Breloy, Arnaud; Ollila, Esa; Pascal, Frederic

Spectral Shrinkage of Tyler's M-Estimator of Covariance Matrix

Published in:
2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, CAMSAP 2019 - Proceedings

DOI:
10.1109/CAMSAP45676.2019.9022652

Published: 01/12/2019

Document Version
Peer reviewed version

Please cite the original version:
Breloy, A., Ollila, E., & Pascal, F. (2019). Spectral Shrinkage of Tyler's M-Estimator of Covariance Matrix. In *2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, CAMSAP 2019 - Proceedings* (pp. 535-538). [9022652] IEEE. https://doi.org/10.1109/CAMSAP45676.2019.9022652

Spectral Shrinkage of Tyler's M-Estimator of Covariance Matrix

Arnaud Breloy
LEME (EA 4416)
University Paris Nanterre, France

Esa Ollila
Department of Signal Processing and Acoustics,
Aalto University, Finland

Frédéric Pascal
L2S / CentraleSupelec
University Paris-Saclay, France

***Abstract*—Covariance matrices usually exhibit specific spectral structures, such as low-rank ones in the case of factor models. In order to exploit this prior knowledge in a robust estimation process, we propose a new regularized version of Tyler's M-estimator of covariance matrix. This estimator is expressed as the minimizer of a robust M-estimating cost function plus a penalty that is unitary invariant (i.e., that only applies on the eigenvalue) that shrinks the estimated spectrum toward a fixed target. The structure of the estimate is expressed through an interpretable fixed-point equation. A majorization-minimization (MM) algorithm is derived to compute this estimator, and the g-convexity of the objective is also discussed. Several simulation studies illustrate the interest of the approach and also explore a method to automatically choose the target spectrum through an auxiliary estimator.**

I. Introduction

Covariance matrix estimation is generally a first step in the analysis of multivariate data sets. In this scope, M-estimators of the scatter [1] offer an interesting alternative to the traditional sample covariance matrix, thanks to their robustness properties among the family of elliptical distributions [2]. Notably, Tyler's M-estimator [3, 4] is a popular choice due to its invariance and robustness properties on this whole family.

Nevertheless, low/insufficient sample support ("small n, large p" problems) are still challenging, as M-estimators cannot be computed for $n < p$. A potential solution to this issue is to use shrinkage (also referred to as regularization) methods [5, 6]. Thus, regularized M-estimators have been proposed as minimizers of a penalized M-estimation objective, which involves a regularization penalty that shrinks the estimate towards a given target matrix $\mathbf{T}$ [7–10]. For this class of estimators, several works were also conducted on bias-variance trade-off for regularization parameter selection (see e.g., [11] and references therein).

In many applications, covariance matrices exhibit particular spectral structures [12–15], such as low-rank ones. Exploiting this structure through regularized M-estimators requires an appropriate setting of the target $\mathbf{T}$, that raises the following dilemma: $i)$ using $\mathbf{T} = \mathbf{I}$ (or $c\mathbf{I}$ with c an estimate of the mean of the eigenvalues) shrinks the eigenvalues of the estimate toward identity (or grand mean), which does not reflect most spectral structure; $ii)$ using another arbitrary matrix (e.g., diagonal) can account for a specific spectral structure, but also shrinks inherently the eigenvectors of the estimate, which may not be desired in general.

To alleviate this problem, we consider here the use of a penalty that is unitary invariant (i.e., that only applies on the eigenvalues [16]). A shrinkage penalty is proposed and applied to derive a new regularized version of Tyler's M-estimator, which is expressed through an interpretable fixed-point equation. A majorization-minimization (MM) algorithm [17] is derived to compute this estimator, and the g-convexity [18] of the objective is also discussed. Several simulations illustrate the interest of the approach and also explore a method to automatically select the target spectrum with the eFusion algorithm [15].

II. Background

A. Tyler's M-estimator

Consider a n-sample of p-dimensional vectors $\{\mathbf{z}_k\}_{k=1}^n$ (simply denoted $\{\mathbf{z}_k\}$ afterwards). Tyler's M-estimator [3] is the minimizer of the cost function

$$\mathcal{L}_{\mathrm{Ty}}(\mathbf{\Sigma}|\{\mathbf{z}_k\}) = \frac{p}{n}\sum_{k=1}^{n} \ln\left(\mathbf{z}_k^H \mathbf{\Sigma}^{-1}\mathbf{z}_k\right) + \ln|\mathbf{\Sigma}| \tag{1}$$

on the set of positive definite matrices, denoted $\mathcal{H}_p^{++}$. For $n > p$ and when the samples span the whole space, this estimator exists and is unique (up to a scale factor). Moreover, it satisfies the following fixed-point equation:

$$\hat{\mathbf{\Sigma}}_{\mathrm{Ty}} = \frac{p}{n}\sum_{k=1}^{n} \frac{\mathbf{z}_k\mathbf{z}_k^H}{\mathbf{z}_k^H \hat{\mathbf{\Sigma}}_{\mathrm{Ty}}^{-1}\mathbf{z}_k} \triangleq \mathcal{H}_{\mathrm{Ty}}(\hat{\mathbf{\Sigma}}_{\mathrm{Ty}}), \tag{2}$$

and can be computed with the fixed-point algorithm [2–4]

$$\mathbf{\Sigma}_{t+1} = \mathcal{H}_{\mathrm{Ty}}(\mathbf{\Sigma}_t). \tag{3}$$

In many applications, the number of variables measured in a single observation can be considerably larger than the number of available samples. For example, in portfolio optimization problems in finance, the number of assets often exceeds the number of historical returns. Thus, the condition $n > p$ can be restrictive in modern applications, which motivated the development of shrinkage methods.

B. Regularized Tyler's M-estimator

For "small n, large p" problems, several formulations (yielding the same estimate structure) have been proposed to compute a regularized counterpart of Tyler's M-estimator [7–10]. Below we present the one used in [8][Sec. IV.B] that is

Work of A. Breloy and F. Pascal has been partially supported by DGA under grant ANR-17-ASTR-0015. Work of E. Ollila was supported by the Academy of Finland grant No. 298118.

most compatible with our derivations of the spectrum regularized Tyler's (SRTy) M-estimator proposed herein. Consider the function

$$\mathcal{L}_{\mathrm{R}}(\boldsymbol{\Sigma}) = \mathrm{Tr}\left\{\boldsymbol{\Sigma}^{-1}\mathbf{T}\right\} + \ln|\boldsymbol{\Sigma}| \tag{4}$$

for a given target matrix $\mathbf{T} \in \mathcal{H}_p^{++}$, i.e, the Kullback-Leibler divergence from $\boldsymbol{\Sigma}$ to $\mathbf{T}$. For a real $\alpha > 0$, regularized Tyler's M-estimators are defined as the minimizers of the objective

$$\mathcal{L}^{\alpha}_{\mathrm{RTy}}(\boldsymbol{\Sigma}) = \mathcal{L}_{\mathrm{Ty}}(\boldsymbol{\Sigma}|\{\mathbf{z}_k\}) + \alpha\mathcal{L}_{\mathrm{R}}(\boldsymbol{\Sigma}) \tag{5}$$

on $\mathcal{H}_p^{++}$. Denote $\beta = \alpha/(1+\alpha)$. When $\beta \in (\min(0, 1 - n/p), 1]$, this estimator exists and is unique [8–10]. Moreover, it satisfies the following fixed-point equation:

$$\hat{\boldsymbol{\Sigma}}_{\mathrm{RTy}} = \frac{p(1-\beta)}{n}\sum_{k=1}^{n}\frac{\mathbf{z}_k\mathbf{z}_k^H}{\mathbf{z}_k^H\hat{\boldsymbol{\Sigma}}_{\mathrm{RTy}}^{-1}\mathbf{z}_k} + \beta\mathbf{T}, \tag{6}$$

where the shrinkage of the estimate towards the target $\mathbf{T}$ becomes clearly apparent. Similarly to $\hat{\boldsymbol{\Sigma}}_{\mathrm{Ty}}$, $\hat{\boldsymbol{\Sigma}}_{\mathrm{RTy}}$ can be computed using the fixed-point iterations

$$\boldsymbol{\Sigma}_{t+1} = (1-\beta)\mathcal{H}_{\mathrm{Ty}}(\boldsymbol{\Sigma}_t) + \beta\mathbf{T}. \tag{7}$$

Note that for β sufficiently large, $\hat{\boldsymbol{\Sigma}}_{\mathrm{RTy}}$ exists in the problematic settings $n < p$ (at the cost of introducing some bias [19]). The use of shrinkage is also beneficial in the cases $p > n$, as it can improve the estimation accuracy (or bias-variance trade-off). However, as discussed in the introduction, the formulation of $\hat{\boldsymbol{\Sigma}}_{\mathrm{RTy}}$ cannot account for a specific spectral structure without additionally shrinking the eigenvectors of the estimate, which motivated the present work.

III. Spectral shrinkage of Tyler's M-estimator

A. Problem formulation

The eigenvalue decomposition (EVD) of a given Hermitian matrix is denoted

$$\boldsymbol{\Sigma} \overset{\mathrm{EVD}}{=} \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H \quad \text{with} \quad \begin{array}{l} \mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_p] \in \mathcal{U}_p^p, \\ \boldsymbol{\Lambda} = \mathrm{diag}(\boldsymbol{\lambda}), \\ \boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_p]. \end{array} \tag{8}$$

Above $\mathcal{U}_p^p$ denotes the set of $p \times p$ unitary matrices and $\mathrm{diag}(\boldsymbol{\lambda})$ a diagonal matrix with elements in $\boldsymbol{\lambda}$ as its diagonal elements. In order to avoid any ambiguity in this definition, we assume ordered eigenvalues as $\lambda_1 \geq \ldots \geq \lambda_p > 0$, and an arbitrary element of each $\mathbf{v}_j$ (e.g., the first entry) can be assumed to be real positive. Define $\boldsymbol{\phi}$ as the operator that extracts the diagonal matrix of eigenvalues from a given Hermitian matrix (with preserved order):

$$\boldsymbol{\phi} : \boldsymbol{\Sigma} \overset{\mathrm{EVD}}{=} \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H \longmapsto \boldsymbol{\phi}(\boldsymbol{\Sigma}) = \boldsymbol{\Lambda}. \tag{9}$$

For a given target spectrum $\boldsymbol{\Lambda}_T$ (a real positive diagonal matrix, with ordered diagonal), we propose the following penalty function or spectral regularizer,

$$\mathcal{L}_{\mathrm{SR}}(\boldsymbol{\Sigma}) = \mathrm{Tr}\left\{\boldsymbol{\Lambda}_T\left(\boldsymbol{\phi}(\boldsymbol{\Sigma})\right)^{-1}\right\} + \ln|\boldsymbol{\phi}(\boldsymbol{\Sigma})|. \tag{10}$$

Note that $\mathcal{L}_{\mathrm{SR}}$ is unitary invariant, i.e., it is only a function of the eigenvalues of the entry. We propose a *spectral regularized Tyler's (SRTy) M-estimator* as the minimizer of the objective

$$\mathcal{L}^{\alpha}_{\mathrm{SRTy}}(\boldsymbol{\Sigma}) = \mathcal{L}_{\mathrm{Ty}}(\boldsymbol{\Sigma}|\{\mathbf{z}_k\}) + \alpha\mathcal{L}_{\mathrm{SR}}(\boldsymbol{\Sigma}) \tag{11}$$

on $\mathcal{H}_p^{++}$ where $\alpha > 0$.

B. Structure of the solution

The following proposition explicits the structure of the proposed regularized estimator.

Proposition 1 *The minimizers $\mathcal{L}^{\alpha}_{\mathrm{SRTy}}$, denoted $\hat{\boldsymbol{\Sigma}}_{\mathrm{SRTy}}$, satisfy the following fixed-point equation:*

$$\begin{cases} \hat{\boldsymbol{\Sigma}}_{\mathrm{SRTy}} = \dfrac{p}{n(1+\alpha)}\displaystyle\sum_{k=1}^{n}\frac{\mathbf{z}_k\mathbf{z}_k^H}{\mathbf{z}_k^H\hat{\boldsymbol{\Sigma}}_{\mathrm{SRTy}}^{-1}\mathbf{z}_k} + \dfrac{\alpha}{1+\alpha}\hat{\mathbf{V}}\boldsymbol{\Lambda}_T\hat{\mathbf{V}}^H \\ \hat{\boldsymbol{\Sigma}}_{\mathrm{SRTy}} \overset{\mathrm{EVD}}{=} \hat{\mathbf{V}}\hat{\boldsymbol{\Lambda}}\hat{\mathbf{V}}^H. \end{cases} \tag{12}$$

Proof: First, remark that $\ln|\boldsymbol{\phi}(\boldsymbol{\Sigma})| = \ln|\boldsymbol{\Sigma}|$, so we have the relation

$$\frac{\partial \ln|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\Sigma}} = \frac{\partial \ln|\boldsymbol{\phi}(\boldsymbol{\Sigma})|}{\partial \boldsymbol{\Sigma}} = \boldsymbol{\Sigma}^{-1}. \tag{13}$$

Moreover

$$\frac{\partial \ln\left(\mathbf{z}_k^H\boldsymbol{\Sigma}^{-1}\mathbf{z}_k\right)}{\partial \boldsymbol{\Sigma}} = -\frac{\boldsymbol{\Sigma}^{-1}\mathbf{z}_k\mathbf{z}_k^H\boldsymbol{\Sigma}^{-1}}{\mathbf{z}_k^H\boldsymbol{\Sigma}^{-1}\mathbf{z}_k}, \tag{14}$$

and for $\boldsymbol{\Sigma} \overset{\mathrm{EVD}}{=} \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H$, it can be shown that

$$\frac{\partial \mathrm{Tr}\left\{\boldsymbol{\Lambda}_T\left(\boldsymbol{\phi}(\boldsymbol{\Sigma})\right)^{-1}\right\}}{\partial \boldsymbol{\Sigma}} = -\mathbf{V}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Lambda}_T\boldsymbol{\Lambda}^{-1}\mathbf{V}^H. \tag{15}$$

Thus, solving $\partial\mathcal{L}^{\alpha}_{\mathrm{SRTy}}(\boldsymbol{\Sigma})/\partial\boldsymbol{\Sigma} = \mathbf{0}$ leads to the solution expressed in (12). ■

Note that when $\boldsymbol{\Lambda}_T = \mathbf{I}$, the proposed estimator coincides with the regularized estimator of [8–10] in (6). Otherwise, the proposed method only shrinks the eigenvalues towards the target $\boldsymbol{\Lambda}_T$ and does not apply any explicit shrinkage on the eigenvectors, which was the desired outcome. Another property of $\hat{\boldsymbol{\Sigma}}_{\mathrm{SRTy}}$ is that it naturally satisfies the constraint

$$\mathrm{Tr}\left\{\boldsymbol{\Lambda}_T\hat{\boldsymbol{\Lambda}}^{-1}\right\} = p, \tag{16}$$

which can be checked by multiplying (12) by $\hat{\boldsymbol{\Sigma}}_{\mathrm{SRTy}}^{-1}$ and taking the trace of the obtained relation.

C. Geodesic convexity

The concept of geodesic convexity (g-convexity) has been used in [7, 10, 18] to study the uniqueness of (regularized) M-estimators. Define the geodesic path between two matrices $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_1$ in $\mathcal{H}_p^{++}$ as

$$\boldsymbol{\Sigma}_x = \boldsymbol{\Sigma}_0^{1/2}\left(\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{-1/2}\right)^x\boldsymbol{\Sigma}_0^{1/2}, \quad x \in [0,1]. \tag{17}$$

A function f is said to be g-convex on $\mathcal{H}_p^{++}$ if

$$f(\boldsymbol{\Sigma}_x) \leq (1-x)f(\boldsymbol{\Sigma}_0) + xf(\boldsymbol{\Sigma}_1), \; x \in (0,1). \tag{18}$$

The g-convexity enjoys properties similar to those of the convexity in the standard Euclidean case. Specifically, if a function is g-convex, then any local minimum is a global minimum (cf. Proposition 1 in [18]). For the considered cost function, we have the following proposition:

Proposition 2 *The objective $\mathcal{L}^{\alpha}_{\mathrm{SRTy}}$ is g-convex in $\boldsymbol{\Sigma}$ on $\mathcal{H}_p^{++}$.*

Algorithm 1 MM algorithm to compute $\hat{\boldsymbol{\Sigma}}_{\mathrm{SRTy}}$

Initialize $\boldsymbol{\Sigma}_0 = \mathbf{I}$
repeat
  Compute $\mathcal{H}_{\mathrm{Ty}}(\boldsymbol{\Sigma}_t) \stackrel{\mathrm{EVD}}{=} \mathbf{U}_t \mathbf{D}_t \mathbf{U}_t^H$ from (2).
  Update $\boldsymbol{\Sigma}_{t+1} = \mathbf{U}_t \left(\frac{1}{1+\alpha} \mathbf{D}_t + \frac{\alpha}{1+\alpha} \boldsymbol{\Lambda}_T \right) \mathbf{U}_t^H$.
  $t = t + 1$
until convergence criterion is met

Proof: The g-convexity of $\mathcal{L}_{\mathrm{Ty}}$ is established in [10]. The function $\mathcal{L}_{\mathrm{SR}}$ is orthogonally invariant, i.e. $\mathcal{L}_{\mathrm{SR}}(\boldsymbol{\Sigma}) = \mathcal{L}_{\mathrm{SR}}(\mathbf{Q}\boldsymbol{\Sigma}\mathbf{Q}^H)$ for any rotation matrix $\mathbf{Q}$. Moreover, $\mathcal{L}_{\mathrm{Ty}}$ is convex in $\{\ln(\lambda_j)\}$ (from the EVD of $\boldsymbol{\Sigma}$ in (8)). Therefore it is g-convex in $\boldsymbol{\Sigma}$ (Theorem 3.1. of [16]). ■

Hence, if a solution satisfying (12) is obtained in the feasible set $\mathcal{H}_p^{++}$, it is a global minimum. Due to space limitations, the conditions for existence and uniqueness of such solution will be discussed in a forthcoming communication.

D. Majorization-minimization algorithm

In this section, we derive an MM algorithm [20] in order to compute the proposed estimator. At each iteration, the MM algorithm operates by minimizing a surrogate (upper bound) of the objective function at the given point. This ensures a monotonic decrement of the objective at each iterations.

Thanks to the concavity of ln, the function $\mathcal{L}_{\mathrm{Ty}}$ can be upperbounded at the point $\boldsymbol{\Sigma}_t$ by the surrogate

$$g(\boldsymbol{\Sigma}|\boldsymbol{\Sigma}_t) = \mathrm{Tr}\left\{\boldsymbol{\Sigma}^{-1}\mathbf{Z}_t\right\} + \ln|\boldsymbol{\Sigma}| + \mathrm{const.} \tag{19}$$

with $\mathbf{Z}_t = \mathcal{H}_{\mathrm{Ty}}(\boldsymbol{\Sigma}_t)$ (defined in (2)) with equality achieved at $\boldsymbol{\Sigma}_t$ (cf. Proposition 1 of [20]). The update of the variable $\boldsymbol{\Sigma}_{t+1}$ is then obtained by solving the following sub-problem:

$$\underset{\boldsymbol{\Sigma}}{\mathrm{minimize}} \quad g(\boldsymbol{\Sigma}|\boldsymbol{\Sigma}_t) + \alpha\mathcal{L}_{\mathrm{SR}}(\boldsymbol{\Sigma}). \tag{20}$$

With the change of variable $\boldsymbol{\Sigma} \stackrel{\mathrm{EVD}}{=} \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H$, this problem is equivalent to

$$\begin{aligned} \underset{\mathbf{V}\in\mathcal{U}_p^p, \boldsymbol{\Lambda}\in\mathcal{D}_p}{\mathrm{minimize}} \quad & \mathrm{Tr}\left\{\mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{V}^H\mathbf{Z}_t\right\} + \alpha\mathrm{Tr}\left\{\boldsymbol{\Lambda}_T\boldsymbol{\Lambda}^{-1}\right\} \\ & + (1+\alpha)\ln|\boldsymbol{\Lambda}|. \end{aligned} \tag{21}$$

Denote $\mathbf{Z}_t \stackrel{\mathrm{EVD}}{=} \mathbf{U}_t\mathbf{D}_t\mathbf{U}_t^H$. The solution for the update of the eigenvectors is given as $\mathbf{V}_{t+1} = \mathbf{U}_t$ (see e.g., [21] and references therein). For this fixed solution, the update of $\boldsymbol{\Lambda}$ is obtained by solving

$$\underset{\boldsymbol{\Lambda}\in\mathcal{D}_p}{\mathrm{minimize}} \quad \mathrm{Tr}\left\{\boldsymbol{\Lambda}^{-1}(\mathbf{D}_t + \alpha\boldsymbol{\Lambda}_T)\right\} + (1+\alpha)\ln|\boldsymbol{\Lambda}|, \tag{22}$$

where $\mathcal{D}_p$ denotes the set of positive diagonal matrices. This then leads to $\boldsymbol{\Lambda}_{t+1} = \{1/(1+\alpha)\}\mathbf{D}_t + \{\alpha/(1+\alpha)\}\boldsymbol{\Lambda}_T$. Eventually, the resulting algorithm take the form of fixed-point iterations, summarized in the table Algorithm 1.

IV. Simulations

A. Validation

For these validation simulations the data follows a multivariate Student t-distribution [2] with the $d = 3$ degrees of freedom (thus, with considerably heavier tails than the normal distribution). The scatter matrix, denoted $\boldsymbol{\Sigma}(\rho)$, is Toeplitz with entries $\boldsymbol{\Sigma}_{i,j} = \rho^{(j-i)}$, with $\rho = \xi(1+\sqrt{-1})/\sqrt{2}$ and its EVD is denoted by $\boldsymbol{\Sigma}(\rho) \stackrel{\mathrm{EVD}}{=} \mathbf{V}(\rho)\boldsymbol{\Lambda}(\rho)\mathbf{V}^H(\rho)$. Figure 1 illustrate the convergence of Algorithm 1 and the convergence of the estimate towards the fixed-point equation in (12), which are empirical validations of Proposition A and the proposed MM algorithm. Figure 2 presents the normalized mean squared error (NMSE) on the shape matrix (namely, trace normalized covariance matrix) and the spectrum for $\hat{\boldsymbol{\Sigma}}_{\mathrm{SCM}}$, $\hat{\boldsymbol{\Sigma}}_{\mathrm{Ty}}$, $\hat{\boldsymbol{\Sigma}}_{\mathrm{RTy}}$, and $\hat{\boldsymbol{\Sigma}}_{\mathrm{SRTy}}$ (with target spectrum $\boldsymbol{\Lambda}_T = \boldsymbol{\Lambda}(\rho_T)$ for various values of $|\rho_T|$). First we confirm that, for $\boldsymbol{\Lambda}_T = \mathbf{I}$ (i.e. $|\rho| = 0$), $\hat{\boldsymbol{\Sigma}}_{\mathrm{RTy}}$ and $\hat{\boldsymbol{\Sigma}}_{\mathrm{SRTy}}$ are equivalent. Clearly, $\mathbf{I}$ (or $\boldsymbol{\Lambda}_T = \mathbf{I}$) is an appropriate target when $|\rho|$ tends closer to 0, which explains the good performance of these estimators on the left panel of Figure 2. However, this is no longer true when $|\rho|$ tends to 1, and $\hat{\boldsymbol{\Sigma}}_{\mathrm{SRTy}}$ using an appropriate target can offer a better bias-variance trade-off. Of course establishing a target spectrum depends on available prior knowledge and the application. In the next example, we test an application where $\boldsymbol{\Lambda}_T$ is automatically selected with an auxiliary method.

B. Data-adaptive target spectrum selection

We now consider a spiked model, where the scatter matrix is built as

$$\boldsymbol{\Sigma} = \sum_{i=1}^{r} \gamma\mathbf{v}_i\mathbf{v}_i^H + \sum_{i=r+1}^{p} \sigma^2\mathbf{v}_i\mathbf{v}_i^H \tag{23}$$

where $\gamma = 20$, $\sigma^2 = 1$, and $\mathbf{v}_r$ are the eigenvectors of $\boldsymbol{\Sigma}(0.5)$ (constructed as in the previous section). In this scenario, one can leverage on recent eigenvalues-fusing algorithms [14, 15] to get a first estimate of the spectral structure (rank and plateaus). Figure 3 illustrates the performance of $\hat{\boldsymbol{\Sigma}}_{\mathrm{SRTy}}$ using eFusion [15] (with corresponding estimator denoted $\hat{\boldsymbol{\Sigma}}_{\mathrm{eFusion}}$) to build its target spectrum, that is, the target spectrum was chosen as $\boldsymbol{\Lambda}_T = \phi(\hat{\boldsymbol{\Sigma}}_{\mathrm{eFusion}})$. On the left panel of Figure 3, the samples are close to Gaussian (namely, the degrees of freedom of the t-distribution is $d = 20$). Since $\hat{\boldsymbol{\Sigma}}_{\mathrm{SCM}}$ is close to be the MLE, it is more accurate than $\hat{\boldsymbol{\Sigma}}_{\mathrm{Ty}}$. $\hat{\boldsymbol{\Sigma}}_{\mathrm{RTy}}$ exhibits a bias-variance trade-off that can improve the estimation accuracy. $\hat{\boldsymbol{\Sigma}}_{\mathrm{eFusion}}$ exhibits the best performance as it is able to fuse eigenvalues correctly in this Gaussian factor model case. $\hat{\boldsymbol{\Sigma}}_{\mathrm{SRTy}}$ outperforms $\hat{\boldsymbol{\Sigma}}_{\mathrm{RTy}}$, as it is able to exploit an appropriate prior on the spectral structure (which is far from $\mathbf{I}$ in the factor model). On the right panel of Figure 3, the samples follow a heavy-tailed t-distribution with $d = 3$ degrees of freedom. $\hat{\boldsymbol{\Sigma}}_{\mathrm{Ty}}$ is robust to these distributions so it outperforms $\hat{\boldsymbol{\Sigma}}_{\mathrm{SCM}}$. $\hat{\boldsymbol{\Sigma}}_{\mathrm{eFusion}}$ appears robust to non-Gaussian distributions, but does not offer as much gain as in the previous case. Finally, $\hat{\boldsymbol{\Sigma}}_{\mathrm{SRTy}}$ offers improved MSE in this non-Gaussian scenario for large values of α.

V. Conclusion

This paper proposed a new spectrum regularizing penalty function $\mathcal{L}_{\mathrm{SR}}(\boldsymbol{\Sigma})$ that can be used in covariance estimation problems to shrink the eigenvalues of the covariance matrix estimator towards a fixed target spectrum $\boldsymbol{\Lambda}_T$. The penalty is orthogonally invariant and g-convex, and it was used in

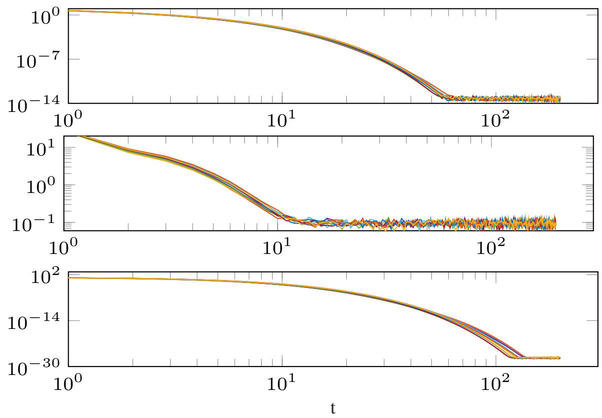

Fig. 1. Convergence criterions w.r.t. iterations for 10 independent Monte Carlo trials: gap to minimal value of objective function (top), MSE between iterates (middle), MSE between left and right terms of fixed-point equation (12) (bottom). $n = 40$, $p = 30$, $|\rho| = \xi = 0.9$. The data is generated from p-variate Student t-distribution with $d = 3$ degrees of freedom.

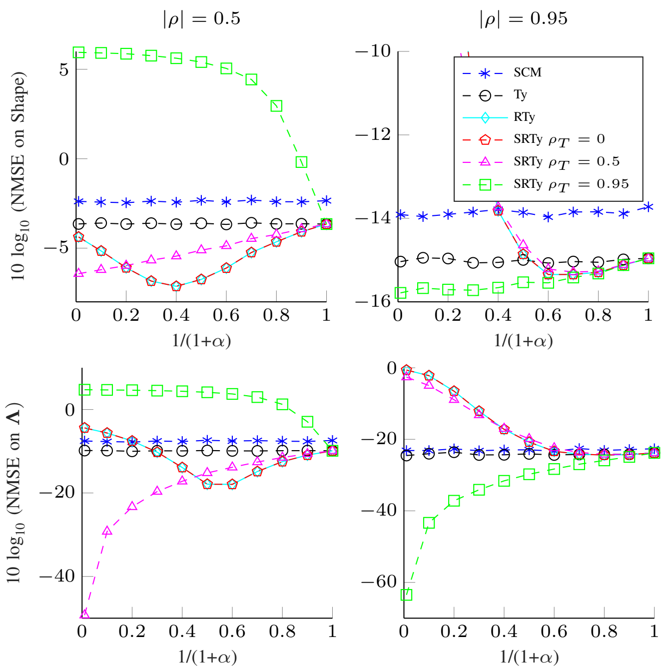

Fig. 2. NMSE on shape matrix (top) and spectrum (bottom) versus $1/(1+\alpha)$ for $p = 12$, $n = 20$, $|\rho| = 0.5$ (left panel) and $|\rho| = 0.95$ (right panel).

this paper to construct a new spectrum-regularized Tyler's M-estimator (SRTy-estimator). An algorithm was proposed to compute this estimator (that satisfies the fixed-point equation (12)) and several simulations illustrated the interest of the approach. The important question of the adaptive selection of the regularization parameters will be discussed in a forthcoming communication by extending [10].

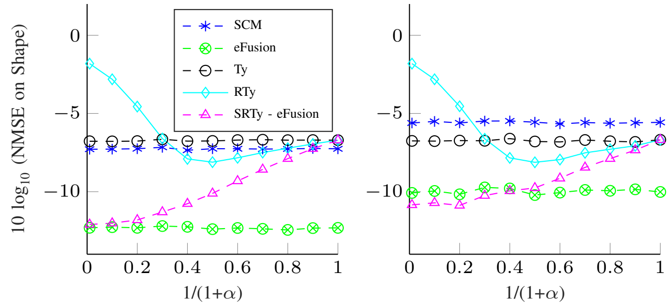

Fig. 3. NMSE on shape matrix versus $1/(1 + \alpha)$ in the spiked covariance model for $p = 12$, $n = 20$, where the data is close to Gaussian ($d = 20$, left) and heavy-tailed ($d = 3$, right).

REFERENCES

[1] R.A. Maronna, "Robust M-estimators of multivariate location and scatter," *The Annals of Statistics*, vol. 4, no. 1, pp. 51–67, 1976.

[2] E. Ollila, D.E. Tyler, V. Koivunen, and H.V. Poor, "Complex elliptically symmetric distributions: Survey, new results and applications," *IEEE Trans. on Sig. Proc.*, vol. 60, no. 11, pp. 5597–5625, 2012.

[3] D.E. Tyler, "A distribution-free M-estimator of multivariate scatter," *The Annals of Statistics*, vol. 15, no. 1, pp. 234–251, 1987.

[4] F. Pascal, Y. Chitour, J-P. Ovarlez, P. Forster, and P. Larzabal, "Existence and characterization of the covariance matrix maximum likelihood estimate in Spherically Invariant Random Processes," *IEEE Trans. on Sig. Proc.*, vol. 56, no. 1, pp. 34 – 48, 2008.

[5] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis*, vol. 88, no. 2, pp. 365 – 411, 2004.

[6] O. Ledoit, M. Wolf, et al., "Nonlinear shrinkage estimation of large-dimensional covariance matrices," *The Annals of Statistics*, vol. 40, no. 2, pp. 1024–1060, 2012.

[7] A. Wiesel, "Unified framework to regularized covariance estimation in scaled Gaussian models," *IEEE Trans. on Sig. Proc.*, vol. 60, no. 1, pp. 29–38, 2012.

[8] Y. Sun, P. Babu, and D.P. Palomar, "Regularized Tyler's Scatter Estimator: Existence, Uniqueness, and Algorithms," *IEEE Trans. on Sig. Proc.*, vol. 62, no. 19, pp. 5143–5156, 2014.

[9] F. Pascal, Y. Chitour, and Y. Quek, "Generalized robust shrinkage estimator and its application to stap detection problem," *IEEE Trans. on Sig. Proc.*, vol. 62, no. 21, pp. 5640–5651, 2014.

[10] E. Ollila and D.E. Tyler, "Regularized M -estimators of scatter matrix," *IEEE Trans. on Sig. Proc.*, vol. 62, no. 22, pp. 6059–6070, 2014.

[11] Q. Hoarau, A. Breloy, G. Ginolhac, A.M. Atto, and J-M. Nicolas, "A subspace approach for shrinkage parameter selection in undersampled configuration for regularised tyler estimators," in *Proc. of IEEE ICASSP*, pp. 3291–3295, 2017.

[12] Y. Sun, A. Breloy, P. Babu, D.P. Palomar, F. Pascal, and G. Ginolhac, "Low-complexity algorithms for low rank clutter parameters estimation in radar systems," *IEEE Trans. on Sig. Proc.*, vol. 64, no. 8, pp. 1986–1998, 2016.

[13] A. Aubry, A. De Maio, and L. Pallotta, "A geometric approach to covariance matrix estimation and its applications to radar problems," *IEEE Trans. on Sig. Proc.*, vol. 66, no. 4, pp. 907–922, 2018.

[14] D.E. Tyler and M. Yi, "Lassoing eigenvalues," *arXiv preprint arXiv:1805.08300*, 2018.

[15] S. Basiri, E. Ollila, G. Drašković, and F. Pascal, "Fusing eigenvalues," in *Proc. of IEEE ICASSP*, pp. 4968–4972, 2019.

[16] D.E. Tyler and M. Yi, "Shrinking the sample covariance matrix using convex penalties on the matrix-log transformation," *arXiv preprint arXiv:1903.08281*, 2019.

[17] Y. Sun, P. Babu, and D.P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. on Sig. Proc.*, vol. PP, no. 99, pp. 1–1, 2016.

[18] A. Wiesel, "Geodesic convexity and covariance estimation," *IEEE Trans. on Sig. Proc.*, vol. 60, no. 12, pp. 6182–6189, 2012.

[19] A. Kammoun, R. Couillet, F. Pascal, and M. Alouini, "Convergence and fluctuations of regularized tyler estimators," *IEEE Trans. on Sig. Proc.*, vol. 64, no. 4, pp. 1048–1060, 2016.

[20] Y. Sun, P. Babu, and D.P. Palomar, "Robust estimation of structured covariance matrix for heavy-tailed elliptical distributions," *IEEE Trans. on Sig. Proc.*, vol. 64, no. 14, pp. 3576–3590, 2016.

[21] B. Kang, V. Monga, and M. Rangaswamy, "Rank-constrained maximum likelihood estimation of structured covariance matrices," *IEEE Trans. on Aero. and Elec. Syst.*, vol. 50, no. 1, pp. 501–515, 2014.