# ROBUST ESTIMATION OF STRUCTURED COVARIANCE MATRIX FOR HEAVY-TAILED DISTRIBUTIONS

*Ying Sun, Prabhu Babu, and Daniel P. Palomar*

Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong

## ABSTRACT

In this paper, we consider the robust covariance estimation problem in the non-Gaussian set-up. In particular, Tyler's $M$-estimator is adopted for samples drawn from a heavy-tailed elliptical distribution. For some applications, the covariance matrix naturally possesses certain structure. Therefore, incorporating the prior structure information in the estimation procedure is beneficial to improving estimation accuracy. The problem is formulated as a constrained minimization of the Tyler's cost function, where the structure is characterized by the constraint set. A numerical algorithm based on majorization-minimization is derived for general structures that can be characterized as a convex set, where a sequence of convex programming is solved. For the set of matrices that can be decomposed as the sum of rank one positive semidefinite matrices, which has a wide range of applications, the algorithm is modified with much lower complexity. Simulation results demonstrate that the proposed structure-constrained Tyler's estimator achieves smaller estimation error than the unconstrained case.

***Index Terms*—** Robust estimation, Tyler's scatter estimator, structure constraint, majorization-minimization.

## 1. INTRODUCTION

Estimating the covariance matrix of random variables from observed samples has been a ubiquitous problem that arises in many applications in signal processing, and closely related to fundamental problems such as coherence estimation, component analysis, time series analysis, etc. It has been realized that the sample covariance estimator, which coincides with the maximum likelihood estimator (MLE) under an independent and identically distributed (i.i.d.) Gaussian noise assumption, is not suitable in many real-world applications. A main cause of the failure of the Gaussian assumption is that real-world data are often corrupted by noise and outliers, and the sample covariance matrix is sensitive to abnormal data in the sense that erroneous observations can severely decrease the estimation accuracy. To address the aforementioned problem, robust $M$-estimators were proposed to limit the influence of abnormal samples. As a particular case, Tyler's $M$-estimator for scatter matrix has received considerable attention recently and demonstrated to work effectively in various kinds of applications with proper modifications [1–6].

Apart from the heavy-tailed empirical distribution of the samples and outlier contamination, another problem that modern applications frequently encounter is the insufficient number of samples compared to the number of parameters being estimated. Consequently, many traditional estimators, including Tyler's estimator, fail to achieve satisfactory estimation accuracy. A popular way to tackle this problem in the literature is through regularization, some commonly used regularization techniques include shrinkage to a given prior [1, 6–9], imposing sparsity assumption [10], and thresholding [11], each designed for different applicable scenarios.

In this paper we are interested in applications where the covariance matrix is known to have certain structure. Instead of estimating the covariance matrix with the blanket assumption that it is Hermitian positive semidefinite, a natural idea is to impose additional structure assumption on the estimator [12–15], which typically reduces the number of parameters to be estimated. We formulate the problem as minimizing the Tyler's cost function under the structure constraint. Instead of trying to find the global optimal directly, which becomes a challenging task due to the non-convexity of the objective function, algorithms are derived based on the majorization-minimization framework that converge to a stationary point. We first work out the algorithm for a general structure that can be characterized as a convex set, where a sequence of convex programming is solved. Then, we consider a specific structure that consists of matrices can be decomposed as the sum of rank one Hermitian positive semidefinite matrices. Many applications in practice, such as the direction-of-arrival (DOA) estimation problem, involve this type of covariance. In such a case the algorithm can be modified to significantly reduce the computational load. Numerical results are provided in the end and the performance of the proposed estimator is compared with the COCA estimator [12].

## 2. TYLER'S ESTIMATOR WITH CONVEX STRUCTURE CONSTRAINT

Consider a number of $N$ samples $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ with $\mathbf{x}_i \in \mathbb{C}^K$ drawn independently from an elliptical underlying distribution with probability density function (pdf) as follows:

$$f(\mathbf{x}) = \det(\mathbf{R}_0)^{-1} g\left(\mathbf{x}^H \mathbf{R}_0^{-1} \mathbf{x}\right),$$

where $\mathbf{R}_0 \succ \mathbf{0}$ (Hermitian positive definite) is the scatter parameter that is proportional to the covariance matrix if it exists, and $g(\cdot)$ characterizes the shape of the distribution. Tyler's estimator for $\mathbf{R}_0$ is defined as the solution to the fixed-point equation

$$\mathbf{R} = \frac{K}{N} \sum_{i=1}^{N} \frac{\mathbf{x}_i \mathbf{x}_i^H}{\mathbf{x}_i^H \mathbf{R}^{-1} \mathbf{x}_i},$$

which can be interpreted as a weighted sum of rank one matrices $\mathbf{x}_i \mathbf{x}_i^H$ with weight decreasing as $\mathbf{x}_i$ gets farther from the center [2].[1] An alternative interpretation is that Tyler's estimator can be viewed

This work was supported by the Hong Kong RGC 16207814 research grant.

[1] In the original paper [2] $\mathbf{x}_i$ is real-valued, but can be generalized to the complex-valued case.

as the maximum likelihood estimator of $\mathbf{R}_0$ for an angular central Gaussian distribution

$$f(\mathbf{s}) \propto \det(\mathbf{R}_0)^{-1} \left(\mathbf{s}^H \mathbf{R}_0^{-1} \mathbf{s}\right)^{-K}. \tag{1}$$

It is known that if $\mathbf{x}$ is elliptically distributed, then the pdf of the normalized sample $\mathbf{s} = \frac{\mathbf{x}}{\|\mathbf{x}\|_2}$ will take the form of (1) [3]. Notice that $f(\mathbf{s})$ does not depend on $g(\cdot)$ , which indicates that Tyler's estimator works for all elliptical distributions. Also the normalization procedure eliminates the influence of the magnitude of an outlier on the estimator. As a tradeoff, Tyler's estimator only estimates $\mathbf{R}_0$ up to a positive scale factor.

Despite of the attractive properties, as most of the classical covariance estimators, Tyler's estimator requires sufficiently many samples to achieve satisfactory estimation accuracy, which limits its scope of application in high dimensional estimation problems. A way to tackle this problem is to introduce some prior information of the covariance matrix into the estimation. The prior information that we are interested in herein is that $\mathbf{R}_0$ takes certain structure that can be characterized by a set $\mathcal{C}$ that is the intersection of a closed convex set and the set of Hermitian positive semidefinite matrices.

Mathematically, the following problem is solved:

$$\begin{array}{ll} \underset{\mathbf{R}}{\text{minimize}} & \log\det(\mathbf{R}) + \frac{K}{N}\sum_{i=1}^{N} \log\left(\mathbf{x}_i^H \mathbf{R}^{-1} \mathbf{x}_i\right) \\ \text{subject to} & \mathbf{R} \in \mathcal{C}. \end{array} \tag{2}$$

As the objective function, denoted by $L(\mathbf{R})$, is non-convex, finding the global optimal of problem (2) is challenging. In this work we focus on design algorithms that find a stationary point instead, whose complexity is more tractable. Although the limit point generated by the algorithm may not be globally optimal, the estimator usually achieves satisfactory performance as shown in the numerical section.

The rest of the paper is based on the assumption that $L(\mathbf{R}) \to +\infty$ as $\mathbf{R}$ converges to a singular limit ($N > K$ is sufficient to meet this assumption if $f(\mathbf{x})$ is continuous), under which a solution of (2) exists. To solve problem (2), we refer to the idea of majorization-minimization [17]. Specifically, since the $\log(\cdot)$ and $\log\det(\cdot)$ functions are concave, the objective function $L(\mathbf{R})$ can be upper bounded by the convex function

$$g(\mathbf{R}|\mathbf{R}_t) = \text{Tr}\left(\mathbf{R}_t^{-1}\mathbf{R}\right) + \frac{K}{N}\sum_{i=1}^{N} \frac{\mathbf{x}_i^H \mathbf{R}^{-1} \mathbf{x}_i}{\mathbf{x}_i^H \mathbf{R}_t^{-1} \mathbf{x}_i} + \text{const.} \tag{3}$$

with equality achieved at $\mathbf{R}_t$. At the $(t+1)$-th iteration, the variable is updated as

$$\mathbf{R}_{t+1} = \arg\min_{\mathbf{R}\in\mathcal{C}} g(\mathbf{R}|\mathbf{R}_t). \tag{4}$$

For notation simplicity, define matrix $\mathbf{M}_t = \frac{K}{N}\sum_{i=1}^{N} \frac{\mathbf{x}_i \mathbf{x}_i^H}{\mathbf{x}_i^H \mathbf{R}_t^{-1} \mathbf{x}_i}$. The inner minimization problem takes the form:

$$\begin{array}{ll} \underset{\mathbf{R}}{\text{minimize}} & \text{Tr}\left(\mathbf{R}_t^{-1}\mathbf{R}\right) + \text{Tr}\left(\mathbf{M}_t\mathbf{R}^{-1}\right) \\ \text{subject to} & \mathbf{R} \in \mathcal{C}, \end{array} \tag{5}$$

which is a convex problem. The minimizer $\mathbf{R}_{t+1}$ is nonsingular, since $L(\mathbf{R})$ is upper-bounded by $g(\mathbf{R}|\mathbf{R}_t)$ and $L(\mathbf{R}) \to +\infty$ as $\mathbf{R}$ converges to a singular limit by assumption.

Following the standard majorization-minimization convergence result, we conclude that every limit point of the sequence $\{\mathbf{R}_t\}$ is a stationary point of the problem (2). Note that for most of the structure constraints that arise in the applications, the set $\mathcal{C}$ possesses the property that $\mathbf{R} \in \mathcal{C}$ iff $r\mathbf{R} \in \mathcal{C}$ for all $r > 0$ , in this case we can add a normalization step after the minimization:

$$\begin{aligned} \tilde{\mathbf{R}}_{t+1} &= \arg\min_{\mathbf{R}\in\mathcal{C}} g(\mathbf{R}|\mathbf{R}_t) \\ \mathbf{R}_{t+1} &= \tilde{\mathbf{R}}_{t+1}/\text{Tr}\left(\tilde{\mathbf{R}}_{t+1}\right), \end{aligned}$$

the sequence $\{\mathbf{R}_t\}$ then converges to the set of stationary points of the equivalent problem

$$\begin{array}{ll} \underset{\mathbf{R}\in\mathcal{C}}{\text{minimize}} & \log\det(\mathbf{R}) + \frac{K}{N}\sum_{i=1}^{N} \log\left(\mathbf{x}_i^H \mathbf{R}^{-1} \mathbf{x}_i\right) \\ \text{subject to} & \text{Tr}(\mathbf{R}) = 1. \end{array}$$

The algorithm described above can be tuned with higher efficiency if more detailed information of $\mathcal{C}$ is available. For instance, if we can further assume that the structure constraint is linear, which includes important examples such as the set of Toeplitz matrices, banded matrices, persymmetric matrices and so on, then applying Schur complement for a positive definite matrix reveals that solving (5) can be reformulated as the following equivalent semidefinite programming (SDP):

$$\begin{array}{ll} \underset{\mathbf{S},\mathbf{R}\in\mathcal{C}}{\text{minimize}} & \text{Tr}\left(\mathbf{R}_t^{-1}\mathbf{R}\right) + \text{Tr}(\mathbf{M}_t\mathbf{S}) \\ \text{subject to} & \begin{bmatrix} \mathbf{S} & \mathbf{I} \\ \mathbf{I} & \mathbf{R} \end{bmatrix} \succeq \mathbf{0}. \end{array}$$

In the next section, we are going to restrict further to a specific family of linear constraint that has a wide range of application in signal processing, and show that simple and fast algorithms can be derived by properly exploiting the structure.

## 3. TYLER'S ESTIMATOR FOR DECOMPOSABLE COVARIANCE

In this section, we focus on the class of structure $\mathcal{C}$ that can be described as

$$\mathcal{C} = \left\{\mathbf{R} | \mathbf{R} = \sum_{j=1}^{L} p_j \mathbf{a}_j \mathbf{a}_j^H\right\}, \tag{6}$$

where the $\mathbf{a}_j$'s are complex-valued given "basis vectors" and the $p_j$'s are real-valued nonnegative coefficient variables. Clearly, $\mathbf{R}$ is a Hermitian positive semidefinite matrix. Define the nonnegative vector $\mathbf{p} \triangleq [p_1, \ldots, p_L]$, $\mathbf{R}$ can be expressed compactly as $\mathbf{R} = \mathbf{A}\mathbf{P}\mathbf{A}^H$, where $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_L]$, and $\mathbf{P} = \text{diag}(p_1, \ldots, p_L)$. Notice that the signal plus noise model $\mathbf{R} = \mathbf{A}\mathbf{P}\mathbf{A}^H + \boldsymbol{\Sigma}$ with $\boldsymbol{\Sigma} = \text{diag}(\sigma_1, \ldots, \sigma_K)$, can also be rewritten as $\mathbf{R} = \tilde{\mathbf{A}}\tilde{\mathbf{P}}\tilde{\mathbf{A}}^H$ by defining the augmented matrices $\tilde{\mathbf{A}} = [\mathbf{A}, \mathbf{I}_K]$, $\tilde{\mathbf{P}} = \text{diag}(\tilde{p}_1, \ldots, \tilde{p}_{L+K})$ with $\tilde{p}_j = p_j$ for $j = 1, \ldots, L$ and $\tilde{p}_j = \sigma_{j-L}$ for $j = L+1, \ldots, L+K$.

Similar to the general convex structure, we majorize both parts of $L(\mathbf{R})$. At the $t$-th iteration, the surrogate function $g(\mathbf{R}|\mathbf{R}_t)$ is defined as (3) and the inner minimization problem takes the form:

$$\begin{array}{ll} \underset{\mathbf{R},\mathbf{p}}{\text{minimize}} & \mathbf{w}_t^H \mathbf{p} + \sum_{i=1}^{N} \tilde{\mathbf{x}}_i^H \mathbf{R}^{-1} \tilde{\mathbf{x}}_i \\ \text{subject to} & \mathbf{R} = \mathbf{A}\mathbf{P}\mathbf{A}^H \\ & \mathbf{p} \geq \mathbf{0}, \end{array} \tag{7}$$

with

$$\mathbf{w}_t = \operatorname{diag}\left(\mathbf{A}^H\mathbf{R}_t^{-1}\mathbf{A}\right),$$

$$\tilde{\mathbf{x}}_i = \sqrt{\frac{K}{N}}\frac{\mathbf{x}_i}{\sqrt{\mathbf{x}_i^H\mathbf{R}_t^{-1}\mathbf{x}_i}}.$$

The problem above, which is convex, is well studied in the literature. A cyclic algorithm similar to LIKES [18] can be derived, which is more computationally efficient than simply calling a convex solver. The procedure is stated in Alg. 1.

**Algorithm 1** Robust LIKES for decomposable covariance estimation

```
Set t = 0, initialize p_t to be any positive vector.
repeat
    R_t = A P_t A^H
    w_t = diag(A^H R_t^{-1} A); x̃_i = sqrt(K/N) x_i / sqrt(x_i^H R_t^{-1} x_i)
    Set r = 0, p^r = p^t
    repeat
        (see (8))
        r <- r + 1
    until some convergence criterion is met
    p_{t+1} = p_{r-1}
    t <- t + 1
until some convergence criterion is met
```

$$\begin{aligned}(\boldsymbol{\beta}_i)_{r+1} &= \mathbf{P}_r\mathbf{A}^H\left(\mathbf{A}\mathbf{P}_r\mathbf{A}^H\right)^{-1}\tilde{\mathbf{x}}_i\\ (p_j)_{r+1} &= \sqrt{\left(\sum_i \left|(\beta_{ij})_{r+1}\right|^2\right)/\,(w_j)_t}\end{aligned}\tag{8}$$

Alg. 1 requires a double loop, where the outer loop updates the surrogate function $g(\mathbf{R}|\mathbf{R}_t)$ and the inner loop iteration (8) solves (7). In this section, we propose a single loop algorithm that treat $\mathbf{P}$ as variable directly instead of $\mathbf{R}$ and find a surrogate function of

$$L(\mathbf{P}) = \log\det\left(\mathbf{A}\mathbf{P}\mathbf{A}^H\right) + \frac{K}{N}\sum_{i=1}^{N}\log\left(\mathbf{x}_i^H\left(\mathbf{A}\mathbf{P}\mathbf{A}^H\right)^{-1}\mathbf{x}_i\right).$$

To be precise, at point $\mathbf{P}_t$, for the first term we have $\log\det\left(\mathbf{A}\mathbf{P}\mathbf{A}^H\right) \le \mathbf{w}_t^H\mathbf{p}$ as before. For the second term, recall that the concave property of $\log(\cdot)$ function leads to the following inequality

$$\frac{K}{N}\sum_{i=1}^{N}\log\left(\mathbf{x}_i^H\mathbf{R}^{-1}\mathbf{x}_i\right) \le \operatorname{Tr}\left(\mathbf{M}_t\mathbf{R}^{-1}\right) + \text{const.}$$

with equality achieved at $\mathbf{R} = \mathbf{R}_t$, or equivalently, $\mathbf{P} = \mathbf{P}_t$. In the next step, we are going to find an upper bound for the quantity $\operatorname{Tr}\left(\mathbf{M}_t\mathbf{R}^{-1}\right)$ with equality achieved at $\mathbf{P} = \mathbf{P}_t$. We claim that

$$\operatorname{Tr}\left(\mathbf{M}_t\mathbf{R}^{-1}\right) \le \operatorname{Tr}\left(\mathbf{M}_t\mathbf{R}_t^{-1}\mathbf{A}\mathbf{P}_t\mathbf{P}^{-1}\mathbf{P}_t\mathbf{A}^H\mathbf{R}_t^{-1}\right).\tag{9}$$

To see this, first of all since $\mathbf{R} = \mathbf{A}\mathbf{P}\mathbf{A}^H$, it is easy to check equality holds at $\mathbf{P} = \mathbf{P}_t$. Next, from the identity

$$\begin{aligned}\mathbf{S} &= \begin{bmatrix}\mathbf{R}_t^{-1}\mathbf{A}\mathbf{P}_t\mathbf{P}^{-1}\mathbf{P}_t\mathbf{A}^H\mathbf{R}_t^{-1} & \mathbf{I}\\ \mathbf{I} & \mathbf{A}\mathbf{P}\mathbf{A}^H\end{bmatrix}\\ &= \begin{bmatrix}\mathbf{R}_t^{-1}\mathbf{A}\mathbf{P}_t\mathbf{P}^{-1/2}\\ \mathbf{A}\mathbf{P}^{1/2}\end{bmatrix}\begin{bmatrix}\mathbf{P}^{-1/2}\mathbf{P}_t\mathbf{A}^H\mathbf{R}_t^{-1} & \mathbf{P}^{1/2}\mathbf{A}^H\end{bmatrix}\end{aligned}$$

we know the matrix $\mathbf{S}$ is positive semidefinite. By Schur complement, if $\mathbf{R} = \mathbf{A}\mathbf{P}\mathbf{A}^H \succ \mathbf{0}$ , $\mathbf{S} \succeq \mathbf{0}$ is equivalent to

$$\mathbf{R}_t^{-1}\mathbf{A}\mathbf{P}_t\mathbf{P}^{-1}\mathbf{P}_t\mathbf{A}^H\mathbf{R}_t^{-1} \succeq \left(\mathbf{A}\mathbf{P}\mathbf{A}^H\right)^{-1},$$

the inequality (9) follows as $\mathbf{M}_t$ is Hermitian positive semidefinite. Therefore, ignoring the constant term, the function $L(\mathbf{P})$ is majorized by

$$g(\mathbf{P}|\mathbf{P}_t) = \mathbf{w}_t^H\mathbf{p} + \operatorname{Tr}\left(\mathbf{M}_t\mathbf{R}_t^{-1}\mathbf{A}\mathbf{P}_t\mathbf{P}^{-1}\mathbf{P}_t\mathbf{A}^H\mathbf{R}_t^{-1}\right),$$

which is not only convex in each $p_j$ , but also separable. Denote the diagonal elements of matrix $\mathbf{P}_t\mathbf{A}^H\mathbf{R}_t^{-1}\mathbf{M}_t\mathbf{R}_t^{-1}\mathbf{A}\mathbf{P}_t$ by $\mathbf{d}_t$, which is real-valued, the update of $\mathbf{P}$ can be found as follows:

$$(p_j)_{t+1} = \sqrt{(d_j)_t\,/\,(w_j)_t}.\tag{10}$$

The procedure is stated in Alg. 2.

A closer examination at the two algorithms reveals that compared to Alg. 1, Alg. 2 just iterate the inner loop (8) for one time instead of iterate until the sequence $\{\mathbf{p}_r\}$ converge, thus is expected to converge within a fewer number of iterations. We note that Alg. 2 also applies to the Gaussian maximum likelihood fitting problem that LIKES solves with a faster speed.

**Algorithm 2** Single loop majorization-minimization for robust decomposable covariance estimation

```
Set t = 0, initialize p_t to be any positive vector.
repeat
    R_t = A P_t A^H
    w_t = diag(A^H R_t^{-1} A)
    M_t = (K/N) sum_{i=1}^{N} x_i x_i^H / (x_i^T R_t^{-1} x_i)
    d_t = diag(P_t A^H R_t^{-1} M_t R_t^{-1} A P_t)
    (p_j)_{t+1} = sqrt((d_j)_t / (w_j)_t)
    t <- t + 1
until some convergence criterion is met
```

Despite problems such as the DOA estimation where Alg. 2 can be directly used, we introduce its application to the estimation of a real-valued Toeplitz covariance matrix. Consider a class of positive semidefinite Toeplitz matrices $T_K$ parameterized by its first row $[t_0, t_1, \ldots, t_{K-1}]$, and each element $t_{ij}$ satisfies $t_{ij} = t_{|i-j|}$. The Topelitz constraint is clearly linear, therefore the sequential SDP algorithm for general convex constraint applies. However, by exploring the structure of a Toeplitz matrix, a more efficient algorithm can be derived as stated below.

Inspired by the idea of [19], we consider Toeplitz matrices by embedding it as the upper-left part of a circulant matrix of larger size. The advantage is that by circulant embedding, $\mathbf{R}$ can be written in the form of (6), and Alg. 2 can be adopted.

Specifically, the feasible set is restricted to be a subset of $K \times K$ real-valued symmetric Toeplitz matrices, denoted by $T_L$, that can be embedded as the upper-left part of a real-valued symmetric circulant matrix of size $L \times L$ with $L > K$. If $\mathbf{R} \in T_L$, it can be written as

$$\mathbf{R} = \mathbf{A}\operatorname{diag}(p_0, \ldots, p_{L-1})\,\mathbf{A}^H,$$

where

$$\mathbf{A} = \begin{bmatrix}\mathbf{I}_K & \mathbf{0}\end{bmatrix}\mathbf{F}_L,$$

with $\mathbf{F}_L$ being a Fourier transform matrix of size $L \times L$. By the previous argument, at the $t$-th iteration $\mathbf{P}_{t+1}$ is the solution of the following problem

$$\begin{array}{ll} \underset{\mathbf{p} \geq \mathbf{0}}{\text{minimize}} & \mathbf{w}_t^H \mathbf{p} + \text{Tr}\left(\mathbf{M}_t \mathbf{R}_t^{-1} \mathbf{A} \mathbf{P}_t \mathbf{P}^{-1} \mathbf{P}_t \mathbf{A}^H \mathbf{R}_t^{-1}\right) \\ \text{subject to} & p_j = p_{L-j}, \ \forall j = 1, \ldots, L-1. \end{array}$$

Denote the diagonal elements of matrix $\mathbf{P}_t \mathbf{A}^H \mathbf{R}_t^{-1} \mathbf{M}_t \mathbf{R}_t^{-1} \mathbf{A} \mathbf{P}_t$ by $\mathbf{d}_t$, the update for $\mathbf{p}$ is given in closed-form as

$$(p_0)_{t+1} = \sqrt{(d_0)_t / (w_0)_t}$$
$$(p_j)_{t+1} = \sqrt{\left((d_j)_t + (d_{L-j})_t\right) / \left((w_j)_t + (w_{L-j})_t\right)}.$$

## 4. NUMERICAL RESULTS

In this section we present numerical results that demonstrate the advantage of incorporating a prior structure information into the estimator in reducing the estimation error. We also compare the proposed estimator with the estimator named COCA derived in [12] that deals with the Tyler's estimator with structure constraint. The estimation error is evaluated by the normalized mean-square error, namely

$$\text{NMSE}\left(\hat{\mathbf{R}}\right) = \text{E}\left\|\hat{\mathbf{R}} - \mathbf{R}_0\right\|_F^2 / \left\|\mathbf{R}_0\right\|_F^2,$$

where all matrices are normalized by their trace. The expected value is approximated by 100 times Monte Carlo simulations. The samples in all the simulations of this section are drawn independently from an elliptically distributed random variable $\mathbf{x} = \sqrt{\tau}\mathbf{u}$, where $\tau \sim \chi^2$ and $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_0)$, with $K = 15$.

In the simulation, $\mathbf{R}$ is chosen to be a Toeplitz matrix of the form

$$\mathbf{R}(\beta)_{ij} = \beta^{|i-j|}. \tag{11}$$

The model attracts a great deal of interest as it corresponds to the autocovariance matrix of an AR(1) process. Figs. 1 and 2 show the NMSE of different estimators as $N$ increases from 20 to 200 with $\beta$ being 0.4 and 0.8, respectively. The size of the circulant matrix $L$ is set to be $2K - 1$ (minimal embedding). We can see that in both cases, circulant embedding approximation achieves almost the same estimation error as imposing Toeplitz constraint and solve via the sequential SDP algorithm. While for $\beta = 0.4$ the proposed estimator achieves roughly the same estimation error as COCA, it outperforms the COCA estimator in the $\beta = 0.8$ case.

Fig. 3 plots the number of iterations versus the objective value that the robust LIKES and the proposed Alg. 2 require for the $\beta = 0.8$ case. Clearly Alg. 2 converges much faster than the double loop robust LIKES. We do not compare with the COCA and sequential SDP algorithms here as they are SDP based methods, whose computational complexity are much higher.

## 5. CONCLUSION

In this paper we have considered the robust covariance estimation problem with structure prior information. The problem has been formulated as a constrained optimization problem and algorithms based on majorization-minimization have been proposed to solve the problem efficiently. Numerical results have demonstrated that including structure information helps in improving the estimation accuracy.

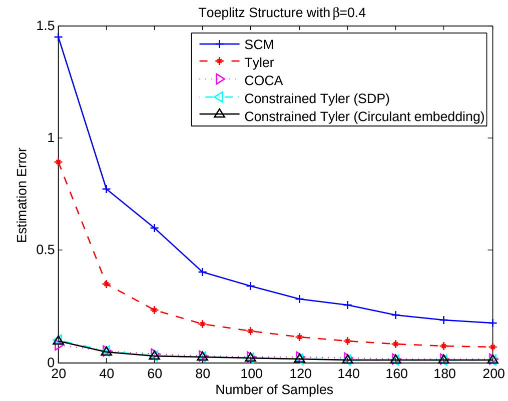


**Fig. 1**. Performance comparison of different estimators with $\mathbf{R}_0 = \mathbf{R}(0.4)$

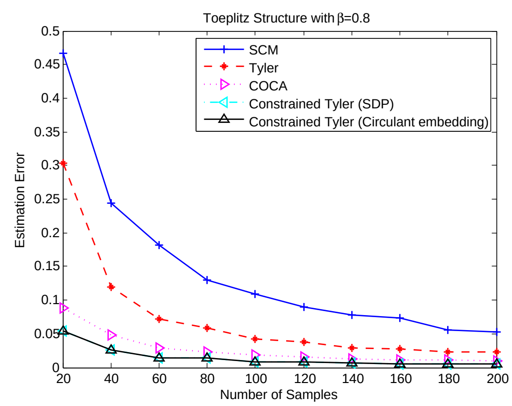


**Fig. 2**. Performance comparison of different estimators with $\mathbf{R}_0 = \mathbf{R}(0.8)$

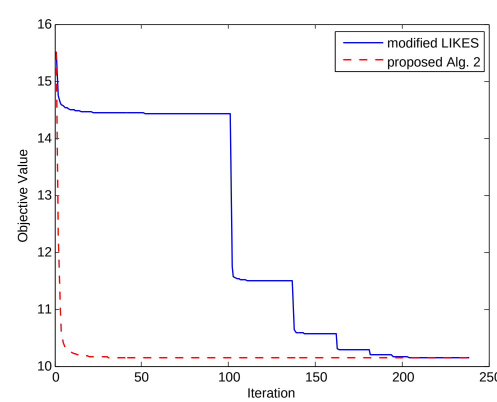


**Fig. 3**. Convergence comparison of robust LIKES and Alg. 2 for constrained Tyler's estimator with $\mathbf{R}_0 = \mathbf{R}(0.8)$

## 6. REFERENCES


[1] Y. I. Abramovich and N. K. Spencer, “Diagonally loaded normalised sample matrix inversion (LNSMI) for outlier-resistant adaptive filtering,” in *Proc. IEEE Int Conf. Acoust., Speech, Signal Process. (ICCASP), 2007.*, Honolulu, HI, 2007, vol. 3, pp. III–1105–III–1108.

[2] D. E. Tyler, “A distribution-free $M$-estimator of multivariate scatter,” *Ann. Statist.*, vol. 15, no. 1, pp. 234–251, 03 1987.

[3] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, “Complex elliptically symmetric distributions: Survey, new results and applications,” *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5597–5625, Nov 2012.

[4] E. Ollila and D. E. Tyler, “Distribution-free detection under complex elliptically symmetric clutter distribution,” in *Proc. IEEE 7th Sensor Array and Multichannel Signal Process. Workshop (SAM)*, Hoboken, NJ, June 2012, pp. 413–416.

[5] Yilun Chen, A. Wiesel, and A. O. Hero, “Robust shrinkage estimation of high-dimensional covariance matrices,” *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4097–4107, 2011.

[6] A. Wiesel, “Unified framework to regularized covariance estimation in scaled Gaussian models,” *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 29–38, 2012.

[7] O. Ledoit and M. Wolf, “A well-conditioned estimator for large-dimensional covariance matrices,” *J. Multivar. Anal.*, vol. 88, no. 2, pp. 365 – 411, 2004.

[8] Ying Sun, P. Babu, and D. P. Palomar, “Regularized robust estimation of mean and covariance matrix under heavy tails and outliers,” in *Proc. IEEE 8th Sensor Array and Multichannel Signal Process. Workshop (SAM)*, A Coruna, 2014, IEEE, pp. 125–128.

[9] Ying Sun, P. Babu, and D. P. Palomar, “Regularized Tyler’s scatter estimator: Existence, uniqueness, and algorithms,” *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 5143–5156, 2014.

[10] Mengyi Zhang, F. Rubio, D. P. Palomar, and X. Mestre, “Finite-sample linear filter optimization in wireless communications and financial systems,” *IEEE Trans. Signal Process.*, vol. 61, no. 20, pp. 5014–5025, 2013.

[11] P. J. Bickel and E. Levina, “Regularized estimation of large covariance matrices,” *Ann. Statist.*, pp. 199–227, 2008.

[12] I. Soloveychik and A. Wiesel, “Tyler’s covariance matrix estimator in elliptical models with convex structure,” *IEEE Trans. Signal Process.*, vol. 62, pp. 5251–5259, 2014.

[13] I. Soloveychik and A. Wiesel, “Group symmetry and non-Gaussian covariance estimation,” *arXiv preprint arXiv:1306.4103*, 2013.

[14] A. Aubry, A. De Maio, L. Pallotta, and A. Farina, “Maximum likelihood estimation of a structured covariance matrix with a condition number constraint,” *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 3004–3021, June 2012.

[15] P. Wirfalt and M. Jansson, “On Kronecker and linearly structured covariance matrix estimation,” *IEEE Trans. Signal Process.*, vol. 62, no. 6, pp. 1536–1547, March 2014.

[16] J. T. Kent, “The complex Bingham distribution and shape analysis,” *J. Roy. Statist. Soc. B*, pp. 285–299, 1994.

[17] M. Razaviyayn, M. Hong, and Z. Luo, “A unified convergence analysis of block successive minimization methods for nonsmooth optimization,” *SIAM J. Optim.*, vol. 23, no. 2, pp. 1126–1153, 2013.

[18] P. Stoica and P. Babu, “SPICE and LIKES: Two hyperparameter-free methods for sparse-parameter estimation,” *Signal Processing*, vol. 92, no. 7, pp. 1580–1590, 2012.

[19] A. Dembo, C. Mallows, and L. Shepp, “Embedding nonnegative definite Toeplitz matrices in nonnegative definite circulant matrices, with application to covariance estimation,” *IEEE Trans. Inf. Theory*, vol. 35, no. 6, pp. 1206–1212, 1989.