Affine Equivariant Tyler's M-Estimator Applied to Tail Parameter Learning of Elliptical Distributions

Esa Ollila, *Senior Member, IEEE*, Daniel P. Palomar, *Fellow, IEEE*, and Frédéric Pascal, *Senior Member, IEEE*

***Abstract*—We propose estimating the scale parameter (mean of the eigenvalues) of the scatter matrix of an unspecified elliptically symmetric distribution using weights obtained by solving Tyler's M-estimator of the scatter matrix. The proposed Tyler's weights-based estimate (TWE) of scale is then used to construct an affine equivariant Tyler's M-estimator as a weighted sample covariance matrix using normalized Tyler's weights. We then develop a unified framework for estimating the unknown tail parameter of the elliptical distribution (such as the degrees of freedom (d.o.f.) ν of the multivariate t (MVT) distribution). Using the proposed TWE of scale, a new robust estimate of the d.o.f. parameter of MVT distribution is proposed with excellent performance in heavy-tailed scenarios, outperforming other competing methods. R-package is available that implements the proposed method.**

***Index Terms*—Covariance matrix, elliptical distributions, scatter matrix, Tyler's M-estimator.**

I. Introduction

We MODEL the observed p-variate observations $\mathbf{x}_1, \ldots, \mathbf{x}_n$ as independent and identically distributed (i.i.d.) random samples from an unspecified centered (i.e., symmetric around the origin) elliptically symmetric (ES) distribution [1], [2]. A continuous random vector $\mathbf{x} \in \mathbb{R}^p$ has centered ES distribution if it possesses a probability density function (pdf) of the form

$$f(\mathbf{x}) = C_{p,g}|\boldsymbol{\Sigma}|^{-1/2} g(\mathbf{x}^\top \boldsymbol{\Sigma}^{-1}\mathbf{x}),$$

where $g : \mathbb{R}_{\geq 0} \to \mathbb{R}_{>0}$ is called the density generator, $\boldsymbol{\Sigma} \succ 0$ is the positive definite symmetric matrix parameter, called the *scatter matrix*, and $C_{p,g}$ is a normalizing constant ensuring that $f(\mathbf{x})$ integrates to 1. We let $\mathbf{x} \sim \mathcal{E}_p(\mathbf{0}, \boldsymbol{\Sigma}, g)$ to denote this case. For example, the centered multivariate normal (MVN) distribution $\mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$ is obtained when $g(t) = \exp(-t/2)$ while the multivariate t (MVT) distribution with $\nu > 0$ degrees of freedom (d.o.f.) is obtained when

$$g(t) = (1 + t/\nu)^{-(p+\nu)/2}. \tag{1}$$

Parameter $\nu > 0$ is a tail parameter of the density. For $\nu \to \infty$, the MVT distribution reduces to the MVN distribution, while $\nu = 1$ corresponds to the multivariate Cauchy distribution. Also, many other subclasses of ES distributions are parametrized by a density generator indexed by an additional tail parameter that is unknown in practice and needs to be estimated. Learning this unknown parameter is also one of the goals of this letter.

We are interested in estimating the scale parameter of the scatter matrix, defined as the mean of its eigenvalues, $\eta = \mathrm{tr}(\boldsymbol{\Sigma})/p = (\lambda_1 + \cdots + \lambda_p)/p$, where $\lambda_i > 0$ denotes the ith eigenvalue of $\boldsymbol{\Sigma}$. Formally, $\eta \equiv \eta(\boldsymbol{\Sigma})$ is a *scale* parameter if it verifies $\eta(\mathbf{I}) = 1$ and $\eta(a\boldsymbol{\Sigma}) = a\eta(\boldsymbol{\Sigma})$ for all $a > 0$ [3].

Tyler's M-estimator [4] is a popular robust M-estimator of the scatter matrix that has been extensively studied both in signal processing and statistics literature (e.g, [5], [6], [7], [8], [9], [10], [11], [12], [13]). It is minimax bias-robust [14] and has the best possible breakdown point attainable by an M-estimator [15]. Tyler's M-estimator is defined as the solution to the fixed-point equation

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n} \frac{p}{\mathbf{x}_i^\top \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{x}_i} \mathbf{x}_i\mathbf{x}_i^\top \triangleq \mathcal{H}(\hat{\boldsymbol{\Sigma}}; \{\mathbf{x}_i\}). \tag{2}$$

Note that me way also write the map $\mathcal{H}(\cdot;\cdot)$ in the form

$$\mathcal{H}(\hat{\boldsymbol{\Sigma}}; \{\mathbf{x}_i\}) = \frac{1}{n}\sum_{i=1}^{n} \hat{w}_i \mathbf{x}_i\mathbf{x}_i^\top \quad \text{with} \quad \hat{w}_i = \frac{p}{\mathbf{x}_i^\top \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{x}_i},$$

where $\hat{w}_i, i = 1, \ldots, n$, are referred to as *Tyler's weights*. Tyler's M-estimator is unique only up to a scaling factor, and therefore a common convention is to consider a solution that verifies $\mathrm{tr}(\hat{\boldsymbol{\Sigma}}) = p$. Thus, Tyler's M-estimator is actually an estimator of a *shape matrix* (normalized scatter matrix) $\boldsymbol{\Lambda}$, defined by $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}/\eta = p\boldsymbol{\Sigma}/\mathrm{tr}(\boldsymbol{\Sigma})$, and verifying $\mathrm{tr}(\boldsymbol{\Lambda}) = p$.

In this letter, we propose an estimator of the scale η based on Tyler's weights $\hat{w}_i$. The proposed scale estimate along with Tyler's M-estimator $\hat{\boldsymbol{\Sigma}}$ are then jointly used for constructing affine equivariant robust estimates of the scatter matrix $\boldsymbol{\Sigma}$ and the covariance matrix $\mathbf{R} = \mathrm{cov}(\mathbf{x})$ (or their shrinkage versions). These developments are described in Section II. Then, in Section III, we propose a unified framework allowing to estimate the tail parameter of the elliptical distribution using the proposed scale statistic $\hat{\eta}$. In the case of the MVT distribution, this leads to a new estimate of the d.o.f. parameter based on Tyler's weights. Finally, Section IV demonstrates the relevance of the proposed approach on simulated data, with concluding remarks in Section V. In the R package `fitHeavyTail` [16], the function `fit_Tyler` implements this method.

Manuscript received 7 May 2023; revised 27 July 2023; accepted 31 July 2023. Date of publication 3 August 2023; date of current version 9 August 2023. This work was supported by Hong Kong GRF under Grant 16207820. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Guang Hua. *(Corresponding author: Esa Ollila.)*

Esa Ollila is with the Department of Information and Communications Engineering, Aalto University, FI-00076 Aalto, Finland (e-mail: esa.ollila@aalto.fi).

Daniel P. Palomar is with the Hong Kong University of Science and Technology, Hong Kong (e-mail: palomar@ust.hk).

Frédéric Pascal is with CNRS Université Paris-Saclay, CentraleSupélec, 91190 Rennes, France (e-mail: frederic.pascal@centralesupelec.fr).

Digital Object Identifier 10.1109/LSP.2023.3301341

II. Estimate of Scale, Scatter, and Covariance Matrix Based on Tyler's Weights

Assuming that $\mathbf{x} \sim \mathcal{E}_p(\mathbf{0}, \boldsymbol{\Sigma}, g)$ has finite 2nd-order moments, then its covariance matrix, $\mathbf{R} = \mathbb{E}[\mathbf{x}\mathbf{x}^\top]$ satisfies

$$\mathbf{R} = \theta \cdot \boldsymbol{\Sigma} \quad \text{for} \quad \theta = \frac{\mathbb{E}[r^2]}{p}, \tag{3}$$

where $r^2 = \|\boldsymbol{\Sigma}^{-1/2}\mathbf{x}\|^2$ is the squared Mahalanobis distance of $\mathbf{x}$ w.r.t. $\boldsymbol{\Sigma}$, whose pdf is given by

$$f_{r^2}(t) = Ct^{p/2-1}g(t), \tag{4}$$

where $C^{-1} = \int_0^\infty t^{p/2-1}g(t)\mathrm{d}t$. Hence pdf of r^2 has one-to-one correspondence with density generator g. From (3), we notice that the scatter matrix $\boldsymbol{\Sigma}$ is proportional to the covariance matrix $\mathbf{R}$ (assuming $\mathbf{R}$ exists). In the MVN case, $\theta = 1$, while for the MVT distribution with density generator as in (1) one obtains $\theta = \nu/(\nu - 2)$ for all $\nu > 2$.

A. Estimate of Scale

As mentioned earlier, Tyler's M-estimator $\hat{\boldsymbol{\Sigma}}$ loses information of the scale η. However, it is yet possible to construct an estimate of η from Tyler's weights. Our Tyler's weights-based estimate (TWE) of scale is defined as the harmonic mean of reciprocal of weights, $1/\hat{w}_i$'s, that is,

$$\hat{\eta}_{\text{TWE}} = \left(\frac{1}{n}\sum_{i=1}^n \hat{w}_i\right)^{-1} = \left(\frac{p}{n}\sum_{i=1}^n [\mathbf{x}_i^\top \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{x}_i]^{-1}\right)^{-1}. \tag{5}$$

One can esily verify that this is a proper scale estimate in the sense that if $\hat{\eta}^*_{\text{TWE}}$ is computed on scaled observations, $\mathbf{x}_i^* = c \cdot \mathbf{x}_i$, $i = 1, \ldots, n$, then $\hat{\eta}^*_{\text{TWE}} = c^2 \cdot \hat{\eta}_{\text{TWE}}$. This follows because Tyler's M-estimator $\hat{\boldsymbol{\Sigma}}$ in (2) with trace constraint is invariant to scaling the data, so $\hat{\boldsymbol{\Sigma}}^* = \hat{\boldsymbol{\Sigma}}$.

The proposed estimate (5) can also be motivated from the following result derived in the high-dimensional random matrix theory (RMT) regime, where $p, n \to \infty$ with $n > p$ and their ratio tending to constant: $p/n \to c \in (0, 1)$. Namely, let $\hat{\boldsymbol{\Sigma}}$ be Tyler's M-estimator in (2) verifying $\mathrm{tr}(\hat{\boldsymbol{\Sigma}}) = p$. Then, it was shown in [11], [13] that $\max_\ell |\eta\hat{w}_\ell - 1| \to 0$ almost surely. The authors in [11] derived this result for the case that data is i.i.d. Gaussian $\mathcal{N}_p(\mathbf{0}, \mathbf{I})$ while [13] extended these results for more general distributions. Thus since $1/\hat{w}_i$ concentrate on η, (5) is a natural robust estimator of scale. Many other robust scale statistics could be constructed from $1/\hat{w}_i$'s, such as the median, trimmed mean, etc. In the next subsection, we illustrate why the proposed harmonic mean (5) is the most natural.

B. Affine Equivariant Estimate of Scatter Matrix

Using the scale estimate $\hat{\eta}_{\text{TWE}}$ and Tyler's M-estimator $\hat{\boldsymbol{\Sigma}}$ (with $\mathrm{tr}(\hat{\boldsymbol{\Sigma}}) = p$), we can form an estimate of the scatter matrix

$$\hat{\boldsymbol{\Sigma}}_{\text{TWE}} = \hat{\eta}_{\text{TWE}} \cdot \hat{\boldsymbol{\Sigma}} \tag{6}$$

referred to as *TWE of scatter matrix*. Thus $\hat{\eta}_{\text{TWE}}$ is scale statistic derived from $\hat{\boldsymbol{\Sigma}}_{\text{TWE}}$ since $\hat{\eta}_{\text{TWE}} = \mathrm{tr}(\hat{\boldsymbol{\Sigma}}_{\text{TWE}})/p$. Equivalently, the trace of $\hat{\boldsymbol{\Sigma}}_{\text{TWE}}$ can be easily computed as the harmonic mean of Tyler's quadratic form:

$$\mathrm{tr}(\hat{\boldsymbol{\Sigma}}_{\text{TWE}}) = \left(\frac{1}{n}\sum_{i=1}^n [\mathbf{x}_i^\top \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{x}_i]^{-1}\right)^{-1}.$$

Recalling (5) we can write (6) in the following more intuitive form:

$$\hat{\boldsymbol{\Sigma}}_{\text{TWE}} = \frac{1}{n}\sum_{i=1}^n \hat{v}_i \mathbf{x}_i\mathbf{x}_i^\top, \tag{7a}$$

$$\hat{v}_i = \frac{\hat{w}_i}{\frac{1}{n}\sum_{\ell=1}^n \hat{w}_\ell} = \frac{[\mathbf{x}_i^\top \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{x}_i]^{-1}}{\frac{1}{n}\sum_{\ell=1}^n [\mathbf{x}_\ell^\top \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{x}_\ell]^{-1}} \tag{7b}$$

where $\hat{v}_1, \ldots, \hat{v}_n$ are normalized Tyler's weights that verify $\frac{1}{n}\sum_{i=1}^n \hat{v}_i = 1$. (7a) and (7b) illustrate that $\hat{\boldsymbol{\Sigma}}_{\text{TWE}}$ is a *weighted sample covariance matrix (SCM)* with weights $\hat{v}_i$.

Finally, we draw the parallel of Tyler's M-estimating equation and our estimator (7). First, note that Tyler's M-estimating (2) verifies broader invariance than just invariance with respect to scaling of the data matrix. Namely, denoting the unit-norm normalized data by $\tilde{\mathbf{x}}_i = \mathbf{x}_i/\|\mathbf{x}_i\|$, $i = 1, \ldots, n$, one can easily verify that the fixed-point equation in (2) can be rewritten as $\hat{\boldsymbol{\Sigma}} = \mathcal{H}(\hat{\boldsymbol{\Sigma}}; \{\tilde{\mathbf{x}}_i\})$, so based on normalized data. Furthermore, since $\mathrm{tr}(\hat{\boldsymbol{\Sigma}}) = p$, one has:

$$\mathrm{tr}(\hat{\boldsymbol{\Sigma}}) = \frac{1}{n}\sum_{i=1}^n \frac{p}{\tilde{\mathbf{x}}_i^\top \hat{\boldsymbol{\Sigma}}^{-1}\tilde{\mathbf{x}}_i}\mathrm{tr}\left(\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^\top\right) = p,$$

or equivalently $\frac{1}{n}\sum_{i=1}^n [\tilde{\mathbf{x}}_i^\top \hat{\boldsymbol{\Sigma}}^{-1}\tilde{\mathbf{x}}_i]^{-1} = 1$ since $\mathrm{tr}(\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^\top) = 1$. It follows that Tyler's M-estimator $\hat{\boldsymbol{\Sigma}}$ with $\mathrm{tr}(\hat{\boldsymbol{\Sigma}}) = p$ is the solution to the following fixed-point equation:

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^n \tilde{w}_i \tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^\top \quad \text{with} \quad \tilde{w}_i = \frac{p\,[\tilde{\mathbf{x}}_i^\top \hat{\boldsymbol{\Sigma}}^{-1}\tilde{\mathbf{x}}_i]^{-1}}{\frac{1}{n}\sum_{\ell=1}^n [\tilde{\mathbf{x}}_\ell^\top \hat{\boldsymbol{\Sigma}}^{-1}\tilde{\mathbf{x}}_\ell]^{-1}}, \tag{8}$$

where $\tilde{w}_i$'s are the normalized Tyler's weights computed on normalized (unit norm) observations $\tilde{\mathbf{x}}_i$'s. Thus, while Tyler's M-estimator $\hat{\boldsymbol{\Sigma}}$ can be interpreted as a weighted SCM based on normalized data $\{\tilde{\mathbf{x}}_i\}$ as shown in (8), TWE of scatter $\hat{\boldsymbol{\Sigma}}_{\text{TWE}}$ in (7) can be viewed as weighted SCM of actual (non-normalized) data $\{\mathbf{x}_i\}$.

It is worthwhile to point out that most robust estimators of scatter are affine equivariant in the sense that an affine transformation on the data $\mathbf{x}_i \mapsto \mathbf{A}\mathbf{x}_i$, $i = 1, \ldots, n$, induces following transformation on the estimate:

$$\hat{\boldsymbol{\Sigma}}(\{\mathbf{A}\mathbf{x}_i\}) = \mathbf{A}\hat{\boldsymbol{\Sigma}}(\{\mathbf{x}_i\})\mathbf{A}^\top,\ \forall \mathbf{A} \in \mathbb{R}^{p\times p} \text{ invertible.} \tag{9}$$

For example, robust Maronna's [17] M-estimators, S-estimators [18], or MM-estimators [19] are affine equivariant. However, Tyler's scatter matrix $\hat{\boldsymbol{\Sigma}}$ is not affine equivariant since (9) only holds up to multiplicative scalar factor as shown in (10). Affine equivariance is desirable since if $\mathbf{x} \sim \mathcal{E}_p(\mathbf{0}, \boldsymbol{\Sigma}, g)$, then $\mathbf{A}\mathbf{x} \sim \mathcal{E}_p(\mathbf{0}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top, g)$. Hence the transformed data shares the same elliptical distribution, but the scatter matrix parameter is mapped to $\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top$. Thus a natural requirement to be imposed on any scatter matrix estimator is that it should verify this same equivariance principle under transformations $\mathbf{x}_i \mapsto \mathbf{A}\mathbf{x}_i$. This is shown next.

Lemma 1: TWE of scatter matrix $\hat{\boldsymbol{\Sigma}}_{\text{TWE}}$ is affine equivariant: $\hat{\boldsymbol{\Sigma}}_{\text{TWE}}(\{\mathbf{A}\mathbf{x}_i\}) = \mathbf{A}\hat{\boldsymbol{\Sigma}}_{\text{TWE}}(\{\mathbf{x}_i\})\mathbf{A}^\top$, $\forall$ invertible $\mathbf{A} \in \mathbb{R}^{p \times p}$.

Proof: It is straightforward to verify that Tyler's M-estimator (with $\text{tr}(\hat{\boldsymbol{\Sigma}}) = p$) is equivariant in the sense that if $\hat{\boldsymbol{\Sigma}}^* = \hat{\boldsymbol{\Sigma}}(\{\mathbf{x}_i^*\})$ denotes Tyler's M-estimator (verifying $\text{tr}(\hat{\boldsymbol{\Sigma}}^*) = p$) computed on data $\mathbf{x}_i^* = \mathbf{A}\mathbf{x}_i$, $i = 1, \ldots, n$, then

$$\hat{\boldsymbol{\Sigma}}^* = c\mathbf{A}\hat{\boldsymbol{\Sigma}}\mathbf{A}^\top, \; c = \frac{p}{\text{tr}(\mathbf{A}\hat{\boldsymbol{\Sigma}}\mathbf{A}^\top)} \tag{10}$$

for all invertible $\mathbf{A} \in \mathbb{R}^{p \times p}$. Now let $\hat{w}_i^*$, $i = 1, \ldots, n$ denote the corresponding Tyler's weights. Then the scale estimate (5) computed as the harmonic mean of reciprocals of weights $1/\hat{w}_i^* = [\mathbf{x}_i^*]^\top(\hat{\boldsymbol{\Sigma}}^*)^{-1}\mathbf{x}_i^*$ is

$$\begin{aligned}\hat{\eta}_{\text{TWE}}^* &= \left(\frac{1}{n}\sum_{i=1}^n \hat{w}_i^*\right)^{-1} = \left(\frac{p}{n}\sum_{i=1}^n [\mathbf{x}_i^\top \mathbf{A}^\top (\hat{\boldsymbol{\Sigma}}^*)^{-1}\mathbf{A}\mathbf{x}_i]^{-1}\right)^{-1} \\ &= c^{-1}\left(\frac{p}{n}\sum_{i=1}^n [\mathbf{x}_i^\top \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{x}_i]^{-1}\right)^{-1} = c^{-1}\hat{\eta}_{\text{TWE}} \end{aligned}\tag{11}$$

where in the 2nd to last identity we simply utilized (10). Thus it follows that

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}_{\text{TWE}}^* &= \hat{\eta}_{\text{TWE}}^* \cdot \hat{\boldsymbol{\Sigma}}^* = c^{-1}\hat{\eta}_{\text{TWE}} \cdot c\mathbf{A}\hat{\boldsymbol{\Sigma}}\mathbf{A}^\top \\ &= \mathbf{A}(\hat{\eta}_{\text{TWE}}\hat{\boldsymbol{\Sigma}})\mathbf{A}^\top = \mathbf{A}\hat{\boldsymbol{\Sigma}}_{\text{TWE}}\mathbf{A}^\top.\end{aligned}$$

□

In the case of large dimensional data, one can also consider a shrinkage Tyler's M-estimator of the scatter matrix as

$$\hat{\boldsymbol{\Sigma}}_{\text{TWE},\beta} = \beta\hat{\boldsymbol{\Sigma}}_{\text{TWE}} + (1-\beta)\hat{\eta}_{\text{TWE}}\mathbf{I}, \tag{12}$$

where the data adaptive shrinkage parameter $\beta \in [0, 1]$ is computed as described in [20, Sect. IV.C]. However, unlike the estimator in [20], the shrinkage TWE in (12) provides an estimator of scatter instead of the shape matrix.

C. An Estimator of Covariance Matrix

If the density generator g (and hence the underlying ES distribution) is specified, then the value of θ in (3) can be determined, and we can use relationship (3) to obtain a covariance matrix estimator as $\hat{\mathbf{R}}_{\text{TWE}} = \theta \cdot \hat{\boldsymbol{\Sigma}}_{\text{TWE}}$. For example, if the data has an MVN distribution, then $\theta = 1$ while $\theta = \nu/(\nu - 2)$ in the case of an MVT distribution with ν d.o.f. However, often the underlying parametric family is known, but the underlying tail parameter, say ν, indexing the density generator is unknown. As is shown in Section III, we can form an estimate of ν, denoted $\hat{\nu}_{\text{TWE}}$, using Tyler's weights. Since $\theta = h(\nu)$ (*cf.* (14) below), a *TWE of covariance matrix* can be computed as $\hat{\mathbf{R}}_{\text{TWE}} = \hat{\theta}_{\text{TWE}} \cdot \hat{\boldsymbol{\Sigma}}_{\text{TWE}}$, where $\hat{\theta}_{\text{TWE}} = h(\hat{\nu}_{\text{TWE}})$. This estimator can be useful e.g., in financial applications, where stock return data often exhibit heavy-tails [21], [22].

III. Estimating the Tail Parameter of ES Distribution

From (3) we can induce the following relationship between the scale parameter $\eta_{\text{cov}} = p^{-1}\text{tr}(\mathbf{R})$ of the covariance matrix

---

Algorithm 1: Distribution Tail Parameter Learning.

Input: Data $\{\mathbf{x}_i\}_{i=1}^n$
Output: Estimated tail parameter ν

1: Compute Tyler's M-estimator $\hat{\boldsymbol{\Sigma}}$ and weights $\hat{w}_i$'s in (2);
2: Compute $\hat{\eta}_{\text{TWE}}$ in (5), and set $\hat{\theta}_{\text{TWE}} = p^{-1}\text{tr}(\mathbf{S})/\hat{\eta}_{\text{TWE}}$;
3: Using (14), estimate ν as $\hat{\nu}_{\text{TWE}} = h^{-1}(\hat{\theta}_{\text{TWE}})$.

---

and scale $\eta = p^{-1}\text{tr}(\boldsymbol{\Sigma})$ of the scatter matrix:

$$\eta_{\text{cov}} = \theta\eta \Leftrightarrow \theta = \eta_{\text{cov}}/\eta. \tag{13}$$

Note that a natural estimate of η_{cov} is $p^{-1}\text{tr}(\mathbf{S})$, where $\mathbf{S} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top$ denotes the sample covariance matrix (SCM). On the other hand, if density generator g is specified up to unknown tail parameter ν, thus indexed by $g_\nu(\cdot)$, then θ in (3) is a following function of the tail parameter ν:

$$\theta = \frac{1}{p}\int_0^\infty t f_{r^2}(t; \nu)\text{d}t \triangleq h(\nu) \tag{14}$$

where the pdf $f_{r^2}(\cdot)$, defined in (4), is one-to-one with $g_\nu(\cdot)$. We do not need numerical integration in most practical cases as often closed-form expression for $h(\nu)$ can be derived. Then, after solving the inverse mapping, $\nu = h^{-1}(\theta) = h^{-1}(\eta_{\text{cov}}/\eta)$, Algorithm 1 offers a unified approach for estimating the tail parameter of an ES distribution:

As an example, if $\mathbf{x}$ follows an MVT distribution with $\nu > 2$ d.o.f., one has that $\theta = h(\nu) = \frac{\nu}{\nu-2}$, which unfolds the relation:

$$\nu = h^{-1}(\theta) = \frac{2\theta}{\theta - 1} \text{ for } \theta > 1. \tag{15}$$

Note that $\nu > 2$ is required for the covariance matrix $\mathbf{R}$ to exist. The obtained estimator $\hat{\nu}_{\text{TWE}}$ is closely related to estimator in [20, Alg. 1], referred to as **OPP estimator** for short. OPP is an iterative approach that iteratively (re-)computes the maximum likelihood estimator (MLE) $\hat{\boldsymbol{\Sigma}}$ of the MVT distribution with ν given by the current estimate of d.o.f. parameter $\nu^{(k)}$. It then computes $\hat{\theta} = \text{tr}(\mathbf{S})/\text{tr}(\hat{\boldsymbol{\Sigma}})$ which provides an update $\nu^{(k+1)} = h^{-1}(\hat{\theta})$ via (15). The algorithm iterates for $k = 0, 1, 2 \ldots$ until convergence, starting from an initial start $\nu^{(0)} = \hat{\nu}_{\text{kurt}}$, where $\hat{\nu}_{\text{kurt}}$ is an estimate of ν based on elliptical kurtosis, proposed in [20], and referred to as **kurtosis estimator**. We also proposed an improved version of OPP estimator in [23], which, however, is impractical for large n and p. ML estimation of ν via the Expectation-Maximization (EM) approach is considered in [24]. This method is unfortunately rather unstable [25].

IV. Simulation Studies

We first consider the case where scatter matrix $\boldsymbol{\Sigma}$ has an autoregressive model (**AR(1)**) structure, $(\boldsymbol{\Sigma})_{ij} = \eta\varrho^{|i-j|}$, where $\eta = \text{tr}(\boldsymbol{\Sigma})/p$ is the scale parameter and ϱ is the correlation parameter, $\varrho \in (-1, 1)$. Since Tyler's M-estimator $\hat{\boldsymbol{\Sigma}}$ is invariant to the data scaling, we can set $\eta = 1$ without favoring any estimator over the other. The number of Monte-Carlo runs is 5,000, and samples are generated from an MVT distribution with different choices of d.o.f. parameter ν.

First, we investigate how the TWE of d.o.f. parameter ν compares against OPP and kurtosis estimators. Fig. 1 displays

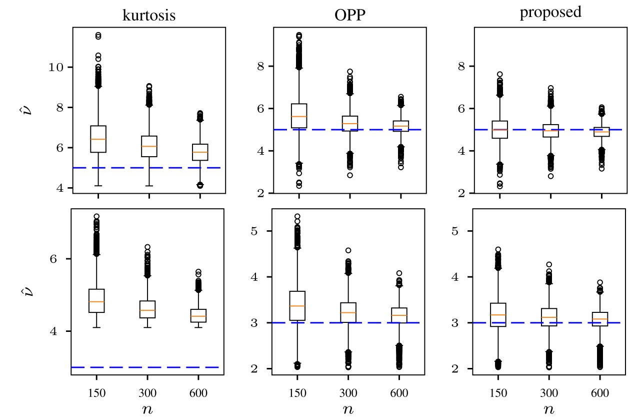

Fig. 1. Boxplots d.o.f. estimates $\hat{\nu}$ as a function of n when $\nu = 5$ (top) and $\nu = 3$ (bottom); $p = 100$, $\varrho = 0.6$.

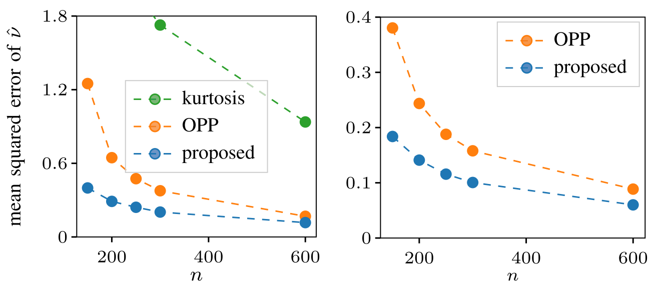

Fig. 2. MSE of d.o.f estimates $\hat{\nu}$ as a function of n when $\nu = 5$ (left panel) and $\nu = 3$ (right panel); $p = 100$, $\varrho = 0.6$.

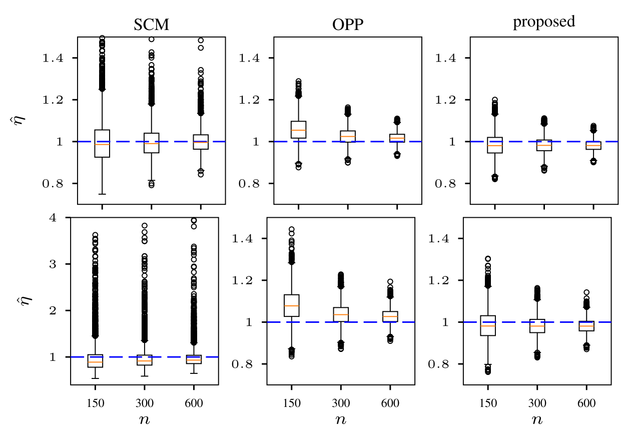

Fig. 3. Estimation errors of scale estimates $\hat{\eta}$ as a function of n when $\nu = 5$ (top) and $\nu = 3$ (bottom); $p = 100$, $\varrho = 0.6$.

the boxplots in the case that $p = 100$ and varying sample lengths when $\nu = 5$ or $\nu = 3$. As can be noted, the proposed TWE attains the best accuracy as well as the smallest variability. Moreover, for $\nu = 5$, its median values are right on the spot. The kurtosis estimator obviously performs poorly when $\nu = 3$ since the 4th-order moment does not exist in this case. Fig. 2 shows the average mean squared error (MSE), $(\hat{\nu} - \nu)^2$, which further illustrates the benefits and high accuracy of the proposed TWE against its competitors.

Fig. 3 displays the boxplots of different estimates of scale η. Here we compare $\hat{\eta}_{\text{TWE}}$ to OPP estimate of scale, defined as $\hat{\eta}_{\text{OPP}} = \text{tr}(\hat{\mathbf{\Sigma}})/p$, where $\hat{\mathbf{\Sigma}}$ is the MLE of scatter based on $\nu = \hat{\nu}_{\text{OPP}}$. We also compare with the scale estimate provided by the SCM, defined as $\hat{\eta}_{\text{cov}} = \text{tr}(\mathbf{S})/p$, but multiplied by $\theta^{-1} = (\nu - 2)/\nu$ to obtain an estimate of η; recall (13). We can notice from Fig. 3 that TWE slightly underestimates the true scale while OPP is overestimating. We also notice that the SCM estimator is clearly unbiased for $\nu = 5$, but has huge variability. Fig. 4 displays the median values of $\hat{\nu}$ for range of ν values when $n = 150$ (and $p = 100$ and $\varrho = 0.6$ as earlier). The proposed TWE estimator significantly outperforms the other estimators for all d.o.f. $\nu \in [3, 8]$.

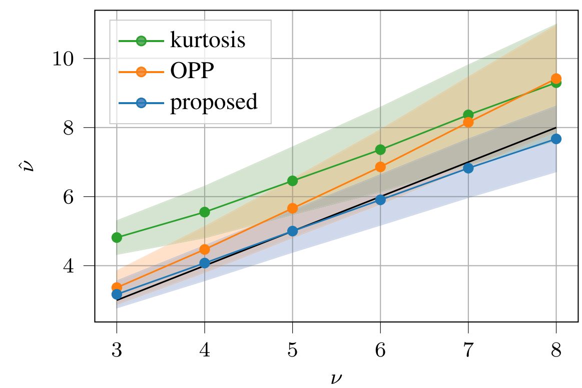

Fig. 4. Estimated values of ν; $p = 100$, $\varrho = 0.6$, $n = 150$. The black line indicates the true value.

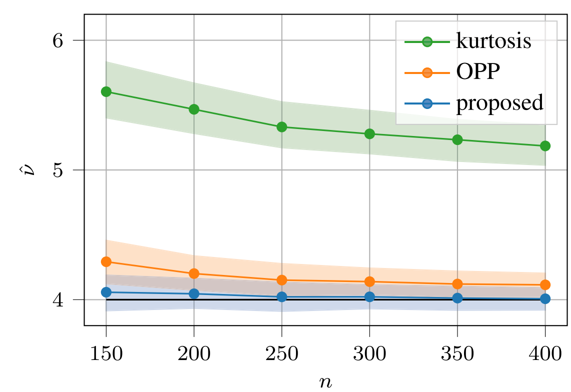

Fig. 5. Estimated values of ν for synthetic stock market data; $p = 100$, $\nu = 4$.

We now consider an example based on stock market data. We generate synthetic data ($p = 100$ assets) with heavy tails following MVT distribution with d.o.f. $\nu = 4$ and covariance matrix as measured from stocks of S&P 500 index. Fig. 5 compares the estimated value of ν versus the number of observations for the following methods: kurtosis estimator, OPP estimator [20], and the proposed estimator, with the latter being clearly superior, illustrating its promising performance for real-world financial data.

V. Concluding Remarks

We proposed a new robust estimator of scale parameter of an elliptical distribution based on the weights from Tyler's M-estimator, which was further used to construct an affine equivariant Tyler's M-estimator. We then proposed a unified framework to estimate the tail parameter of an elliptical distribution. Finally, it should be noted that this method generalizes to complex-valued data in a straightforward manner.

REFERENCES

[1] K.-T. Fang, S. Kotz, and K.-W. Ng, *Symmetric Multivariate and Related Distributions*. London, U.K.: Chapman and Hall, 1990.

[2] 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.

[3] D. Paindaveine, "A canonical definition of shape," *Statist. Probability Lett.*, vol. 78, no. 14, pp. 2240–2247, 2008.

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

[5] Y. 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, Oct. 2014.

[6] F. Pascal, Y. Chitour, and Y. Quek, "Generalized robust shrinkage estimator and its application to STAP detection problem," *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp. 5640–5651, Nov. 2014.

[7] E. Ollila and D. E. Tyler, "Regularized M-estimators of scatter matrix," *IEEE Trans. Signal Process.*, vol. 62, no. 22, pp. 6059–6070, Nov. 2014.

[8] Y. 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, Sep. 2011.

[9] R. Couillet and M. McKay, "Large dimensional analysis and optimization of robust shrinkage covariance matrix estimators," *J. Multivariate Anal.*, vol. 131, pp. 99–120, 2014.

[10] I. Soloveychik and A. Wiesel, "Performance analysis of Tyler's covariance estimator," *IEEE Trans. Signal Process.*, vol. 63, no. 2, pp. 418–426, Jan. 2015.

[11] T. Zhang, X. Cheng, and A. Singer, "Marčenko–Pastur law for Tyler's M-estimator," *J. Multivariate Anal.*, vol. 149, pp. 114–123, 2016.

[12] J. Goes, G. Lerman, and B. Nadler, "Robust sparse covariance estimation by thresholding Tyler's M-estimator," *Ann. Statist.*, vol. 48, no. 1, pp. 86–110, 2020.

[13] E. Romanov, G. Kur, and B. Nadler, "Tyler's and Maronna's M-estimators: Non-asymptotic concentration results," *J. Multivariate Anal.*, vol. 196, 2023, Art. no. 105184.

[14] J. G. Adrover, "Minimax bias-robust estimation of the dispersion matrix of a multivariate distribution," *Ann. Statist.*, vol. 26, no. 6, pp. 2301–2320, 1998.

[15] L. Dümbgen and D. E. Tyler, "On the breakdown properties of some multivariate m-functionals," *Scand. J. Statist.*, vol. 32, no. 2, pp. 247–264, 2005.

[16] D. P. Palomar, R. Zhou, X. Wang, F. Pascal, and E. Ollila, "fitHeavy-Tail: Mean and covariance matrix estimation under heavy tails," 2023, R package version 0.2.0. [Online]. Available: https://CRAN.R-project.org/package=fitHeavyTail

[17] R. A. Maronna, "Robust M-estimators of multivariate location and scatter," *Ann. Statist.*, vol. 5, no. 1, pp. 51–67, 1976.

[18] P. L. Davies, "Asymptotic behaviour of S-estimates of multivariate location parameters and dispersion matrices," *Ann. Statist.*, pp. 1269–1292, 1987.

[19] K. S. Tatsuoka and D. E. Tyler, "On the uniqueness of S-functionals and M-functionals under nonelliptical distributions," *Ann. Statist.*, vol. 28, pp. 1219–1243, 2000.

[20] E. Ollila, D. P. Palomar, and F. Pascal, "Shrinking the eigenvalues of M-estimators of covariance matrix," *IEEE Trans. Signal Process.*, vol. 69, pp. 256–269, 2021.

[21] A. J. McNeil, R. Frey, and P. Embrechts, *Quantitative Risk Management: Concepts, Techniques and Tools-Revised Edition*. Princeton, NJ, USA: Princeton Univ. Press, 2005.

[22] Y. Feng and D. P. Palomar, *A Signal Processing Perspective on Financial Engineering*, vol. 9. Boston, MA, USA: Now Publishers, 2016.

[23] F. Pascal, E. Ollila, and D. P. Palomar, "Improved estimation of the degree of freedom parameter of multivariate t-distribution," in *Proc. 29th Eur. Signal Process. Conf.*, 2021, pp. 860–864.

[24] C. Liu and D. B. Rubin, "ML estimation of the t-distribution using EM and its extensions, ECM and ECME," *Statistica Sinica*, vol. 5, no. 1, pp. 19–39, 1995.

[25] C. Fernandez and M. F. J. Steel, "Multivariate student-t regression models: Pitfalls and inference," *Biometrika*, vol. 86, no. 1, pp. 153–167, 1999.