RMSE of the MT-MUSIC obtained from (12), and the empirical average RMSE of the optimized MT-MUSIC as a function of (b) GSNR ($q = 5,\ p = 16,\ N = 1000,\ d = 0.25$), (c) sample size N ($q = 5,\ p = 16,\ \mathrm{GSNR} = -13$ [dB], $d = 0.25$), (d) degrees of freedom d ($q = 5,\ p = 16,\ \mathrm{GSNR} = -13$ [dB], $N = 1000$), and (e) separation angle ϕ ($q = 2,\ p = 16,\ \mathrm{GSNR} = -13$ [dB], $N = 1000,\ d = 0.25$), as compared to the empirical average RMSEs of the MT-MUSIC$_{\mathrm{sub}}$ [16], TYLER-MUSIC, SGN-MUSIC, RG-MUSIC, AMLE, and the standard SCM-MUSIC. The averaged RMSE associated with the EMCB (21) is used here as a performance benchmark.

tional complexity at the possible expense of less accurate approximation of the minimum point of $J_u(\omega)$.

For each type of comparison, we also report the optimal asymptotic average RMSE of the MT-MUSIC for the considered MT-function (22). This quantity was obtained by minimizing the arithmetic mean of the square roots of the diagonal terms comprising (12) w.r.t. the power parameter $\omega \in \Omega$ in (22). Furthermore, the averaged RMSE associated with the EMCB matrix (21) was used as a performance benchmark. This quantity was obtained by taking the arithmetic mean over the square roots of the diagonal terms comprising (12).

In the following we detail the specific settings under each type of comparison. In the GSNR analysis, the sample size was set to $N = 1000$, the shape parameter of the K-distributed noise was set to $\nu = 1$, and a t-distributed noise with $d = 0.25$ degrees of freedom was considered. In the sample size analysis, the GSNR was set to -11 [dB], -19 [dB], and -13 [dB] for the Gaussian, K-distributed noise and t-distributed noise, respectively. The shape parameter ν of the K-distributed noise and the degrees of freedom d of the t-distributed were set to the same values as in the GSNR analysis. The GSNR and sample size in the localization performance versus the shape parameter ν of the K-distributed noise were set to -19 [dB] and $N = 1000$, respectively. The localization performance versus the degrees of freedom d of the t-distributed noise were evaluated for GSNR $= -13$ [dB] and sample size $N = 1000$.

Observing Fig. 2b and c one can notice that for the Gaussian noise, all compared estimators attain similar performance. Observing Figs. 3b–d and 4b–d, one sees that for the heavy-tailed K-distributed and t-distributed noise, the optimized MT-MUSIC can gain significant performance advantage over the non-robust SCM-MUSIC and the other robust alternatives, and can considerably reduce the gap towards the EMCB. This performance advantage follows from the following properties. First, as discussed in Section 4.3, the empirical MT-covariance associated with the MT-function (22) can gain enhanced robustness to large-norm outliers as compared to Tyler's and the sign-covariance estimators. Second, unlike all compared methods, the proposed approach involves optimization of a consistent estimate of the asymptotic average RMSE of the DOAs estimates.

4.4.3. Simulation example 3

In this example, we perform a source resolution analysis for $q = 2$ correlated 4-QAM signals with covariance matrix

$$\mathbf{\Sigma_S} \triangleq \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \sigma_S^2,$$

where $\rho = 0.4$ is the Pearson correlation coefficient of the two signals. Similarly to the two previous simulation examples, we considered a 16-elements ULA and the same three types of noise distributions, namely, Gaussian noise, K-distributed noise with shape parameter set to $\nu = 1$ and a t-distributed noise with $d = 0.25$ degrees of freedom. The DOAs were set to $\theta_1 = 30^\circ$ and $\theta_2 = \theta_1 + \phi$, where the separation angle parameter ϕ is used to index the localization performance measured in terms of empirical average RMSEs that were obtained via 10^4 Monte-Carlo trials. The optimal MT-function parameter ω_{opt} was selected as in the previous example. Furthermore, similarly to the previous example, the optimal asymptotic average RMSE of the MT-MUSIC (under the MT-function (22)) and the EMCB are also reported. Here, the sample size was set to $N = 1000$, and the GSNRs were set to -10 [dB], -19 [dB] and -13 [dB] for the Gaussian, K-distributed and t-distributed noise, respectively. Observing Fig. 2d one can notice that for the Gaussian noise all compared estimators perform similarly. In Figs. 3e and 4e one sees that when the noise is heavy-tailed, the optimized MT-MUSIC outperforms all other compared algorithms. The performance advantage follows from the same reasons detailed in the previous example.

4.5. Computational load

A general asymptotic computational load (ACL) analysis is reported in Table 1. Observing Table 1, one can notice that the ACL of the optimized MT-MUSIC is linear in the number of grid points K_Ω of the parameter space associated with the non-Gaussian MT-function (22). Furthermore, one sees that similarly to the non-robust SCM-MUSIC the ACL is cubic in the dimension p, and linear in the sample size N as well as in the number of grid points K_Θ of the parameter space Θ.

5. Conclusion

This paper provides asymptotic MSE analysis of the MT-MUSIC algorithm [16]. Based on this analysis, a data-driven procedure for optimal selection of the MT-function parameters was developed that minimizes an empirical estimate of the asymptotic average RMSE. By specifying the MT-function in a class of non-Gaussian outlier-suppressing MT-functions, the optimized MT-MUSIC demonstrated significant performance advantage over the suboptimal version [16] and other state-of-the-art robust MUSIC alternatives.

Conflict of interest

None.

Acknowledgment

This work was supported in part by the United States-Israel Binational Science Foundation under Grant 2014334.

Appendix

A. Proof of Theorem 1

The proof of Theorem 1 is based on the following auxiliary definition and Lemmas:

Definition 2. Let $\lambda_1^{(u)} \geq \ldots \geq \lambda_p^{(u)}$ and $\hat{\lambda}_1^{(u)} \geq \ldots \geq \hat{\lambda}_p^{(u)}$ denote the eigenvalues of $\mathbf{\Sigma}_{\mathbf{x}}^{(u)}$ (3) and $\hat{\mathbf{\Sigma}}_{\mathbf{x}}^{(u)}$ (4), respectively. Furthermore, let $\mathbf{v}_1^{(u)}, \ldots, \mathbf{v}_p^{(u)}$ and $\hat{\mathbf{v}}_1^{(u)}, \ldots, \hat{\mathbf{v}}_p^{(u)}$ denote the unit-norm eigenvectors of $\mathbf{\Sigma}_{\mathbf{x}}^{(u)}$ and $\hat{\mathbf{\Sigma}}_{\mathbf{x}}^{(u)}$, respectively, corresponding to $\lambda_1^{(u)}, \ldots, \lambda_p^{(u)}$ and $\hat{\lambda}_1^{(u)}, \ldots, \hat{\lambda}_p^{(u)}$. We define the following matrices:

$$\mathbf{G}^{(u)} \triangleq \left[\mathbf{v}_{q+1}^{(u)}, \ldots, \mathbf{v}_p^{(u)}\right] \quad \text{and} \quad \hat{\mathbf{G}}^{(u)} \triangleq \left[\hat{\mathbf{v}}_{q+1}^{(u)}, \ldots, \hat{\mathbf{v}}_p^{(u)}\right]. \tag{23}$$

Lemma 1 (Strong consistency of the MT-MUSIC)**.** *Assume that the expectation* $\mathrm{E}[\|\mathbf{x}\|^2 u(\mathbf{x}); P_{\mathbf{x}}] < \infty$ *and condition B-2 is satisfied. Then, the MT-MUSIC estimator is strongly consistent, i.e.,*

$$\hat{\boldsymbol{\theta}}_u \xrightarrow{\text{w.p.1}} \boldsymbol{\theta} \quad \text{as} \quad N \to \infty. \tag{24}$$

Proof of Lemma 1. Since the observations $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are i.i.d., then by Proposition 2 in [16], and the assumption that $\mathrm{E}[\|\mathbf{x}\|^2 u(\mathbf{x}); P_{\mathbf{x}}] < \infty$ we conclude that $\hat{\mathbf{\Sigma}}_{\mathbf{x}}^{(u)} \xrightarrow[N\to\infty]{\text{w.p.1}} \mathbf{\Sigma}_{\mathbf{x}}^{(u)}$. Furthermore, by Theorem 8 in [44, page 183] the eigenvectors comprising $\hat{\mathbf{G}}^{(u)}$ (23), define continuous mappings of $\hat{\mathbf{\Sigma}}_{\mathbf{x}}^{(u)}$. Hence, by Mann–Wald's Theorem [45], $\hat{\mathbf{G}}^{(u)} \xrightarrow[N\to\infty]{\text{w.p.1}} \mathbf{G}^{(u)}$, where $\mathbf{G}^{(u)}$ is defined in (23). Therefore, using the matrix version of Cauchy–Schwarz inequality [50] and

assumption B-2, according to which $\|\mathbf{a}(\vartheta)\| \le c$ for some $c \in \mathbb{R}_{++}$ we obtain that

$$\begin{aligned}&\sup_{\vartheta\in\Theta}\left|\mathbf{a}^H(\vartheta)\hat{\mathbf{G}}^{(u)}\hat{\mathbf{G}}^{(u)H}\mathbf{a}(\vartheta)-\mathbf{a}^H(\vartheta)\mathbf{G}^{(u)}\mathbf{G}^{(u)H}\mathbf{a}(\vartheta)\right|\\&=\sup_{\vartheta\in\Theta}\left|\mathrm{trace}\left\{\left(\hat{\mathbf{G}}^{(u)}\hat{\mathbf{G}}^{(u)H}-\mathbf{G}^{(u)}\mathbf{G}^{(u)H}\right)\mathbf{a}(\vartheta)\mathbf{a}^H(\vartheta)\right\}\right|\\&\le c\left\|\hat{\mathbf{G}}^{(u)}\hat{\mathbf{G}}^{(u)H}-\mathbf{G}^{(u)}\mathbf{G}^{(u)H}\right\|_{\mathrm{F}}\xrightarrow{\text{w.p.1}}0,\quad\text{as}\quad N\to\infty,\end{aligned}\tag{25}$$

where $\mathrm{trace}\{\cdot\}$ and $\|\cdot\|_{\mathrm{F}}$ denote the trace operator and the Frobenius norm [51], respectively. Eq. (25) implies that the pseudo-spectrum of the MT-MUSIC is uniformly consistent. Hence, the proof is complete by performing the same steps below Eq. (5) in the consistency proof of the standard MUSIC algorithm provided in [52]. □

Lemma 2 (Weak convergence of the empirical MT-covariance)**.** *Let* $\boldsymbol{\sigma} \triangleq \mathrm{vec}\left\{\boldsymbol{\Sigma}_{\mathbf{x}}^{(u)}\right\}$ *and* $\hat{\boldsymbol{\sigma}} \triangleq \mathrm{vec}\left\{\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{(u)}\right\}$*, where* $\mathrm{vec}\{\cdot\}$ *denotes the vectorization operator. Assume that the observations* $\mathbf{x}_1,\ldots,\mathbf{x}_n$ *are i.i.d and the expectation* $\mathrm{E}[\|\mathbf{x}\|^4u^2(\mathbf{x});\,P_{\mathbf{x}}]$ *is finite. Then,*

$$\sqrt{N}\left(\hat{\boldsymbol{\sigma}}-\boldsymbol{\sigma}\right)\xrightarrow{\mathrm{D}}\mathcal{CN}(\mathbf{0},\boldsymbol{\Gamma},\mathbf{C})\quad\text{as}\quad N\to\infty,\tag{26}$$

where the covariance $\boldsymbol{\Gamma}\triangleq\mathbf{A}_{1,1}+\mathbf{A}_{2,2}+i(\mathbf{A}_{2,1}-\mathbf{A}_{1,2})$*, the pseudo-covariance* $\mathbf{C}\triangleq\mathbf{A}_{1,1}-\mathbf{A}_{2,2}+i(\mathbf{A}_{2,1}+\mathbf{A}_{1,2})$*, and* $\mathbf{A}_{k,l}\triangleq\mathbf{D}_k\boldsymbol{\Sigma}_{\mathbf{y}}\mathbf{D}_j^T$*, $k,l=1,2$. The matrix* $\boldsymbol{\Sigma}_{\mathbf{y}}$ *is the covariance matrix of a real-valued random vector* $\mathbf{y}\triangleq u(\mathbf{x})[(\mathbf{x}_R\otimes\mathbf{x}_R)^T,(\mathbf{x}_I\otimes\mathbf{x}_I)^T,(\mathbf{x}_I\otimes\mathbf{x}_R)^T,(\mathbf{x}_R\otimes\mathbf{x}_I)^T,\mathbf{x}_R^T,\mathbf{x}_I^T,1]^T$*, where* $\mathbf{x}_R\triangleq\mathrm{Re}\{\mathbf{x}\}$*,* $\mathbf{x}_I\triangleq\mathrm{Im}\{\mathbf{x}\}$ *and* $\otimes$ *denotes the Kronecker product. The matrices* $\mathbf{D}_1$ *and* $\mathbf{D}_2$ *are block-matrices with* $\mathbf{D}_1\triangleq[\mathbf{D}_{1,1},\ \mathbf{D}_{1,2},\ \mathbf{D}_{1,3},\ \mathbf{D}_{1,4},\ \mathbf{D}_{1,5},\ \mathbf{D}_{1,6},\ \mathbf{D}_{1,7}]$ *and* $\mathbf{D}_2\triangleq[\mathbf{D}_{2,1},\ \mathbf{D}_{2,2},\ \mathbf{D}_{2,3},\ \mathbf{D}_{2,4},\ \mathbf{D}_{2,5},\ \mathbf{D}_{2,6},\ \mathbf{D}_{2,7}]$*, respectively. The blocks comprising* $\mathbf{D}_1$ *are given by* $\mathbf{D}_{1,1}=\mathbf{D}_{1,2}\triangleq c\mathbf{I}_{p^2}$*,* $\mathbf{D}_{1,3}=\mathbf{D}_{1,4}\triangleq\mathbf{0}_{p^2}$*,* $\mathbf{D}_{1,5}\triangleq-c(\mathbf{I}_p\otimes\mathrm{Re}\{\boldsymbol{\mu}_{\mathbf{x}}^{(u)}\}+\mathrm{Re}\{\boldsymbol{\mu}_{\mathbf{x}}^{(u)}\}\otimes\mathbf{I}_p)$*,* $\mathbf{D}_{1,6}\triangleq-c(\mathbf{I}_p\otimes\mathrm{Im}\{\boldsymbol{\mu}_{\mathbf{x}}^{(u)}\}+\mathrm{Im}\{\boldsymbol{\mu}_{\mathbf{x}}^{(u)}\}\otimes\mathbf{I}_p)$*,* $\mathbf{D}_{1,7}\triangleq-c(\mathrm{vec}\{\mathrm{Re}\{\boldsymbol{\Sigma}_{\mathbf{x}}^{(u)}\}\}-\mathrm{Re}\{\boldsymbol{\mu}_{\mathbf{x}}^{(u)}\}\otimes\mathrm{Re}\{\boldsymbol{\mu}_{\mathbf{x}}^{(u)}\}-\mathrm{Im}\{\boldsymbol{\mu}_{\mathbf{x}}^{(u)}\}\otimes\mathrm{Im}\{\boldsymbol{\mu}_{\mathbf{x}}^{(u)}\})$*, where* $\mathbf{I}_p$ *denotes a* $p\times p$ *identity matrix,* $\mathbf{0}_p$ *is a* $p\times p$ *matrix with zero entries, and* $c\triangleq\mathrm{E}^{-1}[u(\mathbf{x});P_{\mathbf{x}}]$*. The blocks of* $\mathbf{D}_2$ *take the forms* $\mathbf{D}_{2,1}=\mathbf{D}_{2,2}\triangleq\mathbf{0}_{p^2}$*,* $\mathbf{D}_{2,3}\triangleq-c\mathbf{I}_{p^2}$*,* $\mathbf{D}_{2,4}\triangleq c\mathbf{I}_{p^2}$*,* $\mathbf{D}_{2,5}\triangleq-c(\mathbf{I}_p\otimes\mathrm{Im}\{\boldsymbol{\mu}_{\mathbf{x}}^{(u)}\}-\mathrm{Im}\{\boldsymbol{\mu}_{\mathbf{x}}^{(u)}\}\otimes\mathbf{I}_p)$*,* $\mathbf{D}_{2,6}\triangleq-c(\mathbf{I}_p\otimes\mathrm{Re}\{\boldsymbol{\mu}_{\mathbf{x}}^{(u)}\}-\mathrm{Re}\{\boldsymbol{\mu}_{\mathbf{x}}^{(u)}\}\otimes\mathbf{I}_p)$*,* $\mathbf{D}_{2,7}\triangleq-c(\mathrm{vec}\{\mathrm{Im}\{\boldsymbol{\Sigma}_{\mathbf{x}}^{(u)}\}\}+\mathrm{Im}\{\boldsymbol{\mu}_{\mathbf{x}}^{(u)}\}\otimes\mathrm{Re}\{\boldsymbol{\mu}_{\mathbf{x}}^{(u)}\}-\mathrm{Re}\{\boldsymbol{\mu}_{\mathbf{x}}^{(u)}\}\otimes\mathrm{Im}\{\boldsymbol{\mu}_{\mathbf{x}}^{(u)}\})$*.*

Proof of Lemma 2. Define $\boldsymbol{\xi}\triangleq[\mathrm{Re}\{\boldsymbol{\sigma}\}^T,\mathrm{Im}\{\boldsymbol{\sigma}\}^T]^T$ and $\hat{\boldsymbol{\xi}}\triangleq[\mathrm{Re}\{\hat{\boldsymbol{\sigma}}\}^T,\mathrm{Im}\{\hat{\boldsymbol{\sigma}}\}^T]^T$. Assume that

$$\sqrt{N}\left(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi}\right)\xrightarrow{\mathrm{D}}\mathcal{N}(\mathbf{0},\mathbf{D}\boldsymbol{\Sigma}_{\mathbf{y}}\mathbf{D}^T)\quad\text{as}\quad N\to\infty,\tag{27}$$

where $\mathbf{D}\triangleq[\mathbf{D}_1^T,\mathbf{D}_2^T]^T$. Then (26) directly follows from (27) by the relation between the pseudo-covariance and covariance matrices of a complex random vector and the covariance matrix of its complex-to-real mapping [41]. Therefore, in-order to complete the proof, we need to show that (27) holds.

By the definition of the empirical MT-covariance (4), it follows that

$$\mathrm{Re}\{\hat{\boldsymbol{\sigma}}\}=\frac{\mathbf{z}_1}{z_7}-\frac{\mathbf{z}_5\otimes\mathbf{z}_5}{z_7^2}+\frac{\mathbf{z}_2}{z_7}-\frac{\mathbf{z}_6\otimes\mathbf{z}_6}{z_7^2}\triangleq\mathbf{h}_R(\mathbf{z})\tag{28}$$

and

$$\mathrm{Im}\{\hat{\boldsymbol{\sigma}}\}=\frac{\mathbf{z}_4}{z_7}-\frac{\mathbf{z}_5\otimes\mathbf{z}_6}{z_7^2}-\frac{\mathbf{z}_3}{z_7}+\frac{\mathbf{z}_6\otimes\mathbf{z}_5}{z_7^2}\triangleq\mathbf{h}_I(\mathbf{z}),\tag{29}$$

where the statistics $\mathbf{z}_1\triangleq\frac{1}{N}\sum_{n=1}^N u(\mathbf{x}_n)\mathbf{x}_{Rn}\otimes\mathbf{x}_{Rn}$, $\mathbf{z}_2\triangleq\frac{1}{N}\sum_{n=1}^N u(\mathbf{x}_n)\mathbf{x}_{In}\otimes\mathbf{x}_{In}$, $\mathbf{z}_3\triangleq\frac{1}{N}\sum_{n=1}^N u(\mathbf{x}_n)\mathbf{x}_{In}\otimes\mathbf{x}_{Rn}$, $\mathbf{z}_4\triangleq\frac{1}{N}\sum_{n=1}^N u(\mathbf{x}_n)\mathbf{x}_{Rn}\otimes\mathbf{x}_{In}$, $\mathbf{z}_5\triangleq\frac{1}{N}\sum_{n=1}^N u(\mathbf{x}_n)\mathbf{x}_{Rn}$, $\mathbf{z}_6\triangleq\frac{1}{N}\sum_{n=1}^N u(\mathbf{x}_n)\mathbf{x}_{In}$, $z_7\triangleq\frac{1}{N}\sum_{n=1}^N u(\mathbf{x}_n)$, $\mathbf{x}_{Rn}\triangleq\mathrm{Re}\{\mathbf{x}_n\}$, $\mathbf{x}_{In}\triangleq\mathrm{Im}\{\mathbf{x}_n\}$ and $\mathbf{z}\triangleq[\mathbf{z}_1^T,\mathbf{z}_2^T,\mathbf{z}_3^T,\mathbf{z}_4^T,\mathbf{z}_5^T,\mathbf{z}_6^T,z_7]^T$. Therefore, by (28) and (29) we conclude that $\hat{\boldsymbol{\xi}}$ can be represented as a differential vector function of the statistic $\mathbf{z}$, i.e.,

$$\hat{\boldsymbol{\xi}}=[\mathbf{h}_R^T(\mathbf{z}),\mathbf{h}_I^T(\mathbf{z})]^T\triangleq\mathbf{h}(\mathbf{z}).\tag{30}$$

Furthermore, using the definition of the MT-covariance (3) one can verify that

$$\boldsymbol{\xi}=\mathbf{h}(\boldsymbol{\mu}_{\mathbf{z}}),\tag{31}$$

where $\boldsymbol{\mu}_{\mathbf{z}}\triangleq E[\mathbf{z};P_{\mathbf{z}}]$.

In the following, we find the asymptotic distribution of $\sqrt{N}(\mathbf{z}-\boldsymbol{\mu}_{\mathbf{z}})$. By the definition of $\mathbf{z}$ above, it follows that $\mathbf{z}=\frac{1}{N}\sum_{n=1}^N\mathbf{y}_n$, where $\mathbf{y}_n\triangleq u(\mathbf{x}_n)[(\mathbf{x}_{Rn}\otimes\mathbf{x}_{Rn})^T,(\mathbf{x}_{In}\otimes\mathbf{x}_{In})^T,(\mathbf{x}_{In}\otimes\mathbf{x}_{Rn})^T,(\mathbf{x}_{Rn}\otimes\mathbf{x}_{In})^T,\mathbf{x}_{Rn}^T,\mathbf{x}_{In}^T,1]^T$. Notice that $\{\mathbf{y}_n\}$ is a sequence of real-valued i.i.d. random vectors with mean $\boldsymbol{\mu}_{\mathbf{y}}\triangleq E[\mathbf{y};P_{\mathbf{y}}]=\boldsymbol{\mu}_{\mathbf{z}}$ and covariance $\boldsymbol{\Sigma}_{\mathbf{y}}$. Since $\mathrm{E}[\|\mathbf{x}\|^4u^2(\mathbf{x});\,P_{\mathbf{x}}]<\infty$ it can be shown using Hölder's inequality [25] that $\boldsymbol{\Sigma}_{\mathbf{y}}$ must take finite values. Therefore, by the central limit theorem [46] we conclude that

$$\sqrt{N}(\mathbf{z}-\boldsymbol{\mu}_{\mathbf{z}})\xrightarrow{\mathrm{D}}\mathcal{N}(\mathbf{0},\boldsymbol{\Sigma}_{\mathbf{y}})\quad\text{as}\quad N\to\infty.\tag{32}$$

Hence, by (28)–(32) and [43, Theorem 11.2.14] (delta-method) is follows that (27) holds, with $\mathbf{D}\triangleq[\mathbf{D}_1^T,\mathbf{D}_2^T]^T$, $\mathbf{D}_1=\frac{d\mathbf{h}_R(\mathbf{z})}{d\mathbf{z}}\Big|_{\mathbf{z}=\boldsymbol{\mu}_{\mathbf{z}}}$ and $\mathbf{D}_2=\frac{d\mathbf{h}_I(\mathbf{z})}{d\mathbf{z}}\Big|_{\mathbf{z}=\boldsymbol{\mu}_{\mathbf{z}}}$. □

Lemma 3. *Let* $\mathbf{a}_1,\ldots,\mathbf{a}_q,\mathbf{b}_1,\ldots,\mathbf{b}_q\in\mathbb{C}^p$ *such that* $\mathbf{a}_k^H\boldsymbol{\Sigma}_{\mathbf{x}}^{(u)}\mathbf{b}_k=0$*,* $\forall k=1,\ldots,q$*. Define a random vector* $\mathbf{y}\in\mathbb{C}^q$ *such that* $[\mathbf{y}]_k\triangleq\mathrm{Re}\{\mathbf{a}_k^H\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{(u)}\mathbf{b}_k\}$*. Furthermore, assume that* $\boldsymbol{\mu}_{\mathbf{x}}^{(u)}=\mathbf{0}$ *and* $\mathrm{E}[\|\mathbf{x}\|^4u^2(\mathbf{x});\,P_{\mathbf{x}}]$ *is finite. Then,*

$$\sqrt{N}\mathbf{y}\xrightarrow{\mathrm{D}}\mathcal{N}(\mathbf{0},\mathbf{D}_{\mathbf{y}})\quad\text{as}\quad N\to\infty,\tag{33}$$

where $\mathbf{D}_{\mathbf{y}}\triangleq\mathrm{E}[\varphi_u^2(\mathbf{x})\mathbf{d}(\mathbf{x})\mathbf{d}^T(\mathbf{x});P_{\mathbf{x}}]$*,* $\varphi_u(\cdot)$ *is defined below (2) and* $[\mathbf{d}(\mathbf{x})]_k\triangleq\mathrm{Re}\{\mathbf{a}_k^H\mathbf{x}\mathbf{x}^H\mathbf{b}_k\}$*.*

Proof of Lemma 3. According to Eq. (4), the random vector $\mathbf{y}$ can be written as

$$\mathbf{y}=\frac{\frac{1}{N}\sum_{n=1}^N\mathbf{z}_n}{\frac{1}{N}\sum_{n=1}^N u(\mathbf{x}_n)}-\mathbf{q},\tag{34}$$

where the k-th entries of $\mathbf{z}_n$ and $\mathbf{q}$ are defined as $[\mathbf{z}_n]_k\triangleq u(\mathbf{x}_n)\,\mathrm{Re}\{\mathbf{a}_k^H\mathbf{x}_n\mathbf{x}_n^H\mathbf{b}_k\}$ and $[\mathbf{q}]_k\triangleq\mathrm{Re}\{\mathbf{a}_k^H(\hat{\boldsymbol{\mu}}_{\mathbf{x}}^{(u)}\hat{\boldsymbol{\mu}}_{\mathbf{x}}^{(u)H})\mathbf{b}_k\}$. Furthermore, by the fact that $\mathbf{z}_n$, $n=1,\ldots,N$ are real and i.i.d, the assumptions that $\mathrm{E}[\|\mathbf{x}\|^4u^2(\mathbf{x});\,P_{\mathbf{x}}]<\infty$ and $\mathbf{a}_k^H\boldsymbol{\Sigma}_{\mathbf{x}}^{(u)}\mathbf{b}_k=0$, Lemma 4 below (implying that $\mathbf{z}_n$ has a finite covariance), and the central limit theorem (C.L.T) [46] we conclude that the $\sqrt{N}$ scaled version of the numerator in (34) satisfies:

$$\sqrt{N}\frac{1}{N}\sum_{n=1}^N\mathbf{z}_n\xrightarrow{\mathrm{D}}\mathcal{N}(\mathbf{0},\mathbf{D}_{\mathbf{z}})\quad\text{as}\quad N\to\infty,\tag{35}$$

where $\mathbf{D}_{\mathbf{z}}\triangleq\mathrm{E}[u^2(\mathbf{x})\mathbf{d}(\mathbf{x})\mathbf{d}^T(\mathbf{x});P_{\mathbf{x}}]$. By Definition 2, the MT-function $u(\mathbf{x})$ is non-negative and $0<\mathrm{E}[u(\mathbf{x});\,P_{\mathbf{x}}]<\infty$. Thus, since $\{u(\mathbf{x}_n)\}$ is a sequence of real i.i.d. variables, then by Khinchine's strong law of large numbers [24], the denominator in (34) satisfies:

$$\frac{1}{N}\sum_{n=1}^N u(\mathbf{x}_n)\xrightarrow[N\to\infty]{\text{w.p.1}}\mathrm{E}[u(\mathbf{x});P_{\mathbf{x}}].\tag{36}$$

Since $\mathrm{E}[\|\mathbf{x}\|^4u^2(\mathbf{x});\,P_{\mathbf{x}}]$ is assumed to be finite, then by Hölder's inequality [25] $\mathrm{E}[\|\mathbf{x}\|^2u(\mathbf{x});P_{\mathbf{x}}]\le\sqrt{\mathrm{E}[\|\mathbf{x}\|^4u^2(\mathbf{x});P_{\mathbf{x}}]}<\infty$. Therefore, by Proposition 2 in [16], the assumption that $\boldsymbol{\mu}_{\mathbf{x}}^{(u)}=\mathbf{0}$, relation (36), the C.L.T, Slutskey's theorem [42], and [42, Theorem 2(iii)] it follows that

$$\sqrt{N}\mathbf{q}\xrightarrow{\mathrm{p}}\mathbf{0}\quad\text{as}\quad N\to\infty,\tag{37}$$

where "$\xrightarrow{p}$" denotes convergence in probability [25]. Finally, by (34)–(36), (37) and Slutsky's theorem we conclude that (33) holds. □

Lemma 4. *Assume that* $\mathrm{E}[\|\mathbf{x}\|^4 u^2(\mathbf{x}); P_{\mathbf{x}}]$ *is finite. Then,* $\mathrm{E}[|\mathrm{Re}\{\mathbf{a}^H\mathbf{x}\mathbf{x}^H\mathbf{b}\}\mathrm{Re}\{\mathbf{c}^H\mathbf{x}\mathbf{x}^H\mathbf{d}\}|u^2(\mathbf{x})]$ *is finite for any* $\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d}\in\mathbb{C}^p$.

Proof of Lemma 4. Let $c\triangleq\|\mathbf{a}\|\|\mathbf{b}\|\|\mathbf{c}\|\|\mathbf{d}\|$. By the Cauchy–Schwarz inequality we have

$$|\mathrm{Re}\{\mathbf{a}^H\mathbf{x}\mathbf{x}^H\mathbf{b}\}\mathrm{Re}\{\mathbf{c}^H\mathbf{x}\mathbf{x}^H\mathbf{d}\}| \le c\|\mathbf{x}\|^4.$$

Hence,

$$\mathrm{E}[|\mathrm{Re}\{\mathbf{a}^H\mathbf{x}\mathbf{x}^H\mathbf{b}\}\mathrm{Re}\{\mathbf{c}^H\mathbf{x}\mathbf{x}^H\mathbf{d}\}|u^2(\mathbf{x})] \le c\mathrm{E}[\|\mathbf{x}\|^4 u^2(\mathbf{x}); P_{\mathbf{x}}] < \infty.$$

□

Proof of the Theorem. Let $\boldsymbol{\vartheta}\triangleq[\vartheta_1,\ldots,\vartheta_q]^T$ denote an arbitrary vector parameter of DOAs. Define

$$\beta(\boldsymbol{\vartheta}) \triangleq \sum_{m=1}^{q} J(\vartheta_m), \tag{38}$$

where $J(\vartheta)\triangleq\mathbf{a}^H(\vartheta)\hat{\mathbf{G}}^{(u)}\hat{\mathbf{G}}^{(u)H}\mathbf{a}(\vartheta)$ and $\hat{\mathbf{G}}^{(u)}$ is defined in (23). By Hölder's inequality [25] and assumption B-1,

$$\mathrm{E}[\|\mathbf{x}\|^2 u(\mathbf{x}); P_{\mathbf{x}}] \le \sqrt{\mathrm{E}[\|\mathbf{x}\|^4 u^2(\mathbf{x}); P_{\mathbf{x}}]} < \infty. \tag{39}$$

Therefore, by assumption B-2 it is follows from Lemma 1, stated above, that $\hat{\boldsymbol{\theta}}_u$ is a strongly consistent estimate of $\boldsymbol{\theta}$. Hence, by assumption B-4, we conclude that for sufficiently large N there exists an open set that contains the line segment connecting $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}_u$. Thus, under assumption B-3 the MT-MUSIC estimator $\hat{\boldsymbol{\theta}}_u$ satisfies $\nabla\beta(\boldsymbol{\vartheta})|_{\boldsymbol{\vartheta}=\hat{\boldsymbol{\theta}}_u}=\mathbf{0}$. Furthermore, by assumption B-3 and the mean-value theorem [40]

$$\mathbf{0} = \nabla\beta(\boldsymbol{\vartheta})|_{\boldsymbol{\vartheta}=\hat{\boldsymbol{\theta}}_u} = \mathbf{g}(\boldsymbol{\theta}) + \mathbf{F}(\bar{\boldsymbol{\theta}}_u)(\hat{\boldsymbol{\theta}}_u - \boldsymbol{\theta}), \tag{40}$$

where

$$\mathbf{g}(\boldsymbol{\vartheta}) \triangleq \nabla\beta(\boldsymbol{\vartheta}) = [\dot{J}(\vartheta_1),\ldots,\dot{J}(\vartheta_q)]^T, \tag{41}$$

$$\mathbf{F}(\boldsymbol{\vartheta}) \triangleq \nabla^2\beta(\boldsymbol{\vartheta}) = \mathrm{diag}\{\ddot{J}(\vartheta_1),\ldots,\ddot{J}(\vartheta_q)\}, \tag{42}$$

$$\dot{J}(\vartheta) \triangleq \nabla J(\vartheta) = 2\mathrm{Re}\{\dot{\mathbf{a}}^H(\vartheta)\hat{\mathbf{G}}^{(u)}\hat{\mathbf{G}}^{(u)H}\mathbf{a}(\vartheta)\}, \tag{43}$$

$$\begin{aligned}\ddot{J}(\vartheta) &\triangleq \nabla^2 J(\vartheta)\\ &= 2\mathrm{Re}\left\{\ddot{\mathbf{a}}^H(\vartheta)\hat{\mathbf{G}}^{(u)}\hat{\mathbf{G}}^{(u)H}\mathbf{a}(\vartheta) + \dot{\mathbf{a}}^H(\vartheta)\hat{\mathbf{G}}^{(u)}\hat{\mathbf{G}}^{(u)H}\dot{\mathbf{a}}(\vartheta)\right\},\end{aligned} \tag{44}$$

$\dot{\mathbf{a}}(\vartheta)\triangleq d\mathbf{a}(\vartheta)/d\vartheta$, $\ddot{\mathbf{a}}(\vartheta)\triangleq d^2\mathbf{a}(\vartheta)/d\vartheta^2$, $\mathrm{diag}\{\cdot,\ldots,\cdot\}$ denotes a diagonal matrix with the diagonal entries given by its arguments, and $\bar{\boldsymbol{\theta}}_u$ is a vector that lies in the line segment connecting $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}_u$.

In the following, we analyze strong convergence of $\mathbf{F}(\bar{\boldsymbol{\theta}}_u)$ in (40). Since $\hat{\boldsymbol{\theta}}_u$ is strongly consistent, $\bar{\boldsymbol{\theta}}_u$ must be strongly consistent. By (39) and Proposition 2 in [16], the empirical MT-covariance $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{(u)}$ is a strongly consistent estimator of $\boldsymbol{\Sigma}_{\mathbf{x}}^{(u)}$. Furthermore, by Theorem 8 in [44, p. 183] the eigenvectors comprising $\hat{\mathbf{G}}^{(u)}$ in (23) define continuous mappings of $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{(u)}$. Hence, by the definition of $\mathbf{G}^{(u)}$ and Mann-Wald's Theorem [45], $\hat{\mathbf{G}}^{(u)}\xrightarrow[N\to\infty]{\text{w.p.1}}\mathbf{G}^{(u)}$. Thus, by (42), (44), assumptions A-2, B-3, and Mann–Wald's Theorem [45]

$$\mathbf{F}(\bar{\boldsymbol{\theta}}_u)\xrightarrow[N\to\infty]{\text{w.p.1}}\mathrm{diag}\{h(\theta_1),\ldots,h(\theta_q)\}\triangleq\boldsymbol{\Gamma}, \tag{45}$$

where $h(\vartheta)\triangleq 2\dot{\mathbf{a}}^H(\vartheta)\mathbf{G}^{(u)}\mathbf{G}^{(u)H}\dot{\mathbf{a}}(\vartheta)$.

We proceed by analyzing convergence in distribution of $\mathbf{g}(\boldsymbol{\theta})$ in (40). To do so, we begin with small perturbation analysis of the eigenvectors of $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{(u)}$ corresponding to the q largest eigenvalues. By assumption B-1 and Lemma 2, it follows that the empirical MT-covariance $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{(u)}$ (4) converges to the MT-covariance $\boldsymbol{\Sigma}_{\mathbf{x}}^{(u)}$ (3) at a rate of $1/\sqrt{N}$. Furthermore, $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{(u)}$ is a Hermitian matrix. These two properties lead to the following stochastic representation of $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{(u)}$:

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{(u)} = \boldsymbol{\Sigma}_{\mathbf{x}}^{(u)} + \epsilon\Delta, \tag{46}$$

where $\epsilon\in\mathbb{R}$ is $O(1/\sqrt{N})$ perturbation factor and $\Delta\in\mathbb{C}^{p\times p}$ is a random Hermitian perturbation matrix. Thus, similarly to [47, Eq. (A.8)] and [48, Eq. (24)], that follow from the small perturbation analysis of an eigenvector corresponding to a simple eigenvalue in [49, pp. 67–70], it can be shown that the unit-norm eigenvectors corresponding to the q largest eigenvalues of $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{(u)}$ are given by:

$$\hat{\mathbf{v}}_i^{(u)} = \mathbf{v}_i^{(u)} + \epsilon\sum_{j=1,j\neq i}^{p} t_{i,j}\mathbf{v}_j^{(u)} + O(\epsilon^2),\quad i=1,\ldots,q, \tag{47}$$

where $\mathbf{v}_i^{(u)}$ denotes the eigenvector of $\boldsymbol{\Sigma}_{\mathbf{x}}^{(u)}$ corresponding to its i-th largest eigenvalue, and

$$t_{i,j} \triangleq -\frac{\mathbf{v}_j^{(u)H}\Delta\mathbf{v}_i^{(u)}}{\lambda_j^{(u)}-\lambda_i^{(u)}}. \tag{48}$$

Notice that by assumption B-5, the denominator of $t_{i,j}$ does not take zero values under the summation indices in (47). Furthermore, note that the complex-conjugate of $t_{i,j}$ satisfy:

$$t_{i,j}^* = -\frac{\mathbf{v}_i^{(u)H}\Delta\mathbf{v}_j^{(u)}}{\lambda_j^{(u)}-\lambda_i^{(u)}} \triangleq -t_{j,i}. \tag{49}$$

Therefore, by (23), (41), (43), (46)–(49) and Assumption A-2, the elements comprising $\mathbf{g}(\boldsymbol{\theta})$ satisfy:

$$\begin{aligned}\dot{J}(\theta_k) &= 2\mathrm{Re}\left\{\dot{\mathbf{a}}^H(\theta_k)\Big(\mathbf{I}-\sum_{i=1}^{q}\hat{\mathbf{v}}_i^{(u)}\hat{\mathbf{v}}_i^{(u)H}\Big)\mathbf{a}(\theta_k)\right\}\\ &= -2\epsilon\mathrm{Re}\left\{\dot{\mathbf{a}}^H(\theta_k)\sum_{i=1}^{q}\sum_{\substack{j=1\\ j\neq i}}^{p}\left(t_{i,j}^*\mathbf{v}_i^{(u)}\mathbf{v}_j^{(u)H} + t_{i,j}\mathbf{v}_j^{(u)}\mathbf{v}_i^{(u)H}\right)\mathbf{a}(\theta_k)\right\}\\ &\quad + O(\epsilon^2)\\ &= -2\epsilon\mathrm{Re}\left\{\dot{\mathbf{a}}^H(\theta_k)\sum_{i=1}^{q}\sum_{j=q+1}^{p}t_{i,j}\mathbf{v}_j^{(u)}\mathbf{v}_i^{(u)H}\mathbf{a}(\theta_k)\right\} + O(\epsilon^2)\\ &= -2\epsilon\mathrm{Re}\left\{\dot{\mathbf{a}}^H(\theta_k)\mathbf{P}_{\mathbf{A}}^{\perp}\Delta\mathbf{E}^{(u)}\mathbf{a}(\theta_k)\right\} + O(\epsilon^2)\\ &= -2\mathrm{Re}\left\{\dot{\mathbf{a}}^H(\theta_k)\mathbf{P}_{\mathbf{A}}^{\perp}\big(\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{(u)}-\boldsymbol{\Sigma}_{\mathbf{x}}^{(u)}\big)\mathbf{E}^{(u)}\mathbf{a}(\theta_k)\right\} + O(\epsilon^2)\\ &= -2\mathrm{Re}\left\{\dot{\mathbf{a}}^H(\theta_k)\mathbf{P}_{\mathbf{A}}^{\perp}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{(u)}\mathbf{E}^{(u)}\mathbf{a}(\theta_k)\right\} + O(\epsilon^2),\quad k=1,\ldots,q,\end{aligned} \tag{50}$$

where $\mathbf{P}_{\mathbf{A}}^{\perp}=\sum_{j=q+1}^{p}\mathbf{v}_j^{(u)}\mathbf{v}_j^{(u)H}$, $\mathbf{E}^{(u)}\triangleq\sum_{i=1}^{q}\frac{\mathbf{v}_i^{(u)}\mathbf{v}_i^{(u)H}}{\lambda_i^{(u)}-\gamma^{(u)}} = \mathbf{A}(\mathbf{A}^H(\boldsymbol{\Sigma}_{\mathbf{x}}^{(u)}-\gamma^{(u)}\mathbf{I}_p)\mathbf{A})^{-1}\mathbf{A}^H$, $\gamma^{(u)}$ is defined in Eq. (6) and $\mathbf{I}_p$ denotes a $p\times p$ identity matrix.

Hence, by (41), (50), the property that by Assumption A-2 $\dot{\mathbf{a}}^H(\theta_k)\mathbf{P}_{\mathbf{A}}^{\perp}\boldsymbol{\Sigma}_{\mathbf{x}}^{(u)}\mathbf{E}^{(u)}\mathbf{a}(\theta_k)=0\ \forall k=1,\ldots,q$, assumption B-1, and Lemma 3 stated above, we conclude that

$$\sqrt{N}\mathbf{g}(\boldsymbol{\theta})\xrightarrow{D}\mathcal{N}\big(\mathbf{0},\mathbf{D}(\boldsymbol{\theta})\big)\quad\text{as}\quad N\to\infty, \tag{51}$$

where $\mathbf{D}(\boldsymbol{\theta})\triangleq\mathrm{E}[\varphi_u^2(\mathbf{x})\tilde{\boldsymbol{\psi}}_u(\mathbf{x},\boldsymbol{\theta})\tilde{\boldsymbol{\psi}}_u^T(\mathbf{x},\boldsymbol{\theta});P_{\mathbf{x}}]$, $\varphi_u(\cdot)$ is defined below Eq. (2), and $[\tilde{\boldsymbol{\psi}}_u(\mathbf{x},\boldsymbol{\theta})]_k\triangleq 2\mathrm{Re}\{\dot{\mathbf{a}}^H(\theta_k)\mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{x}\mathbf{x}^H\mathbf{E}^{(u)}\mathbf{a}(\theta_k)\}$.

Finally, by (40), (45), (51) and Slutsky's theorem we conclude that

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_u - \boldsymbol{\theta}) = -\sqrt{N}\mathbf{F}^{-1}(\bar{\boldsymbol{\theta}})\mathbf{g}(\boldsymbol{\theta}) \xrightarrow[N\to\infty]{\mathrm{D}} \mathcal{N}(\mathbf{0}, \mathbf{R}^{(u)}(\boldsymbol{\theta})),$$

where $\mathbf{R}^{(u)}(\boldsymbol{\theta}) \triangleq \mathrm{E}[\varphi_u^2(\mathbf{x})\boldsymbol{\psi}_u(\mathbf{x},\boldsymbol{\theta})\boldsymbol{\psi}_u^T(\mathbf{x},\boldsymbol{\theta}); P_{\mathbf{x}}]$ and $\boldsymbol{\psi}_u(\mathbf{x},\boldsymbol{\theta}) \triangleq \boldsymbol{\Gamma}^{-1}\tilde{\boldsymbol{\psi}}_u(\mathbf{x},\boldsymbol{\theta})$. □

B. Proof of Proposition 1

The proof of Proposition 1 is based on the following auxiliary definition and Lemma.

Definition 3. Define the following vectors and variables:

$$\mathbf{c}_i \triangleq \mathbf{P}_{\mathbf{A}}^{\perp}\dot{\mathbf{a}}(\theta_i), \qquad \mathbf{d}_i \triangleq \mathbf{E}^{(u)}\mathbf{a}(\theta_i),$$
$$\hat{\mathbf{c}}_i \triangleq \mathbf{P}_{\hat{\mathbf{A}}}^{\perp}\dot{\mathbf{a}}(\hat{\theta}_i), \qquad \hat{\mathbf{d}}_i \triangleq \hat{\mathbf{E}}^{(u)}\mathbf{a}(\hat{\theta}_i),$$

$$\hat{\psi}_n^{(i)} \triangleq \mathrm{Re}\{\hat{\mathbf{c}}_i^H\mathbf{x}_n\mathbf{x}_n^H\hat{\mathbf{d}}_i\} \text{ and } \psi_n^{(i)} \triangleq \mathrm{Re}\{\mathbf{c}_i^H\mathbf{x}_n\mathbf{x}_n^H\mathbf{d}_i\}, \tag{52}$$

$i = 1,\ldots,q,\ n = 1,\ldots,N$, where $\mathbf{E}^{(u)}$ and $\hat{\mathbf{E}}^{(u)}$ are defined in Eqs. (11) and (16), respectively.

Lemma 5. *Assume that conditions A-2, B-1-B-3 are satisfied. Then, the random variable*

$$A_1 \triangleq \left|\frac{1}{N}\sum_{n=1}^{N} u^2(\mathbf{x}_n)\hat{\psi}_n^{(i)}\hat{\psi}_n^{(j)} - \frac{1}{N}\sum_{n=1}^{N} u^2(\mathbf{x}_n)\psi_n^{(i)}\psi_n^{(j)}\right| \xrightarrow{\text{w.p.1}} 0$$
as $N \to \infty$.

Proof of Lemme 5. We begin by proving strong consistency of $\hat{\mathbf{c}}_i$ and $\hat{\mathbf{d}}_i$ defined in (52). The proof is based on the following claims:

(a) $\mathbf{a}(\hat{\theta}_i) \xrightarrow[N\to\infty]{\text{w.p.1}} \mathbf{a}(\theta_i)$ and $\dot{\mathbf{a}}(\hat{\theta}_i) \xrightarrow[N\to\infty]{\text{w.p.1}} \dot{\mathbf{a}}(\theta_i)$, $i = 1,\ldots,q$,

(b) $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{(u)} \xrightarrow[N\to\infty]{\text{w.p.1}} \boldsymbol{\Sigma}_{\mathbf{x}}^{(u)}$,

(c) $\hat{\gamma}^{(u)} \xrightarrow[N\to\infty]{\text{w.p.1}} \gamma^{(u)}$, where $\gamma^{(u)}$ and $\hat{\gamma}^{(u)}$ are defined in (6) and below (16), respectively.

Claim (a) follows directly from assumptions B-1-B-3, (39), Lemma 1 (stated above), and Mann-Wald's Theorem [45]. Claim (b) is a consequence of assumption B-1, (39) and Proposition 2 in [16]. Claim (c) follows from claim (b), the fact that the eigenvalues of $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{(u)}$ define continuous mappings on $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{(u)}$, assumption A-2 and Mann–Wald's Theorem. Hence, by (11), (16), (52), Claims (a)–(c), and Mann–Wald's Theorem,

$$\hat{\mathbf{c}}_i \xrightarrow{\text{w.p.1}} \mathbf{c}_i \quad \text{and} \quad \hat{\mathbf{d}}_i \xrightarrow{\text{w.p.1}} \mathbf{d}_i \quad \text{as} \quad N \to \infty. \tag{53}$$

Next, we upper bound A_1. By the triangle inequality

$$A_1 \le B_1 + B_2, \tag{54}$$

where $B_1 \triangleq |\frac{1}{N}\sum_{n=1}^{N} u^2(\mathbf{x}_n)\hat{\psi}_n^{(j)}(\hat{\psi}_n^{(i)} - \psi_n^{(i)})|$ and $B_2 \triangleq |\frac{1}{N}\sum_{n=1}^{N} u^2(\mathbf{x}_n)\psi_n^{(i)}(\hat{\psi}_n^{(j)} - \psi_n^{(j)})|$. Using (52), the triangle inequality and Cauchy–Schwarz inequality we obtain that

$$\begin{aligned}|\hat{\psi}_n^{(i)} - \psi_n^{(i)}| &= \left|\mathrm{Re}\{(\hat{\mathbf{c}}_i - \mathbf{c}_i)^H\mathbf{x}_n\mathbf{x}_n^H\hat{\mathbf{d}}_i + \mathbf{c}_i^H\mathbf{x}_n\mathbf{x}_n^H(\hat{\mathbf{d}}_i - \mathbf{d}_i)\right| \\ &\le \left|(\hat{\mathbf{c}}_i - \mathbf{c}_i)^H\mathbf{x}_n\mathbf{x}_n^H\hat{\mathbf{d}}_i\right| + \left|\mathbf{c}_i^H\mathbf{x}_n\mathbf{x}_n^H(\hat{\mathbf{d}}_i - \mathbf{d}_i)\right| \\ &\le \left(\|\hat{\mathbf{c}}_i - \mathbf{c}_i\|\|\hat{\mathbf{d}}_i\| + \|\hat{\mathbf{d}}_i - \mathbf{d}_i\|\|\mathbf{c}_i\|\right)\|\mathbf{x}_n\|^2 \end{aligned} \tag{55}$$

and

$$|\hat{\psi}_n^{(i)}| \le \|\hat{\mathbf{c}}_i\|\|\hat{\mathbf{d}}_i\|\|\mathbf{x}_n\|^2. \tag{56}$$

Therefore, by (55) and (56) we conclude that

$$\begin{aligned}B_1 \le{}& \|\hat{\mathbf{c}}_i\|\|\hat{\mathbf{d}}_i\|\|\hat{\mathbf{c}}_i - \mathbf{c}_i\|\|\hat{\mathbf{d}}_i\|\frac{1}{N}\sum_{n=1}^{N} u^2(\mathbf{x}_n)\|\mathbf{x}_n\|^4 \\ &+ \|\hat{\mathbf{c}}_i\|\|\hat{\mathbf{d}}_i\|\|\hat{\mathbf{d}}_i - \mathbf{d}_i\|\|\mathbf{c}_i\|\frac{1}{N}\sum_{n=1}^{N} u^2(\mathbf{x}_n)\|\mathbf{x}_n\|^4.\end{aligned} \tag{57}$$

Since $\{u^2(\mathbf{x}_n)\|\mathbf{x}_n\|^4\}$ is a sequence of i.i.d. variables, then by condition B-1 and Khinchine's strong law of large numbers

$$\frac{1}{N}\sum_{n=1}^{N} u^2(\mathbf{x}_n)\|\mathbf{x}_n\|^4 \xrightarrow[N\to\infty]{\text{w.p.1}} \mathrm{E}[\|\mathbf{x}\|^4 u^2(\mathbf{x}); P_{\mathbf{x}}] < \infty. \tag{58}$$

Thus, by (53), (57) and (58) we conclude that $B_1 \xrightarrow[N\to\infty]{\text{w.p.1}} 0$. Similarly, one can verify that $B_2 \xrightarrow[N\to\infty]{\text{w.p.1}} 0$. Finally, by (54), $A_1 \xrightarrow[N\to\infty]{\text{w.p.1}} 0$. □

Proof of the Proposition. Notice that the i,j-th entry of (14) can be written as:

$$\left[\hat{\mathbf{R}}^{(u)}(\hat{\boldsymbol{\theta}}_u)\right]_{i,j} = \frac{\frac{1}{N}\sum_{n=1}^{N} u^2(\mathbf{x}_n)\hat{\psi}_n^{(i)}\hat{\psi}_n^{(j)}}{\left(\frac{1}{N}\sum_{n=1}^{N} u(\mathbf{x}_n)\right)^2\hat{T}_i\hat{T}_j}, \tag{59}$$

where $\hat{T}_i \triangleq \dot{\mathbf{a}}^H(\hat{\theta}_i)\mathbf{P}_{\hat{\mathbf{A}}}^{\perp}\dot{\mathbf{a}}(\hat{\theta}_i)$.

We begin by proving strong convergence of the numerator in (59). By the triangle inequality

$$\left|\frac{1}{N}\sum_{n=1}^{N} u^2(\mathbf{x}_n)\hat{\psi}_n^{(i)}\hat{\psi}_n^{(j)} - \mathrm{E}[u^2(\mathbf{x})\psi_n^{(i)}\psi_n^{(j)}; P_{\mathbf{x}}]\right| \le A_1 + A_2, \tag{60}$$

where

$$A_1 \triangleq \left|\frac{1}{N}\sum_{n=1}^{N} u^2(\mathbf{x}_n)\hat{\psi}_n^{(i)}\hat{\psi}_n^{(j)} - \frac{1}{N}\sum_{n=1}^{N} u^2(\mathbf{x}_n)\psi_n^{(i)}\psi_n^{(j)}\right|,$$

$$A_2 \triangleq \left|\frac{1}{N}\sum_{n=1}^{N} u^2(\mathbf{x}_n)\psi_n^{(i)}\psi_n^{(j)} - \mathrm{E}[u^2(\mathbf{x})\psi_n^{(i)}\psi_n^{(j)}; P_{\mathbf{x}}]\right|.$$

Notice that $\{u^2(\mathbf{x}_n)\psi_n^{(i)}\psi_n^{(j)}\}$ is a sequence of real i.i.d. variables. Therefore, using assumption B-1, Lemma 4 and Khinchines strong law of large numbers [24], one can verify that $A_2 \xrightarrow[N\to\infty]{\text{w.p.1}} 0$. Furthermore, by assumptions A-2, B-1-B-3 and Lemma 5 stated above, $A_1 \xrightarrow[N\to\infty]{\text{w.p.1}} 0$. Therefore, by (60) we conclude that the numerator in (59) satisfies:

$$\frac{1}{N}\sum_{n=1}^{N} u^2(\mathbf{x}_n)\hat{\psi}_n^{(i)}\hat{\psi}_n^{(j)} \xrightarrow[N\to\infty]{\text{w.p.1}} \mathrm{E}[u^2(\mathbf{x})\psi_n^{(i)}\psi_n^{(j)}; P_{\mathbf{x}}] \tag{61}$$

Next, convergence of $\hat{T}_i$ in the denominator of (59) is proved. By assumption B-1 and Eq. (39) $\mathrm{E}[\|\mathbf{x}\|^2 u(\mathbf{x}); P_{\mathbf{x}}] < \infty$. Thus, by Lemma 1, Mann–Wald's Theorem [45] and assumptions B-1 and B-3

$$\hat{T}_i \xrightarrow{\text{w.p.1}} \dot{\mathbf{a}}^H(\theta_i)\mathbf{P}_{\mathbf{A}}^{\perp}\dot{\mathbf{a}}(\theta_i) \quad \text{as} \quad N \to \infty. \tag{62}$$

Finally, by (36), (59), (61), (62), Mann–Wald's Theorem and Eqs. (9), (12)–(14) we conclude that (17) holds. □

C. Implementation of the MT-MUSIC with spherical MT-functions

In [16], the MT-MUSIC algorithm was implemented with a zero-centered spherically contoured Gaussian MT-function. Under the assumption of spherical CG noise, we proved in [16, Theorem 1] that the noise subspace can be determined from the eigendecomposition of the corresponding MT-covariance, i.e., condition A-2 is satisfied. In this appendix, we show that the MT-MUSIC can be implemented with other types of MT-functions that do not necessarily belong to the Gaussian family. In a more specified manner, we show that under the assumption of spherical CG noise (also made in [16]) condition A-2 is satisfied when the MT-function $u(\cdot)$ belongs to the wide class of strictly-positive spherically contoured functions, i.e., we assume that the MT-function takes the form:

$$u(\mathbf{x}) = h(\|\mathbf{x}\|), \tag{63}$$

where $h : \mathbb{R}_+ \to \mathbb{R}_{++}$ is a strictly-positive function.

In the following Theorem, the structure of the resulting MT-covariance is derived. This structure provides a non-trivial extension to the one stated in Theorem 1 in [16] for the specific case of spherically contoured Gaussian MT-function. For simplicity, we shall assume that the MT-mean $\boldsymbol{\mu}_{\mathbf{x}}^{(u)}=\mathbf{0}$. Under the MT-function (63) this assumption is satisfied if the signals are symmetrically distributed about the origin.

Theorem 2. *Under the observation model (1), the MT-function (63) and the CG noise (19), the MT-covariance (3) takes the form:*

$$\boldsymbol{\Sigma}_{\mathbf{x}}^{(u)}=\mathbf{A}\mathbf{J}^{(u)}\mathbf{A}^H+\alpha^{(u)}\mathbf{I}_p, \tag{64}$$

where

$$\mathbf{J}^{(u)}\triangleq \mathrm{E}\left[\boldsymbol{s}\boldsymbol{s}^H\beta_p(\kappa)\frac{\|\mathbf{x}\|^4}{\tau^4\sigma_{\mathbf{w}}^4}\varphi_h(\|\mathbf{x}\|);P_{\mathbf{x},\boldsymbol{S},\tau}\right]$$

is a strictly positive-definite matrix,

$$\alpha^{(u)}\triangleq \mathrm{E}\left[\gamma_p(\kappa)\|\mathbf{x}\|^2\varphi_h(\|\mathbf{x}\|);P_{\mathbf{x},\boldsymbol{S},\tau}\right],$$

$\mathbf{I}_p$ *is a* $p\times p$ *identity matrix,* $\kappa\triangleq 2\|\mathbf{x}\|\|\mathbf{A}\boldsymbol{s}\|/(\tau^2\sigma_{\mathbf{w}}^2)$,

$$\beta_p(\kappa)\triangleq\begin{cases}\frac{4I_{p+1}(\kappa)}{\kappa^2 I_{p-1}(\kappa)}, & \text{if } \kappa>0\\ 0, & \text{if } \kappa=0\end{cases},\quad \gamma_p(\kappa)\triangleq\begin{cases}\frac{2I_p(\kappa)}{\kappa I_{p-1}(\kappa)}, & \text{if } \kappa>0\\ 1/p, & \text{if } \kappa=0\end{cases},$$

$\varphi_h(\|\mathbf{x}\|)\triangleq h(\|\mathbf{x}\|)/\mathrm{E}[h(\|\mathbf{x}\|);P_{\mathbf{x}}]$, *and* $I_k(\cdot)$ *is the* k*th order modified Bessel function of the first kind [53].*

The structure (64), along with the property that the matrix $\mathbf{J}^{(u)}$ is non-singular and the assumption that steering matrix $\mathbf{A}$ in the array model (1) has a full column rank, imply that Condition A-2 is satisfied. Under this key condition, the spanning vectors of the null-space of $\mathbf{A}^H$, also called the noise subspace, can be determined from the eigen-decomposition of the MT-covariance.

The proof of Theorem 2 is based on the following Lemma:

Lemma 6. *Let* $\mathbf{q}\sim\mathcal{CN}(\boldsymbol{\mu},\sigma^2\mathbf{I}_p)$, *where* $\mathbf{q}\in\mathbb{C}^p$. *Then*

$$\mathrm{E}\left[\mathbf{q}\mathbf{q}^H;P_{\mathbf{q}|\|\mathbf{q}\|}\right]=\beta_p(\kappa)\frac{\|\mathbf{q}\|^4}{\sigma^4}\boldsymbol{\mu}\boldsymbol{\mu}^H+\|\mathbf{q}\|^2\gamma_p(\kappa)\mathbf{I}_p,$$

where $P_{\cdot|\cdot}$ *denotes a conditional probability measure,* $\beta_p(\cdot)$ *and* $\gamma_p(\cdot)$ *are defined below (64) and* $\kappa\triangleq 2\|\mathbf{q}\|\|\boldsymbol{\mu}\|/\sigma^2$.

Proof of Lemma 6. First, notice that since $\mathbf{q}$ is a continuous random vector, then $\|\mathbf{q}\|>0$ w.p. 1. Therefore, $\kappa=0$ w.p. 1 if and only if $\boldsymbol{\mu}=\mathbf{0}$. The proof begins for the special case of $\boldsymbol{\mu}=\mathbf{0}$, which is equivalent to $\kappa=0$ w.p. 1. Note that in this case, the random vector $\mathbf{q}$ is spherically symmetric, and therefore, $\mathbf{q}\overset{\mathrm{D}}{=}\mathbf{U}\mathbf{q}$ for any unitary matrix $\mathbf{U}\in\mathbb{C}^{p\times p}$ [5], where "$\overset{\mathrm{D}}{=}$" denotes equality in distribution. Hence, the equality $\mathrm{E}\left[\mathbf{q}\mathbf{q}^H;P_{\mathbf{q}|\|\mathbf{q}\|}\right]=\mathbf{U}\mathrm{E}\left[\mathbf{q}\mathbf{q}^H;P_{\mathbf{q}|\|\mathbf{q}\|}\right]\mathbf{U}^H$ holds for any unitary matrix $\mathbf{U}\in\mathbb{C}^{p\times p}$, which implies that

$$\mathrm{E}\left[\mathbf{q}\mathbf{q}^H;P_{\mathbf{q}|\|\mathbf{q}\|}\right]=\frac{\mathrm{tr}\left\{\mathrm{E}\left[\mathbf{q}\mathbf{q}^H;P_{\mathbf{q}|\|\mathbf{q}\|}\right]\right\}}{p}\mathbf{I}_p=\frac{\|\mathbf{q}\|^2}{p}\mathbf{I}_p, \tag{65}$$

where $\mathrm{tr}\{\cdot\}$ denotes the trace operator.

Now, we assume that $\boldsymbol{\mu}\neq\mathbf{0}$, which is equivalent to $\kappa>0$ w.p. 1. Here, we shall use the relation

$$\mathrm{E}\left[\mathbf{q}\mathbf{q}^H;P_{\mathbf{q}|\|\mathbf{q}\|}\right]=\mathrm{E}\left[\frac{\mathbf{q}\mathbf{q}^H}{\|\mathbf{q}\|^2};P_{\mathbf{q}|\|\mathbf{q}\|}\right]\|\mathbf{q}\|^2 \tag{66}$$

and derive the conditional expectation in the r.h.s. of (66). Let $\mathbf{u}_1\triangleq\mathrm{Re}\{\mathbf{q}\}$, $\mathbf{u}_2\triangleq\mathrm{Im}\{\mathbf{q}\}$ and $\mathbf{u}\triangleq[\mathbf{u}_1^T,\mathbf{u}_2^T]^T$. Notice that

$$\mathrm{E}\left[\frac{\mathbf{u}\mathbf{u}^T}{\|\mathbf{u}\|^2};P_{\mathbf{u}|\|\mathbf{u}\|=\rho}\right]=\mathrm{E}\left[\mathbf{u}'\mathbf{u}'^T;P_{\mathbf{u}'|\|\mathbf{u}'\|=1}\right]\triangleq\mathbf{D}, \tag{67}$$

where $\mathbf{u}'\triangleq\mathbf{u}/\rho\sim\mathcal{N}(c\cdot\boldsymbol{\eta},c\cdot\kappa^{-1}\mathbf{I}_{2p})$, $\boldsymbol{\eta}\triangleq\boldsymbol{\mu}_{\mathbf{u}}/\|\boldsymbol{\mu}_{\mathbf{u}}\|$, $\boldsymbol{\mu}_{\mathbf{u}}\triangleq[\boldsymbol{\mu}_1^T,\boldsymbol{\mu}_2^T]^T$, $\boldsymbol{\mu}_1\triangleq\mathrm{Re}\{\boldsymbol{\mu}\}$, $\boldsymbol{\mu}_2\triangleq\mathrm{Im}\{\boldsymbol{\mu}\}$, $\kappa\triangleq 2\rho\|\boldsymbol{\mu}_{\mathbf{u}}\|/\sigma^2$, and $c\triangleq\|\boldsymbol{\mu}_{\mathbf{u}}\|/\rho$. By [54, p. 379], it follows that the conditional distribution of $\mathbf{u}'$ given that $\|\mathbf{u}'\|=1$ is $M_{2p}(\boldsymbol{\eta},\kappa)$, where $M_{2p}(\boldsymbol{\eta},\kappa)$ is the Von-Mises Fisher distribution [55] on the $2p-1$ dimensional unit sphere $\mathbb{S}_{2p-1}$, with directional mean $\boldsymbol{\eta}$ and concentration parameter κ. Hence, the density function of the conditional probability measure $P_{\mathbf{u}'|\|\mathbf{u}'\|=1}$ is given by [55]:

$$f_{\mathbf{u}'|\|\mathbf{u}'\|=1}(\boldsymbol{t})\triangleq C(\kappa)\exp\left(\kappa\boldsymbol{\eta}^T\boldsymbol{t}\right),\qquad \boldsymbol{t}\in\mathbb{S}_{2p-1}, \tag{68}$$

where $C(\kappa)\triangleq\kappa^{p-1}/((2\pi)^p I_{p-1}(\kappa))$ is a normalization constant. Define the vector $\tilde{\boldsymbol{\eta}}\triangleq\kappa\boldsymbol{\eta}$. Notice that $\kappa=\|\tilde{\boldsymbol{\eta}}\|$ and thus

$$\frac{d\kappa}{d\tilde{\boldsymbol{\eta}}}=\frac{\tilde{\boldsymbol{\eta}}}{\kappa}. \tag{69}$$

Furthermore, by the recursive relation 10.51.6 in [56] one can verify that

$$\left(\frac{1}{\kappa}\frac{d}{d\kappa}\right)^m C^{-1}(\kappa)=C^{-1}(\kappa)\frac{I_{p-1+m}(\kappa)}{\kappa^m I_{p-1}(\kappa)}. \tag{70}$$

Therefore, by (68)–(70) and Theorem 2.40 in [57] for differentiation under the integral sign, we conclude that the conditional autocorrelation matrix in (67) takes the form:

$$\begin{aligned}\mathbf{D}&=C(\kappa)\frac{d^2}{d\tilde{\boldsymbol{\eta}}d\tilde{\boldsymbol{\eta}}^T}\int_{\mathbb{S}_{2p-1}}\exp\left(\tilde{\boldsymbol{\eta}}^T\boldsymbol{t}\right)d\omega_{2p-1}(\boldsymbol{t})\\ &=C(\kappa)\frac{d^2}{d\tilde{\boldsymbol{\eta}}d\tilde{\boldsymbol{\eta}}^T}C^{-1}(\kappa)\;=\;C(\kappa)\frac{d\kappa}{d\tilde{\boldsymbol{\eta}}}\frac{d}{d\kappa}\left(\frac{d\kappa}{d\tilde{\boldsymbol{\eta}}}\frac{d}{d\kappa}C^{-1}(\kappa)\right)^T\\ &=C(\kappa)\left(\left(\frac{1}{\kappa}\frac{d}{d\kappa}\right)C^{-1}(\kappa)\mathbf{I}+\tilde{\boldsymbol{\eta}}\tilde{\boldsymbol{\eta}}^T\left(\frac{1}{\kappa}\frac{d}{d\kappa}\right)^2C^{-1}(\kappa)\right)\\ &=\frac{I_p(\kappa)}{\kappa I_{p-1}(\kappa)}\mathbf{I}+\tilde{\boldsymbol{\eta}}\tilde{\boldsymbol{\eta}}^T\frac{I_{p+1}(\kappa)}{\kappa^2 I_{p-1}(\kappa)}\;=\;\frac{I_p(\kappa)}{\kappa I_{p-1}(\kappa)}\mathbf{I}+\boldsymbol{\eta}\boldsymbol{\eta}^T\frac{I_{p+1}(\kappa)}{I_{p-1}(\kappa)},\end{aligned} \tag{71}$$

where ω_{2p-1} denotes Lebegue's surface measure on $\mathbb{S}_{2p-1}$. Therefore, by (67) and (71)

$$\mathrm{E}\left[\frac{\mathbf{u}_k\mathbf{u}_j^T}{\|\mathbf{u}\|^2};P_{\mathbf{u}|\|\mathbf{u}\|=\rho}\right]=\frac{I_p(\kappa)}{\kappa I_{p-1}(\kappa)}\mathbf{I}\delta_{k,j}+\frac{\boldsymbol{\mu}_k\boldsymbol{\mu}_j^T}{\|\boldsymbol{\mu}_{\mathbf{u}}\|^2}\frac{I_{p+1}(\kappa)}{I_{p-1}(\kappa)}, \tag{72}$$

$k,j=1,2$, where δ_{kj} is the Kronecker delta function. Finally, by (72) and the definition of $\mathbf{u}$, we conclude that

$$\begin{aligned}\mathrm{E}\left[\frac{\mathbf{q}\mathbf{q}^H}{\|\mathbf{q}\|^2};P_{\mathbf{q}|\|\mathbf{q}\|}\right]&=\mathrm{E}\left[\frac{\mathbf{u}_1\mathbf{u}_1^T}{\|\mathbf{u}\|^2};P_{\mathbf{u}|\|\mathbf{u}\|=\rho}\right]+i\mathrm{E}\left[\frac{\mathbf{u}_2\mathbf{u}_1^T}{\|\mathbf{u}\|^2};P_{\mathbf{u}|\|\mathbf{u}\|=\rho}\right]\\ &\quad-i\mathrm{E}\left[\frac{\mathbf{u}_1\mathbf{u}_2^T}{\|\mathbf{u}\|^2};P_{\mathbf{u}|\|\mathbf{u}\|=\rho}\right]+\mathrm{E}\left[\frac{\mathbf{u}_2\mathbf{u}_2^T}{\|\mathbf{u}\|^2};P_{\mathbf{u}|\|\mathbf{u}\|=\rho}\right]\\ &=\frac{2I_p(\kappa)}{\kappa I_{p-1}(\kappa)}\mathbf{I}+\frac{I_{p+1}(\kappa)}{I_{p-1}(\kappa)}\frac{\boldsymbol{\mu}\boldsymbol{\mu}^H}{\|\boldsymbol{\mu}\|^2}\;=\;\frac{2I_p(\kappa)}{\kappa I_{p-1}(\kappa)}\mathbf{I}\\ &\quad+\frac{4I_{p+1}(\kappa)}{\kappa^2 I_{p-1}(\kappa)}\frac{\|\mathbf{q}\|^2}{\sigma^4}\boldsymbol{\mu}\boldsymbol{\mu}^H.\end{aligned} \tag{73}$$

The proof of the Lemma follows directly from (65), (66) and (73). □

Proof of Theorem 2. By the definition of the MT-covariance (3), relation (63), and the assumption of zero MT-mean, the MT-covariance takes the form:

$$\begin{aligned}\boldsymbol{\Sigma}_{\mathbf{x}}^{(u)}&=\mathrm{E}[\mathbf{x}\mathbf{x}^H\varphi_h(\|\mathbf{x}\|);P_{\mathbf{x}}]\\ &=\mathrm{E}\left[\mathrm{E}\left[\mathbf{x}\mathbf{x}^H;P_{\mathbf{x}|\boldsymbol{S},\tau,\|\mathbf{x}\|}\right]\varphi_h(\|\mathbf{x}\|);P_{\boldsymbol{S},\tau,\|\mathbf{x}\|}\right],\end{aligned} \tag{74}$$

where $\varphi_h(\cdot)$ is defined below (64), and $P_{\cdot|\cdot}$ denotes a conditional probability measure. The second equality in (74) follows directly

from the law of total expectation. Notice that by (1) and (19) the random vector $\mathbf{x}|\boldsymbol{s},\tau \sim \mathcal{CN}\left(\mathbf{A}\boldsymbol{s},\tau^2\sigma_{\mathbf{w}}^2\mathbf{I}_p\right)$. Thus, by Lemma 6 stated above, we obtain that the conditional expectation

$$\mathrm{E}\left[\mathbf{x}\mathbf{x}^H;P_{\mathbf{x}|\boldsymbol{S},\tau,\|\mathbf{x}\|}\right]=\beta_p(\kappa)\frac{\|\mathbf{x}\|^4}{\tau^4\sigma_{\mathbf{w}}^4}\mathbf{A}\boldsymbol{s}\boldsymbol{s}^H\mathbf{A}^H+\gamma_p(\kappa)\|\mathbf{x}\|^2\mathbf{I}_p, \tag{75}$$

where $\beta_p(\cdot)$, $\gamma_p(\cdot)$ and κ are defined below (64). Finally, relation (64) is obtained by substituting (75) into (74).

We now prove that the non-negative matrix $\mathbf{J}^{(u)}$ defined below (64) is strictly positive-definite, i.e., non-singular. We show that if $\mathbf{J}^{(u)}$ is singular, then the signal covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{S}}\triangleq\mathrm{E}\left[\boldsymbol{s}\boldsymbol{s}^H;P_{\boldsymbol{S}}\right]$, which under the array model (1) is assumed to be non-singular, must be singular. Under the assumption that $\mathbf{J}^{(u)}$ is singular, there exists a deterministic non-zero vector $\mathbf{v}\in\mathbb{C}^p$ such that

$$\mathbf{v}^H\mathbf{J}^{(u)}\mathbf{v}=\mathrm{E}\left[\left|\mathbf{v}^H\boldsymbol{s}\right|^2\beta_p(\kappa)\frac{\|\mathbf{x}\|^4}{\tau^4\sigma_{\mathbf{w}}^4}\varphi_h(\|\mathbf{x}\|);P_{\mathbf{x},\boldsymbol{S},\kappa}\right]=0. \tag{76}$$

Notice that by (1) and the CG noise (19) the observation vector $\mathbf{x}$ is a continuous random vector, and therefore, $\|\mathbf{x}\|>0$ w.p. 1. Also note that $\varphi_h(\cdot)$, τ and $\sigma_{\mathbf{w}}^4$ are strictly positive. Hence, by (76) and Proposition 2.3.9 in [25], we conclude that $|\mathbf{v}^H\boldsymbol{s}|^2\beta_p(\kappa)=0$ w.p. 1. In the following, we show that the latter equality implies that $|\mathbf{v}^H\boldsymbol{s}|^2=0$ w.p. 1. Clearly, if $\beta_p(\kappa)>0$ then $|\mathbf{v}^H\boldsymbol{s}|^2=0$. Now, if $\beta_p(\kappa)=0$, it follows from the definition of $\beta_p(\kappa)$ that $\kappa=0$. Therefore, by the definition of κ, below (64), we conclude that since $\|\mathbf{x}\|$ and τ are strictly positive w.p. 1 and $\sigma_{\mathbf{w}}^4>0$, then $\|\mathbf{A}\boldsymbol{s}\|=0$ w.p. 1. Thus, since the steering matrix $\mathbf{A}$ has a full column rank, we conclude that when $\kappa=0$ the signal vector $\boldsymbol{s}=\mathbf{0}$ w.p. 1. Therefore,

$$\mathbf{v}^H\boldsymbol{\Sigma}_{\boldsymbol{S}}\mathbf{v}=\mathrm{E}\left[\left|\mathbf{v}^H\boldsymbol{s}\right|^2;P_{\boldsymbol{S}}\right]=0,$$

which implies that $\boldsymbol{\Sigma}_{\boldsymbol{S}}$ is singular. □

References

[1] R.O. Schmidt, Multiple emitter location and signal parameter estimation, IEEE Trans. Antennas Propag. 34 (3) (1986) 276–280.
[2] H. Krim, M. Viberg, Two decades of array signal processing research: the parametric approach, IEEE Signal Process. Mag. 13 (4) (1996) 67–94.
[3] T.W. Anderson, An Introduction to Multivariate Statistical Analysis, Wiley, 2003.
[4] S. Visuri, H. Oja, V. Koivunen, Subspace-based direction-of-arrival estimation using nonparametric statistics, IEEE Trans. Signal Process. 49 (9) (2001) 2060–2073.
[5] E. Ollila, D.E. Tyler, V. Koivunen, H.V. Poor, Complex elliptically symmetric distributions: survey, new results and applications, IEEE Trans. Signal Process. 60 (1) (2012) 5597–5625.
[6] R.A. Maronna, Robust m-estimators of multivariate location and scatter, Ann. Stat. 4 (1) (1976) 51–67.
[7] P.J. Huber, Robust Statistics, Wiley, 1981.
[8] D.E. Tyler, A distribution-free m-estimator of multivariate scatter, Ann. Stat. 15 (1) (1987) 234–251.
[9] R. Couillet, A. Kammoun, Robust g-MUSIC, in: Proceedings of EUSIPCO, 2014, pp. 2155–2159.
[10] R. Couillet, Robust spiked random matrices and a robust g-MUSIC estimator, J. Multivariate Anal. 140 (2015) 139–161.
[11] P. Tsakalides, C.L. Nikias, The robust covariation-based MUSIC (ROC-MUSIC) algorithm for bearing estimation in impulsive noise environments, IEEE Trans. Signal Process. 44 (7) (1996) 1623–1633.
[12] T.H. Liu, J.M. Mendel, A subspace-based direction finding algorithm using fractional lower order statistics, IEEE Trans. Signal Process. 49 (8) (2001) 1605–1613.
[13] A. Swami, B.M. Sadler, On some detection and estimation problems in heavy–tailed noise, Signal Process. 82 (12) (2002) 1829–1846.
[14] C.H. Lim, S.C.-M. See, A.M. Zoubir, B.P. Ng, Robust adaptive trimming for high-resolution direction finding, IEEE Signal Process. Lett. 16 (7) (2009) 580–583.
[15] W.J. Zeng, H.C. So, L. Huang, l_p-MUSIC: robust direction-of-arrival estimator for impulsive noise environments, IEEE Trans. Signal Process. 61 (17) (2013) 4296–4308.
[16] K. Todros, A.O. Hero, Robust multiple signal classification via probability measure transformation, IEEE Trans. Signal Process. 63 (5) (2015) 1156–1170.
[17] K. Todros, A.O. Hero, On measure transformed canonical correlation analysis, IEEE Trans. Signal Process. 60 (9) (2012) 4570–4585.
[18] K. Todros, A.O. Hero, Robust measure transformed MUSIC for DOA estimation, in: Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014, pp. 4190–4194.
[19] K. Todros, A.O. Hero, Measure-transformed quasi-maximum likelihood estimation, IEEE Trans. Signal Process. 65 (3) (2017) 748–763.
[20] N. Halay, K. Todros, A.O. Hero, Binary hypothesis testing via measure transformed quasi likelihood ratio test, IEEE Trans. Signal Process. 65 (24) (2017) 6381–6396.
[21] N. Halay, K. Todros, Plug-in measure-transformed quasi-likelihood ratio test for random signal detection, IEEE Signal Process. Lett. 24 (6) (2017) 838–842.
[22] K. Todros, Measure-transformed Gaussian quasi score test, 2017, Proceedings of EUSIPCO,
[23] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, W.A. Stahel, Robust Statistics: the Approach Based on Influence Functions, John Wiley & Sons, 2011.
[24] G.B. Folland, Real Analysis, John Wiley and Sons, 1984.
[25] K.B. Athreya, S.N. Lahiri, Measure Theory and Probability Theory, Springer-Verlag, 2006.
[26] P. Stoica, A. Nehorai, MUSIC, maximum likelihood, and cramér-rao bound, IEEE Trans. Acoust., Speech, Signal Process. 37 (5) (1989) 720–741.
[27] J.F. Cardoso, E. Moulines, Asymptotic performance analysis of direction-finding algorithms based on fourth-order cumulants, IEEE Trans. Signal Process. 43 (1) (1995) 214–224.
[28] J.P. Delmas, Performance bounds and statistical analysis of DOA estimation, Academic Press Library in Signal Processing, Vol. 3, 1st Edition, Array and Statistical Signal Processing, Elsevier, 2013.
[29] J.P. Delmas, Asymptotic performance of second-order algorithms, IEEE Trans. Signal Process. 50 (1) (2002) 49–57.
[30] S. Boydand, L. Vandenberghe, Convex Optimization, Cambridge Univ. Press, 2004.
[31] E. Conto, M. Longo, Characterization of radar clutter as a spherically invariant random process, IEE Proc. F (Commun. Radar Signal Process.) 134 (2) (1987) 191–197.
[32] F. Gini, A. Farina, Vector subspace detection in compound-gaussian clutter. Part I: survey and new results, IEEE Trans. Aerosp. Electron. Syst. 38 (4) (2002) 1295–1311.
[33] M. Rangaswamy, Spherically invariant random processes for modelling non–gaussian radar clutter, in: Proceedings of Twenty-Seventh Asilomar Conference Signals, Systems and Computers, 1993, pp. 1106–1110.
[34] E. Ollila, D.E. Tyler, V. Koivunen, H.V. Poor, Compound-Gaussian clutter modelling with an inverse gaussian texture distribution, IEEE Signal Process. Lett. 19 (12) (2012) 876–879.
[35] A. De Maio, S.M. Greco, Modern Radar Detection Theory, The Institution of Engineering and Technology, United Kingdom (U.K.), 2016.
[36] 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.
[37] S.M. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory, Prentice-Hall, 1993.
[38] R.W. Miller, C.B. Chang, A modified Cramér–Rao bound and its applications, IEEE Trans, Inf. Theory 24 (1978) 398–400.
[39] O. Besson, Y. Abramovich, B. Johnson, Direction-of-arrival estimation in a mixture of k-distributed and gaussian noise, Signal Process. 128 (2016) 512–520.
[40] C.H. Edwards, Advanced Calculus of Several Variables, Dover, 1973.
[41] P. Schreier, L.L. Scharf, Statistical Signal Processing of Complex-Valued Data, Cambridge University Press, 2010.
[42] A.W.V.d. Vaart, Asymptotic statistics, Cambridge university press, 2000.
[43] E.L. Lehmann, J.P. Romano, Testing Statistical Hypotheses, Springer Texts in Statistics, 2005.
[44] J.R. Magnus, H. Neudecker, Matrix differential calculus with applications in statistics and econometrics, Wiley, 1988.
[45] H.B. Mann, A. Wald, On stochastic limit and order relationships, Ann. Math. Stat. 14 (1943) 217–226.
[46] R.J. Serfling, Approximation Theorems of Mathematical Statistics, John Wiley & Sons, 1980.
[47] M. Kaveh, A.J. Barabell, The statistical performance of the MUSIC and the minimum-norm algorithms in resolving plane waves in noise, IEEE Trans. Acoust., Speech, Signal Proc. 34 (2) (1986) 331–341.
[48] J.P. Delmas, Y. Meurisse, On the second-order statistics of the EVD of sample covariance matrices application to the detection of noncircular or/and nongaussian components, IEEE Trans. Signal Process. 59 (8) (2011) 4017–4023.
[49] J.H. Wilkinson, The Algebraic Eigenvalue Problem, Oxford University Press, 1965.
[50] X. Yang, Some trace inequalities for operators, J. Austr. Math. Soc.-Ser. A 58 (2) (1995) 281.
[51] G.H. Golub, C.F.V. Loan, Matrix Computations, 3rd ed., Johns Hopkins, Baltimore, MD, 1996.
[52] P. Vallet, X. Mestre, P. Loubaton, Performance analysis of an improved MUSIC DoA estimator, IEEE Trans. Signal Process. 63 (23) (2015) 6407–6422.
[53] M. Abramowitz, I.A. Stegun, Handbook of mathematical functions with formulas, graphs, and mathematical tables, Vol. 9, Dover, 1972.
[54] K.V. Mardia, Characterizations of directional distributions, vol. 3, pp. 365–385, Springer, 1975.
[55] K.V. Mardia, P.E. Jupp, Directional Statistics, John Wiley & Sons, 2000.
[56] F.W. Olver, D.W. Lozier, R.F. Boisvert, C.W. Clark, NIST Handbook of Mathematical Functions, Cambridge University Press, 2010.
[57] M. Giaquinta, G. Modica, Mathematical analysis: an introduction to functions of several variables, 2009. Birkhäuser, p. 88, Boston.