Robust estimation of structured scatter matrices in (mis)matched models

Bruno Mériaux, Chengfang Ren, Mohammed Nabil El Korso, Arnaud Breloy, Philippe Forster

Robust Estimation of Structured Scatter Matrices in (Mis)matched Models

Bruno Mériaux[a,*], Chengfang Ren[a], Mohammed Nabil El Korso[b], Arnaud Breloy[b], Philippe Forster[c]

[a] *SONDRA, CentraleSupélec, 91192 Gif-sur-Yvette, France*
[b] *Paris Nanterre University, LEME EA-4416, 92410 Ville d'Avray, France*
[c] *Paris Nanterre University, SATIE, 94230 Cachan, France*

Abstract

Covariance matrix estimation is a ubiquitous problem in signal processing. In most modern signal processing applications, data are generally modeled by non-Gaussian distributions with covariance matrices exhibiting a particular structure. Taking into account this structure and the non-Gaussian behavior improve drastically the estimation accuracy. In this paper, we consider the estimation of structured scatter matrix for complex elliptically distributed observations, where the *assumed* model can differ from the actual distribution of the observations. Specifically, we tackle this problem, in a mismatched framework, by proposing a novel estimator, named StructurEd ScAtter Matrix Estimator (SESAME), which is based on a two-step estimation procedure. We conduct theoretical analysis on the unbiasedness and the asymptotic efficiency and Gaussianity of SESAME. In addition, we derive a recursive estimation procedure that iteratively applies the SESAME method, called Recursive-SESAME (R-SESAME), reaching with improved performance at lower sample support the (Mismatched) Cramér-Rao Bound. Furthermore, we show that some special cases of the proposed method allow to retrieve preexisting methods. Finally, numerical results corroborate the theoretical analysis and assess the usefulness of the proposed algorithms.

Keywords: Structured covariance matrix, scatter matrix estimation, Complex Elliptically Symmetric distributions, mismatched framework.

*Corresponding author
Email address: bruno.meriaux@centralesupelec.fr (Bruno Mériaux)

1. Introduction

Covariance Matrix (CM) estimation turns out to be a crucial step in most existing algorithms used in various domains such as adaptive signal processing, financial engineering, communication systems [1, 2]. By definition, a CM is Hermitian (symmetric for real random variables) and belongs to the set of Positive Semi-Definite (PSD) matrices. However, in many applications, the CM naturally holds also a specific structure such as Toeplitz or Kronecker product [3, 4, 5, 6]. Taking into account this structure in the estimation problem usually implies a smaller parameter of interest vector to be estimated, and thus leads theoretically to a better estimation accuracy. The structured estimation problem has been investigated for various types of structures. For instance, the Toeplitz structure appears in array processing with Uniform Linear Array (ULA) or in time series analysis [3, 4]. In MIMO communications or spatio-temporal noise processes in MEG/EEG data, the CM exhibits a Kronecker structure, where the factor matrices could be themselves structured [5, 6]. In some applications, the CM lies in a low-dimensional subspace [7, 8].

In the afore-mentioned works the CM estimation is unequivocally improved, when the prior structure is considered. However, they usually assume complex Gaussian distributed samples. Then, the structured CM estimation is addressed either by projecting the Sample Covariance Matrix onto a subset describing a structure [9, 10] or by deriving the Maximum Likelihood (ML) estimator under structure constraints. In some practical applications, performance is degraded because the assumption of Gaussian distribution leads to non-robustness, either to heavy-tailed distributed data or to outliers in the sample set. In order to overcome this issue, a wide class of distribution free methods based on the unstructured Tyler's estimate has been proposed [11, 12, 13, 14]. Those methods begin by normalizing the zero mean observations to get rid of the potential power fluctuations. Specifically, in [11], a robust extension of the COvariance Matching Estimation Technique (COMET) procedure is derived. In [12, 13], estimators have been proposed which minimize a constrained version of Tyler's cost function using iterative Majorization-Minimization algorithms. In [14], a COnvexly ConstrAined (COCA) CM estimator is presented, based on the generalized Method

of Moments (MoM) for Tyler's estimate subject to convex constraints. An alternative approach is to model data following a Complex Elliptically Symmetric (CES) distribution [15] in order to take into account prior information on the scaling factor (also called "texture" in radar context). The main reason lies in the fact that the CES distributions encompass a large number of distributions such as Gaussian, Generalized Gaussian, compound Gaussian, t-distribution, W-distribution and K-distribution, just to cite a few examples. Those distributions turn out to fit accurately spiky radar clutter measurements or other heavy-tailed observations [16, 17, 18].

In most estimation problems, a key assumption concerns the "good" knowledge of the data model. In other words, if the *assumed* model to derive an estimation procedure coincides with the *true* model of the observations, it is referred to as the matched case. However, some misspecifications are often unavoidable in practice, due to imperfect knowledge of the *true* data model for instance. In the literature, the misspecified model framework has been investigated, notably the behavior of the ML estimator under mistmached conditions [19, 20]. More recently, some classical tools to conduct asymptotic analysis have been extended to the misspecified context. More specifically for CES distributions, a lower bound on the error covariance matrix of mismatched estimators has been derived, leading to the Misspecified Cramér Rao Bound (MCRB) [21, 22, 23].

In this paper, we introduce a StructurEd ScAtter Matrix Estimator (SESAME), in the mismatched framework, for any given CES distribution whose scatter matrix owns a convex structure. This method is carried out in two steps. SESAME combines an unstructured ML-estimate of the scatter matrix and the minimization of an appropriate criterion from the previous estimate. A theoretical analysis of the proposed SESAME's asymptotic performance (weak consistency, bias, efficiency and asymptotic distribution) is conducted. Also, we propose a recursive generalization of SESAME that leads to a second method, called Recursive-SESAME (R-SESAME), which outperforms SESAME in the sense that the asymptotic behavior is reached for lower sample supports. In addition, we relate SESAME to two special cases. The first one lies on a Gaussian *assumed* model, then the so-called COMET procedure from [10] is a special case of SESAME. The second one is the matched

case, which concurs with the EXtended Invariance Principle (EXIP) [24] applied to CES distributions. In [25] the derivation is made for the particular case of the t-distribution.

The paper is organized as follows. In Section 2, a brief review of the CES distributions and robust scatter matrix estimators is presented. Section 3 focuses on the proposed estimator. The theoretical performance analysis is conducted in Section 4. An improvement of the first proposed algorithm, called R-SESAME, is presented in Section 5. Two special cases of SESAME are considered in Section 6, which are respectively related to the COMET method from [10] and the EXIP from [24]. Numerical results illustrate the previous theoretical analysis in Section 7 and we conclude in Section 8.

In the following, vectors (respectively matrices) are denoted by boldface lowercase letters (respectively uppercase letters). The k-th element of a vector $\mathbf{a}$ is referred to as a_k and $[\mathbf{A}]_{k,\ell}$ stands for the (k,ℓ) element of the matrix $\mathbf{A}$. The notation $\overset{d}{=}$ indicates "has the same distribution as". Convergence in distribution and in probability are, respectively, denoted by $\xrightarrow{d}$ and $\xrightarrow{\mathcal{P}}$. The used concept of consistency refers to the weak consistency, i.e., with a convergence in probability. For a matrix $\mathbf{A}$, $|\mathbf{A}|$ and $\mathrm{Tr}\left(\mathbf{A}\right)$ denote the determinant and the trace of $\mathbf{A}$. $\mathbf{A}^T$ (respectively $\mathbf{A}^H$ and $\mathbf{A}^*$) stands for the transpose (respectively conjugate transpose and conjugate) matrix. $\mathbf{I}_m$ is the identity matrix of size m. The vec-operator $\mathrm{vec}(\mathbf{A})$ stacks all columns of $\mathbf{A}$ into a vector. The operator $\otimes$ refers to Kronecker matrix product and finally, the subscript "e" refers to the true value. The notations $\ker\left(\mathbf{A}\right)$, $\mathrm{rank}\left(\mathbf{A}\right)$ and $\mathrm{span}\left(\mathbf{A}\right)$ denote the null-space, the rank of $\mathbf{A}$ and the subspace spanned by the columns of $\mathbf{A}$. The operator $\mathbb{E}_p\left[\cdot\right]$ refers to the expectation operator w.r.t. the probability density function (p.d.f.) p, which will be omitted if there is no ambiguity. Finally, the set of non-negative real numbers is denoted by $\mathbb{R}^+$.

2. Background and Problem Setup

In this section, we introduce the CES distribution model and some robust scatter matrix estimators. A more substantial survey on CES distribution can be found in [15].

2.1. CES distribution

A m-dimensional random vector (r.v.), $\mathbf{y} \in \mathbb{C}^m$ follows a CES distribution if and only if it admits the following stochastic representation [15]:

$$\mathbf{y} \stackrel{d}{=} \mathbf{m} + \sqrt{Q}\mathbf{A}\mathbf{u}, \tag{1}$$

where $\mathbf{m} \in \mathbb{C}^m$ is the location parameter and the non-negative real-valued random variable Q, called the 2nd-order modular variate, is independent of the complex r.v. $\mathbf{u}$. The latter r.v. is uniformly distributed on the unit k-sphere $\mathbb{C}S^k \triangleq \left\{\mathbf{z} \in \mathbb{C}^k \mid \|\mathbf{z}\| = 1\right\}$ with $k \leq m$, denoted by $\mathbf{u} \sim \mathcal{U}\left(\mathbb{C}S^k\right)$. The matrix $\mathbf{A} \in \mathbb{C}^{m\times k}$ has $\text{rank}(\mathbf{A}) = k$. Furthermore, if they exist, the mean of $\mathbf{y}$ is equal to $\mathbf{m}$ and the covariance matrix of $\mathbf{y}$ is proportional to the scatter matrix, $\mathbf{R} = \mathbf{A}\mathbf{A}^H$, more precisely $\mathbb{E}\left[(\mathbf{y}-\mathbf{m})(\mathbf{y}-\mathbf{m})^H\right] = \frac{1}{k}\mathbb{E}\left[Q\right]\mathbf{R}$. In the following, we assume that $\mathbf{m}$ is known and equal to zero, without loss of generality and that $k = m$, so $\text{rank}(\mathbf{R}) = m$ to belong to the absolutely continuous case. In this case, the p.d.f. of such a vector exists and can be written as [15]:

$$p_{\mathbf{Y}}(\mathbf{y}; \mathbf{R}, g) = C_{m,g}|\mathbf{R}|^{-1} g\left(\mathbf{y}^H\mathbf{R}^{-1}\mathbf{y}\right), \tag{2}$$

in which the function $g : \mathbb{R}^+ \to \mathbb{R}^+$ is called the density generator and satisfies $\delta_{m,g} \triangleq \int_0^{+\infty} t^{m-1}g(t)\mathrm{d}t < \infty$ and $C_{m,g} = \frac{\Gamma(m)}{\pi^m}\delta_{m,g}^{-1}$ is the normalizing constant. In this case, we denote $\mathbf{y} \sim \mathbb{C}\mathcal{ES}_m\left(\mathbf{0}, \mathbf{R}, g\right)$ in short. Furthermore, thanks to (1), the quadratic form $\mathbf{y}^H\mathbf{R}^{-1}\mathbf{y} \stackrel{d}{=} Q$ has the following p.d.f.:

$$p_Q\left(q\right) = \delta_{m,g}^{-1} q^{m-1} g\left(q\right). \tag{3}$$

As already indicated, numerous non-Gaussian distributions (e.g., Generalized Gaussian, t-distribution, K-distribution, etc.) belong to the family of CES distributions. The expression of the density generator function for these commonly used CES distributions is given in Table 1.

2.2. M-estimators

Let us consider N i.i.d. zero mean CES distributed observations, $\mathbf{y}_n \sim \mathbb{CES}_m(\mathbf{0}, \mathbf{R}, g)$, $n = 1, \ldots, N$, with $N > m$. A complex M-estimator of the scatter matrix, $\mathbf{R}$, is defined as the solution to the following fixed-point equation [15]:

$$\widehat{\mathbf{M}} = \frac{1}{N}\sum_{n=1}^{N} u\left(\mathbf{y}_n^H \widehat{\mathbf{M}}^{-1}\mathbf{y}_n\right)\mathbf{y}_n\mathbf{y}_n^H, \quad \text{and} \quad \widehat{\mathbf{R}} \propto \widehat{\mathbf{M}}. \tag{4}$$

In the case of the ML estimator, the function $u(\cdot)$ is given by $u_{\text{ML}}(s) = -\dfrac{g'(s)}{g(s)}$, where $g'(\cdot)$ refers to the derivative of $g(\cdot)$. In the case of unknown density generator function $g(\cdot)$, the previous function $u_{\text{ML}}(\cdot)$ can be replaced by another, $u(\cdot)$, which satisfies a set of general assumptions to ensure the existence and uniqueness of the solution[1] of (4). For convenience, let us consider the following equation, which expresses the limit towards which the equation (4) converges when N tends to infinity

$$\mathbf{M} = \mathbb{E}\left[u\left(\mathbf{y}^H\mathbf{M}^{-1}\mathbf{y}\right)\mathbf{y}\mathbf{y}^H\right], \tag{5}$$

where $\mathbf{y} \sim \mathbb{CES}_m(\mathbf{0}, \mathbf{R}, g)$. Under some mild assumption [15, 26, 27], (5) (respectively (4)) admits a unique solution $\mathbf{M}$ (respectively $\widehat{\mathbf{M}}$) and

$$\mathbf{M} = \sigma^{-1}\mathbf{R},$$

where σ is the solution of $\mathbb{E}\left[\psi(\sigma|\mathbf{t}|^2)\right] = m$ with $\mathbf{t} \sim \mathbb{CES}_m(\mathbf{0}, \mathbf{I}, g)$ and $\psi(s) = su(s)$. In addition, the estimator $\widehat{\mathbf{M}}$ can be easily obtained by an iterative procedure and it has been shown to be consistent w.r.t. $\mathbf{M}$. Finally, the asymptotic distribution of $\widehat{\mathbf{M}}$ has been established for the complex case in [15, 28] and reads:

$$\sqrt{N}\text{vec}\left(\widehat{\mathbf{M}} - \mathbf{M}\right) \xrightarrow{d} \mathcal{GCN}(\mathbf{0}, \mathbf{\Sigma}, \mathbf{\Omega}), \tag{6}$$

[1] These conditions have been first studied in the real case by [26, 27] then generalized for the complex case in [15]. They hold for $N > m$ and the usual distributions reported in Table 1. (See Theorems 6 and 7 and Section V.B. of [15] for more details).

where $\mathcal{GCN}(\mathbf{0}, \mathbf{\Sigma}, \mathbf{\Omega})$ denotes the zero mean non-circular complex Gaussian distribution [29] in which

$$\begin{cases} \mathbf{\Sigma} = \sigma_1 \mathbf{M}^T \otimes \mathbf{M} + \sigma_2 \text{vec}(\mathbf{M}) \text{vec}(\mathbf{M})^H, \\ \\ \mathbf{\Omega} = \sigma_1 \left(\mathbf{M}^T \otimes \mathbf{M}\right) \mathbf{K} + \sigma_2 \text{vec}(\mathbf{M}) \text{vec}(\mathbf{M})^T = \mathbf{\Sigma}\mathbf{K}, \end{cases} \tag{7}$$

and

$$\begin{cases} \sigma_1 = \dfrac{a_1(m+1)^2}{(a_2+m)^2}, \\ \\ \sigma_2 = \dfrac{1}{a_2^2}\left[(a_1-1) - a_1(a_2-1)\dfrac{m+(m+2)a_2}{(a_2+m)^2}\right], \end{cases} \tag{8}$$

where $\mathbf{K}$ is the commutation matrix, which satisfies $\mathbf{K}\text{vec}(\mathbf{A}) = \text{vec}\left(\mathbf{A}^T\right)$ [30], and the coefficients a_1 and a_2 are defined by:

$$\begin{cases} a_1 = \dfrac{1}{m(m+1)} \mathbb{E}\left[\psi^2\left(\sigma|\mathbf{t}|^2\right)\right], \\ a_2 = \dfrac{1}{m} \mathbb{E}\left[\sigma|\mathbf{t}|^2 \psi'\left(\sigma|\mathbf{t}|^2\right)\right]. \end{cases} \tag{9}$$

We remark that $|\mathbf{t}|^2 = \mathbf{t}^H\mathbf{t} \overset{d}{=} Q$, thus in the following, we note $A = \mathbb{E}\left[\psi^2\left(\sigma|\mathbf{t}|^2\right)\right] = \mathbb{E}\left[\psi^2\left(\sigma Q\right)\right]$ and $B = \mathbb{E}\left[\sigma|\mathbf{t}|^2\psi'\left(\sigma|\mathbf{t}|^2\right)\right] = \mathbb{E}\left[\sigma Q \psi'\left(\sigma Q\right)\right]$. The subscript $_{\text{ML}}$ is used when A and B are computed with $\psi_{\text{ML}}(s) = s u_{\text{ML}}(s)$. Regarding the ML-estimator, we easily obtain $\sigma = 1$. Indeed, as $\int_0^{+\infty} t^{m-1} g(t) \mathrm{d}t < \infty$, i.e., $g(t) \underset{+\infty}{\propto} t^{-m-\alpha}$, $\alpha > 0$, we obtain

$$\begin{aligned} \mathbb{E}\left[\psi_{\text{ML}}(Q)\right] &= -\int_{\mathbb{R}^+} q \frac{g'(q)}{g(q)} \delta_{m,g}^{-1} q^{m-1} g(q) \, \mathrm{d}q \\ &= -\delta_{m,g}^{-1}\Big[g(q)q^m\Big]_0^{+\infty} + \delta_{m,g}^{-1} \int_{\mathbb{R}^+} m g(q) q^{m-1} g(q) \, \mathrm{d}q \\ &= \delta_{m,g}^{-1} \int_{\mathbb{R}^+} m g(q) q^{m-1} g(q) \, \mathrm{d}q = m. \end{aligned}$$

Similarly, we can show that $B_{\text{ML}} = A_{\text{ML}} - m^2$, thus the coefficients σ_1 and σ_2, which fully describe the asymptotic (pseudo)-covariance matrices of the scatter ML estimator, can be

	Gaussian $\mathbb{C}\mathcal{N}_m$	Generalized Gaussian $\mathbb{C}\mathcal{GN}_{m,s,b}$	Student $\mathbb{C}t_{m,d}$	W-distribution $\mathbb{C}\mathcal{W}_{m,s,b}$	K-distribution $\mathbb{C}\mathcal{K}_{m,\nu}$
$g(t)$	$\exp(-t)$	$\exp(-t^s/b)\ \ s,b>0$	$(1+t/d)^{-(d+m)}\ \ d>0$	$t^{s-1}\exp(-t^s/b)\ \ s,b>0$	$\sqrt{t}^{\nu-m}K_{\nu-m}\left(2\sqrt{\nu t}\right)\ \ \nu>0$
$C_{m,g}$	π^{-m}	$\frac{s\Gamma(m)b^{-m/s}}{\pi^m\Gamma(m/s)}$	$\frac{\Gamma(m+d)}{\pi^m d^m\Gamma(d)}$	$\frac{s\Gamma(m)b^{-(m+s-1)/s}}{\pi^m\Gamma((m+s-1)/s)}$	$2\frac{\nu^{(\nu+m)/2}}{\pi^m\Gamma(\nu)}$
A_{ML}	$m(m+1)$	$m(m+s)$	$\frac{m(m+1)(m+d)}{d+m+1}$	$s(m+s-1)+m^2$	$\frac{2^{-(m+\nu)}}{\Gamma(\nu)\Gamma(m)}\int_{\mathbb{R}^+}x^{m+\nu+1}\frac{K^2_{\nu-m-1}(x)}{K_{\nu-m}(x)}\mathrm{d}x$
$\sigma_{1,\mathrm{ML}}$	1	$\frac{m+1}{m+s}$	$\frac{d+m+1}{d+m}$	$\frac{m(m+1)}{s(m+s-1)+m^2}$	No closed form, numerical evaluation from (10)
$\sigma_{2,\mathrm{ML}}$	0	$\frac{1-s}{s(m+s)}$	$\frac{d+m+1}{d(d+m)}$	$\frac{m(1-s)(m+s)}{s(m+s-1)(s(m+s-1)+m^2)}$	

$K_\lambda(\cdot)$ refers to the modified Bessel function of the second kind [31].

Table 1: Examples with common CES distributions

simplified by

$$\sigma_{1,\mathrm{ML}}=\frac{m(m+1)}{A_{\mathrm{ML}}}\ \text{ and }\ \sigma_{2,\mathrm{ML}}=\frac{-\sigma_{1,\mathrm{ML}}(1-\sigma_{1,\mathrm{ML}})}{1+m(1-\sigma_{1,\mathrm{ML}})}. \tag{10}$$

For the common CES distributions already mentioned in Table 1, there are explicit expressions for the constant A_{ML} and the coefficients $\sigma_{1,\mathrm{ML}}$ and $\sigma_{2,\mathrm{ML}}$ related to the scatter matrix unstructured ML-estimator. The latter are needed for the derivation of the proposed algorithms. All the results are summed up in Table 1 for a centered m-dimensional complex random vector. The guidelines for the above calculations are given in Appendix A.

2.3. Problem setup

Let us consider N i.i.d. zero mean CES distributed observations, $\mathbf{y}_n \sim \mathbb{C}\mathcal{ES}_m(\mathbf{0},\mathbf{R}_{\mathrm{e}},g)$, $n=1,\ldots,N$, where the function $g(\cdot)$ characterizes the *true* but *unknown* distribution. The *true* p.d.f. is denoted by $p_{\mathbf{Y}}(\mathbf{y}_n;\mathbf{R}_{\mathrm{e}})$. In the following, the only assumption made on the distribution of the data is the belonging to the class of zero mean CES distributions. Consequently, we assume that the observations, $\mathbf{y}_1,\ldots,\mathbf{y}_N$, are sampled from a CES distribution with a density generator $g_{\mathrm{mod}}(t)$, possibly different from $g(t)$ for all $t\in\mathbb{R}^+$, specifically the observations are assumed to follow $\mathbb{C}\mathcal{ES}_m(\mathbf{0},\mathbf{R},g_{\mathrm{mod}})$. The *assumed* p.d.f. is then denoted by $f_{\mathbf{Y}}(\mathbf{y}_n;\mathbf{R})$. Note that the mismatch only concerns the density generator function, in this paper. Furthermore, we assume that the scatter matrix belongs to a convex subset $\mathscr{S}$ of Hermitian matrices (e.g., Toeplitz, persymmetric, banded,...), for which there exists a one-to-one differentiable mapping $\boldsymbol{\mu}\mapsto\boldsymbol{\mathcal{R}}(\boldsymbol{\mu})$ from $\mathbb{R}^P$ to $\mathscr{S}$. Thus, the unknown parameter

of interest is the vector $\boldsymbol{\mu}$ with exact value $\boldsymbol{\mu}_{\mathrm{e}}$, and $\mathbf{R}_{\mathrm{e}} = \boldsymbol{\mathcal{R}}(\boldsymbol{\mu}_{\mathrm{e}})$ corresponds to the true structured scatter matrix. Note that there is no *assumed* misspecification of the scatter matrix parameterization, since the latter is mostly derived from physical considerations on the measurement system (e.g., Toeplitz structure for a uniform linear array). The *assumed* log-likelihood function is given, up to an additive constant, by

$$\mathcal{L}(\mathbf{y}_1, \ldots, \mathbf{y}_N; \boldsymbol{\mu}) \triangleq \sum_{n=1}^{N} \log f_{\mathbf{Y}}(\mathbf{y}_n; \boldsymbol{\mu}) = -N \log |\boldsymbol{\mathcal{R}}(\boldsymbol{\mu})| + \sum_{n=1}^{N} \log g_{\mathrm{mod}}\left(\mathbf{y}_n^H \boldsymbol{\mathcal{R}}(\boldsymbol{\mu})^{-1} \mathbf{y}_n\right). \quad (11)$$

The above function is generally non-convex w.r.t. $\mathbf{R}$, its minimization w.r.t. $\boldsymbol{\mu}$ is therefore a laborious and computationally prohibitive problem. To overcome this issue, we propose in the next section a new estimation method, which is tractable and gives unique estimates. Furthermore, for linear structures, we obtain closed form expressions of these estimates.

3. SESAME : StructurEd ScAtter Matrix Estimator

In this section, we propose a two-step estimation procedure of $\boldsymbol{\mu}$. The first step consists in computing an unstructured estimate of $\mathbf{R}_{\mathrm{e}}$, denoted by $\widehat{\mathbf{R}}_m$. The estimation of $\boldsymbol{\mu}$ is then obtained by minimizing an appropriate designed criterion based on the first-step estimate and inspired from [10, 25]. For notational convenience, we omit the dependence on N for the estimators based on N observations when there is no ambiguity.

3.1. Algorithm

Starting from the *assumed* distribution of the data, $\mathbb{C}\mathcal{ES}_m(\mathbf{0}, \boldsymbol{\mathcal{R}}(\boldsymbol{\mu}), g_{\mathrm{mod}})$, we compute first the related unstructured ML-estimator of the scatter matrix, which is the solution to the following fixed-point equation

$$\widehat{\mathbf{R}}_m = \frac{1}{N} \sum_{n=1}^{N} u_{\mathrm{mod}}\left(\mathbf{y}_n^H \widehat{\mathbf{R}}_m^{-1} \mathbf{y}_n\right) \mathbf{y}_n \mathbf{y}_n^H \triangleq \mathcal{H}_N\left(\widehat{\mathbf{R}}_m\right), \quad (12)$$

with $u_{\mathrm{mod}}(s) = -\dfrac{g'_{\mathrm{mod}}(s)}{g_{\mathrm{mod}}(s)}$. As already mentioned in the background section, the iterative algorithm $\mathbf{R}_{k+1} = \mathcal{H}_N(\mathbf{R}_k)$ converges to $\widehat{\mathbf{R}}_m$ for any initialization point with $N > m$ [15, 27].

Algorithm 1 SESAME

Input: N i.i.d. data, $\mathbf{y}_n \sim \mathbb{C}\mathcal{ES}_m(\mathbf{0}, \mathbf{R}_{\mathrm{e}}, g)$ with $N > m$

1: Compute $\widehat{\mathbf{R}}_m$ from $\mathbf{y}_1, \ldots, \mathbf{y}_N$ with (12)

2: Compute $\widehat{\mathbf{R}}$ from $\mathbf{y}_1, \ldots, \mathbf{y}_N$ (could be any consistent estimator of $\mathbf{R}_{\mathrm{e}}$ up to a scale factor, e.g., $\widehat{\mathbf{R}} = \widehat{\mathbf{R}}_m$)

3: **Return:** $\widehat{\boldsymbol{\mu}}_{\mathrm{SESAME}}$ by minimizing $\mathcal{J}_{\widehat{\mathbf{R}}_m, \widehat{\mathbf{R}}}(\boldsymbol{\mu})$ (See (13) or (14).)

Moreover, the consistency w.r.t. $\sigma^{-1}\mathbf{R}_{\mathrm{e}}$ where σ is the solution of $\mathbb{E}\left[\psi_{\mathrm{mod}}\left(\sigma|\mathbf{t}|^2\right)\right] = m$, in which $\psi_{\mathrm{mod}}(s) = s u_{\mathrm{mod}}(s)$ and $\mathbf{t} \sim \mathbb{C}\mathcal{ES}_m(\mathbf{0}, \mathbf{I}, g)$ as well as the asymptotic Gaussianity, verifying (6), of this estimator are established in [15, 28].

For the second step, we estimate $\boldsymbol{\mu}$ by minimizing the following proposed criterion $\mathcal{J}_{\widehat{\mathbf{R}}_m, \widehat{\mathbf{R}}}(\boldsymbol{\mu})$:

$$\widehat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}} \mathcal{J}_{\widehat{\mathbf{R}}_m, \widehat{\mathbf{R}}}(\boldsymbol{\mu}) \quad \text{with} \tag{13}$$

$$\mathcal{J}_{\widehat{\mathbf{R}}_m, \widehat{\mathbf{R}}}(\boldsymbol{\mu}) = \kappa_1 \mathrm{Tr}\left(\widehat{\mathbf{R}}^{-1}\left(\widehat{\mathbf{R}}_m - \boldsymbol{\mathcal{R}}(\boldsymbol{\mu})\right)\widehat{\mathbf{R}}^{-1}\left(\widehat{\mathbf{R}}_m - \boldsymbol{\mathcal{R}}(\boldsymbol{\mu})\right)\right) + \kappa_2\left[\mathrm{Tr}\left(\widehat{\mathbf{R}}^{-1}\left(\widehat{\mathbf{R}}_m - \boldsymbol{\mathcal{R}}(\boldsymbol{\mu})\right)\right)\right]^2,$$

where $\widehat{\mathbf{R}}$ refers to any consistent estimator of $\mathbf{R}_{\mathrm{e}}$ up to a scale factor, such as for instance $\widehat{\mathbf{R}}_m$, $\kappa_2 = \kappa_1 - 1$ and $\kappa_1 = \dfrac{\mathbb{E}_{f_{\mathbf{Y}}}\left[\psi^2_{\mathrm{mod}}\left(|\mathbf{t}_{\mathrm{mod}}|^2\right)\right]}{m(m+1)} \neq 0$ where $\mathbf{t}_{\mathrm{mod}} \sim \mathbb{C}\mathcal{ES}_m(\mathbf{0}, \mathbf{I}, g_{\mathrm{mod}})$. We recall that $f_{\mathbf{Y}}$ is the *assumed* p.d.f. of $\mathbf{y}_n$. The criterion $\mathcal{J}_{\widehat{\mathbf{R}}_m, \widehat{\mathbf{R}}}(\boldsymbol{\mu})$ is strongly related to the Fisher information metric derived for CES distributions in [32]. Using the relations linking the trace and the vec-operator [33, Tables 2 and 3], we rewrite (13) as

$$\begin{aligned}\widehat{\boldsymbol{\mu}} &= \arg\min_{\boldsymbol{\mu}} \left(\hat{\mathbf{r}}_m - \mathbf{r}(\boldsymbol{\mu})\right)^H \widehat{\mathbf{Y}}\left(\hat{\mathbf{r}}_m - \mathbf{r}(\boldsymbol{\mu})\right) \\ &= \arg\min_{\boldsymbol{\mu}} \left\|\widehat{\mathbf{Y}}^{1/2}\left(\hat{\mathbf{r}}_m - \mathbf{r}(\boldsymbol{\mu})\right)\right\|_2^2,\end{aligned} \tag{14}$$

with $\widehat{\mathbf{Y}} = \kappa_1 \widehat{\mathbf{W}}^{-1} + \kappa_2 \mathrm{vec}\left(\widehat{\mathbf{R}}^{-1}\right)\mathrm{vec}\left(\widehat{\mathbf{R}}^{-1}\right)^H$, $\widehat{\mathbf{W}} = \widehat{\mathbf{R}}^T \otimes \widehat{\mathbf{R}}$, $\hat{\mathbf{r}}_m = \mathrm{vec}\left(\widehat{\mathbf{R}}_m\right)$ and $\mathbf{r}(\boldsymbol{\mu}) = \mathrm{vec}\left(\boldsymbol{\mathcal{R}}(\boldsymbol{\mu})\right)$. Thus $\boldsymbol{\mathcal{R}}(\widehat{\boldsymbol{\mu}})$ leads to a structured estimate of $\mathbf{R}_{\mathrm{e}}$.

The SESAME algorithm is recapped in the box **Algorithm 1**. In practice, we choose $\widehat{\mathbf{R}} = \widehat{\mathbf{R}}_m$ and minimize the cost function $\mathcal{J}_{\widehat{\mathbf{R}}_m, \widehat{\mathbf{R}}_m}(\boldsymbol{\mu})$. However, any other consistent estimator of $\mathbf{R}_{\mathrm{e}}$ up to a scale factor can be used. This property will notably be exploited in Section 5.

Given $\widehat{\mathbf{R}}_m$ and $\widehat{\mathbf{R}}$, the function $\mathcal{J}_{\widehat{\mathbf{R}}_m,\widehat{\mathbf{R}}}(\boldsymbol{\mu})$ is convex w.r.t $\boldsymbol{\mathcal{R}}(\boldsymbol{\mu})$. Therefore, for $\mathbf{R} \in \mathscr{S}$ convex set, the minimization of (14) w.r.t. $\boldsymbol{\mathcal{R}}(\boldsymbol{\mu})$ is a convex problem that admits a unique solution. Consequently, the one-to-one mapping ensures the uniqueness of $\widehat{\boldsymbol{\mu}}$.

3.2. Practical implementation for holding the PSD constraint

Through the mapping $\boldsymbol{\mu} \mapsto \boldsymbol{\mathcal{R}}(\boldsymbol{\mu})$, SESAME ensures that the estimate $\widehat{\mathbf{R}}_S = \boldsymbol{\mathcal{R}}(\widehat{\boldsymbol{\mu}})$ belongs to the desired convex subset $\mathscr{S}$. In order to stress the reliance of the criterion $\mathcal{J}_{\widehat{\mathbf{R}}_m,\widehat{\mathbf{R}}}(\boldsymbol{\mu})$ defined in (13) on this mapping, we introduce the following notation:

$$\mathcal{C}_{\widehat{\mathbf{R}}_m,\widehat{\mathbf{R}}}\left(\boldsymbol{\mathcal{R}}\left(\boldsymbol{\mu}\right)\right) = \mathcal{J}_{\widehat{\mathbf{R}}_m,\widehat{\mathbf{R}}}(\boldsymbol{\mu}),$$

leading to the following reformulation of SESAME

$$\widehat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}} \mathcal{C}_{\widehat{\mathbf{R}}_m,\widehat{\mathbf{R}}}\left(\boldsymbol{\mathcal{R}}\left(\boldsymbol{\mu}\right)\right). \tag{15}$$

Obviously, when it comes to scatter matrix estimation, the PSD constraint has to be complied. Hence the considered subset is expressed $\mathscr{S} = \mathcal{S}_M^+ \cap \mathscr{L}$, where $\mathcal{S}_M^+$ is the cone of PSD matrices, and $\mathscr{L}$ describes the particular structure of the CM (e.g., Toeplitz, banded, persymmetric, etc.). However, in this case, the mapping $\boldsymbol{\mu} \mapsto \boldsymbol{\mathcal{R}}(\boldsymbol{\mu})$ does not usually possess an explicit expression since the PSD constraint cannot be expressed in a simple system of equations. To overcome this issue, let us consider an auxiliary mapping $\boldsymbol{\mu} \mapsto \boldsymbol{\mathcal{R}}_{\mathscr{L}}(\boldsymbol{\mu})$ so that $\boldsymbol{\mathcal{R}}_{\mathscr{L}}(\boldsymbol{\mu}) \in \mathscr{L}$, $\forall \boldsymbol{\mu}$. The most common example (which holds for the aforementioned structure) is the linear structure $\boldsymbol{\mathcal{R}}_{\mathscr{L}}(\boldsymbol{\mu}) = \sum_{p=1}^{P} \mu_p \mathbf{B}_p$, where $\{\mathbf{B}_p\}_{p=1\ldots P}$ form a basis of a linear convex set. We can then recast an equivalent SESAME as

$$\widehat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}} \mathcal{C}_{\widehat{\mathbf{R}}_m,\widehat{\mathbf{R}}}\left(\boldsymbol{\mathcal{R}}_{\mathscr{L}}\left(\boldsymbol{\mu}\right)\right) \quad \text{s.t.} \quad \boldsymbol{\mathcal{R}}_{\mathscr{L}}\left(\boldsymbol{\mu}\right) \in \mathcal{S}_M^+, \tag{16}$$

which can be obtained using standard semi-definite program solvers. Nevertheless, this approach can be computationally demanding. Therefore, it is worth noting that if the constraint is omitted, the resulting SESAME

$$\tilde{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}} \mathcal{C}_{\widehat{\mathbf{R}}_m,\widehat{\mathbf{R}}}\left(\boldsymbol{\mathcal{R}}_{\mathscr{L}}\left(\boldsymbol{\mu}\right)\right), \tag{17}$$

provides usually tractable expressions, and own closed form solutions for linear structures. The major interest of this formulation is that (16) and (17) yield the same estimate if the PSD constraint is naturally satisfied by $\mathbf{R}_{\mathscr{L}}(\tilde{\boldsymbol{\mu}})$ from (17) retrospectively. Moreover, for sufficiently large N, this latter relaxation provides a positive semi-definite matrix $\boldsymbol{\mathcal{R}}_{\mathscr{L}}(\tilde{\boldsymbol{\mu}})$ with probability arbitrarily close to one, thanks to the consistency of SESAME (see **Theorem 1.**).

In Section 7, performance of these two proposed SESAME implementations are compared. It is noted that imposing PSD constraint may be beneficial for a few number of samples, but that performance of both implementations is generally equivalent.

In the following, $\widehat{\boldsymbol{\mu}}$ will be referred to as the SESAME estimate of $\boldsymbol{\mu}$, if there is no ambiguity, and we address the study of its consistency and efficiency in Section 4.

4. Asymptotic Analysis

This section provides a statistical analysis of the proposed estimator SESAME, $\widehat{\boldsymbol{\mu}}$, which is the unique solution minimizing the criterion (13) w.r.t. $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_P)^T \in \mathbb{R}^P$ as already mentioned. Again, it is recalled that we consider a mismatched scenario, i.e. the function g_{mod} may differ from the *true* one, g.

4.1. Pseudo-parameter and consistency

Theorem 1. The SESAME estimate, $\widehat{\boldsymbol{\mu}}$, given by (13), is a consistent estimator of $\boldsymbol{\mu}_c$ such that $\text{vec}\left(\boldsymbol{\mathcal{R}}(\boldsymbol{\mu}_c)\right) = \mathbf{r}_c \triangleq \sigma^{-1}\mathbf{r}_e = \sigma^{-1}\text{vec}\left(\boldsymbol{\mathcal{R}}(\boldsymbol{\mu}_e)\right)$. Likewise, $\boldsymbol{\mathcal{R}}(\widehat{\boldsymbol{\mu}})$ is a consistent estimator of $\sigma^{-1}\boldsymbol{\mathcal{R}}(\boldsymbol{\mu}_e)$.

Proof. Using the consistency of $\widehat{\mathbf{R}}_m$ w.r.t. $\sigma^{-1}\mathbf{R}_e$ [15], we have $\widehat{\mathbf{r}}_m \xrightarrow{\mathcal{P}} \sigma^{-1}\mathbf{r}_e = \mathbf{r}_c = \text{vec}\left(\boldsymbol{\mathcal{R}}(\boldsymbol{\mu}_c)\right)$. Then, since $\widehat{\mathbf{R}}$ is any consistent estimator of $\mathbf{R}_e$ up to a scale factor, i.e., $\widehat{\mathbf{R}} \xrightarrow{\mathcal{P}} \alpha\mathbf{R}_e$, $\alpha > 0$, we obtain, for large N, $\widehat{\mathbf{Y}} \xrightarrow{\mathcal{P}} \mathbf{Y}_\infty \triangleq \kappa_1\alpha^{-2}\left(\mathbf{R}_e^T \otimes \mathbf{R}_e\right)^{-1} + \kappa_2\alpha^{-2}\text{vec}\left(\mathbf{R}_e^{-1}\right)\text{vec}\left(\mathbf{R}_e^{-1}\right)^H$. Consequently from (14), we obtain $\widehat{\boldsymbol{\mu}} \xrightarrow{\mathcal{P}} \boldsymbol{\mu}_\infty$ where $\boldsymbol{\mu}_\infty$ is the solution of the asymptotic

criterion (14) for $N \to \infty$, i.e.,

$$\boldsymbol{\mu}_\infty = \arg\min_{\boldsymbol{\mu}} \left\| \mathbf{Y}_\infty^{1/2} \left(\mathbf{r}_c - \mathbf{r}(\boldsymbol{\mu}) \right) \right\|_2^2 . \tag{18}$$

Furthermore, we have

$$\begin{aligned} \operatorname{rank}\left(\mathbf{Y}_\infty\right) &= \operatorname{rank}\left(\left(\mathbf{R}_e^T \otimes \mathbf{R}_e \right)^{-1/2} \left(\mathbf{I}_{m^2} + \left(1 - \frac{1}{\kappa_1} \right) \operatorname{vec}\left(\mathbf{I}_m\right) \operatorname{vec}\left(\mathbf{I}_m\right)^H \right) \left(\mathbf{R}_e^T \otimes \mathbf{R}_e \right)^{-1/2} \right) \\ &= \operatorname{rank}\left(\mathbf{I}_{m^2} + m \left(1 - \frac{1}{\kappa_1} \right) \mathbf{n}\mathbf{n}^H \right), \end{aligned}$$

where $\left(\mathbf{R}_e^T \otimes \mathbf{R}_e \right)^{-1/2}$ is full-rank and $\mathbf{n} = \dfrac{1}{\sqrt{m}} \operatorname{vec}\left(\mathbf{I}_m\right)$ is a unit vector. The matrix $\mathbf{Y}_\infty$ is full-rank if and only if

$$\left| \mathbf{I}_{m^2} + m \left(1 - \frac{1}{\kappa_1} \right) \mathbf{n}\mathbf{n}^H \right| \equiv 1 + m \left(1 - \frac{1}{\kappa_1} \right) \neq 0 \;\Leftrightarrow\; \mathbb{E}_{f_{\mathbf{Y}}} \left[\psi_{\text{mod}}^2 \left(|\mathbf{t}_{\text{mod}}|^2 \right) \right] \neq m^2 .$$

Since $\mathbb{E}_{f_{\mathbf{Y}}} \left[\psi_{\text{mod}}^2 \left(|\mathbf{t}_{\text{mod}}|^2 \right) \right] > m^2$, $\mathbf{Y}_\infty^{1/2}$ is a full-rank matrix. Moreover, since the mapping is one-to-one, the unique solution to the above problem is $\boldsymbol{\mu}_\infty = \boldsymbol{\mu}_c$ with probability one, which establishes the consistency of $\widehat{\boldsymbol{\mu}}$ w.r.t. $\boldsymbol{\mu}_c$. Finally, the continuous mapping theorem [34] implies $\boldsymbol{\mathcal{R}}\left(\widehat{\boldsymbol{\mu}}\right) \xrightarrow{\mathcal{P}} \boldsymbol{\mathcal{R}}\left(\boldsymbol{\mu}_c\right) = \sigma^{-1} \boldsymbol{\mathcal{R}}\left(\boldsymbol{\mu}_{\text{e}}\right)$. ■

With a potential model misspecification, the so-called *pseudo-true* parameter vector, $\boldsymbol{\mu}_0$, is classically introduced for an asymptotic analysis [20, 22, 23]. The latter is defined as the minimizer of the Kullback-Leibler divergence (KLD) between the *true* and the *assumed* models, i.e.,

$$\boldsymbol{\mu}_0 = \arg\min_{\boldsymbol{\mu}} \mathcal{D}\left(p_{\mathbf{Y}} \| f_{\boldsymbol{\mu}} \right) = \arg\max_{\boldsymbol{\mu}} \mathbb{E}_{p_{\mathbf{Y}}} \left[\log f_{\mathbf{Y}}\left(\mathbf{y}_n ; \boldsymbol{\mu} \right) \right], \tag{19}$$

where $\mathcal{D}\left(p_{\mathbf{Y}} \| f_{\boldsymbol{\mu}} \right) \triangleq \mathbb{E}_{p_{\mathbf{Y}}} \left[\log \dfrac{p_{\mathbf{Y}}\left(\mathbf{y}_n ; \boldsymbol{\mu}_e \right)}{f_{\mathbf{Y}}\left(\mathbf{y}_n ; \boldsymbol{\mu} \right)} \right]$. In the following, we always assume the existence and the uniqueness of the *pseudo-true* parameter vector, $\boldsymbol{\mu}_0$ (the reader is referred to [22] for necessary and sufficient conditions).

Corollary 1. The *pseudo-true* parameter vector, $\boldsymbol{\mu}_0$, is equal to $\boldsymbol{\mu}_c$. Thus, the SESAME estimate, $\widehat{\boldsymbol{\mu}}$, given by (13), is a consistent estimator of $\boldsymbol{\mu}_0$ such that $\boldsymbol{\mu}_0 = \arg\min_{\boldsymbol{\mu}} \mathcal{D}\left(p_{\mathbf{Y}} \| f_{\boldsymbol{\mu}}\right)$.

Proof. For the derivation of the *pseudo-true* parameter vector, it is easier to work on $\boldsymbol{\mathcal{R}}\left(\boldsymbol{\mu}\right)$ than directly on $\boldsymbol{\mu}$. To this end, let us introduce

$$\mathbf{R}_0 = \arg\min_{\boldsymbol{\mathcal{R}}(\boldsymbol{\mu})} \mathcal{D}\left(p_{\mathbf{Y}} \| f_{\boldsymbol{\mu}}\right) = \arg\max_{\boldsymbol{\mathcal{R}}(\boldsymbol{\mu})} \mathbb{E}_{p_{\mathbf{Y}}}\left[\log f_{\mathbf{Y}}\left(\mathbf{y}_n; \boldsymbol{\mathcal{R}}\left(\boldsymbol{\mu}\right)\right)\right]. \tag{20}$$

The differential of the KLD w.r.t. $\boldsymbol{\mathcal{R}}\left(\boldsymbol{\mu}\right)$ is given by [22]

$$\begin{aligned}
\partial\mathcal{D}\left(p_{\mathbf{Y}} \| f_{\boldsymbol{\mu}}\right) &= -\mathbb{E}_{p_{\mathbf{Y}}}\left[\partial \log f_{\mathbf{Y}}\left(\mathbf{y}_n; \boldsymbol{\mathcal{R}}\left(\boldsymbol{\mu}\right)\right)\right] = -\mathbb{E}_{p_{\mathbf{Y}}}\left[\partial \log |\boldsymbol{\mathcal{R}}\left(\boldsymbol{\mu}\right)|^{-1} + \partial \log g_{\text{mod}}\left(\mathbf{y}_n^H \boldsymbol{\mathcal{R}}\left(\boldsymbol{\mu}\right)^{-1} \mathbf{y}_n\right)\right] \\
&= \operatorname{Tr}\left(\boldsymbol{\mathcal{R}}\left(\boldsymbol{\mu}\right)^{-1} \partial\boldsymbol{\mathcal{R}}\left(\boldsymbol{\mu}\right)\right) + \operatorname{Tr}\left(\mathbb{E}_{p_{\mathbf{Y}}}\left[\frac{g'_{\text{mod}}(Q_R)}{g_{\text{mod}}(Q_R)} \boldsymbol{\mathcal{R}}\left(\boldsymbol{\mu}\right)^{-1} \mathbf{y}_n \mathbf{y}_n^H \boldsymbol{\mathcal{R}}\left(\boldsymbol{\mu}\right)^{-1} \partial\boldsymbol{\mathcal{R}}\left(\boldsymbol{\mu}\right)\right]\right) \\
&= \operatorname{Tr}\left(\boldsymbol{\mathcal{R}}\left(\boldsymbol{\mu}\right)^{-1} \partial\boldsymbol{\mathcal{R}}\left(\boldsymbol{\mu}\right)\right) + \operatorname{Tr}\left(\boldsymbol{\mathcal{R}}\left(\boldsymbol{\mu}\right)^{-1} \mathbb{E}_{p_{\mathbf{Y}}}\left[\frac{g'_{\text{mod}}(Q_R)}{g_{\text{mod}}(Q_R)} \mathbf{y}_n \mathbf{y}_n^H\right] \boldsymbol{\mathcal{R}}\left(\boldsymbol{\mu}\right)^{-1} \partial\boldsymbol{\mathcal{R}}\left(\boldsymbol{\mu}\right)\right),
\end{aligned}$$

with $Q_R \triangleq \mathbf{y}_n^H \boldsymbol{\mathcal{R}}\left(\boldsymbol{\mu}\right)^{-1} \mathbf{y}_n$. By following the standard rules of matrix calculus [35], the derivative of the KLD w.r.t. $\boldsymbol{\mathcal{R}}\left(\boldsymbol{\mu}\right)$ is then given by

$$\frac{\partial\mathcal{D}\left(p_{\mathbf{Y}} \| f_{\boldsymbol{\mu}}\right)}{\partial\boldsymbol{\mathcal{R}}\left(\boldsymbol{\mu}\right)} = \boldsymbol{\mathcal{R}}\left(\boldsymbol{\mu}\right)^{-1} - \boldsymbol{\mathcal{R}}\left(\boldsymbol{\mu}\right)^{-1} \mathbb{E}_{p_{\mathbf{Y}}}\left[u_{\text{mod}}(Q_R) \mathbf{y}_n \mathbf{y}_n^H\right] \boldsymbol{\mathcal{R}}\left(\boldsymbol{\mu}\right)^{-1}. \tag{21}$$

Finally the matrix, which cancels (21), denoted by $\mathbf{R}_0$, is solution to the fixed-point equation

$$\mathbf{R}_0 = \mathbb{E}_{p_{\mathbf{Y}}}\left[u_{\text{mod}}\left(\mathbf{y}_n^H \mathbf{R}_0^{-1} \mathbf{y}_n\right) \mathbf{y}_n \mathbf{y}_n^H\right],$$

which coincides with (5). Thus, by uniqueness of the solution of (5), we obtain $\mathbf{R}_0 = \sigma^{-1}\mathbf{R}_e = \mathbf{R}_c = \boldsymbol{\mathcal{R}}\left(\boldsymbol{\mu}_c\right)$. Therefore, there exists $\boldsymbol{\mu}_0$ such that $\mathbf{R}_0 = \boldsymbol{\mathcal{R}}\left(\boldsymbol{\mu}_0\right)$ and $\boldsymbol{\mu}_c = \boldsymbol{\mu}_0 = \arg\min_{\boldsymbol{\mu}} \mathcal{D}\left(p_{\mathbf{Y}} \| f_{\boldsymbol{\mu}}\right)$. Applying **Theorem 1.** concludes the proof. ■

4.2. Asymptotic distribution

Theorem 2. Let $\widehat{\boldsymbol{\mu}}_N$ be the SESAME estimate obtained with **Algorithm 1** from N i.i.d. observations, $\mathbf{y}_n \sim \mathbb{CES}_m\left(\mathbf{0}, \boldsymbol{\mathcal{R}}\left(\boldsymbol{\mu}_{\text{e}}\right), g\right)$ but with an *assumed* model of $\mathbb{CES}_m\left(\mathbf{0}, \boldsymbol{\mathcal{R}}\left(\boldsymbol{\mu}\right), g_{\text{mod}}\right)$. The estimate $\widehat{\boldsymbol{\mu}}_N$ is asymptotically unbiased and Gaussian distributed w.r.t. $\boldsymbol{\mu}_0$, such that

$\boldsymbol{\mu}_0 = \arg\min_{\boldsymbol{\mu}} \mathcal{D}\left(p_{\mathbf{Y}} \| f_{\boldsymbol{\mu}}\right)$. Specifically,

$$\sqrt{N}\left(\widehat{\boldsymbol{\mu}}_N - \boldsymbol{\mu}_0\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Gamma}_{\boldsymbol{\mu}}\right), \tag{22}$$

with

$$\boldsymbol{\Gamma}_{\boldsymbol{\mu}} = \left(\kappa_1 \mathbf{C} + \kappa_2 \mathbf{D}\right)^{-1} \left(\beta_1 \mathbf{C} + \beta_2 \mathbf{D}\right) \left(\kappa_1 \mathbf{C} + \kappa_2 \mathbf{D}\right)^{-1}, \tag{23}$$

where

$$\begin{cases} \mathbf{C} = \boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right)^H \mathbf{W}_0^{-1} \boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right), \\ \mathbf{D} = \boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right)^H \mathbf{U}_0 \boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right), \end{cases} \tag{24}$$

in which $\mathbf{W}_0 = \mathbf{R}_0^T \otimes \mathbf{R}_0$, $\mathbf{U}_0 = \operatorname{vec}\left(\mathbf{R}_0^{-1}\right) \operatorname{vec}\left(\mathbf{R}_0^{-1}\right)^H$, $\beta_1 = \sigma_1 \kappa_1^2$, $\beta_2 = \sigma_1 \kappa_2 \left(2\kappa_1 + m\kappa_2\right) + \sigma_2 \left(\kappa_1 + m\kappa_2\right)^2$ and $\left.\dfrac{\partial \mathbf{r}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}\right|_{\boldsymbol{\mu}} \triangleq \boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}\right)$ refers to the Jacobian matrix of $\mathbf{r}(\boldsymbol{\mu})$ evaluated in $\boldsymbol{\mu}$.

Proof. The estimate $\widehat{\boldsymbol{\mu}}_N$ is given by minimizing the function $\mathcal{J}_{\widehat{\mathbf{R}}_m, \widehat{\mathbf{R}}}(\boldsymbol{\mu})$. The consistency of $\widehat{\boldsymbol{\mu}}_N$ (cf. **Corollary 1.**) allows us to write the following Taylor expansion around $\boldsymbol{\mu}_0$:

$$0 = \left.\frac{\partial \mathcal{J}_{\widehat{\mathbf{R}}_m, \widehat{\mathbf{R}}}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}\right|_{\boldsymbol{\mu}=\widehat{\boldsymbol{\mu}}_N} = \left.\frac{\partial \mathcal{J}_{\widehat{\mathbf{R}}_m, \widehat{\mathbf{R}}}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}\right|_{\boldsymbol{\mu}=\boldsymbol{\mu}_0} + \left(\left.\frac{\partial^2 \mathcal{J}_{\widehat{\mathbf{R}}_m, \widehat{\mathbf{R}}}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu} \partial \boldsymbol{\mu}^T}\right|_{\boldsymbol{\mu}=\boldsymbol{\xi}_N}\right) \left(\widehat{\boldsymbol{\mu}}_N - \boldsymbol{\mu}_0\right),$$

with $\boldsymbol{\xi}_N$ on the line segment connecting $\boldsymbol{\mu}_0$ and $\widehat{\boldsymbol{\mu}}_N$, i.e., $\exists\, c \in]0,1[$ such that $\boldsymbol{\xi}_N = c\boldsymbol{\mu}_0 + (1-c)\widehat{\boldsymbol{\mu}}_N$ [36, Theorem 5.4.8], leading to

$$\sqrt{N}\left(\widehat{\boldsymbol{\mu}}_N - \boldsymbol{\mu}_0\right) = -\left(\left.\frac{\partial^2 \mathcal{J}_{\widehat{\mathbf{R}}_m, \widehat{\mathbf{R}}}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu} \partial \boldsymbol{\mu}^T}\right|_{\boldsymbol{\mu}=\boldsymbol{\xi}_N}\right)^{-1} \sqrt{N} \mathbf{g}_N(\boldsymbol{\mu}_0),$$

subject to invertibility, with $\mathbf{g}_N(\boldsymbol{\mu}) = \dfrac{\partial \mathcal{J}_{\widehat{\mathbf{R}}_m, \widehat{\mathbf{R}}}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}$.

- First, the consistency of $\widehat{\boldsymbol{\mu}}_N$ implies $\boldsymbol{\xi}_N \xrightarrow{\mathcal{P}} \boldsymbol{\mu}_0$. Moreover, by consistency of $\widehat{\mathbf{R}}_m$ w.r.t.

$\mathbf{R}_0 = \sigma^{-1}\mathbf{R}_\mathrm{e}$ and $\widehat{\mathbf{R}}$ w.r.t. $\alpha\mathbf{R}_\mathrm{e}$, the continuous mapping theorem yields to

$$\left.\frac{\partial^2 \mathcal{J}_{\widehat{\mathbf{R}}_m,\widehat{\mathbf{R}}}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}\partial\boldsymbol{\mu}^T}\right|_{\boldsymbol{\mu}=\boldsymbol{\xi}_N} \xrightarrow{\mathcal{P}} \left.\frac{\partial^2 \mathcal{J}_{\mathbf{R}_0,\alpha\mathbf{R}_\mathrm{e}}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}\partial\boldsymbol{\mu}^T}\right|_{\boldsymbol{\mu}=\boldsymbol{\mu}_0} = \alpha^{-2}\left.\frac{\partial^2 \mathcal{J}_{\mathbf{R}_0,\mathbf{R}_\mathrm{e}}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}\partial\boldsymbol{\mu}^T}\right|_{\boldsymbol{\mu}=\boldsymbol{\mu}_0} \triangleq \alpha^{-2}\mathbf{H}(\boldsymbol{\mu}_0). \tag{25}$$

After some calculus, we obtain from (14)

$$\mathbf{H}(\boldsymbol{\mu}_0) = 2\left.\frac{\partial \mathbf{r}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}}\right|^H_{\boldsymbol{\mu}=\boldsymbol{\mu}_0} \mathbf{Y}_\mathrm{e} \left.\frac{\partial \mathbf{r}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}}\right|_{\boldsymbol{\mu}=\boldsymbol{\mu}_0} = 2\boldsymbol{\mathcal{J}}(\boldsymbol{\mu}_0)^H \mathbf{Y}_\mathrm{e}\boldsymbol{\mathcal{J}}(\boldsymbol{\mu}_0),$$

where $\mathbf{Y}_\mathrm{e} = \kappa_1\mathbf{W}_\mathrm{e}^{-1} + \kappa_2\mathbf{U}_\mathrm{e}$, with $\mathbf{W}_\mathrm{e} = \mathbf{R}_\mathrm{e}^T \otimes \mathbf{R}_\mathrm{e}$ and $\mathbf{U}_\mathrm{e} = \mathrm{vec}\left(\mathbf{R}_\mathrm{e}^{-1}\right)\mathrm{vec}\left(\mathbf{R}_\mathrm{e}^{-1}\right)^H$.

- Second, the gradient $\mathbf{g}_N(\boldsymbol{\mu})$ is equal to

$$\begin{aligned}\mathbf{g}_N(\boldsymbol{\mu}) &= -\frac{\partial \mathbf{r}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}}^H \widehat{\mathbf{Y}}\left(\widehat{\mathbf{r}}_m - \mathbf{r}(\boldsymbol{\mu})\right) - \left(\widehat{\mathbf{r}}_m - \mathbf{r}(\boldsymbol{\mu})\right)^H \widehat{\mathbf{Y}}\frac{\partial \mathbf{r}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}} \\ &= -2\Re\left(\frac{\partial \mathbf{r}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}}^H \widehat{\mathbf{Y}}\left(\widehat{\mathbf{r}}_m - \mathbf{r}(\boldsymbol{\mu})\right)\right).\end{aligned} \tag{26}$$

- Finally, using the asymptotic distribution of $\widehat{\mathbf{r}}_m$ given by (6), we can show that (see Appendix B for details)

$$-\sqrt{N}\mathbf{g}_N(\boldsymbol{\mu}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{R}_\infty), \tag{27}$$

with $\mathbf{R}_\infty = 4\alpha^{-4}\sigma^{-4}\boldsymbol{\mathcal{J}}(\boldsymbol{\mu}_0)^H\left(\beta_1\mathbf{W}_0^{-1} + \beta_2\mathbf{U}_0\right)\boldsymbol{\mathcal{J}}(\boldsymbol{\mu}_0) = 4\alpha^{-4}\sigma^{-4}(\beta_1\mathbf{C} + \beta_2\mathbf{D})$

According to Slutsky's lemma with (25) and (27) [37, Chapters 2], we finally obtain

$$\sqrt{N}(\widehat{\boldsymbol{\mu}}_N - \boldsymbol{\mu}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_{\boldsymbol{\mu}}), \tag{28}$$

with

$$\begin{aligned}\boldsymbol{\Gamma}_{\boldsymbol{\mu}} &= \alpha^4\mathbf{H}(\boldsymbol{\mu}_0)^{-1}\mathbf{R}_\infty\mathbf{H}(\boldsymbol{\mu}_0)^{-H} \\ &= (\kappa_1\mathbf{C} + \kappa_2\mathbf{D})^{-1}(\beta_1\mathbf{C} + \beta_2\mathbf{D})(\kappa_1\mathbf{C} + \kappa_2\mathbf{D})^{-1},\end{aligned} \tag{29}$$

which concludes the proof on asymptotic unbiasedness and asymptotic Gaussianity of SESAME estimator w.r.t. the *pseudo-true* parameter $\boldsymbol{\mu}_0$. ■

4.3. Efficiency in the mismatched framework

In the matched context, the Cramér-Rao Bound (CRB) is a lower bound of the variance of any unbiased estimator (which corresponds then to the Mean Square Error) of a deterministic parameter. Such an estimator is said to be (asymptotically) efficient if its variance reaches the CRB for an (in)finite number of samples.

Likewise, under misspecified models, the Misspecified Cramér-Rao Bound (MCRB) is defined as a lower bound of the variance of any unbiased estimator $\widehat{\boldsymbol{\mu}}_g$ of $\boldsymbol{\mu}_0$, where $\boldsymbol{\mu}_0$ is actually the *pseudo-true* parameter vector [22, 23]. Specifically, we have

$$\mathrm{Var}\left(\widehat{\boldsymbol{\mu}}_g\right) \succeq \frac{1}{N}\mathbf{A}^{-1}\left(\boldsymbol{\mu}_0\right)\mathbf{B}\left(\boldsymbol{\mu}_0\right)\mathbf{A}^{-1}\left(\boldsymbol{\mu}_0\right) \triangleq \frac{1}{N}\mathbf{MCRB}, \tag{30}$$

where, for all $k,\ell = 1,\ldots,P$, $\left[\mathbf{B}\left(\boldsymbol{\mu}_0\right)\right]_{k,\ell} = \mathbb{E}_{p_{\mathbf{Y}}}\left[\left.\frac{\partial \log f_{\mathbf{Y}}\left(\mathbf{y}_n;\boldsymbol{\mu}\right)}{\partial \mu_k}\right|_{\boldsymbol{\mu}=\boldsymbol{\mu}_0}\left.\frac{\partial \log f_{\mathbf{Y}}\left(\mathbf{y}_n;\boldsymbol{\mu}\right)}{\partial \mu_\ell}\right|_{\boldsymbol{\mu}=\boldsymbol{\mu}_0}\right]$ and $\left[\mathbf{A}\left(\boldsymbol{\mu}_0\right)\right]_{k,\ell} = \mathbb{E}_{p_{\mathbf{Y}}}\left[\left.\frac{\partial^2 \log f_{\mathbf{Y}}\left(\mathbf{y}_n;\boldsymbol{\mu}\right)}{\partial \mu_k \partial \mu_\ell}\right|_{\boldsymbol{\mu}=\boldsymbol{\mu}_0}\right]$ and we define the m-efficiency property by

Definition (Property of m-efficiency)**.** In the mismatched framework, an unbiased estimator $\widehat{\boldsymbol{\mu}}_g$ of $\boldsymbol{\mu}_0$ is said to be (asymptotically) m-efficient if its variance reaches the MCRB in the (in)finite-sample regime.

Theorem 3. The SESAME estimate, $\widehat{\boldsymbol{\mu}}_N$ obtained with **Algorithm 1**, is asymptotically m-efficient, i.e.,

$$\sqrt{N}\left(\widehat{\boldsymbol{\mu}}_N - \boldsymbol{\mu}_0\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{MCRB}\right), \tag{31}$$

with

$$\mathbf{MCRB} = \sigma_1\mathbf{C}^{-1} + \sigma_2\mathbf{C}^{-1}\mathbf{D}\mathbf{C}^{-1} = \left(\sigma_1^{-1}\mathbf{C} - \frac{\sigma_2}{\sigma_1(\sigma_1 + m\sigma_2)}\mathbf{D}\right)^{-1}, \tag{32}$$

in which $\mathbf{C}$ and $\mathbf{D}$ are given in (24).

Proof. The study of the m-efficiency of $\widehat{\boldsymbol{\mu}}_N$ is reduced to prove the equality between $\boldsymbol{\Gamma}_{\boldsymbol{\mu}}$ and **MCRB**, since **Theorem 2.** holds. First, we give the expression of **MCRB** from (30). The

derivation of the MCRB for the considered problem has already been conducted in [23] and is only summarized here

$$\mathbf{MCRB} = \left(-\frac{a_2+m}{m+1}\mathbf{C} + \frac{1-a_2}{m+1}\mathbf{D}\right)^{-1} (a_1\mathbf{C} + (a_1-1)\mathbf{D}) \left(-\frac{a_2+m}{m+1}\mathbf{C} + \frac{1-a_2}{m+1}\mathbf{D}\right)^{-1}. \quad (33)$$

Then, let us recall that the matrix $\mathbf{D} = \boldsymbol{\mathcal{J}}(\boldsymbol{\mu}_0)^H \operatorname{vec}\left(\mathbf{R}_0^{-1}\right) \operatorname{vec}\left(\mathbf{R}_0^{-1}\right)^H \boldsymbol{\mathcal{J}}(\boldsymbol{\mu}_0)$ has a rank lower or equal to 1. Then, since the mapping $\boldsymbol{\mu} \mapsto \boldsymbol{\mathcal{R}}(\boldsymbol{\mu})$ is one-to-one, the matrix $\boldsymbol{\mathcal{J}}(\boldsymbol{\mu}_0) \in \mathbb{C}^{m^2 \times P}$ is full-rank column. Let $\boldsymbol{\mathcal{B}} \in \mathbb{C}^{m^2 \times m^2 - P}$ be the matrix, whose columns form an orthonormal basis for the nullspace of $\boldsymbol{\mathcal{J}}(\boldsymbol{\mu}_0)^H$, that is $\boldsymbol{\mathcal{J}}(\boldsymbol{\mu}_0)^H \boldsymbol{\mathcal{B}} = \mathbf{0}$ and $\boldsymbol{\mathcal{B}}^H \boldsymbol{\mathcal{B}} = \mathbf{I}_{m^2-P}$.

- Case $\operatorname{rank}(\mathbf{D}) = 0$

If $\operatorname{rank}(\mathbf{D}) = 0$, i.e., $\mathbf{D} = 0$, we obviously have from (29) and (33)

$$\boldsymbol{\Gamma}_{\boldsymbol{\mu}} = \frac{\beta_1}{\kappa_1^2}\mathbf{C}^{-1} = \sigma_1\mathbf{C}^{-1} = a_1\left(\frac{m+1}{a_2+m}\right)^2 \mathbf{C}^{-1} = \mathbf{MCRB}.$$

This occurs when $\operatorname{vec}(\mathbf{R}_0) \in \operatorname{span}(\boldsymbol{\mathcal{B}})$, i.e., $\operatorname{vec}(\mathbf{R}_0)^H \boldsymbol{\mathcal{B}} \neq \mathbf{0}$.

- Case $\operatorname{rank}(\mathbf{D}) = 1$

This more interesting case happens when $\operatorname{vec}(\mathbf{R}_0) \notin \operatorname{span}(\boldsymbol{\mathcal{B}})$, i.e., $\operatorname{vec}(\mathbf{R}_0)^H \boldsymbol{\mathcal{B}} = \mathbf{0}$. In order to compare the matrices $\boldsymbol{\Gamma}(\boldsymbol{\mu})$ and $\mathbf{MCRB}$, we develop their expression using the Sherman-Morrison formula [38]. On one hand, we obtain

$$\boldsymbol{\Gamma}_{\boldsymbol{\mu}} = \sigma^{-2}\left(\delta_1\mathbf{C}^{-1} + \delta_2\mathbf{C}^{-1}\mathbf{D}\mathbf{C}^{-1} + \delta_3\mathbf{C}^{-1}\mathbf{D}\mathbf{C}^{-1}\mathbf{D}\mathbf{C}^{-1} + \delta_4\mathbf{C}^{-1}\mathbf{D}\mathbf{C}^{-1}\mathbf{D}\mathbf{C}^{-1}\mathbf{D}\mathbf{C}^{-1}\right) \quad (34)$$

$$\text{with} \quad \begin{cases} \delta_1 &= \dfrac{\beta_1}{\kappa_1^2} = \sigma_1, \\ \delta_2 &= \dfrac{1}{\kappa_1^2}\left(\beta_2 - \dfrac{2\beta_1\kappa_2}{\kappa_1 + \varepsilon_e\kappa_2}\right), \\ \delta_3 &= \dfrac{\kappa_2}{\kappa_1^2(\kappa_1+\varepsilon_e\kappa_2)}\left(\dfrac{\beta_1\kappa_2}{\kappa_1+\varepsilon_e\kappa_2} - 2\beta_2\right), \\ \delta_4 &= \dfrac{\kappa_2^2\beta_2}{\kappa_1^2(\kappa_1+\varepsilon_e\kappa_2)^2}. \end{cases}$$

On the other hand, we obtain

$$\mathbf{MCRB} = \sigma^{-2}\left(\gamma_1\mathbf{C}^{-1} + \gamma_2\mathbf{C}^{-1}\mathbf{D}\mathbf{C}^{-1} + \gamma_3\mathbf{C}^{-1}\mathbf{D}\mathbf{C}^{-1}\mathbf{D}\mathbf{C}^{-1} + \gamma_4\mathbf{C}^{-1}\mathbf{D}\mathbf{C}^{-1}\mathbf{D}\mathbf{C}^{-1}\mathbf{D}\mathbf{C}^{-1}\right) \quad (35)$$

$$\text{with} \quad \begin{cases} \gamma_1 &= \sigma_1, \\ \gamma_2 &= \left(\dfrac{2\sigma_1(1-a_2)}{a_2+m+\varepsilon_e(a_2-1)} + \dfrac{a_1-1}{a_1}\sigma_1 \right), \\ \gamma_3 &= \dfrac{1-a_2}{a_2+m+\varepsilon_e(a_2-1)} \left(\dfrac{2\sigma_1(a_1-1)}{a_1} + \dfrac{(1-a_2)\sigma_1}{a_2+m+\varepsilon_e(a_2-1)} \right), \\ \gamma_4 &= \dfrac{(1-a_2)^2(a_1-1)\sigma_1}{(a_2+m+\varepsilon_e(a_2-1))^2 a_1}, \end{cases}$$

where $\varepsilon_e = \mathrm{Tr}\left(\mathbf{D}\mathbf{C}^{-1}\right)$. It can be also noted that $\mathbf{C}^{-1}\mathbf{D}\mathbf{C}^{-1}\mathbf{D}\mathbf{C}^{-1} = \varepsilon_e \mathbf{C}^{-1}\mathbf{D}\mathbf{C}^{-1}$ and $\mathbf{C}^{-1}\mathbf{D}\mathbf{C}^{-1}\mathbf{D}\mathbf{C}^{-1}\mathbf{D}\mathbf{C}^{-1} = \varepsilon_e^2 \mathbf{C}^{-1}\mathbf{D}\mathbf{C}^{-1}$. Finally, we obtain

$$\begin{aligned} \boldsymbol{\Gamma}_{\boldsymbol{\mu}} &= \sigma^{-2}\sigma_1 \mathbf{C}^{-1} + \sigma^{-2}\left(\delta_2 + \delta_3\varepsilon_e + \delta_4\varepsilon_e^2\right)\mathbf{C}^{-1}\mathbf{D}\mathbf{C}^{-1} \\ &= \sigma^{-2}\sigma_1 \mathbf{C}^{-1} + \sigma^{-2}\left(\sigma_1 \frac{\kappa_2^2\,(m-\varepsilon_e)}{(\kappa_1+\varepsilon_e\kappa_2)^2} + \sigma_2\left(\frac{\kappa_1+m\kappa_2}{\kappa_1+\varepsilon_e\kappa_2}\right)^2\right)\mathbf{C}^{-1}\mathbf{D}\mathbf{C}^{-1}, \end{aligned} \tag{36}$$

and

$$\begin{aligned} \mathbf{MCRB} &= \sigma^{-2}\sigma_1 \mathbf{C}^{-1} + \sigma^{-2}\left(\gamma_2 + \gamma_3\varepsilon_e + \gamma_4\varepsilon_e^2\right)\mathbf{C}^{-1}\mathbf{D}\mathbf{C}^{-1} \\ &= \sigma^{-2}\sigma_1 \mathbf{C}^{-1} + \sigma^{-2}\left(\sigma_1 \frac{(a_2-1)^2\,(m-\varepsilon_e)}{(m+1)^2} + \sigma_2\left(\frac{a_2(m+1)}{a_2(1+\varepsilon_e)+m-\varepsilon_e}\right)^2\right)\mathbf{C}^{-1}\mathbf{D}\mathbf{C}^{-1}. \end{aligned} \tag{37}$$

Consequently, $\boldsymbol{\Gamma}_{\boldsymbol{\mu}}$ will be equal to $\mathbf{MCRB}$, if $\varepsilon_e = m$. In the following we compute ε_e. Let us note that,

$$\varepsilon_e = \mathrm{Tr}\left(\mathbf{D}\mathbf{C}^{-1}\right) = \mathrm{vec}\left(\mathbf{R}_0\right)^H \mathbf{W}_0^{-1}\boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right)\left(\boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right)^H \mathbf{W}_0^{-1}\boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right)\right)^{-1}\boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right)^H \mathbf{W}_0^{-1}\mathrm{vec}\left(\mathbf{R}_0\right). \tag{38}$$

According to Corollary 1 of [39], the following equality holds

$$\boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right)\left(\boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right)^H \mathbf{W}_0^{-1}\boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right)\right)^{-1}\boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right)^H = \mathbf{W}_0 - \mathbf{W}_0\boldsymbol{\mathcal{B}}\left(\boldsymbol{\mathcal{B}}^H\mathbf{W}_0\boldsymbol{\mathcal{B}}\right)^{-1}\boldsymbol{\mathcal{B}}^H\mathbf{W}_0.$$

Thus, (38) can be rewritten as

$$\varepsilon_e = \mathrm{vec}\left(\mathbf{R}_0\right)^H \mathbf{W}_0^{-1}\left(\mathbf{W}_0 - \mathbf{W}_0\boldsymbol{\mathcal{B}}\left(\boldsymbol{\mathcal{B}}^H\mathbf{W}_0\boldsymbol{\mathcal{B}}\right)^{-1}\boldsymbol{\mathcal{B}}^H\mathbf{W}_0\right)\mathbf{W}_0^{-1}\mathrm{vec}\left(\mathbf{R}_0\right)$$

$$= \operatorname{vec}(\mathbf{R}_0)^H \mathbf{W}_0^{-1} \operatorname{vec}(\mathbf{R}_0) - \underbrace{\operatorname{vec}(\mathbf{R}_0)^H \boldsymbol{\mathcal{B}}}_{\mathbf{0}} \left(\boldsymbol{\mathcal{B}}^H \mathbf{W}_0 \boldsymbol{\mathcal{B}}\right)^{-1} \underbrace{\boldsymbol{\mathcal{B}}^H \operatorname{vec}(\mathbf{R}_0)}_{\mathbf{0}}$$

$$= \operatorname{Tr}\left(\mathbf{R}_0^{-1}\mathbf{R}_0\right) - 0 = m, \qquad \text{which concludes the proof.}$$

■

5. Recursive SESAME

In this section, we propose a recursive procedure for SESAME based on an iterative refinement of the cost function $\mathcal{J}_{\widehat{\mathbf{R}}_m,\widehat{\mathbf{R}}}(\boldsymbol{\mu})$, which preserves the same asymptotic performance as SESAME and turns out to provide an estimation accuracy improvement in most cases. The SESAME algorithm provides an estimate $\widehat{\boldsymbol{\mu}}$ by minimizing $\mathcal{J}_{\widehat{\mathbf{R}}_m,\widehat{\mathbf{R}}}(\boldsymbol{\mu})$, where $\mathcal{J}_{\widehat{\mathbf{R}}_m,\widehat{\mathbf{R}}}(\boldsymbol{\mu})$ is defined by either (13) or (14):

$$\widehat{\boldsymbol{\mu}}^{\text{SESAME}} = \arg\min_{\boldsymbol{\mu}} \mathcal{J}_{\widehat{\mathbf{R}}_m,\widehat{\mathbf{R}}}(\boldsymbol{\mu}),$$

where $\widehat{\mathbf{R}}_m$ is an unstructured estimator of $\mathbf{R}_{\text{e}}$ obtained with (12) and $\widehat{\mathbf{R}}$ is any consistent estimator of $\mathbf{R}_{\text{e}}$ up to a scale factor. According to **Theorem 1.**, $\boldsymbol{\mathcal{R}}\left(\widehat{\boldsymbol{\mu}}^{\text{SESAME}}\right)$ is a consistent estimator of $\mathbf{R}_{\text{e}}$ up to a scale factor. Intuitively, the better the estimator $\widehat{\mathbf{R}}$ is, the better the solution $\widehat{\boldsymbol{\mu}}$ should be. This leads naturally to a recursive procedure, where the minimized norm is refined at each step by updating $\widehat{\mathbf{R}}$ with the previously computed $\boldsymbol{\mathcal{R}}\left(\widehat{\boldsymbol{\mu}}^{\text{SESAME}}\right)$. For a finite number of steps, N_{it}, we obtain the Recursive SESAME (R-SESAME) for $\boldsymbol{\mu}$, denoted $\widehat{\boldsymbol{\mu}}^{\text{R-SESAME}}$ and achieved at the k-th stage by solving

$$\widehat{\boldsymbol{\mu}}^{(k+1)} = \arg\min_{\boldsymbol{\mu}} \mathcal{J}_{\widehat{\mathbf{R}}_m,\widehat{\mathbf{R}}^{(k)}}(\boldsymbol{\mu}) \quad \text{with } \widehat{\mathbf{R}}^{(k)} = \boldsymbol{\mathcal{R}}\left(\widehat{\boldsymbol{\mu}}^{(k)}\right), \quad \text{for } k = 1, \ldots, N_{\text{it}}, \tag{39}$$

with $\widehat{\boldsymbol{\mu}}^{\text{R-SESAME}} = \widehat{\boldsymbol{\mu}}^{(N_{\text{it}}+1)}$. Since $N_{\text{it}} < \infty$, the existence of R-SESAME estimate is always ensured. The R-SESAME algorithm is recapped in the box **Algorithm 2.**

By applying **Theorems 1.** and **2.** at each iteration, the next theorem follows immediately.

Theorem 4. Let $\widehat{\boldsymbol{\mu}}_N^{\text{R-SESAME}}$ be the R-SESAME estimate given by **Algorithm 2** and based on N i.i.d. observations, $\mathbf{y}_n \sim \mathbb{CES}_m\left(\mathbf{0}, \boldsymbol{\mathcal{R}}\left(\boldsymbol{\mu}_{\text{e}}\right), g\right)$ but with an *assumed* model

Algorithm 2 Recursive-SESAME

Input: N i.i.d. data, $\mathbf{y}_n \sim \mathbb{CES}_m(\mathbf{0}, \mathbf{R}_{\mathrm{e}}, g)$ with $N > m$

1: Compute $\widehat{\mathbf{R}}_m$ from $\mathbf{y}_1, \ldots, \mathbf{y}_N$ with (12)
2: Initialize $\widehat{\boldsymbol{\mu}}^{(1)}$ by minimizing $\mathcal{J}_{\widehat{\mathbf{R}}_m, \widehat{\mathbf{R}}_m}(\boldsymbol{\mu})$
3: **for** $k = 1$ to N_{it} **do**
4: Compute $\widehat{\boldsymbol{\mu}}^{(k+1)}$ from (39)
5: **end for**
6: **Return:** $\widehat{\boldsymbol{\mu}}^{\text{R-SESAME}} = \widehat{\boldsymbol{\mu}}^{(N_{\mathrm{it}}+1)}$

of $\mathbb{CES}_m(\mathbf{0}, \boldsymbol{\mathcal{R}}(\boldsymbol{\mu}), g_{\mathrm{mod}})$. Therefore $\widehat{\boldsymbol{\mu}}_N^{\text{R-SESAME}}$ is a consistent estimator of $\boldsymbol{\mu}_0$. Likewise, $\boldsymbol{\mathcal{R}}\left(\widehat{\boldsymbol{\mu}}_N^{\text{R-SESAME}}\right)$ is a consistent estimator of $\sigma^{-1}\boldsymbol{\mathcal{R}}(\boldsymbol{\mu}_{\mathrm{e}})$. Moreover, it is asymptotically unbiased, m-efficient and Gaussian distributed w.r.t. $\boldsymbol{\mu}_0$, such that $\boldsymbol{\mu}_0 = \arg\min_{\boldsymbol{\mu}} \mathcal{D}(p_{\mathbf{Y}} \| f_{\boldsymbol{\mu}})$. Specifically,

$$\sqrt{N}\left(\widehat{\boldsymbol{\mu}}_N^{\text{R-SESAME}} - \boldsymbol{\mu}_0\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{MCRB}). \tag{40}$$

R-SESAME improves SESAME in the sense that it empirically reaches its asymptotic regime for lower sample supports. In practice, we can use a more flexible implementation rather than imposing a fixed number of iterations. For example, the stopping criteria can combine a maximal number of iterations, $N_{\max}$, and a relative gap between the estimates of two successive iterations below a defined threshold, $\varepsilon_{\mathrm{tol}}$.

6. Special Cases

In this section, we consider some special cases of the SESAME algorithm, notably under a (wrongly) *assumed* Gaussian distribution and perfectly CES matched model.

6.1. Case 1: misspecification with assumed Gaussian distributed data

A common practice in array processing applications is to assume a Gaussian distributed data. This prevalent choice leads generally to the simplest derivation and for a real-time implementation of the algorithm. However, as already mentioned in the introduction, the performance of Gaussian-based models can be strongly degraded with the presence of outliers or heterogeneities in the data set.

Let us assume a Gaussian model, i.e., $g_{\text{mod}}(t) = \exp(-t)$; whereas, the data are CES distributed, i.e., $\mathbf{y}_n \sim \mathbb{CES}_m(\mathbf{0}, \boldsymbol{\mathcal{R}}(\boldsymbol{\mu}_{\text{e}}), g)$, $n = 1, \ldots, N$ with $g_{\text{mod}} \neq g$. In this case, the proposed SESAME algorithm reads

$$\begin{aligned} \widehat{\boldsymbol{\mu}} &= \arg\min_{\boldsymbol{\mu}} \mathcal{J}^{\text{SESAME-G}}_{\widehat{\mathbf{R}}_{\text{SCM}}, \widehat{\mathbf{R}}}(\boldsymbol{\mu}) \quad \text{with} \\ \mathcal{J}^{\text{SESAME-G}}_{\widehat{\mathbf{R}}_{\text{SCM}}, \widehat{\mathbf{R}}}(\boldsymbol{\mu}) &= \text{Tr}\left(\widehat{\mathbf{R}}^{-1}\left(\widehat{\mathbf{R}}_{\text{SCM}} - \boldsymbol{\mathcal{R}}(\boldsymbol{\mu})\right)\widehat{\mathbf{R}}^{-1}\left(\widehat{\mathbf{R}}_{\text{SCM}} - \boldsymbol{\mathcal{R}}(\boldsymbol{\mu})\right)\right), \end{aligned} \tag{41}$$

with the Sample Covariance Matrix (SCM) $\widehat{\mathbf{R}}_{\text{SCM}} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{y}_n \mathbf{y}_n^H$, $\widehat{\mathbf{R}}$ refers to any consistent estimator of $\mathbf{R}_{\text{e}}$ up to a scale factor. It is worth mentioning that replacing $\widehat{\mathbf{R}}$ by $\widehat{\mathbf{R}}_{\text{SCM}}$ leads to the well-known COMET procedure [10].

Corollary 2. Let $\widehat{\boldsymbol{\mu}}_N$ be the SESAME-G estimate obtained by minimizing (41) from N i.i.d. observations, $\mathbf{y}_n \sim \mathbb{CES}_m(\mathbf{0}, \boldsymbol{\mathcal{R}}(\boldsymbol{\mu}_{\text{e}}), g)$. Therefore $\widehat{\boldsymbol{\mu}}_N$ is consistent, asymptotically unbiased, m-efficient and Gaussian distributed w.r.t. $\boldsymbol{\mu}_0$, such that $\boldsymbol{\mathcal{R}}(\boldsymbol{\mu}_0) = \sigma_{\boldsymbol{\mathcal{C}}}^{-1}\boldsymbol{\mathcal{R}}(\boldsymbol{\mu}_{\text{e}})$ with $\sigma_{\boldsymbol{\mathcal{C}}}^{-1} = \dfrac{\mathbb{E}_{p_{\mathbf{Y}}}[Q]}{m}$. Specifically,

$$\sqrt{N}\left(\widehat{\boldsymbol{\mu}}_N - \boldsymbol{\mu}_0\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Gamma}_{\boldsymbol{\mu}}^{\text{SESAME-G}}\right), \tag{42}$$

with

$$\boldsymbol{\Gamma}_{\boldsymbol{\mu}}^{\text{SESAME-G}} = \mathbf{C}^{-1}\left((\kappa+1)\mathbf{C} + \kappa\mathbf{D}\right)\mathbf{C}^{-1}, \tag{43}$$

where $\kappa = \dfrac{m\mathbb{E}_{p_{\mathbf{Y}}}[Q^2]}{(m+1)\mathbb{E}_{p_{\mathbf{Y}}}[Q]^2} - 1$ is the so-called elliptical kurtosis parameter [15, 40] and $Q \overset{d}{=} \mathbf{y}^H \boldsymbol{\mathcal{R}}(\boldsymbol{\mu}_{\text{e}})^{-1}\mathbf{y}$ with $\mathbf{y} \sim \mathbb{CES}_m(\mathbf{0}, \boldsymbol{\mathcal{R}}(\boldsymbol{\mu}_{\text{e}}), g)$.

Proof. We have $\psi_{\text{mod}}(s) = s$, then $m = \mathbb{E}_{p_{\mathbf{Y}}}[\psi_{\text{mod}}(\sigma_{\boldsymbol{\mathcal{C}}}Q)] = \sigma_{\boldsymbol{\mathcal{C}}}\mathbb{E}_{p_{\mathbf{Y}}}[Q]$. According to Table 1, we have $\kappa_1 = 1$ and $\kappa_2 = 0$. Moreover, we easily obtain $a_1 = \kappa+1$ and $a_2 = 1$, thus $\sigma_1 = \kappa+1$ and $\sigma_2 = \kappa$ and finally $\beta_1 = \kappa + 1$ and $\beta_2 = \kappa$. Applying **Theorems 1.** and **2** concludes the proof. ∎

6.2. Case 2: SESAME in the perfectly matched model

In the situation where the *assumed* model coincides with the *true* one, it refers to the matched case. Then the SESAME procedure is equivalent to the EXIP [24] derived for the CES distributions and the structure $\boldsymbol{\mathcal{R}}(\boldsymbol{\mu})$ [25]. This estimator is referred to SESAME-E.

Corollary 3. Let $\widehat{\boldsymbol{\mu}}_N$ be the SESAME-E estimate obtained with **Algorithm 1** from N i.i.d. observations, $\mathbf{y}_n \sim \mathbb{CES}_m(\mathbf{0}, \boldsymbol{\mathcal{R}}(\boldsymbol{\mu}_{\mathrm{e}}), g)$, where the function $g(\cdot) = g_{\mathrm{mod}}(\cdot)$ is assumed to be known. Therefore $\widehat{\boldsymbol{\mu}}_N$ is consistent, asymptotically unbiased, efficient and Gaussian distributed w.r.t. $\boldsymbol{\mu}_{\mathrm{e}}$. Specifically,

$$\sqrt{N}\left(\widehat{\boldsymbol{\mu}}_N - \boldsymbol{\mu}_{\mathrm{e}}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{F}^{-1}\right), \tag{44}$$

where $\mathbf{F}$ is the Fisher Information Matrix (FIM), which is expressed by $\mathbf{F}^{-1} = \boldsymbol{\mathcal{J}}(\boldsymbol{\mu}_{\mathrm{e}})^H \mathbf{Y}_{\mathrm{e}} \boldsymbol{\mathcal{J}}(\boldsymbol{\mu}_{\mathrm{e}})$ in which $\mathbf{Y}_{\mathrm{e}} = \kappa_1 \mathbf{W}_{\mathrm{e}}^{-1} + \kappa_2 \mathbf{U}_{\mathrm{e}}$.

Proof. The above expression of the FIM follows straightforwardly from [41]. Moreover, we have obviously $\boldsymbol{\mu}_0 = \boldsymbol{\mu}_{\mathrm{e}}$ and $\psi_{\mathrm{mod}}(\cdot) = \psi_{\mathrm{ML}}(\cdot)$, so $\sigma = 1$. According to **Theorem 1.**, we have then $\widehat{\boldsymbol{\mu}}_N \xrightarrow{\mathcal{P}} \boldsymbol{\mu}_{\mathrm{e}}$. Furthermore, we have $\kappa_1 = \sigma_{1,\mathrm{ML}}^{-1}$ and $\sigma_{2,\mathrm{ML}} = \dfrac{-\sigma_{1,\mathrm{ML}}(1-\sigma_{1,\mathrm{ML}})}{1+m(1-\sigma_{1,\mathrm{ML}})}$, thus $\beta_1 = \kappa_1$ and $\beta_2 = \kappa_2$, which yields $\boldsymbol{\Gamma}_{\boldsymbol{\mu}} = (\kappa_1 \mathbf{C} + \kappa_2 \mathbf{D})^{-1} = \mathbf{F}^{-1}$. Finally, according to **Theorem 2.**, we obtain

$$\sqrt{N}\left(\widehat{\boldsymbol{\mu}}_N - \boldsymbol{\mu}_{\mathrm{e}}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{F}^{-1}\right). \tag{45}$$

■

7. Applications and Numerical Results

This section presents the SESAME in the particular case of linear structure under both matched and mismatched models.

7.1. SESAME with linear parameterization

In the case of linear parameterization for the scatter matrix, where the PSD constraint has been relaxed, SESAME yields a closed form expression. Indeed, there exists a basis of

Hermitian matrices $\{\mathbf{B}_p\}_{p=1\ldots P} \in \mathbb{C}^{m\times m}$ such that

$$\boldsymbol{\mathcal{R}}(\boldsymbol{\mu}) = \sum_{p=1}^{P} \mu_p \mathbf{B}_p \quad \text{with} \quad \boldsymbol{\mu} = [\mu_1, \ldots, \mu_P]^T \in \mathbb{R}^P,$$

with $P \leq m^2$. By using the vec-operator, we obtain a matrix, $\mathbf{P} \in \mathbb{C}^{m^2\times P}$, which relates the vectorized matrix $\boldsymbol{\mathcal{R}}(\boldsymbol{\mu})$ to $\boldsymbol{\mu}$ according to:

$$\mathbf{r}(\boldsymbol{\mu}) = \text{vec}\left(\boldsymbol{\mathcal{R}}(\boldsymbol{\mu})\right) = \mathbf{P}\boldsymbol{\mu}.$$

In this case, the PSD constraint is not taken into account. Considering this linear structure, the SESAME criterion (14) reads

$$\widehat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}} \left\| \widehat{\mathbf{Y}}^{1/2} \widehat{\mathbf{r}}_m - \widehat{\mathbf{Y}}^{1/2} \mathbf{P}\boldsymbol{\mu} \right\|^2.$$

The well known analytical solution gives

$$\widehat{\boldsymbol{\mu}} = \left(\mathbf{P}^H \widehat{\mathbf{Y}} \mathbf{P}\right)^{-1} \mathbf{P}^H \widehat{\mathbf{Y}} \widehat{\mathbf{r}}_m. \tag{46}$$

As an example, we consider the Toeplitz structure, which is frequently exhibited in array processing. Let $\mathbf{R}_{\text{e}} = \boldsymbol{\mathcal{R}}(\boldsymbol{\mu}_{\text{e}}) \in \mathbb{C}^{m\times m}$ belong to $\mathcal{S}$, the convex subset of Hermitian matrices with Toeplitz structure. A natural parameterization is as follows:

$$\boldsymbol{\mathcal{R}}(\boldsymbol{\mu}) = \begin{pmatrix} R_1 & R_2 & \cdots & R_m \\ R_2^* & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & R_2 \\ R_m^* & \cdots & R_2^* & R_1 \end{pmatrix} \text{ and } \boldsymbol{\mu} = \begin{pmatrix} R_1 \\ \Re(R_2) \\ \Im(R_2) \\ \vdots \\ \Re(R_m) \\ \Im(R_m) \end{pmatrix} \in \mathbb{R}^{2m-1}.$$

7.2. Validation of the statistical analysis

First, we illustrate the previous theoretical analysis on SESAME performance under misspecifications. For the simulations, we choose a scatter matrix with an Hermitian Toeplitz structure. For $m = 5$, the *true* scatter matrix has its first row $[2, \rho, \rho^2, \ldots, \rho^{m-1}]$, with $\rho = 0.8 + 0.3i$. The coefficients κ_1 and κ_2 needed in SESAME algorithm are connected to $\sigma_{1,\mathrm{ML}}$ w.r.t. the *assumed* p.d.f., whose explicit expressions can be found in Table 1. The others coefficients σ, σ_1 and σ_2 are numerically computed. We consider two scenarios of misspecifications:

S1: the *true* and *assumed* p.d.f. are t-distributions with different degrees of freedom, specifically we have

$$\begin{cases} g(t) = \left(1 + \dfrac{t}{d}\right)^{-(d+m)} & , \quad \text{with} \quad d = 3, \\ g_{\mathrm{mod}}(t) = \left(1 + \dfrac{t}{d_{\mathrm{mod}}}\right)^{-(d_{\mathrm{mod}}+m)} & , \quad \text{with} \quad d_{\mathrm{mod}} = 5. \end{cases}$$

S2: the *true* p.d.f is a W-distribution and we assume a Gaussian model, which would be the most common hypothesis as already mentioned in the introduction. Thus, we have

$$\begin{cases} g(t) = t^{s-1} \exp\left(-\dfrac{t^s}{b}\right) & , \quad \text{with} \quad b = 2 \quad \text{and} \quad s = 0.8, \\ g_{\mathrm{mod}}(t) = \exp(-t). \end{cases}$$

The estimate obtained by SESAME under misspecification is generally referred to as $\widehat{\boldsymbol{\mu}}_{\text{Mis-SESAME}}$ whereas the one computed in the matched case is denoted by $\widehat{\boldsymbol{\mu}}_{\text{SESAME-E}}$. For Gaussian *assumed* models, we consider the particular COMET estimate, designated as $\widehat{\boldsymbol{\mu}}_{\text{COMET}}$. We also compare the performance of SESAME under both matched and mismatched models with the CRB and the MCRB. To draw the comparison, we define the Pseudo Mean Square Error (PMSE) w.r.t the *pseudo*-parameter $\boldsymbol{\mu}_0$ by

$$\mathbf{PMSE}_{\boldsymbol{\mu}_0}(\widehat{\boldsymbol{\mu}}) = \mathbb{E}\left[(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0)(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0)^T\right].$$

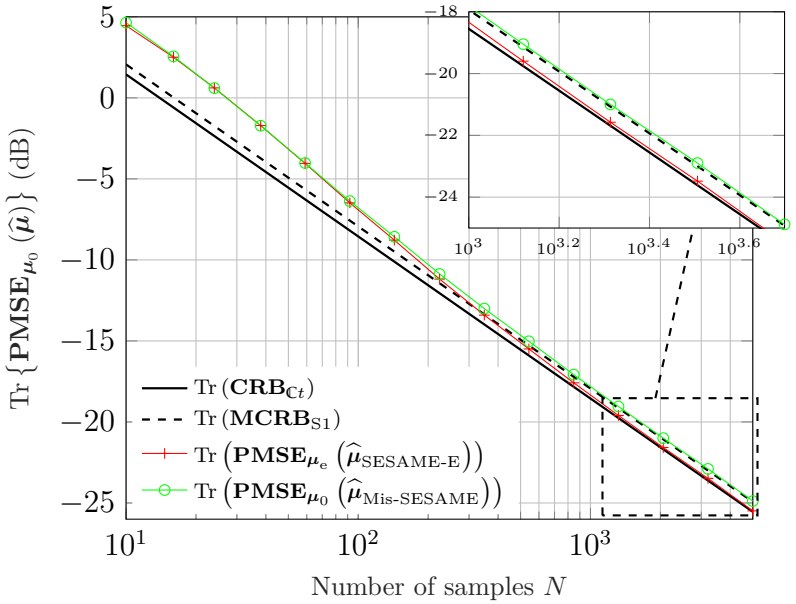

Figure 1: PMSE of SESAME procedures, *true* and *assumed* models are t-distribution with $d = 3$ and $d_{\text{mod}} = 5$.

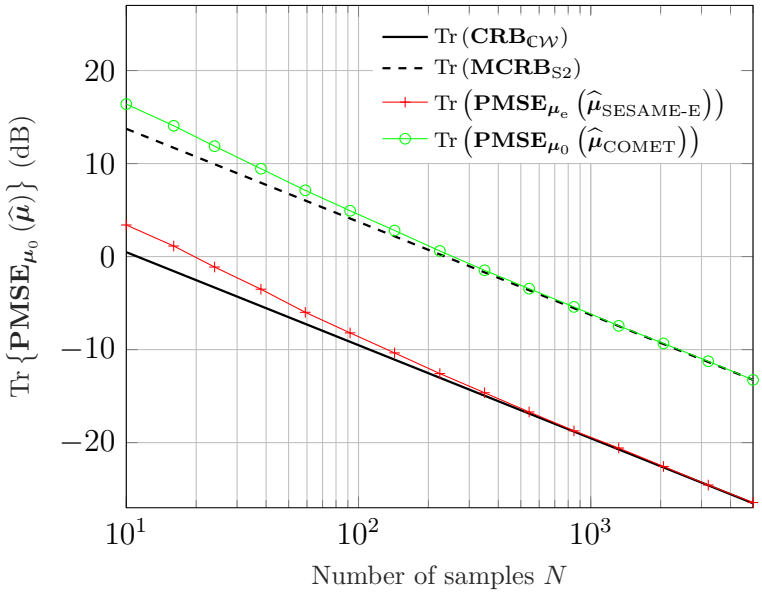

Figure 2: PMSE of SESAME procedures, *true* p.d.f. is W-distribution with $b = 2$ and $s = 0.8$ and *assumed* model is Gaussian.

In Fig. 1 and Fig. 2, the asymptotic covariance of the SESAME estimates under both matched and mismatched models reach the corresponding CRB derived in either matched or mismatched scenarios, i.e., the (m-)efficiency of the algorithm is verified. The unbiasedness

as well as the consistency can be also indirectly observed in Fig. 1 and Fig. 2.

7.3. On implementations

In the following, we consider the matched case only and we desire to illustrate the discussion on the PSD constraint conducted in the Subsection 3.2 and the performance improvement brought by R-SESAME compared to SESAME. For these numerical results, we consider an Hermitian Toeplitz scatter matrix with t-distributed (cf Section 2.1 and Table 1) observations. For $m = 4$, the Toeplitz scatter matrix is defined with its first row: $[1, -0.83 - 0.20i, 0.78 + 0.37i, -0.66 - 0.70i]$. We generate 5000 sets of N independent m-dimensional t-distributed samples, $\mathbf{y}_n \sim \mathbb{C}t_{m,d}(\mathbf{0}, \mathbf{R}_{\mathrm{e}})$, $n = 1, \ldots, N$ with $d = 5$ degrees of freedom.

We compare the performance of the proposed algorithms to the state of the art and the CRB. Furthermore, we display the performance of SESAME estimation scheme by replacing the first step by the joint-algorithm proposed in [42] to deal with the possibility of unknown parameter d (a possible relaxation of the assumption on the knowledge of the *true* model). For the performance of R-SESAME, we only consider the unstructured ML estimator as first step at each iteration. Our algorithms are compared to RCOMET from [11], COCA from [14] and Constrained Tyler from [12]. The three methods are based on the Tyler's scatter estimator [43] using normalized observations $\mathbf{z}_n = \mathbf{y}_n / \|\mathbf{y}_n\|$. It should be noted that, for Constrained Tyler, the Algorithm 3 in [12] derived for real-valued PSD Toeplitz matrices can not be directly applied. However, the Vandermonde factorization of PSD Toeplitz matrices [44] allows us to use the Algorithm 2 of [12]. In this algorithm, the set of PSD Toeplitz matrices is parameterized by $\mathcal{S} = \left\{\mathbf{R} \mid \mathbf{R} = \mathbf{A}\mathbf{P}\mathbf{A}^H\right\}$ through the unknown diagonal matrix $\mathbf{P} \succeq \mathbf{0}$ and with $\mathbf{A} = [\mathbf{a}(-90°), \mathbf{a}(-88°), \cdots, \mathbf{a}(86°), \mathbf{a}(88°)]$, where

$$\mathbf{a}(\theta) = \left[1, e^{-j\pi \sin(\theta)}, \cdots, e^{-j\pi(m-1)\sin(\theta)}\right]^T.$$

Finally, we compare to the empirical estimate $\boldsymbol{\mu}$ obtained by averaging the real and imaginary parts of diagonals of the unstructured ML estimator, which corresponds to the Euclidean projection onto the Toeplitz set.

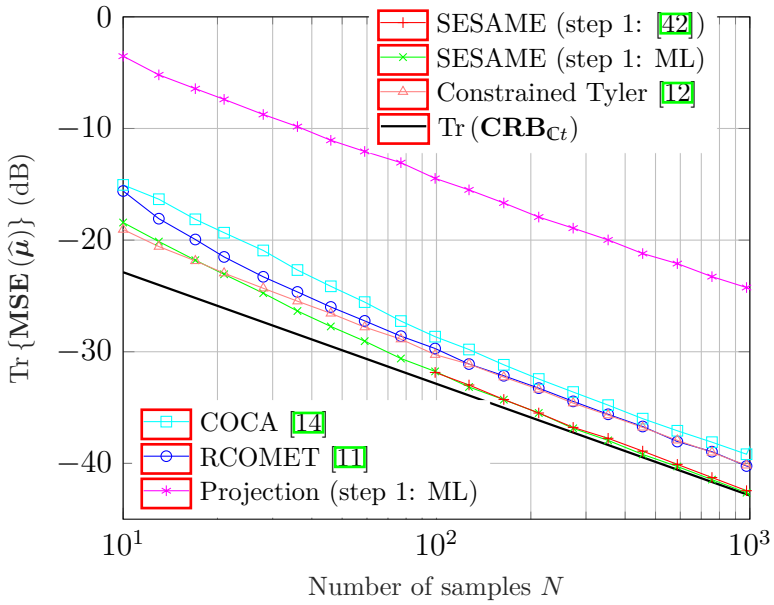

Figure 3: Efficiency simulation

The asymptotic efficiency of our estimator is checked in Fig. 3 : its MSE reaches the CRB as N increases. RCOMET, Constrained Tyler and COCA do not reach this bound since they do not take into account the underlying distribution of the data. Despite the absence of convergence proof for the joint-algorithm in [42], we notice that optimal asymptotic performance for $\boldsymbol{\mu}$ may be approached with unknown d thanks to an appropriate first estimation step. Performance of the proposed estimation scheme with the joint-algorithm as first step are not displayed for small N, since the joint-algorithm of [42] did not converge for part of the 5000 runs. In addition, the asymptotic unbiasedness of SESAME as well as those of the other algorithms can be indirectly observed on the Fig. 3.

In Fig. 4, the comparison is drawn between SESAME and R-SESAME with various numbers of iterations through the MSE. As already stated, we observe that the CRB is reached faster with R-SESAME than SESAME.

In Fig. 5 we analyse the influence of the PSD constraint on the performance of SESAME and R-SESAME. Working with a linear parameterization, the solution of SESAME without the PSD constraint is given by the closed form (46). We can check the positive semidefiniteness a posteriori. If the estimate $\mathbf{R}(\widehat{\boldsymbol{\mu}})$ is not PSD, then we solve the problem (16)

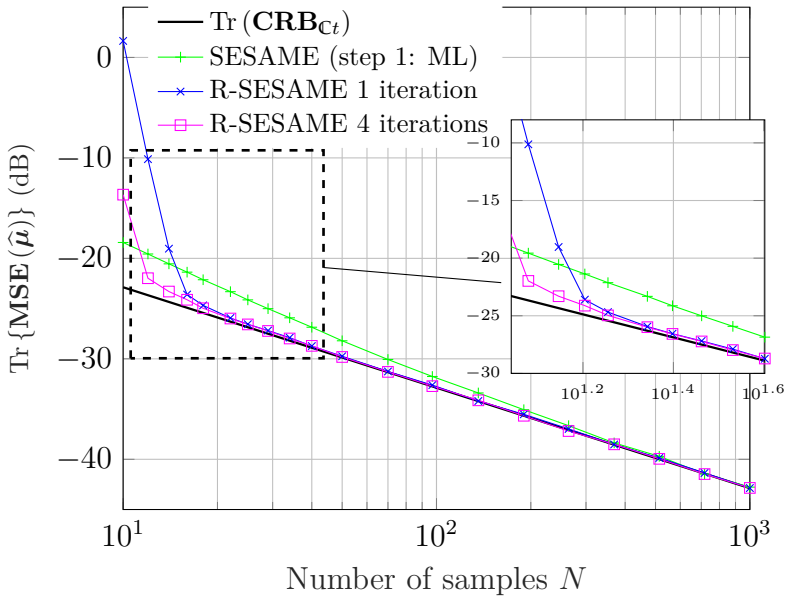

Figure 4: Efficiency simulation R-SESAME

N	RCOMET [11]	Projection ①: ML	SESAME ①: ML	SESAME PSD ①: ML	R-SESAME 1 iteration	R-SESAME PSD 1 it.	R-SESAME 4 it.	R-SESAME PSD 4 it.	Constr. Tyl. [12]	SESAME ①: [42]	COCA [14]
100	0.012s	0.018s	0.018s	0.018s	0.018s	0.018s	0.019s	0.019s	0.21s	0.46s	2.18s
500	0.048s	0.086s	0.086s	0.086s	0.087s	0.087s	0.087s	0.088s	0.76s	1.94s	15.10s
1000	0.090s	0.17s	0.173s	0.173s	0.173s	0.173s	0.174s	0.174s	1.44s	3.63s	50.40s

Table 2: Average calculation time
(Matlab R2017 a, CPU E3-1270 v5 @ 3.60 GHz, 32 GB RAM)

with CVX [45, 46]. Thus, we obtain SESAME PSD and R-SESAME PSD. For R-SESAME PSD, we draw the results for both stopping criteria: a fixed number of iterations and a combination of a maximal number of iterations and a relative gap between two successive estimates sufficiently small.

According to Fig. 5, the performance of SESAME with or without the PSD constraint appears almost identical, even for small sample support. However for small N, R-SESAME is strongly improved when the PSD constraint is considered.

Table 2 summarizes the average calculation time of the different algorithms. The proposed algorithms appear to be an interesting and time-efficient alternative to the current state of the art. The estimation scheme with the joint-algorithm is slower than the one using the exact ML-estimator, which makes sense since the degree of freedom of the t-distribution

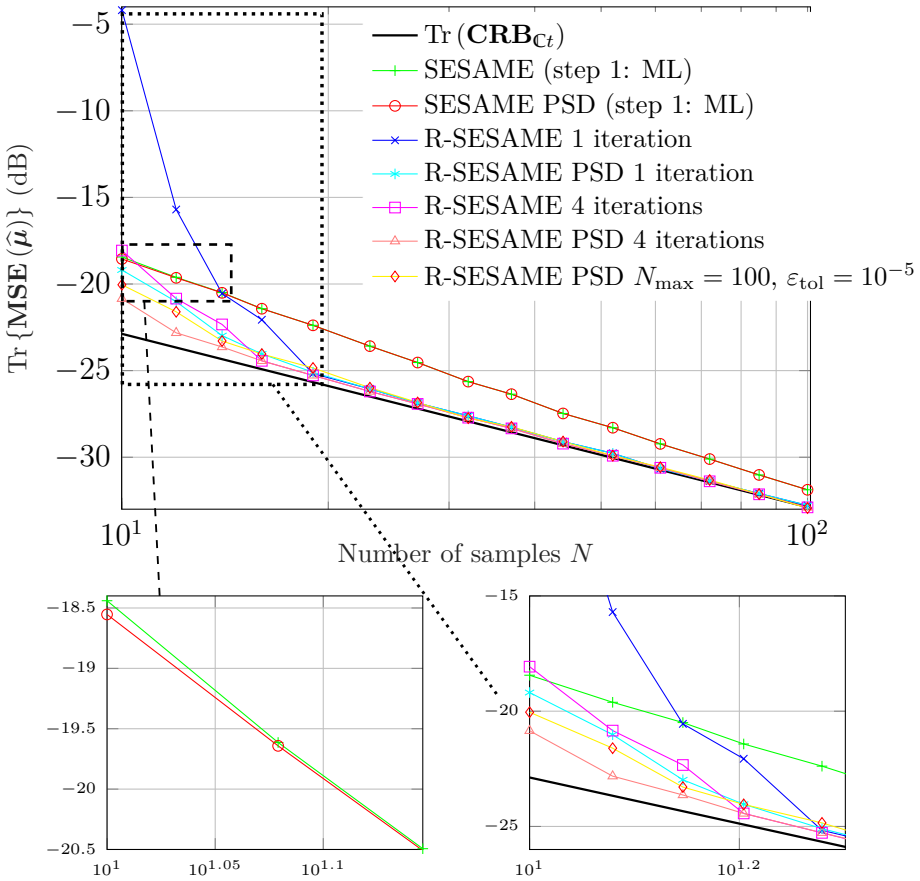

Figure 5: Comparison SESAME/R-SESAME with or without the PSD constraint

is also estimated. The COCA estimator suffers from heavy computational cost, since the number of constraints grows linearly in N.

8. Conclusion

In this paper, we have introduced structured scatter matrix estimation algorithms for centered CES distributions. Firstly, we have derived a two-step estimation procedure for the structured scatter matrix of CES distributions referred to as SESAME, taking into account possible misspecifications on the statistical model of the observations. Then, we have conducted a theoretical asymptotic analysis to demonstrate the consistency, the unbiasedness, the asymptotic Gaussianity and m-efficiency of SESAME method in the mismatched framework. In addition, we have analytically expressed the coefficients appearing in the criterion for common CES distributions. Secondly, we have proposed a recursive applica-

tion of the SESAME algorithm, which procedure leads to R-SESAME, possessing the same asymptotic performance but leading empirically to a faster convergence to the (M)CRB than SESAME. Thirdly, we have shown that some established procedures corresponds to special cases of SESAME. More specifically, with Gaussian *assumed* model, we retrieve the COMET procedure as a special case whereas in the matched case, SESAME coincides with the EXIP approach for centered CES distributions. Finally, numerical results corroborated the theoretical analysis and assessed the interest of the proposed algorithms.

9. Appendices

9.1. Appendix A: detail of calculus for Table 1

9.1.1. Complex Generalized Gaussian distribution

The density generator function is given by:

$$g(t) = \exp\left(-\frac{t^s}{b}\right) \text{ with } s > 0,\ b > 0 \text{ and } C_{m,g} = \frac{s\Gamma(m)b^{-m/s}}{\pi^m\Gamma(m/s)}.$$

We obtain $u_{\text{ML}}(t) = \frac{s}{b}t^{s-1}$ and $\psi_{\text{ML}}(t) = tu(t) = \frac{s}{b}t^s$. To ensure the convergence of the ML scatter matrix estimator, the parameter s has to be strictly lower than 1. Thus, by recognizing a generalized Gamma distribution, we have

$$\begin{aligned} A_{\text{ML}} = \mathbb{E}\left[\psi_{\text{ML}}^2(Q)\right] &= \frac{s^3}{\Gamma(m/s)b^{2+m/s}} \int_{\mathbb{R}^+} x^{m-1+2s} e^{-\frac{x^s}{b}}\,\mathrm{d}x \\ &= \frac{s^3}{\Gamma(m/s)b^{2+m/s}} \frac{\Gamma(m/s+2)b^{2+m/s}}{s} = m(m+s). \end{aligned}$$

9.1.2. Complex W-distribution

The density generator function can be written by:

$$g(t) = t^{s-1}e^{-\frac{t^s}{b}} \text{ with } s > 0,\ b > 0 \text{ and } C_{m,g} = \frac{s\Gamma(m)b^{-(s+m-1)/s}}{\pi^m\Gamma\left(\frac{s+m-1}{s}\right)}.$$

We obtain $u_{\mathrm{ML}}(t) = \frac{s}{b}t^{s-1} - (s-1)t^{-1}$ and $\psi_{\mathrm{ML}}(t) = tu(t) = \frac{s}{b}t^{s} - (s-1)$. To satisfy the Maronna's condition defined in [26], the parameter is necessarily $s < 1$. Again, the generalized Gamma distribution allows us to write

$$\begin{aligned} A_{\mathrm{ML}} &= \frac{s}{\Gamma\left(\frac{s+m-1}{s}\right) b^{\frac{s+m-1}{s}}} \int\limits_{\mathbb{R}^+} \left(\frac{s}{b}x^s - s + 1\right)^2 x^{s+m-2} e^{-\frac{x^s}{b}} \mathrm{d}x \\ &= (m+2s-1)(m+s-1) - 2(s-1)(m+s-1) + (s-1)^2 \\ &= (m+s-1)s + m^2. \end{aligned}$$

9.1.3. Complex K-distribution

The density generator function is

$$g(t) = \sqrt{t}^{\nu-m} \mathrm{K}_{\nu-m}\left(2\sqrt{\nu t}\right) \text{ with } \nu > 0 \text{ and } C_{m,g} = 2\frac{\nu^{(\nu+m)/2}}{\pi^m \Gamma(\nu)},$$

where $K_\lambda(\cdot)$ denotes the modified Bessel function of the second kind. Straightforward, we get $u_{\mathrm{ML}}(t) = \frac{\sqrt{\nu}}{t}\frac{K_{\nu-m-1}\left(2\sqrt{\nu t}\right)}{K_{\nu-m}\left(2\sqrt{\nu t}\right)}$ and $\psi_{\mathrm{ML}}(t) = \sqrt{\nu t}\frac{K_{\nu-m-1}\left(2\sqrt{\nu t}\right)}{K_{\nu-m}\left(2\sqrt{\nu t}\right)}$. Therefore, we obtain

$$\begin{aligned} A_{\mathrm{ML}} = \mathbb{E}\left[\psi^2_{\mathrm{ML}}(Q)\right] &= \frac{2\nu^{(\nu+m)/2+1}}{\Gamma(\nu)\Gamma(m)} \int\limits_{\mathbb{R}^+} \sqrt{x}^{m+\nu} \frac{K^2_{\nu-m-1}\left(2\sqrt{\nu x}\right)}{K_{\nu-m}\left(2\sqrt{\nu x}\right)} \mathrm{d}x \\ &= \frac{1}{2^{m+\nu}\Gamma(\nu)\Gamma(m)} \int\limits_{\mathbb{R}^+} x^{m+\nu+1} \frac{K^2_{\nu-m-1}(x)}{K_{\nu-m}(x)} \mathrm{d}x. \end{aligned}$$

The integral expression can not be simplified into an explicit form.

9.1.4. Complex t-distribution

The reader is referred to [42].

9.2. Appendix B: proof of equation (27)

The starting point is the expression of the gradient evaluated in $\boldsymbol{\mu} = \boldsymbol{\mu}_0$, i.e.,

$$\mathbf{g}_N(\boldsymbol{\mu}_0) = -\left.\frac{\partial \mathbf{r}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}\right|^H_{\boldsymbol{\mu}=\boldsymbol{\mu}_0} \widehat{\mathbf{Y}}\left(\widehat{\mathbf{r}}_m - \mathbf{r}_0\right) - \left(\widehat{\mathbf{r}}_m - \mathbf{r}_0\right)^H \widehat{\mathbf{Y}} \left.\frac{\partial \mathbf{r}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}\right|_{\boldsymbol{\mu}=\boldsymbol{\mu}_0} \tag{47}$$

$$= -2\Re\left(\left.\frac{\partial \mathbf{r}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}\right|_{\boldsymbol{\mu}=\boldsymbol{\mu}_0}^{H} \widehat{\mathbf{Y}}\left(\widehat{\mathbf{r}}_m - \mathbf{r}_0\right)\right).$$

Consistency of $\widehat{\mathbf{Y}}$ w.r.t. $\mathbf{Y}_\infty = \alpha^{-2}\mathbf{Y}_{\mathrm{e}}$ together with the asymptotic distribution of $\sqrt{N}\left(\widehat{\mathbf{r}}_m - \mathbf{r}_0\right)$ (cf (6) and (7)) yield, by using Slutsky's lemma [37, Chapters 2], to

$$-\sqrt{N}\mathbf{g}_N(\boldsymbol{\mu}_0) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{R}_\infty\right),$$

where the asymptotic covariance $\mathbf{R}_\infty$ reads [47]

$$\mathbf{R}_\infty = 2\left[\mathbf{D}_h \boldsymbol{\Sigma} \mathbf{D}_h^H + \Re\left(\mathbf{D}_h \boldsymbol{\Omega} \mathbf{D}_h^T\right)\right], \tag{48}$$

with

$$\mathbf{D}_h = \left.\frac{\partial \mathbf{r}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}\right|_{\boldsymbol{\mu}=\boldsymbol{\mu}_0}^{H} \alpha^{-2}\mathbf{Y}_{\mathrm{e}} = \alpha^{-2}\boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right)^H \mathbf{Y}_{\mathrm{e}} \text{ and } \boldsymbol{\Omega} = \boldsymbol{\Sigma}\mathbf{K} \text{ (cf (7))}.$$

Moreover, we can easily check that

$$\mathbf{K}\mathbf{Y}_{\mathrm{e}}^T = \mathbf{Y}_{\mathrm{e}}\mathbf{K} \text{ and } \mathbf{K}\left.\frac{\partial \mathbf{r}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}\right|_{\boldsymbol{\mu}_0}^{*} = \left.\frac{\partial \mathbf{r}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}\right|_{\boldsymbol{\mu}_0},$$

where $\mathbf{K}$ is the commutation matrix satisfying $\mathbf{K}\mathrm{vec}\left(\mathbf{A}\right) = \mathrm{vec}\left(\mathbf{A}^T\right)$ and $\mathbf{A}^*$ refers to the conjugate matrix. Then, we obtain

$$\mathbf{R}_\infty = 2\alpha^{-4}\left[\boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right)^H \mathbf{Y}_{\mathrm{e}} \boldsymbol{\Sigma} \mathbf{Y}_{\mathrm{e}} \boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right) + \Re\left(\boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right)^H \mathbf{Y}_{\mathrm{e}} \boldsymbol{\Sigma} \mathbf{Y}_{\mathrm{e}} \boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right)\right)\right]. \tag{49}$$

Furthermore, we have $\boldsymbol{\Sigma} = \sigma^{-2}\left(\sigma_1 \mathbf{W}_{\mathrm{e}} + \sigma_2 \mathrm{vec}\left(\mathbf{R}_{\mathrm{e}}\right)\mathrm{vec}\left(\mathbf{R}_{\mathrm{e}}\right)^H\right)$ with coefficients σ_1 and σ_2 defined by (10). To lighten the notations, we introduce $\mathbf{V}_{\mathrm{e}} = \mathrm{vec}\left(\mathbf{R}_{\mathrm{e}}\right)\mathrm{vec}\left(\mathbf{R}_{\mathrm{e}}\right)^H$. Moreover, we have $\mathrm{vec}\left(\mathbf{R}_{\mathrm{e}}\right)^H \mathrm{vec}\left(\mathbf{R}_{\mathrm{e}}^{-1}\right) = m$. After some calculus, we obtain:

$$\begin{aligned}
\mathbf{Y}_{\mathrm{e}}\mathbf{W}_{\mathrm{e}}\mathbf{Y}_{\mathrm{e}} &= \kappa_1^2 \mathbf{W}_{\mathrm{e}}^{-1} + \kappa_2\left(2\kappa_1 + m\kappa_2\right)\mathbf{U}_{\mathrm{e}} \\
\mathbf{Y}_{\mathrm{e}}\mathbf{V}_{\mathrm{e}}\mathbf{Y}_{\mathrm{e}} &= \left(\kappa_1 + m\kappa_2\right)^2 \mathbf{U}_{\mathrm{e}} \\
\mathbf{Y}_{\mathrm{e}}\boldsymbol{\Sigma}\mathbf{Y}_{\mathrm{e}} &= \sigma^{-2}\left[\sigma_1\kappa_1^2 \mathbf{W}_e^{-1} + \sigma_1\kappa_2\left(2\kappa_1 + m\kappa_2\right)\mathbf{U}_e + \sigma_2\left(\kappa_1 + m\kappa_2\right)^2 \mathbf{U}_e\right]
\end{aligned}$$

$$= \sigma^{-2} \underbrace{\sigma_1 \kappa_1^2}_{\beta_1} \mathbf{W}_e^{-1} + \sigma^{-2} \underbrace{\left[\sigma_1 \kappa_2 \left(2\kappa_1 + m\kappa_2\right) + \sigma_2 \left(\kappa_1 + m\kappa_2\right)^2\right]}_{\beta_2} \mathbf{U}_e.$$

Then, (49) becomes

$$\begin{aligned}
\mathbf{R}_\infty &= 2\alpha^{-4}\sigma^{-2}\left[\boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right)^H \left(\beta_1 \mathbf{W}_{\mathrm{e}}^{-1} + \beta_2 \mathbf{U}_{\mathrm{e}}\right) \boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right) + \Re\left(\boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right)^H \left(\beta_1 \mathbf{W}_{\mathrm{e}}^{-1} + \beta_2 \mathbf{U}_{\mathrm{e}}\right) \boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right)\right)\right] \\
&= 4\alpha^{-4}\sigma^{-2}\boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right)^H \left(\beta_1 \mathbf{W}_{\mathrm{e}}^{-1} + \beta_2 \mathbf{U}_{\mathrm{e}}\right) \boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right) = 4\alpha^{-4}\sigma^{-4}\boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right)^H \left(\beta_1 \mathbf{W}_0^{-1} + \beta_2 \mathbf{U}_0\right) \boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right),
\end{aligned}$$

since

$$\begin{aligned}
\left[\boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right)^H \mathbf{W}_{\mathrm{e}}^{-1} \boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right)\right]^* &= \left(\boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right)^*\right)^H \mathbf{W}_{\mathrm{e}}^{-1^*} \boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right)^* = \left(\mathbf{K}\boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right)\right)^H \mathbf{K}\mathbf{W}_{\mathrm{e}}^{-1}\mathbf{K}\mathbf{K}\boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right) \\
&= \boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right)^H \mathbf{W}_{\mathrm{e}}^{-1} \boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right), \\
\left[\boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right)^H \mathbf{U}_{\mathrm{e}} \boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right)\right]^* &= \left(\boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right)^*\right)^H \left(\mathbf{W}_{\mathrm{e}}^{-1/2} \mathrm{vec}\left(\mathbf{I}_m\right) \mathrm{vec}\left(\mathbf{I}_m\right)^H \mathbf{W}_{\mathrm{e}}^{-1/2}\right)^* \boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right)^* \\
&= \boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right)^H \mathbf{K}\mathbf{K}\mathbf{W}_{\mathrm{e}}^{-1/2}\mathbf{K}\mathrm{vec}\left(\mathbf{I}_m\right) \mathrm{vec}\left(\mathbf{I}_m\right)^H \mathbf{K}\mathbf{W}_{\mathrm{e}}^{-1/2}\mathbf{K}\mathbf{K}\boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right) \\
&= \boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right)^H \mathbf{W}_{\mathrm{e}}^{-1/2} \mathrm{vec}\left(\mathbf{I}_m\right) \mathrm{vec}\left(\mathbf{I}_m\right)^H \mathbf{W}_{\mathrm{e}}^{-1/2} \boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right) = \boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right)^H \mathbf{U}_{\mathrm{e}} \boldsymbol{\mathcal{J}}\left(\boldsymbol{\mu}_0\right).
\end{aligned}$$

Finally, we obtain

$$-\sqrt{N}\mathbf{g}_N(\boldsymbol{\mu}_0) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{R}_\infty\right).$$

Acknowledgements

This research work is partially funded by the Direction Générale de l'Armement (D.G.A.) as well as the ANR ASTRID Project D.G.A. MARGARITA referenced ANR-17-ASTR-0015.

References

[1] J. Fan, Y. Fan, J. Lv, High dimensional covariance matrix estimation using a factor model, Journal of Econometrics 147 (1) (2008) 186–197.

[2] F. Dietrich, Robust Signal Processing for Wireless Communications, Vol. 2, Springer Science & Business Media, 2008.

[3] M. Haardt, M. Pesavento, F. Röemer, M. N. El Korso, Subspace methods and exploitation of special

array structures, in: Array and Statistical Signal Processing, Vol. 3 of Academic Press Library in Signal Processing, Elsevier, 2014, Ch. 15, pp. 651–717.

[4] D. R. Fuhrmann, M. I. Miller, On the existence of positive-definite maximum-likelihood estimates of structured covariance matrices, IEEE Transactions on Information Theory 34 (4) (1988) 722–729.

[5] P. Wirfält, M. Jansson, On kronecker and linearly structured covariance matrix estimation, IEEE Transactions on Signal Processing 62 (6) (2014) 1536–1547.

[6] K. Werner, M. Jansson, P. Stoica, On estimation of covariance matrices with kronecker product structure, IEEE Transactions on Signal Processing 56 (2) (2008) 478–491.

[7] P. Forster, Generalized rectification of cross spectral matrices for arrays of arbitrary geometry, IEEE Transactions on Signal Processing 49 (5) (2001) 972–978.

[8] A. Combernoux, G. Ginolhac, P. Forster, Generalized rectification in the l1-norm with application to robust array processing, in: Proc. of European Signal Processing Conference (EUSIPCO), 2011, pp. 619–623.

[9] K. Greenewald, E. Zelnio, A. H. Hero, Robust SAR STAP via Kronecker decomposition, IEEE Transactions on Aerospace and Electronic Systems 52 (6) (2016) 2612–2625.

[10] B. Ottersten, P. Stoica, R. Roy, Covariance matching estimation techniques for array signal processing applications, Elsevier Digital Signal Processing 8 (3) (1998) 185–210.

[11] B. Mériaux, C. Ren, M. N. El Korso, A. Breloy, P. Forster, Robust-COMET for covariance estimation in convex structures: algorithm and statistical properties, in: Proc. of IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2017, pp. 1–5.

[12] Y. Sun, P. Babu, D. P. Palomar, Robust estimation of structured covariance matrix for heavy-tailed elliptical distributions, IEEE Transactions on Signal Processing 14 (64) (2016) 3576–3590.

[13] A. Breloy, Y. Sun, P. Babu, G. Ginolhac, D. P. Palomar, Robust rank constrained kronecker covariance matrix estimation, in: Proc. of Asilomar Conference on Signals, Systems and Computers (ASILOMAR), 2016, pp. 810–814.

[14] I. Soloveychik, A. Wiesel, Tyler's covariance matrix estimator in elliptical models with convex structure, IEEE Transactions on Signal Processing 62 (20) (2014) 5251–5259.

[15] E. Ollila, D. E. Tyler, V. Koivunen, H. V. Poor, Complex elliptically symmetric distributions: Survey, new results and applications, IEEE Transactions on Signal Processing 60 (11) (2012) 5597–5625.

[16] K. D. Ward, C. J. Baker, S. Watts, Maritime surveillance radar. part 1: Radar scattering from the ocean surface, IEE Proceedings F (Communications, Radar and Signal Processing) 137 (2) (1990) 51–62.

[17] K. J. Sangston, F. Gini, M. S. Greco, Adaptive detection of radar targets in compound-Gaussian clutter, in: Proc. of IEEE Radar Conference, May, 2015, pp. 587–592.

[18] X. Zhang, M. N. El Korso, M. Pesavento, MIMO radar target localization and performance evaluation

under SIRP clutter, Elsevier Signal Processing 130 (2017) 217–232.

[19] P. J. Huber, The behavior of maximum likelihood estimates under nonstandard conditions, Proceedings of the fifth Berkeley symposium on mathematical statistics and probability 1 (1) (1967) 221–233.

[20] H. White, Maximum likelihood estimation of misspecified models, Econometrica 50 (1) (1980) 1–25.

[21] C. D. Richmond, L. L. Horowitz, Parameter bounds on estimation accuracy under model misspecification, IEEE Transactions on Signal Processing 63 (9) (2015) 2263–2278.

[22] S. Fortunati, F. Gini, M. S. Greco, The misspecified Cramér-Rao bound and its application to scatter matrix estimation in complex elliptically symmetric distributions, IEEE Transactions on Signal Processing 64 (9) (2016) 2387–2399.

[23] A. Mennad, S. Fortunati, M. N. El Korso, A. Younsi, A. M. Zoubir, A. Renaux, Slepian-Bangs-type formulas and the related misspecified Cramér-Rao bounds for complex elliptically symmetric distributions, Elsevier Signal Processing 142 (2017) 320–329.

[24] P. Stoica, T. Söderström, On reparametrization of loss functions used in estimation and the invariance principle, Elsevier Signal Processing 17 (4) (1989) 383–387.

[25] B. Mériaux, C. Ren, M. N. El Korso, A. Breloy, P. Forster, Efficient estimation of scatter matrix with convex structure under t-distribution, in: Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2018, pp. 4474–4478.

[26] R. A. Maronna, Robust M-estimators of multivariate location and scatter, The Annals of Statistics 4 (1) (1976) 51–67.

[27] J. T. Kent, D. E. Tyler, Redescending M-estimates of multivariate location and scatter, The Annals of Statistics 19 (4) (1991) 2102–2119.

[28] M. Mahot, F. Pascal, P. Forster, J.-P. Ovarlez, Asymptotic properties of robust complex covariance matrix estimates, IEEE Transactions on Signal Processing 61 (13) (2013) 3348–3356.

[29] A. Van den Bos, The multivariate complex normal distribution - A generalization, IEEE Transactions on Information Theory 41 (2) (1995) 537–539.

[30] J. R. Magnus, H. Neudecker, The commutation matrix: some properties and applications, The Annals of Statistics 7 (2) (1979) 381–394.

[31] M. Abramowitz, I. A. Stegun, Handbook of mathematical functions: with formulas, graphs, and mathematical tables, Vol. 55, Courier Corporation, 1964.

[32] A. Breloy, G. Ginolhac, A. Renaux, F. Bouchard, Intrinsic Cramér–Rao bounds for scatter and shape matrices estimation in CES distributions, IEEE Signal Processing Letters 26 (2) (2019) 262–266.

[33] J. Brewer, Kronecker products and matrix calculus in system theory, IEEE Transactions on Circuits and Systems 25 (9) (1978) 772–781.

[34] H. B. Mann, A. Wald, On stochastic limit and order relationships, The Annals of Mathematical Statis-

tics 14 (3) (1943) 217–226.

[35] J. R. Magnus, H. Neudecker, Matrix differential calculus with applications to simple, hadamard, and kronecker products, Journal of Mathematical Psychology 29 (4) (1985) 474–492.

[36] W. F. Trench, Introduction to Real Analysis, Prentice Hall, 2002.

[37] A. W. Van der Vaart, Asymptotic Statistics (Cambridge Series in Statistical and Probabilistic Mathematics), Vol. 3, Cambridge University Press, 2000.

[38] K. S. Miller, On the inverse of the sum of matrices, Mathematics Magazine 54 (2) (1981) 67–72.

[39] P. Stoica, B. C. Ng, On the Cramér-Rao bound under parametric constraints, IEEE Signal Processing Letters 5 (7) (1998) 177–179.

[40] R. Muirhead, Aspects of Multivariate Statistical Theory, Wiley Series in Probability and Statistics, Wiley, 2009.

[41] O. Besson, Y. I. Abramovich, On the fisher information matrix for multivariate elliptically contoured distributions, IEEE Signal Processing Letters 20 (11) (2013) 1130–1133.

[42] S. Fortunati, F. Gini, M. S. Greco, Matched, mismatched, and robust scatter matrix estimation and hypothesis testing in complex t-distributed data, EURASIP Journal on Advances in Signal Processing 2016 (1) (2016) 123.

[43] D. E. Tyler, A distribution-free M-estimator of multivariate scatter, The Annals of Statistics 15 (1) (1987) 234–251.

[44] T. Bäckström, Vandermonde factorization of toeplitz matrices and applications in filtering and warping, IEEE Transactions on Signal Processing 61 (24) (2013) 6257–6263.

[45] M. Grant, S. Boyd, Graph implementations for nonsmooth convex programs, in: V. Blondel, S. Boyd, H. Kimura (Eds.), Recent Advances in Learning and Control, Lecture Notes in Control and Information Sciences, Springer-Verlag Limited, 2008, pp. 95–110, http://stanford.edu/~boyd/graph_dcp.html.

[46] M. Grant, S. Boyd, CVX: Matlab software for disciplined convex programming, version 2.1, http://cvxr.com/cvx (Mar. 2014).

[47] J.-P. Delmas, H. Abeida, Survey and some new results on performance analysis of complex-valued parameter estimators, Elsevier Signal Processing 111 (2015) 210–221.