

Tyler's Covariance Matrix Estimator in Elliptical Models with Convex Structure

Ilya Soloveychik, Ami Wiesel

Abstract—We address structured covariance estimation in elliptical distributions by assuming that the covariance is a priori known to belong to a given convex set, e.g., the set of Toeplitz or banded matrices. We consider the General Method of Moments (GMM) optimization applied to robust Tyler's scatter M-estimator subject to these convex constraints. Unfortunately, GMM turns out to be non-convex due to the objective. Instead, we propose a new COCA estimator - a convex relaxation which can be efficiently solved. We prove that the relaxation is tight in the unconstrained case for a finite number of samples, and in the constrained case asymptotically. We then illustrate the advantages of COCA in synthetic simulations with structured compound Gaussian distributions. In these examples, COCA outperforms competing methods such as Tyler's estimator and its projection onto the structure set.

Index Terms—Elliptical distribution, Tyler's scatter estimator, Generalized Method of Moments, robust covariance estimation.

I. INTRODUCTION

Covariance matrix estimation is a fundamental problem in the field of statistical signal processing. Many algorithms for detection and inference rely on accurate covariance estimators [1, 2]. The problem is well understood in the Gaussian unstructured case. But becomes significantly harder when the underlying distribution is non-Gaussian, for example in elliptical distributions, and when there is prior knowledge on the structure. In this paper, we propose a unified framework for covariance estimation in elliptical distributions with general convex structure.

Over the last years there was a great interest in covariance estimation with known structure. The motivation to these works is that in many modern applications the dimension of the underlying distribution is large and there are not enough samples to estimate it precisely without additional structure hypotheses. The prior information on the structure reduces the number of degrees of freedom in the model and allows accurate estimation with a small number of samples. This is clearly true when the structure is exact, but also when it is approximate due to the well known bias-variance tradeoff. Prior knowledge on the structure can originate from the

physics of the underlying phenomena or from similar datasets, e.g., [3, 4, 5, 6, 7].

Many covariance structures are easily represented in convex form. Probably the most popular structure is the Toeplitz model. It arises naturally in the analysis of stationary time series which are used in a wide range of applications including radar imaging, target detection, speech recognition, and communication systems, [3, 4, 8]. Toeplitz matrices are also used to model the correlation of cyclostationary processes in periodic time series [9]. In other settings the number of parameters can be reduced by assuming that the covariance matrix is sparse [10, 11]. A popular sparse model is the banded covariance, which is associated with time-varying moving average models [11]. Another important example of a convex structure is the SPICE estimator, which was proposed in [6] to treat high-dimensional array processing problems, where the covariance structure is approximated by a low-dimensional linear combination of known rank one matrices. In the last decade, all of these structures have been successfully considered in the Gaussian case.

In a different line of works, there is an increasing interest in robust covariance estimation for non-Gaussian distributions [12, 13, 14, 15, 16]. Significant attention is being paid to the family of elliptical and generalized elliptical (GE) distributions [17, 18], which include as particular cases Generalized Gaussian distribution (GG), Compound Gaussian (CG) and many others [17]. Elliptical models are commonly used to measure radar clutter [19], noise and interference in indoor and outdoor mobile communication channels [20] and other applications. For these purposes robust covariance estimators were developed including Maronna's famous scatter M-estimator [21]. Later Tyler [22] proposed a particular kind of scatter M-estimator which has become widely used [12, 13, 23]. Although, generally M-estimators are given as solutions to optimization programs, Tyler showed that his M-estimator can be obtained as a solution to a simple fixed point equation. One of the most prominent disadvantage of these methods is that the optimization programs appearing from them are non-convex, thus making imposition of additional constraints rather difficult. In some cases it is possible to cure this obstacle by changing the metric of the underlying manifold, thus appealing to the notion of geodesic (g -) convexity. It has been recently shown that some of the popular M-estimators are in fact g -convex, which significantly simplifies their treatment [14]. However, the imposition of additional constraints on the scatter matrix requires to define them as g -convex sets, rather than classical convex sets. Different improvements were achieved in this direction [16], but this field is not

Copyright (c) 2014 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org

This work was partially supported by the Intel Collaboration Research Institute for Computational Intelligence, the Kaete Klausner Scholarship and ISF Grant 786/11.

The authors are with the Rachel and Selim Benin School of Computer Science and Engineering, the Hebrew University of Jerusalem, Israel (e-mail: ilya.soloveychik@mail.huji.ac.il)

The preliminary results were presented in ICASSP 2014.

still developed enough in order to solve the problems under consideration.

In the present work we derive COCA - CONvexly CONstrAined Covariance Matching estimator. The underlying idea is based on the concept of Generalized Method of Moments (GMM) [24]. COCA searches for a covariance possessing the given convex structure that minimizes the norm of a sample moment's identity. This identity is in fact the optimality condition of an M-estimator. COCA tries to simultaneously satisfy this condition while constraining the structure. Unfortunately, this requires solving a high dimensional non-convex minimization program. Instead, we propose a convex relaxation and express COCA as a standard convex optimization with linear matrix inequalities which can be computed using off-the-shelf numerical solvers, such as CVX, [25, 26]. In terms of its analysis, we prove two promising results. First, in the unconstrained case, COCA is tight and identical to Tyler's estimator. This result basically "convexifies" Tyler's estimator. Second, in the structured case, COCA is asymptotically tight and hence consistent. Finally, we demonstrate the finite sample advantages of COCA over existing methods using numerical simulations.

The paper is organized in the following way. First, we formulate the problem, derive its Cramer-Rao performance bound and briefly describe the existing solutions: the sample covariance, Tyler's estimator and the projection method. We then introduce GMM, derive its convex relaxation named COCA and show that it coincides with Tyler's estimator in the unconstrained case. Then we prove that adding convex structure does not affect asymptotic consistency of the COCA estimator. Finally, we provide numerical examples and applications demonstrating the performance advantages of COCA.

We denote by $\mathcal{P}(p)$ the closed cone of hermitian positive semi-definite $p \times p$ matrices. We write $\mathbf{M} \succeq 0$ if $\mathbf{M} \in \mathcal{P}(p)$ for some $p \in \mathbb{N}$ and $\mathbf{M} \succ 0$ if in addition all the eigenvalues of \mathbf{M} are positive. When convergence of random entities is considered *a.s.* denotes the almost sure with respect to the probability measure convergence and \xrightarrow{P} denotes convergence in probability. For a matrix \mathbf{M} , $\|\mathbf{M}\|_F$ and $\|\mathbf{M}\|_2$ denote correspondingly its Frobenius and spectral norms; $\|\mathbf{M}\|$ stands for a not specified norm. $|\mathbf{M}|$ denotes the determinant of the matrix; \mathbf{M}^T stands for the transpose and \mathbf{M}^H for the conjugate transpose matrix. Given a matrix \mathbf{M} , the operator $\text{vec}(\mathbf{M})$ stacks all its columns into a one tall column. The sample measurements $\mathbf{x}_i \in \mathbb{C}^p$ are assumed to be independent and identically distributed (i.i.d.). We write $\mathbf{x} \sim \mathcal{N}(0, \mathbf{M})$ for a centered complex circularly symmetric normally distributed random vector with covariance matrix \mathbf{M} . We use j for the imaginary unit and avoid using it as an ordering index. \mathbf{I} denotes the identity matrix of a proper dimension. Matrices are denoted by bold Capital letters \mathbf{M} , column vectors by bold non-capital \mathbf{v} and scalars by non-capital r .

II. MODEL AND PROBLEM FORMULATION

A. Complex Angular Elliptical distribution

Consider a p dimensional complex zero mean Generalized Elliptically (GE) distributed random vector $\mathbf{s} \in \mathbb{C}^p$, [17, 18].

Such a vector can be defined as [17]

$$\mathbf{s} = r\mathbf{\Lambda}\mathbf{u},$$

where \mathbf{u} is a q dimensional random vector, uniformly distributed over the unit complex hypersphere, r is a nonnegative random variable, $\mathbf{\Lambda} \in \mathbb{C}^{p \times q}$. The random variable r is called the generating variate of \mathbf{s} , we assume r has no atom at 0. If, in addition, we require r to be stochastically independent of \mathbf{u} the distribution becomes elliptical. Below, we will normalize the random vectors and eliminate their generating variate. This will allow us to treat both GE and elliptical families in a similar way. The parameter $\mathbf{\Theta}_0 = \mathbf{\Lambda}\mathbf{\Lambda}^H$ is referred to as the dispersion or shape matrix of \mathbf{x} and coincides with its covariance matrix (up to a scaling factor) when the latter exists. We assume $\mathbf{\Theta}_0 \succ 0$. The topic of this paper is the estimation of this shape matrix.

A closely related distribution is the Complex Angular Elliptically (CAE), [27], denoted by $\mathbf{x} \sim \mathcal{U}(\mathbf{\Theta}_0)$. This distribution can be obtained by normalizing a GE random vector \mathbf{s} :

$$\mathbf{x} = \frac{\mathbf{s}}{\|\mathbf{s}\|}, \mathbf{s} \neq 0,$$

The CAE probability density function is given by [27]

$$p(\mathbf{x}) = \frac{(p-1)!}{\pi^p} \frac{1}{|\mathbf{\Theta}_0|(\mathbf{x}^H \mathbf{\Theta}_0^{-1} \mathbf{x})^p}, \quad (1)$$

where \mathbf{x} belongs to a complex unit p -dimensional sphere. Ignoring additive constants, the negative log-likelihood function of CAE distribution is given by

$$l(\mathbf{\Theta}; \mathbf{x}) = \log|\mathbf{\Theta}| + p \log(\mathbf{x}^H \mathbf{\Theta}^{-1} \mathbf{x}). \quad (2)$$

The GE class of distributions includes proper complex Gaussian, compound Gaussian, elliptical, skew-elliptical, CAE and other distributions, [18]. An important property of the GE family is that the shape matrix of a population does not change whenever the random vector is divided by its Euclidean norm [17, 18]. As explained above, after normalization any GE vector becomes CAE distributed. This allows us to treat the shape matrices of all the mentioned distributions using a single robust estimator.

Note that the negative log-likelihood (2) is insensitive to multiplication of the shape matrix by a positive constant, thus we are only interested in the estimation of the shape matrix up to a positive scalar factor. There are different approaches of fixing the scale; below we fix the trace of the estimator to get rid of this ambiguity.

B. Structure

In many applications, it is common to assume prior information on the structure of $\mathbf{\Theta}_0$. In particular, we can assume that it belongs to a known closed convex subset $\mathcal{S} \subset \mathcal{P}(p)$. In many applications the role of \mathcal{S} is played by a part of an affine hyperplane laying inside $\mathcal{P}(p)$. For simplicity, we consider the case of an affine set \mathcal{S} , but most of the results can be generalized to an arbitrary closed convex set separated from zero. Specifically, we assume that

$$\mathcal{S} = \mathcal{L} \cap \mathcal{P}(p), \quad (3)$$

where the hyperplane \mathcal{L} is parametrized as

$$\mathcal{L} = \mathbf{B}_0 + \left\{ \sum_{i=1}^k a_i \mathbf{B}_i, \mathbf{B}_i \in \mathbb{C}^{p \times p}, a_i \in \mathbb{R} \right\}, \quad (4)$$

where $\mathbf{B}_0 \succeq 0$ and all \mathbf{B}_i are hermitian, $k = \dim \mathcal{L} \leq p^2$. Note that \mathcal{L} is parametrized by the real numbers $a_i \in \mathbb{R}, i = 1, \dots, k$, this is necessary to preserve the hermitian structure of the matrices in the affine hyperplane - if complex numbers are allowed the resulting sum may not be hermitian.

Typical examples of such affine structures are:

- **Toeplitz:** In stationary time series, the covariance between the i -th and the h -th components depend only on the the difference $|i - h|$. This kind of processes is encountered very often in many engineering areas including statistical signal processing, radar imaging, target detection, speech recognition, and communications systems, [3, 4, 8, 9, 28, 29]. The hyperplane \mathcal{L} forms a $k = 2p - 1$ dimensional affine subspace. Using the notations in (4) we define the basis matrices as

$$\mathbf{B}_0 = 0, \mathbf{B}_1 = \mathbf{I},$$

$$\mathbf{B}_i = \begin{pmatrix} 0 & \dots & 1 & \dots & \dots & 0 \\ \vdots & \ddots & \vdots & 1 & \ddots & 0 \\ 1 & \dots & 0 & \dots & \dots & 0 \\ 0 & 1 & \dots & 0 & \dots & 1 \\ \vdots & \ddots & \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \dots & 1 & \dots & 0 \end{pmatrix}, i = 2, \dots, p,$$

where the both $i - 1$ -th subdiagonals consist of ones, and analogously

$$\mathbf{B}_i = \begin{pmatrix} 0 & \dots & j & \dots & \dots & 0 \\ \vdots & \ddots & \vdots & j & \ddots & 0 \\ -j & \dots & 0 & \dots & \dots & 0 \\ 0 & -j & \dots & 0 & \dots & j \\ \vdots & \ddots & \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \dots & -j & \dots & 0 \end{pmatrix},$$

$i = p + 1, \dots, 2p - 1$.

- **Banded:** A natural approach to covariance modeling is to quantify the statistical relation using the notion of independence or correlation, which corresponds to sparsity in the covariance matrix [11]. Assuming that i -th element of the random vector is uncorrelated with the h -th if $|i - h| > b$ leads to b -banded structure, also known as time varying moving average models. Using the definition in (4) we have symmetric matrices

$$\mathbf{B}_0 = 0, \mathbf{B}_m = \mathbf{E}_{ih} + \mathbf{E}_{hi}, \quad (5)$$

where \mathbf{E}_{ih} are the unit matrices, i runs from 1 to p , h from i to $\min(i + b, p)$ and m runs from 1 to $m_s = \frac{(2p-b)(b+1)}{2}$ and insures linear ordering, and the hermitian ones

$$\mathbf{B}_m = j\mathbf{E}_{ih} - j\mathbf{E}_{hi}, \quad (6)$$

where i runs from 1 to $p - 1$, h from $i + 1$ to $\min(i + b, p)$ and m runs from $m_s + 1$ to $2m_s - p = p(2b + 1) - b(b + 1)$.

- **Direction of Arrival Problem:** The problem of finding the direction of arrivals (DOA's) of k plane waves impinging on a passive array of p narrow-banded sensors can be reduced to that of estimating the parameters in the following model [6]

$$\mathbf{x}_i = \mathbf{B}(\boldsymbol{\theta})\mathbf{y}_i + \boldsymbol{\omega}_i, \quad i = 1, \dots, n, \quad (7)$$

where $\mathbf{x}_i \in \mathbb{C}^p$ are the noisy observation vectors, $\mathbf{y}_i \in \mathbb{C}^k$ are the unknown signal vectors, and $\boldsymbol{\omega}_i \in \mathbb{C}^p$ are hidden noise processes. The vector $\boldsymbol{\theta} = [\theta_1 \dots \theta_k]$ consists of unknown real parameters, and the matrix $\mathbf{B}(\boldsymbol{\theta}) \in \mathbb{C}^{p \times k}$ has the following special structure:

$$\mathbf{B}(\boldsymbol{\theta}) = [\mathbf{b}(\theta_1) \dots \mathbf{b}(\theta_k)], \quad (8)$$

where $\mathbf{b}(\theta_h)$ is the so-called steering vector or transfer vector (between the h -th signal source and the array output i). The exact form of the $\mathbf{b}(\theta)$ vectors depend on the array configuration. For example in a uniform and linear array we have

$$\mathbf{b}(\theta) = [1 \ e^{j\theta} \ e^{2j\theta} \ \dots \ e^{(p-1)j\theta}]^T \quad (9)$$

Assuming that the elements of \mathbf{y}_i are statistically independent of the noise, the covariance matrix of \mathbf{x}_i can be decomposed as

$$\boldsymbol{\Theta}_0 = \sum_{i=1}^k v_i^2 \mathbf{b}(\theta_i) \mathbf{b}(\theta_i)^H + \sigma^2 \mathbf{I}, \quad (10)$$

where v_i^2 denote the signal sources powers and σ^2 stands for the power of the additive white noise. The goal is to estimate $\boldsymbol{\theta}$ from the measurements $\mathbf{x}_i, i = 1, \dots, n$. In particular, a standard approach is to estimate the covariance with a structure that satisfies (10) and solve for the corresponding $\boldsymbol{\theta}$. For this purpose, we generate a dense grid of N points $\bar{\theta}_i$ over the interval of possible angles $[\theta_l; \theta_u]$ and fit the true covariance matrix by the linear model

$$\boldsymbol{\Theta} = \sum_{i=1}^N a_i \mathbf{b}(\bar{\theta}_i) \mathbf{b}(\bar{\theta}_i)^H + \sigma^2 \mathbf{I},$$

where $a_i \geq 0$. If necessary, the l_1 norm of the parameter vector $\mathbf{a} = \{a_1, \dots, a_k\}$ can be constrained to ensure sparsity and linear independence. Returning to the structure notations in (4) we have

$$\mathbf{B}_0 = \sigma^2 \mathbf{I}, \mathbf{B}_i = \mathbf{b}(\bar{\theta}_i) \mathbf{b}(\bar{\theta}_i)^H, i = 1, \dots, N.$$

C. Problem

We can now state the problem addressed in this paper: let $\mathbf{x}_i \sim \mathcal{U}(\boldsymbol{\Theta}_0), i = 1, \dots, n$, $\boldsymbol{\Theta}_0 \in \mathcal{S}$ and assume the prior knowledge on the true covariance matrix is given in the form of an affine set. We are interested in estimation of the unknown shape matrix $\boldsymbol{\Theta}_0$.

III. PERFORMANCE BOUNDS

Before addressing the possible solutions for the above covariance estimation problem, it is instructive to examine the inherent performance bounds. For this purpose, we consider the Cramer-Rao Bound (**CRB**) on the variance of an unbiased estimators. Under mild regularity conditions, the **CRB** is asymptotically achievable by the Maximum Likelihood Estimator (MLE) and is therefore an important benchmark.

A straight forward approach to the **CRB** is to use the structured parameterization in (4) and compute the Fisher Information Matrix **FIM**(**a**) associated with the parameter vector $\mathbf{a} = \{a_1, \dots, a_k\}$. The **CRB**(**a**) matrix would then be obtained by inverting this matrix. Below we use the explicit dependence of the estimator $\Theta(\mathbf{a})$ on its parameter vector **a** to calculate the **CRB**(Θ). We then bound the Mean Squared Error **MSE**(Θ) of any unbiased estimator by the trace of the **CRB**(Θ) matrix.

As we have already mentioned above the negative log-likelihood (2) of the CAE population is not sensitive to the scaling of the shape matrix. Thus, the **FIM** is singular and cannot be inverted to obtain the **CRB**, this phenomenon is known as non-identifiability of parameters, see e.g. [30] and references therein for an extensive treatment of this issue. Indeed, it is impossible to estimate the scaling of the covariance due to the normalization in our model. Instead, we need an alternative parameterization which eliminates this scale invariance. Specifically, we add the constraint $\text{Tr}(\Theta) = p$ and reparameterize the structure of \mathcal{L} while lowering its dimension to $k' = k - 1$:

$$\mathcal{L}' = \mathbf{D}_0 + \left\{ \sum_{i=1}^{k-1} a_i \mathbf{D}_i, \mathbf{D}_i \in \mathbb{C}^{p \times p}, a_i \in \mathbb{R} \right\}, \mathbf{D}_0 \succeq 0. \quad (11)$$

From now on we denote

$$\mathcal{S}' = \mathcal{L}' \cap \mathcal{P}(p).$$

For example, in the Toeplitz and banded examples discussed above we have:

- **Toeplitz:** The coefficient a_1 in \mathcal{L} is no longer needed since the main diagonal becomes known, and

$$\mathbf{D}_0 = \mathbf{I}, \mathbf{D}_i = \mathbf{B}_{i+1}, i = 1, \dots, 2p - 2.$$

- **Banded:** Of the p diagonal elements, the first $p - 1$ are chosen as independent, and we obtain

$$\mathbf{D}_0 = \mathbf{I}, \mathbf{D}_i = \mathbf{B}_i - \mathbf{B}_p, i = 1, \dots, p - 1,$$

$$\mathbf{D}_m = \mathbf{B}_{m+1}, m = p, \dots, p(2b + 1) - b(b + 1) - 1.$$

- **Direction of Arrival Problem:** In the DOA case it is more convenient to set $\text{Tr}(\Theta) = \sigma^2 p$, thus

$$\mathbf{D}_0 = \sigma^2 \mathbf{I}.$$

All the matrices $\mathbf{B}_i, i = 1, \dots, N$ satisfy $\text{Tr}(\mathbf{B}_i) = \mathbf{b}(\bar{\theta}_i)^H \mathbf{b}(\bar{\theta}_i) = p$, thus we set

$$\mathbf{D}_i = \mathbf{B}_i - \mathbf{I}, i = 1, \dots, N.$$

Note that \mathbf{D}_i are linearly independent due to the specific choice of \mathbf{B}_i as above.

Given this scale dependent parametrization, the **FIM**(**a**) computed element-wise reads as

$$\begin{aligned} \mathbf{FIM}_{hm}(\mathbf{a}) &= \mathbb{E} \left(\frac{\partial \ln p(\mathbf{x}; \mathbf{a}_0)}{\partial a_h} \frac{\partial \ln p(\mathbf{x}; \mathbf{a}_0)}{\partial a_m} \right) \\ &= -\mathbb{E} \left(\frac{\partial^2 \ln p(\mathbf{x}; \mathbf{a}_0)}{\partial a_h \partial a_m} \right), \end{aligned} \quad (12)$$

where \mathbf{a}_0 corresponds to the parametrization of the true covariance matrix. We follow the calculations of [27, 31] to obtain

$$\mathbf{FIM}(\mathbf{a})_{hm} = \frac{p \text{Tr}(\Theta_0^{-1} \mathbf{D}_m \Theta_0^{-1} \mathbf{D}_h) - \text{Tr}(\Theta_0^{-1} \mathbf{D}_h) \text{Tr}(\Theta_0^{-1} \mathbf{D}_m)}{p + 1},$$

$h, m = 1, \dots, k'.$

The unknown covariance Θ depends on **a** linearly:

$$\text{vec}(\Theta) = \mathbf{J} \mathbf{a},$$

where

$$\mathbf{J} = \left\{ \frac{\partial \text{vec}(\Theta)}{\partial \mathbf{a}} \right\} = \{ \text{vec}(\mathbf{D}_1), \dots, \text{vec}(\mathbf{D}_k) \}$$

is the Jacobian $p^2 \times k'$ matrix. Thus, the **CRB** of the covariance error reads as

$$\mathbf{CRB}(\Theta) = \mathbf{J} \mathbf{CRB}(\mathbf{a}) \mathbf{J}^H = \mathbf{J} \mathbf{FIM}^{-1}(\mathbf{a}) \mathbf{J}^H,$$

and the total **MSE** over all the elements in the covariance estimator is bounded as

$$\mathbf{MSE}(\Theta^{\text{Est}}) = \mathbb{E} \left[\|\Theta^{\text{Est}} - \Theta\|_F^2 \right] \geq \text{Tr}(\mathbf{CRB}(\Theta)).$$

IV. EXISTING SOLUTIONS

In this section, we review the existing solutions to the covariance estimation problem with and without structure.

A. Sample Covariance

The classical solution to the above covariance estimation problem is the sample covariance matrix defined by

$$\Theta^{\text{SC}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^H. \quad (13)$$

The sample covariance estimator is unbiased, always exists and is asymptotically consistent in any distribution with bounded second moments by the Law of Large Numbers. In the Gaussian case when $n \geq p$, it also maximizes the likelihood and is asymptotically efficient. In the elliptical case it converges to a scaled shape matrix. Sample covariance has been extensively studied so far and is generally suboptimal. A broad exposition on sample covariance performance for a large class of distributions was performed in [32, 33]. An additional disadvantage of this estimator is its ignorance to the prior structure.

In the recent years there have been proposed a number of covariance matrix estimators for Gaussian models with convex structure based on sample covariance, see e.g. [34, 35].

B. Tyler's M-estimator

The most popular approach to the shape matrix estimation in elliptical distributions is due to Tyler [22]. Tyler's M-estimator is defined as the fixed point solution to the equation:

$$\Theta^{\text{Tyler}} = \frac{p}{n} \sum_{i=1}^n \frac{\mathbf{x}_i \mathbf{x}_i^H}{\mathbf{x}_i^H [\Theta^{\text{Tyler}}]^{-1} \mathbf{x}_i}. \quad (14)$$

This equation defines Θ^{Tyler} up to a scaling factor, so the scale has to be fixed by some additional constraint. Two popular choices are to fix $\text{Tr}(\Theta^{\text{Tyler}})$ or $|\Theta^{\text{Tyler}}|$ as constant. When $n > p$, it has been proven that the fixed point iteration converges to the unique solution with probability one [21]. This estimator is asymptotically consistent in all elliptical distributions. In fact, it maximizes the likelihood of the CAE population (2). The advantages of Tyler's estimator are its simplicity and robustness. Its most significant drawbacks are that it does not necessarily exist if $n < p$ and can hardly exploit known structure since the optimization problem obtained by minimizing the empirical likelihood (2) is not convex. In [23] knowledge based variants of the fixed point iteration were proposed without convergence analysis. Recently, regularized and structured versions of Tyler's estimator were proposed in [13, 14, 36, 37] based on the theories of concave Perron Frobenius and geodesic convexity. Another approach of imposing linear symmetry structure on Tyler's estimator, making extensive use of the g -convexity of the problem, was proposed by [16], where the constraint set is given as a set of fixed points of certain isometries over the manifold $\mathcal{P}(p)$. Unfortunately, these approaches are limited in their modeling capabilities and cannot deal with general convex models as described above.

C. Convex Projection

A natural approach for introducing convex structure into covariance estimation is via projection. In our settings the projection is made onto a convex set \mathcal{S}' defined above, e.g. [38, 39]. Given any unstructured estimator Θ^{Est} , e.g., the sample covariance or Tyler's estimator, its projection onto the closed convex set \mathcal{S}' is defined as

$$\mathcal{P}_{\mathcal{S}'}(\Theta^{\text{Est}}) = \underset{\mathbf{M} \in \mathcal{S}'}{\text{argmin}} \|\mathbf{M} - \Theta^{\text{Est}}\|, \quad (15)$$

where $\|\cdot\|$ is some norm. For convex structures as described above, the projection is a convex optimization problem which can be efficiently solved using standard numerical packages, e.g., CVX, [25, 26].

The main advantage of the projection method is that, when $\Theta_0 \in \mathcal{S}'$, the projection $\mathcal{P}_{\mathcal{S}'}(\Theta^{\text{Est}})$ is closer to Θ_0 than Θ^{Est} . The main disadvantage is that it requires a two-step solution which does not couple the distribution properties and the structure information simultaneously and is therefore suboptimal.

V. COCA ESTIMATOR

A. Definition

In this section we propose COCA - the CONvexly ConStrAined covariance estimator for GE distributions. Unlike the

existing solutions, COCA exploits both the elliptical nature and the structure of the underlying distribution. COCA is based on the GMM [24] together with an asymptotically tight convex relaxation.

The underlying principle behind COCA is the following identity [17, 40]:

$$\mathbf{E} \left(p \frac{\mathbf{x}_i \mathbf{x}_i^H}{\mathbf{x}_i^H \Theta_0^{-1} \mathbf{x}_i} \right) = \Theta_0, \quad (16)$$

holding for all GE and, in particular, CAE populations. Indeed, Tyler's estimator is just the sample based solution that satisfies this identity. When the number of samples is small, even without any structural assumptions, the solution to this equation does not necessarily exist. Instead, we propose the GMM approach which seeks an approximate solution to

$$\min_{\Theta \in \mathcal{S}'} \left\| \Theta - \frac{p}{n} \sum_{i=1}^n \frac{\mathbf{x}_i \mathbf{x}_i^H}{\mathbf{x}_i^H \Theta^{-1} \mathbf{x}_i} \right\|, \quad (17)$$

where $\|\cdot\|$ is some norm. Intuitively, this optimization tries to simultaneously solve Tyler's program and project it onto the set of prior structure. By choosing an adaptive weighted norm, an optimal solution to (17) would result in an asymptotically consistent and accurate estimator [24, 38]. Unfortunately, the objective is non-convex and it is not clear how to find its global solution in a tractable manner.

In what follows, we propose a convex relaxation of (17) that allows a computationally efficient solution. First, let us introduce auxiliary variables $d_i, i = 1, \dots, n$:

$$\begin{aligned} \min_{\Theta \in \mathcal{S}', d_i} & \left\| \Theta - \frac{1}{n} \sum_{i=1}^n d_i \mathbf{x}_i \mathbf{x}_i^H \right\| \\ \text{subject to} & d_i = \frac{p}{\mathbf{x}_i^H \Theta^{-1} \mathbf{x}_i}, i = 1 \dots n. \end{aligned} \quad (18)$$

This problem is not convex due to the equality constraints. We suggest to relax them to the inequalities:

$$\begin{aligned} \min_{\Theta \in \mathcal{S}', d_i} & \left\| \Theta - \frac{1}{n} \sum_{i=1}^n d_i \mathbf{x}_i \mathbf{x}_i^H \right\| \\ \text{subject to} & d_i \leq \frac{p}{\mathbf{x}_i^H \Theta^{-1} \mathbf{x}_i}, i = 1 \dots n, \\ & d_i > 0, i = 1 \dots n. \end{aligned} \quad (19)$$

This relaxed problem is actually a convex minimization program. In order to show this we use

Proposition 1. (Schur's Complement [41]) For any hermitian matrix \mathbf{X} of the form

$$\mathbf{X} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^H & \mathbf{C} \end{pmatrix},$$

if \mathbf{A} and \mathbf{C} are invertible then the following properties hold:

- 1) $\mathbf{X} \succeq 0$ iff $\mathbf{C} \succ 0$ and $\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^H \succeq 0$,
- 2) $\mathbf{X} \succeq 0$ iff $\mathbf{A} \succ 0$ and $\mathbf{C} - \mathbf{B}^H\mathbf{A}^{-1}\mathbf{B} \succeq 0$.

As a corollary we obtain that for $\Theta \in \mathcal{P}(p)$, $\Theta \succ 0$, $\mathbf{x} \in \mathbb{C}^p$ and $\alpha > 0$ the following conditions are equivalent:

- 1) $\Theta \succeq \frac{1}{\alpha} \mathbf{x} \mathbf{x}^H$,
- 2) $\alpha \geq \mathbf{x}^H \Theta^{-1} \mathbf{x}$.

Thus, we rewrite the inequalities $d_i \leq \frac{p}{\mathbf{x}_i^H \Theta^{-1} \mathbf{x}_i}, i = 1 \dots n$ as linear matrix inequalities (LMI):

$$\Theta^{\text{COCA}} = \arg \begin{cases} \min_{\Theta \in \mathcal{S}', d_i} & \left\| \Theta - \frac{1}{n} \sum_{i=1}^n d_i \mathbf{x}_i \mathbf{x}_i^H \right\| \\ \text{subject to} & \Theta \succeq \frac{1}{p} d_i \mathbf{x}_i \mathbf{x}_i^H, \forall i = 1 \dots n, \\ & d_i > 0, \forall i = 1 \dots n. \end{cases} \quad (20)$$

In this form COCA can be efficiently computed by standard semi-definite program solvers, e.g., CVX, [25, 26].

B. Consistency

The non-relaxed version of COCA in (17) is clearly a reasonable approach for structured covariance estimation in elliptical models. The interesting question is how tight is the relaxation. We now provide two promising results in this direction.

Theorem 1. *In the unstructured case $\mathcal{S} = \mathcal{P}(p)$ with $n \geq p + 1$, the COCA estimator is unique up to a positive scaling factor and coincides with Tyler's estimator.*

Proof. It is known that when $n \geq p + 1$, (20) has at least one solution which results in a zero objective value. It is Tyler's estimator which satisfies

$$d_i^* = \frac{p}{\mathbf{x}_i^H [\Theta^*]^{-1} \mathbf{x}_i}, i = 1 \dots n.$$

It remains to show that there are no other feasible solutions which result in a zero objective and is not a scaled version of this one. Indeed, assume in contradiction that there is such an additional solution, and for it $\Theta = \frac{1}{n} \sum_{i=1}^n d_i \mathbf{x}_i \mathbf{x}_i^H$. Multiply each inequality $d_i \leq \frac{p}{\mathbf{x}_i^H \Theta^{-1} \mathbf{x}_i}$ by the matrix $\mathbf{x}_i \mathbf{x}_i^H$ for $i = 1 \dots n$ and sum up to obtain

$$\Theta = \frac{1}{n} \sum_{i=1}^n d_i \mathbf{x}_i \mathbf{x}_i^H \preceq \frac{p}{n} \sum_{i=1}^n \frac{\mathbf{x}_i \mathbf{x}_i^H}{\mathbf{x}_i^H \Theta^{-1} \mathbf{x}_i} = f(\Theta). \quad (21)$$

The inequality (21) reads now as $\Theta \preceq f(\Theta)$. As stated in the Corollary VI from [12] (the Corollary VI is formulated there for the real case, but it remains valid in the complex case as explained there by the authors), this implies that Θ is the fixed point of f : $\Theta = f(\Theta)$, which is exactly the definition of Tyler's estimator in (14). Thus proving that it is the only solution to (20) up to a positive scaling factor. \square

Theorem 1 proves that the unconstrained COCA estimator performs as a classical MLE and is, thus, efficient with the asymptotically normal distribution:

$$\sqrt{n} (\Theta^{\text{COCA}} - \Theta_0) \xrightarrow{P} \mathcal{N}(0, \mathbf{CRB}).$$

In the constrained case, for a general convex set \mathcal{S}' , the analysis is more difficult but we still have promising asymptotic results.

Theorem 2. *In the structured case, COCA is an asymptotically consistent estimator of the true shape matrix $\Theta_0 \in \mathcal{S}'$.*

Proof. Is provided in the Appendix. \square

The efficiency of the constrained COCA estimator remains an open question. We do not expect the constrained estimator to be statistically efficient. However, we believe that it can get quite close by adaptively tuning the norm as detailed in the next subsection.

C. Choice of norm

The definition of the COCA estimator depends on the choice of the norm in the objective of (20). The consistency result in Theorem 2 is invariant to this choice, but the finite sample performance may change significantly. Natural choices are the Frobenius, trace and spectral norms. In addition, it is well known from the theory of GMM that adaptive weighted norms can enhance the performance of estimators. Similar ideas were applied in [38] to develop covariance matching estimators.

VI. COMPUTATIONAL COMPLEXITY

The COCA estimator developed in this paper can significantly change the approach to Tyler's shape matrix estimator, as it transforms the involved optimization into a convex problem, making the imposition of affine constraints possible. Unfortunately, it suffers from a significant drawback since it is computationally complex and involves solution of high-dimensional non-linear programs. In fact, this is also the drawback in the Gaussian case, where the projection onto the structure set requires a comparable computational effort (see [39] and references therein). The most appropriate general class of methods usually applied in such SDP programs is known as interior-point algorithms, [42], which are polynomial in the dimension of the problem. In practice, interior-point optimizers like MOSEK, SeDuMi and SDPT3 solve problems in a fixed number of iterations between about 10 and 100. Each iteration has polynomial complexity typically $O((k+n)^3)$. The exact power and additional logarithmic multipliers depend on the norm involved in the COCA-optimization and other specific details of the program at hand. To enjoy the advantages of the COCA fully, the proposed algorithms should be tuned to exploit the specific structure of the problem, e.g. Toeplitz or banded and the norm. Because of this and lack of space such algorithms are outside the scope of the current paper. One of the main directions of our future research is the development of less demanding COCA solvers.

VII. NUMERICAL RESULTS

In this section we demonstrate the advantages of COCA using numerical simulations. We investigated the performance benefits of COCA when the true shape matrix was either Toeplitz, banded or constructed based on the DOA grid. We compared the following estimators: Θ^{SC} in (13), Θ^{Tyler} in (14), Θ^{Proj} in (15) and Θ^{COCA} in (20). In Θ^{Proj} we projected Tyler's estimator when it existed and the sample covariance otherwise.

For each number of samples n we generated 1000 sets of independent, compound proper normally distributed p -dimensional samples and calculated the empirical MSE for all the estimators. The samples were generated as $\mathbf{x} = \sqrt{\tau} \mathbf{v}$, where the random variable $\tau \sim \chi^2$ and the random vector \mathbf{v} was zero-mean circularly symmetric normally distributed

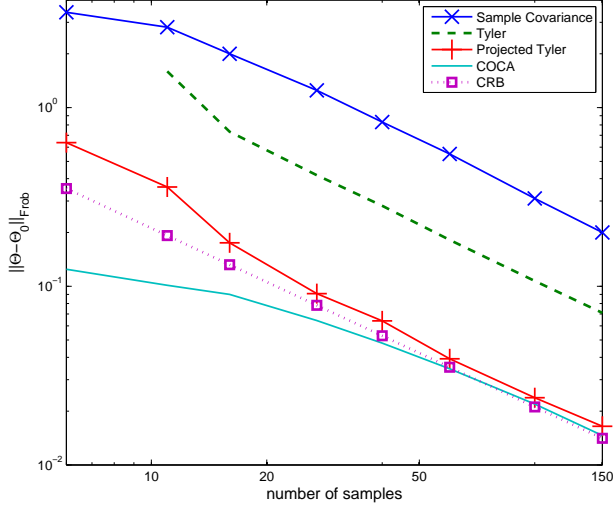


Fig. 1. COCA in the Toeplitz case.

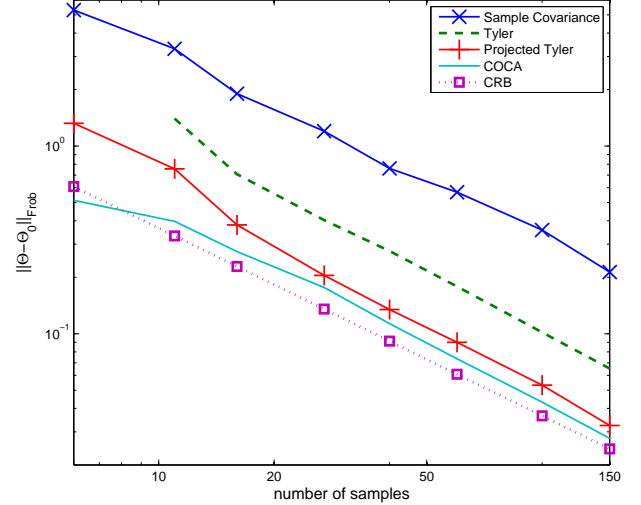


Fig. 2. COCA in the banded case.

with the covariance matrix Θ_0 . To compare the performance of COCA and all the other estimators to the **CRB** we used the Frobenius norm in the optimization problem and when calculating the empirical **MSE**.

A. Toeplitz Covariance Matrix

For $p = 10$ the Toeplitz shape matrix was chosen to have 1-s on the main diagonal and $\frac{1}{5} \pm \frac{j}{5}, \frac{1}{25} \pm \frac{j}{25}$ on the first two sub-diagonals correspondingly. The results are reported in Fig. 1. It is easy to see the performance advantage of COCA over all the other estimators. For convenience we also put the constrained **CRB** for this case on the same plot. As the COCA estimator performs better in the sense of **MSE** we can imply that it is biased.

B. Banded Covariance Matrix

As an example of banded structure we took a matrix having the numbers $20, 40, \dots, 20p$ on the main diagonal for $p = 10$, $12 \pm 3j, \dots, (12 \pm 3j)(p - 1)$ and $2 \pm 2j, \dots, (2 \pm 2j)(p - 2)$ on the first two sub-diagonals correspondingly and scaled it to have trace p . The band width is 2 in this case. The averaged errors and the **CRB** are reported in Fig. 2.

C. DOA Covariance Matrix

In the DOA experiments we took $n = 5$ signal sources uniformly localized in the interval $[0, \pi]$. The noise added was white of energy $\sigma = 1/100$. The number of sensors was $p = 10$. The grid was constructed by dividing the range $[0, \pi]$ into p equal subintervals. The convergence rates for different estimators are present in Fig. 3. We also provide the **CRB** for comparison.

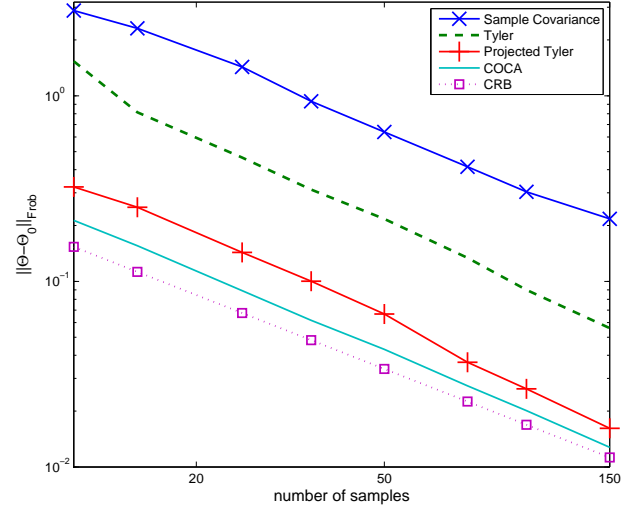


Fig. 3. COCA in the DOA case.

D. Different Norms Comparison

As we have already mentioned, the formulation of the COCA-estimator (20) leaves freedom for the choice of the norm. In the examples above we used the Frobenius norm. In this section we compare the performance of COCA with different norms. In particular we took the spectral, the Frobenius and the trace norms and compared the COCA performance with the **CRB**. The numerical results for the banded ($b = 2$) type of constraints are provided in Figure 4. As we can see the choice of the norm affects the results quite slightly and the one making the optimization problem easier to solve should be picked.

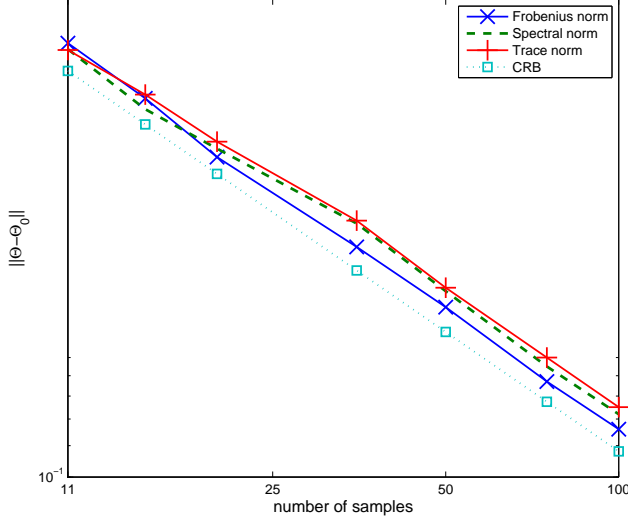


Fig. 4. Performance of the COCA with different norms in the banded case.

E. Discussion

As we can see with the examples considered above, the COCA estimator outperforms the benchmarks used. Its MSE is actually quite close to the projection estimator, which can be considered as a one-step approximation to the COCA. We must also note that the MSE of COCA is less than the CRB for small values of n in the first two figures. This is indeed possible, since COCA is only asymptotically unbiased and may turn out to be biased in finite samples.

VIII. CONCLUSION

In this paper we address structured covariance estimation in GE distributions. In particular, we assume that the covariance is a priori known to belong to a given convex set, e.g., the set of Toeplitz or banded matrices. We utilize the MLE of the shape matrix of normalized population which is a solution to a non-convex program and propose its convex relaxation based on the GMM technique. It is shown that the relaxed program (COCA) is tight in the unconstrained case and asymptotically tight in the constrained settings. Numerical simulations show that COCA performs better than other comparable techniques, such as unconstrained Tyler's estimator and its projection.

Our future work will first of all address the performance properties of the COCA estimator and its generalizations based on M-estimators. In addition, as we have already mentioned, when treated using general purpose numerical packages COCA may become a resource demanding program. Our second aim is to develop a more specific algorithm to make the COCA estimator computationally scalable.

APPENDIX

A. Proof of Theorem 2

Proof. We assume that the set $\mathcal{S}' \subset \mathcal{P}(p)$ is a compact convex set separated from zero. Otherwise, we take a large enough ball \mathcal{B} centered at zero and replace the constraint \mathcal{S}' by the

intersection $\mathcal{S}' \cap \mathcal{B}$. Denote $d_i^0 = d_i^0(\mathbf{x}) = \frac{p}{\mathbf{x}_i^H \Theta_0^{-1} \mathbf{x}_i}$. For the sake of convenience we will consider the realizations of the samples as infinite sequences $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n, \dots)$. Consider now a random function

$$h_n(\Theta_0, \mathbf{d}; \mathbf{x}) = \left\| \Theta_0 - \frac{1}{n} \sum_{i=1}^n d_i \mathbf{x}_i \mathbf{x}_i^H \right\|,$$

where $\mathbf{d} = (d_1, d_2, \dots, d_n, \dots)$.

The (strong) Law of Large Numbers implies that

$$\begin{aligned} h_n(\Theta_0, \mathbf{d}^0; \mathbf{x}) &= \left\| \Theta_0 - \frac{1}{n} \sum_{i=1}^n d_i^0 \mathbf{x}_i \mathbf{x}_i^H \right\| \\ &\rightarrow \left\| \Theta_0 - \mathbf{E} \left(\frac{\mathbf{x}_i \mathbf{x}_i^H}{\mathbf{x}_i^H \Theta_0^{-1} \mathbf{x}_i} \right) \right\|, n \rightarrow \infty \text{ a.s.} \end{aligned}$$

For the elliptical distribution $\mathbf{E} \left(\frac{\mathbf{x}_i \mathbf{x}_i^H}{\mathbf{x}_i^H \Theta_0^{-1} \mathbf{x}_i} \right) = \Theta_0$ [17], so we get that

$$h_n(\Theta_0, \mathbf{d}^0; \mathbf{x}) \rightarrow 0, n \rightarrow \infty \text{ a.s.} \quad (22)$$

For now, given a realization \mathbf{x} and a number $n \in \mathbb{N}$ denote by $(\hat{\Theta}^{(n)}, \hat{\mathbf{d}}^{(n)})$ the solution of (19), all the $\hat{d}_i^{(n)} = 0$ for $i > n$. Define a random variable \tilde{h}_n depending on \mathbf{x} :

$$\tilde{h}_n(\mathbf{x}) = \left\| \hat{\Theta}^{(n)} - \frac{1}{n} \sum_{i=1}^n \hat{d}_i^{(n)} \mathbf{x}_i \mathbf{x}_i^H \right\|.$$

For each \mathbf{x} :

$$\tilde{h}_n(\mathbf{x}) \leq h_n(\Theta_0, \mathbf{d}^0; \mathbf{x}),$$

since $(\hat{\Theta}^{(n)}, \hat{\mathbf{d}}^{(n)})$ is the extremum of the target function, thus (22) implies

$$\tilde{h}_n(\mathbf{x}) \rightarrow 0, n \rightarrow \infty \text{ a.s.}$$

Since \mathcal{S}' is compact we can choose a convergent subsequence, and renumber it if needed. We now have:

$$\hat{\Theta}^{(n)} \rightarrow \bar{\Theta} \succ 0, n \rightarrow \infty, \quad (23)$$

$$\frac{1}{n} \sum_{i=1}^n \hat{d}_i^{(n)} \mathbf{x}_i \mathbf{x}_i^H \rightarrow \bar{\Theta}, n \rightarrow \infty. \quad (24)$$

All these events happen with probability one and $\bar{\Theta} = \bar{\Theta}(\mathbf{x})$ depends on the realization. The relation (23) implies that for any $\epsilon > 0$ there exists $n_1 \in \mathbb{N}$ starting from which $\hat{\Theta}^{(n)} \prec (1 + \epsilon)\bar{\Theta}$. Thus,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \hat{d}_i^{(n)} \mathbf{x}_i \mathbf{x}_i^H &\preceq \frac{1}{n} \sum_{i=1}^n \frac{\mathbf{x}_i \mathbf{x}_i^H}{\mathbf{x}_i^H [\hat{\Theta}^{(n)}]^{-1} \mathbf{x}_i} \\ &\preceq \frac{1}{n} \sum_{i=1}^n \frac{\mathbf{x}_i \mathbf{x}_i^H}{\mathbf{x}_i^H \bar{\Theta}^{-1} \mathbf{x}_i} + \frac{\epsilon}{n} \sum_{i=1}^n \frac{\mathbf{x}_i \mathbf{x}_i^H}{\mathbf{x}_i^H \bar{\Theta}^{-1} \mathbf{x}_i}. \end{aligned} \quad (25)$$

In a similar way (24) implies that for the same $\epsilon > 0$ there exists $n_2 \in \mathbb{N}$ starting from which

$$(1 - \epsilon)\bar{\Theta} \preceq \frac{1}{n} \sum_{i=1}^n \hat{d}_i^{(n)} \mathbf{x}_i \mathbf{x}_i^H. \quad (26)$$

Now for $n \geq \max(n_1, n_2)$

$$\begin{aligned} (1 - \epsilon)\bar{\Theta} &\preceq \frac{1}{n} \sum_{i=1}^n \tilde{d}_i^{(n)} \mathbf{x}_i \mathbf{x}_i^H \\ &\preceq \frac{1}{n} \sum_{i=1}^n \frac{\mathbf{x}_i \mathbf{x}_i^H}{\mathbf{x}_i^H \bar{\Theta}^{-1} \mathbf{x}_i} + \frac{\epsilon}{n} \sum_{i=1}^n \frac{\mathbf{x}_i \mathbf{x}_i^H}{\mathbf{x}_i^H \bar{\Theta}^{-1} \mathbf{x}_i}. \end{aligned}$$

Since ϵ was chosen arbitrarily and all the sums here are bounded, this implies that

$$\bar{\Theta} \preceq \frac{1}{n} \sum_{i=1}^n \frac{\mathbf{x}_i \mathbf{x}_i^H}{\mathbf{x}_i^H \bar{\Theta}^{-1} \mathbf{x}_i}, \quad (27)$$

for sufficiently large n . From here the proof continues as in Theorem 1: due to Corollary V.I from [12] this implies that (27) holds with equality and the uniqueness implies that $\bar{\Theta} = \Theta_0$ a.s. up to a scaling factor.

Assume now that the original sequence $\hat{\Theta}^{(n)}$ does not converge, which implies that it has a subsequence $\hat{\Theta}^{(n_i)}$, which converges to a different limit $\bar{\Theta}_{\text{wrong}} \neq \Theta_0$, but this contradicts the previous reasoning for the convergent subsequence, thus showing that the original sequence $\hat{\Theta}^{(n)}$ converges to Θ_0 . \square

REFERENCES

- [1] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, 1996.
- [2] E. R. Dougherty, A. Datta, and C. Sima, "Research issues in genomic signal processing," *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 46–68, 2005.
- [3] D. R. Fuhrmann, "Application of Toeplitz covariance estimation to adaptive beamforming and detection," *IEEE Transactions on Signal Processing*, vol. 39, no. 10, pp. 2194–2198, 1991.
- [4] W. J. Roberts and Y. Ephraim, "Hidden Markov modeling of speech using Toeplitz covariance matrices," *Speech Communication*, vol. 31, no. 1, pp. 1–14, 2000.
- [5] D. S. Pollock, "Circulant matrices and time-series analysis," *International Journal of Mathematical Education in Science and Technology*, vol. 33, no. 2, pp. 213–230, 2002.
- [6] P. Stoica, P. Babu, and J. Li, "SPICE: A sparse covariance-based estimation method for array processing," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 629–638, 2011.
- [7] H. Wang and L. Cai, "On adaptive spatial-temporal processing for airborne surveillance radar systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 30, no. 3, pp. 660–670, 1994.
- [8] D. L. Snyder, J. A. O'Sullivan, and M. I. Miller, "The use of maximum likelihood estimation for forming images of diffuse radar targets from delay-Doppler data," *IEEE Transactions on Information Theory*, vol. 35, no. 3, pp. 536–548, 1989.
- [9] R. Dahlhaus, "Efficient parameter estimation for self-similar processes," *The Annals of Statistics*, pp. 1749–1766, 1989.
- [10] T. T. Cai, Z. Ren, and H. H. Zhou, "Optimal rates of convergence for estimating Toeplitz covariance matrices," *Probability Theory and Related Fields*, pp. 1–43, 2012.
- [11] P. J. Bickel and E. Levina, "Regularized estimation of large covariance matrices," *The Annals of Statistics*, pp. 199–227, 2008.
- [12] F. Pascal, Y. Chitour, J. P. Ovarlez, P. Forster, and P. Larzabal, "Covariance structure maximum-likelihood estimates in compound Gaussian noise: existence and algorithm analysis," *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 34–48, 2008.
- [13] Y. I. Abramovich, N. K. Spencer, and M. D. Turley, "Time-varying autoregressive (TVAR) models for multiple radar observations," *IEEE Transactions on Signal Processing*, vol. 55, no. 4, pp. 1298–1311, 2007.
- [14] A. Wiesel, "Geodesic convexity and covariance estimation," *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6182–6189, 2012.
- [15] T. Zhang, A. Wiesel, and M. S. Greco, "Multivariate generalized gaussian distribution: Convexity and graphical models," *IEEE Transactions on Signal Processing*, vol. 61, no. 16, pp. 4141–4148, 2013.
- [16] I. Soloveychik and A. Wiesel, "Group symmetry and non-Gaussian covariance estimation," *arXiv preprint arXiv:1306.4103*, 2013.
- [17] G. Frahm, "Generalized elliptical distributions: theory and applications," *Universität zu Köln*, 2004.
- [18] G. Frahm and U. Jaekel, "Tyler's M-estimator, random matrix theory, and generalized elliptical distributions with applications to finance," Tech. Rep., 2007.
- [19] E. Conte, M. Longo, and M. Lops, "Modelling and simulation of non-Rayleigh radar clutter," *IEE Proceedings F on Radar and Signal Processing*, vol. 138, no. 2, pp. 121–130, 1991.
- [20] D. Middleton, "Man-made noise in urban environments and transportation systems: models and measurements," *IEEE Transactions on Communications*, vol. 21, no. 11, pp. 1232–1241, 1973.
- [21] R. A. Maronna, "Robust M-estimators of multivariate location and scatter," *The annals of statistics*, pp. 51–67, 1976.
- [22] D. E. Tyler, "A distribution-free M-estimator of multivariate scatter," *The Annals of Statistics*, vol. 15, no. 1, pp. 234–251, 1987.
- [23] F. Bandiera, O. Besson, and G. Ricci, "Knowledge-aided covariance matrix estimation and adaptive detection in compound-Gaussian noise," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5391–5396, 2010.
- [24] L. Mátyás, "Generalized method of moments estimation," *Cambridge University Press*, vol. 5, 1999.
- [25] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.0 beta," <http://cvxr.com/cvx>, Sep. 2013.
- [26] —, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, pp. 95–110.

- [27] M. Greco and F. Gini, "Cramér-Rao lower bounds on covariance matrix estimation for complex elliptically symmetric distributions," *IEEE Transactions on Signal Processing*, vol. 61, no. 24, pp. 6401–6409, 2013.
- [28] J. P. Burg, D. G. Luenberger, and D. L. Wenger, "Estimation of structured covariance matrices," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 963–974, 1982.
- [29] A. Wiesel, O. Bibi, and A. Globerson, "Time varying autoregressive moving average models for covariance estimation," *IEEE Transactions on Signal Processing*, vol. 61, no. 11, pp. 2791–2801, 2013.
- [30] Y.-H. Li and P.-C. Yeh, "An interpretation of the Moore-Penrose generalized inverse of a singular Fisher Information Matrix," *IEEE Transactions on Signal Processing*, vol. 60, no. 10, pp. 5532–5536, 2012.
- [31] O. Besson and Y. Abramovich, "On the Fisher information matrix for multivariate elliptically contoured distributions," *IEEE Signal Processing Letters*, vol. 20, no. 11, pp. 1130–1133, 2013.
- [32] R. Vershynin, "How close is the sample covariance matrix to the actual covariance matrix?" *Journal of Theoretical Probability*, vol. 25, no. 3, pp. 655–686, 2012.
- [33] N. Srivastava and R. Vershynin, "Covariance estimation for distributions with $2+\epsilon$ moments," *arXiv preprint arXiv:1106.2775*, 2011.
- [34] A. J. Rothman, P. J. Bickel, E. Levina, and J. Zhu, "Sparse permutation invariant covariance estimation," *Electronic Journal of Statistics*, vol. 2, pp. 494–515, 2008.
- [35] P. Shah and V. Chandrasekaran, "Group symmetry and covariance regularization," *Electronic Journal of Statistics*, vol. 6, pp. 1600–1640, 2012.
- [36] Y. Chen, A. Wiesel, and A. O. Hero, "Robust shrinkage estimation of high-dimensional covariance matrices," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4097–4107, 2011.
- [37] A. Wiesel, "Unified framework to regularized covariance estimation in scaled gaussian models," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 29–38, 2012.
- [38] B. Ottersten, P. Stoica, and R. Roy, "Covariance matching estimation techniques for array signal processing applications," *Digital Signal Processing*, vol. 8, no. 3, pp. 185–210, 1998.
- [39] D. Henrion and J. Malick, "Projection methods in conic optimization," *Handbook on Semidefinite, Conic and Polynomial Optimization*, pp. 565–600, 2012.
- [40] G. Frahm and U. Jaekel, "A generalization of Tylers M-estimators to the case of incomplete data," *Computational Statistics & Data Analysis*, vol. 54, no. 2, pp. 374–393, 2010.
- [41] F. Zhang, *The Schur complement and its applications*. Springer, 2005, vol. 4.
- [42] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM review*, vol. 38, no. 1, pp. 49–95, 1996.



in 2011.

Ilya Soloveychik (SM'13) received the B.Sc. degree in applied mathematics and physics from the Moscow Institute of Physics and Technology, in 2007, and the M.Sc degree in mathematics from the Hebrew University of Jerusalem, in 2013. He is currently a Ph.D. student with the Rachel and Selim Benin School of Computer Science and Engineering, the Hebrew University of Jerusalem, Israel.

He was the Prize recipient of the Russian Olympiad in Physics in 2003. He received the Klein Prize and the Kaete Klausner Research Scholarship



of Jerusalem, Israel.

Ami Wiesel received the B.Sc. and M.Sc. degrees in Electrical Engineering from Tel-Aviv University, Tel-Aviv, Israel, in 2000 and 2002, respectively, and the Ph.D. degree in Electrical Engineering from the Technion - Israel Institute of Technology, Haifa, Israel, in 2007. He was a postdoctoral fellow with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, in 2007-2009. Since Jan. 2010, he is a faculty member at the Rachel and Selim Benin School of Computer Science and Engineering at the Hebrew University

of Jerusalem, Israel. Dr. Wiesel was a recipient of the Young Author Best Paper Award for a 2006 paper in the *IEEE Transactions on Signal Processing* and a Student Paper Award for a 2005 Workshop on Signal Processing Advances in Wireless Communications (SPAWC) paper. He was awarded the Weinstein Study Prize in 2002, the Intel Award in 2005, the Viterbi Fellowship in 2005 and 2007, and the Marie Curie Fellowship in 2008.