A review of Tyler's shape matrix and its extensions

Sara Taskinen, Gabriel Frahm, Klaus Nordhausen, and Hannu Oja

Abstract In a seminal paper, Tyler (1987a) suggests an M-estimator for shape, which is now known as Tyler's shape matrix. Tyler's shape matrix is increasingly popular due to its nice statistical properties. It is distribution free within the class of generalized elliptical distributions. Further, under very mild regularity conditions, it is consistent and asymptotically normally distributed after the usual standardization. Tyler's shape matrix is still the subject of active research, e.g, in the signal-processing literature, which discusses structured and regularized shape matrices. In this article, we review Tyler's original shape matrix and some recent developments.

Key words: M-estimator, generalized elliptical distribution, high dimension, robust estimator, regularization

1 Introduction

Maronna (1976) and Huber (1981) propose robust M-estimators for location and scatter of multivariate elliptically distributed data. Since their seminal work, we can find many contributions finding new ways to estimate the location vector and scatter matrix. See Maronna et al. (2018) for a nice overview of robust multivariate methods.

Sara Taskinen and Klaus Nordhausen
Department of Mathematics and Statistics, University of Jyväskylä, Jyväskylä, Finland. e-mail: sara.taskinen@jyu.fi e-mail: klaus.k.nordhausen@jyu.fi

Gabriel Frahm
Department of Mathematics and Statistics, Helmut Schmidt University, Hamburg, Germany. e-mail: frahm@hsu-hh.de

Hannu Oja
Department of Mathematics and Statistics, University of Turku, Turku, Finland. e-mail: hannu.oja@utu.fi

In this work, we focus on the robust M-estimator for shape, introduced by Tyler (1987a). We start by fixing some notation. Consider first a location-scatter model. This means that the p-variate observations $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ are independent copies of

$$\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Omega}\mathbf{e},$$

where $\boldsymbol{\mu} \in \mathbb{R}^p$ is a location vector and $\boldsymbol{\Omega} \in \mathbb{R}^{p\times q}$ is a transformation (or mixing) matrix with $\mathrm{rk}(\boldsymbol{\Omega}) = p$. Hence, we have that $p \leq q$ and the symmetric positive-definite matrix $\boldsymbol{\Sigma} := \boldsymbol{\Omega}\boldsymbol{\Omega}^\top \in \mathbb{R}^{p\times p}$ is referred to as the scatter matrix. Without loss of generality, we may choose the decomposition $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\Sigma}^{\frac{1}{2}}$, where $\boldsymbol{\Sigma}^{\frac{1}{2}}$ is the unique symmetric root of $\boldsymbol{\Sigma} > 0$.

Different multivariate models are obtained by making specific assumptions about the q-variate random vector $\mathbf{e}$ (Nordhausen & Oja, 2018b). For example, it is typically assumed that $p = q$ and that $\mathbf{e}$ has a spherically symmetric absolutely continuous distribution on $\mathbb{R}^p$, i.e., the density function of $\mathbf{e}$ is of the form $f(\mathbf{e}) = \exp\{-\rho(||\mathbf{e}||)\}$ for some function $\rho\colon \mathbb{R}_0^+ \to \mathbb{R}$, where $\|\cdot\|$ denotes the Euclidean norm (Fang et al., 1990). Then, we can decompose $\mathbf{e}$ into a radial part and an angular part by $\mathbf{e} = r\mathbf{u}$, where the modulus, i.e., the radius, $r = \|\mathbf{e}\| > 0$ and the direction $\mathbf{u} = \|\mathbf{e}\|^{-1}\mathbf{e}$ are stochastically independent with $\mathbf{u}$ being uniformly distributed on the unit hypersphere in $\mathbb{R}^p$. The density of the modulus is proportional to $r^{p-1}\exp\{-\rho(r)\}$.

For all $\tau > 0$ we have that $\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Omega}r\mathbf{u} = \boldsymbol{\mu} + \boldsymbol{\Upsilon}s\mathbf{u}$ with $\boldsymbol{\Upsilon} := \boldsymbol{\Omega}/\tau$ and $s := \tau r$. Hence, the scatter matrix of $\mathbf{x}$ is defined only up to scale. To fix $\boldsymbol{\Sigma}$ we could assume that $\mathbf{E}(r^2) = p$ or $\mathbf{Med}(r^2) = \chi^2_{p,0.5}$, where $\chi^2_{p,0.5}$ is the median of the χ^2-distribution with p degrees of freedom. The first assumption requires that the second moment of r is finite, whereas the second assumption does not require any moment condition on r at all. If the first assumption is satisfied, we have that $\mathbf{COV}(\mathbf{e}) = \mathbf{I}_p$, where $\mathbf{I}_p$ is the $p \times p$ identity matrix, and $\mathbf{COV}(\mathbf{x}) = \boldsymbol{\Sigma}$. However, it is more common to impose the scaling condition

$$\mathbf{E}\big(\varphi(r^2)\big) = p \tag{1}$$

with $\varphi\big(r^2\big) := w\big(r^2\big)\, r^2$, where w is a real-valued partial function on $\mathbb{R}_0^+$.[1] In fact, this is typically done both in M-estimation and in ML-estimation of scatter (Tyler, 1982; Frahm et al., 2020). The chosen weight function w is considered appropriate if and only if there exists no scaling constant $\tau \neq 1$ such that $\mathbf{E}\big(\varphi\big((\tau r)^2\big)\big) = p$.[2] In the special case of $w : r^2 \mapsto 1$, we obtain the simple scaling condition $\mathbf{E}(r^2) = p$ mentioned above.

Under the above mentioned assumptions, $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ are independent copies of the p-variate random vector $\mathbf{x}$, which follows an elliptically symmetric distribution with density function

$$f(\mathbf{x}) = \det(\boldsymbol{\Sigma})^{-\frac{1}{2}} g\big((\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\big).$$

[1] A partial function $f : D \to C$ is a function from a subset of D to C.

[2] See Frahm (2022) for a detailed explanation.

The function $g : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ is referred to as the density generator of $\mathbf{x}$. Given that the first moment of r is finite, the location vector $\boldsymbol{\mu}$ is the mean vector of $\mathbf{x}$, and if the second moment of r is finite, $\mathbf{COV}(\mathbf{x}) = \mathbf{E}(r^2)/p \cdot \boldsymbol{\Sigma}$ is the covariance matrix of $\mathbf{x}$.

If we allow r to be negative and to depend on $\mathbf{u}$, then $\mathbf{x}$ is generalized elliptically distributed (Frahm & Jaekel, 2010). It is worth noting that, in this case, we can no longer assume that $p = q$ without loss of generality. A particular generalized elliptical distribution, which will be of interest later on, is obtained by setting $\boldsymbol{\mu} = \mathbf{0}$ and $r = \|\boldsymbol{\Omega}\mathbf{u}\|^{-1}$ with $p = q$. The random vector $\mathbf{x} = \boldsymbol{\Omega}\mathbf{u}/\|\boldsymbol{\Omega}\mathbf{u}\|$ follows an angular central Gaussian distribution on the sphere (Tyler, 1987b). In the bivariate case, i.e., $p = 2$, the angular central Gaussian distribution turns into the wrapped Cauchy distribution after angle doubling (Kent & Tyler, 1988).

The last location-scatter model relevant later on is the so-called independent component model where it is assumed that the components of $\mathbf{e}$ are mutually independent. In independent component analysis the goal is to estimate $\mathbf{e}$ based on $\mathbf{x}$ alone (for an overview see for example Nordhausen & Oja, 2018a). If not stated otherwise, in the following we will assume that $\mathbf{x}$ follows an elliptically symmetric distribution.

The scatter matrix $\boldsymbol{\Sigma}$ can be written as $\boldsymbol{\Sigma} = \sigma^2\mathbf{V}$, where $\sigma^2 = \sigma^2(\boldsymbol{\Sigma})$ represents the scale of $\boldsymbol{\Sigma}$. A scale function $\sigma^2(\cdot)$ is such that $\sigma^2(\mathbf{I}_p) = 1$ and $\sigma^2(\tau^2\boldsymbol{\Sigma}) = \tau^2\sigma^2(\boldsymbol{\Sigma})$ for all $\tau > 0$. Further, the matrix $\mathbf{V} = \boldsymbol{\Sigma}/\sigma^2(\boldsymbol{\Sigma})$ is the unique shape matrix associated with $\boldsymbol{\Sigma}$. Classical choices of $\sigma^2(\boldsymbol{\Sigma})$ are $\boldsymbol{\Sigma}_{11}$ (Hettmansperger & Randles, 2002; Hallin & Paindaveine, 2006; Hallin et al., 2006), $\mathrm{tr}(\boldsymbol{\Sigma})/p$ (Tyler, 1987a; Dümbgen, 1998; Frahm & Jaekel, 2015; Taskinen & Oja, 2016) and $\det(\boldsymbol{\Sigma})^{1/p}$ (Tatsuoka & Tyler, 2000; Dümbgen & Tyler, 2005; Salibián-Barrera et al., 2006; Taskinen et al., 2006; Paindaveine, 2008).

Note that $\mathrm{tr}(\boldsymbol{\Sigma})/p$ and $\det(\boldsymbol{\Sigma})^{1/p}$ correspond to the arithmetic and geometric means of the eigenvalues of $\boldsymbol{\Sigma}$, respectively. The use of $\det(\boldsymbol{\Sigma})^{1/p}$ as a scale function yields a canonical definition of shape, meaning that the scale and shape estimators are asymptotically independent if the data are elliptically distributed (Paindaveine, 2008). The scale describes the "size," whereas the shape describes the "orientation" of an elliptical distribution and it is well-known that several multivariate methods, such as principal component analysis, canonical correlation analysis, and multivariate regression, require the shape matrix only (Croux & Haesbroeck, 2000; Taskinen et al., 2006; Salibián-Barrera et al., 2006).

In the robust-statistics literature, several functionals for multivariate distributions are proposed. Let $\mathbf{x}$ be a p-variate random vector with cumulative distribution function $F_{\mathbf{x}}$. Then a functional $\boldsymbol{\mu}(F_{\mathbf{x}}) \in \mathbb{R}^p$ is said to be a location vector if it is affine equivariant in the sense that $\boldsymbol{\mu}(F_{\mathbf{Ax+b}}) = \mathbf{A}\boldsymbol{\mu}(F_{\mathbf{x}}) + \mathbf{b}$ for any nonsingular matrix $\mathbf{A} \in \mathbb{R}^{p\times p}$ and vector $\mathbf{b} \in \mathbb{R}^p$. A symmetric positive-definite functional $\mathbf{S}(F_{\mathbf{x}}) \in \mathbb{R}^{p\times p}$ is called a scatter matrix if $\mathbf{S}(F_{\mathbf{Ax+b}}) = \mathbf{A}\mathbf{S}(F_{\mathbf{x}})\mathbf{A}^\top$. Further, a symmetric positive-definite functional $\mathbf{V}(F_{\mathbf{x}}) \in \mathbb{R}^{p\times p}$ is referred to as a shape matrix if $\mathbf{V}(F_{\mathbf{x}}) = \mathbf{S}(F_{\mathbf{x}})/\sigma^2(\mathbf{S}(F_{\mathbf{x}}))$ and thus

$$\mathbf{V}(F_{\mathbf{Ax+b}}) = \frac{\mathbf{A}\mathbf{V}(F_{\mathbf{x}})\mathbf{A}^\top}{\sigma^2(\mathbf{A}\mathbf{V}(F_{\mathbf{x}})\mathbf{A}^\top)}.$$

Hence, in general, a shape matrix is not affine equivariant and $\mathbf{V}(F_{\mathbf{Ax+b}})$ even is not proportional to $\mathbf{AV}(F_{\mathbf{x}})\mathbf{A}^\top$. However, if we use the canonical scale function $\det(\boldsymbol{\Sigma})^{1/p}$, we have that

$$\mathbf{V}(F_{\mathbf{Ax+b}}) = \frac{\mathbf{AV}(F_{\mathbf{x}})\mathbf{A}^\top}{\sigma^2(\mathbf{AA}^\top)}.$$

Thus, at least for $\sigma^2(\mathbf{AA}^\top) = 1$, i.e., if not the scale, but only the shape of the distribution of $\mathbf{x}$ is affected by the transformation $\mathbf{A}$, the canonical shape matrix remains equivariant (Frahm, 2009).

If the distribution of $\mathbf{x}$ is elliptically symmetric, then $\boldsymbol{\mu}(F_{\mathbf{x}}) = \boldsymbol{\mu}$. This means that all location vectors correspond to the same population quantity $\boldsymbol{\mu}$. By contrast, all scatter matrices are related to each other by $\mathbf{S}(F_{\mathbf{x}}) = \sigma^2(\mathbf{S}(F_{\mathbf{x}}))\mathbf{V}$, where $\mathbf{V}$ is the (unique) shape matrix of $\mathbf{x}$. Put another way, a scatter matrix is always a multiple of the shape matrix $\mathbf{V}$. Finally, if the functionals $\boldsymbol{\mu}(\cdot)$, $\mathbf{S}(\cdot)$, and $\mathbf{V}(\cdot)$ are applied to an empirical distribution function $\hat{F}_{\mathbf{x}}$, i.e., the empirical distribution of a random sample $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$, we obtain the corresponding estimators, which we denote by $\hat{\boldsymbol{\mu}} = \boldsymbol{\mu}(\hat{F}_{\mathbf{x}})$, $\hat{\mathbf{S}} = \mathbf{S}(\hat{F}_{\mathbf{x}})$, and $\hat{\mathbf{V}} = \mathbf{V}(\hat{F}_{\mathbf{x}})$, respectively.

As mentioned above, several multivariate methods can be based on shape matrices only. Such matrices can be easily defined by normalizing any scatter matrix with a scale parameter. On the other hand, sometimes shape matrices arise naturally as a result of some estimation procedure. In this review we discuss Tyler's shape matrix, proposed in the seminal paper by Tyler (1987a), which was initially motivated via estimating equations utilizing spatial sign scores. Recall that spatial sign scores are defined as $\mathbf{U}(\mathbf{x}) = \mathbf{x}/||\mathbf{x}||$, for $\mathbf{x} \neq \mathbf{0}$, and $\mathbf{U}(\mathbf{0}) = \mathbf{0}$ (Möttönen & Oja, 1995). We define Tyler's shape matrix and review its statistical properties in Section 2. Section 3 is devoted to some recent extensions of Tyler's shape matrix and the paper is concluded with some discussion on Section 4.

2 Definition and statistical properties

Assume that $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ with $n > p$ is a random sample from a centered p-variate elliptical distribution, i.e., $\boldsymbol{\mu} = \mathbf{0}$. Further, suppose that r has no atom at $\mathbf{0}$, which means that $\mathrm{P}(r = 0) = 0$. In Tyler (1982, 1983) tests for sphericity and related shape estimators based on Huber's M-estimators were considered. It was noted that it is possible to use in the M-estimation procedure a weight function that yields a distribution-free test and estimate under the elliptical model. This served as a motivation in Tyler (1987a) to propose a shape matrix estimator $\hat{\mathbf{V}}$ as a solution of

$$\hat{\mathbf{V}} = \frac{p}{n} \sum_{i=1}^{n} \frac{\mathbf{x}_i \mathbf{x}_i^\top}{\mathbf{x}_i^\top \hat{\mathbf{V}}^{-1} \mathbf{x}_i}. \tag{2}$$

Tyler (1987a) considers the solution of (2) an M-estimator for scatter, since it can be written as

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n} w\big(r_i^2\big)\mathbf{x}_i\mathbf{x}_i^\top$$

with $r_i^2 = \mathbf{x}_i^\top\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{x}_i$ for $i = 1, 2, \ldots, n$, where the weight function w is $r^2 \mapsto p/r^2$. However, we prefer to call it Tyler's shape matrix. This term is commonly used in the literature, too. The shape matrix can also be seen as the limit of two popular M-estimators for scatter, namely Huber's M-estimator and the ML-estimator under the assumption that the data have a multivariate t-distribution.

More precisely, Huber's weight function is

$$w : r^2 \longmapsto \begin{cases} \gamma\,, & r^2 < \lambda \\ \gamma\lambda/r^2\,, & r^2 \geq \lambda\,, \end{cases}$$

where the parameters $\gamma, \lambda > 0$ are such that $\mathbf{E}\big(\varphi(\chi_p^2)\big) = p$. Now, Tyler's weight function occurs for $\lambda \searrow 0$, i.e., λ approaching zero form above. Alternatively, we obtain Tyler's weight function by setting $\nu = 0$ in the Student-type weight function $r^2 \mapsto (p+\nu)/(r^2+\nu)$ or $\alpha = 1$ in the power M-weight function $r^2 \mapsto (r^2/p)^{-\alpha}$ proposed by Frahm et al. (2020).

Another way to write down the estimation equation in (2) is via spatial sign scores, which are defined in Möttönen & Oja (1995). Then Tyler's shape matrix $\hat{\mathbf{V}}$ solves

$$\frac{p}{n}\sum_{i=1}^{n}\frac{\hat{\mathbf{V}}^{-\frac{1}{2}}\mathbf{x}_i\mathbf{x}_i^\top\hat{\mathbf{V}}^{-\frac{1}{2}}}{||\hat{\mathbf{V}}^{-\frac{1}{2}}\mathbf{x}_i||^2} = \frac{p}{n}\sum_{i=1}^{n}\mathbf{U}(\mathbf{z}_i)\mathbf{U}(\mathbf{z}_i)^\top = \mathbf{I}_p$$

with $\mathbf{z}_i := \hat{\mathbf{V}}^{-\frac{1}{2}}\mathbf{x}_i$ for $i = 1, 2, ..., n$ and $\mathbf{U}(\mathbf{z}_i) := \|\mathbf{z}_i\|^{-1}\mathbf{z}_i$. This means that $\hat{\mathbf{V}}$ is chosen such that the spatial signs of the transformed observations $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n$ are, approximately, uniformly distributed on the unit hypersphere.

In any case, (2) can be re-written, equivalently, as

$$\hat{\mathbf{V}} = \frac{p}{n}\sum_{i=1}^{n}\frac{r_i^2\mathbf{V}^{\frac{1}{2}}\mathbf{u}_i\mathbf{u}_i^\top(\mathbf{V}^{\frac{1}{2}})^\top}{r_i^2\mathbf{u}_i^\top(\mathbf{V}^{\frac{1}{2}})^\top\hat{\mathbf{V}}^{-1}\mathbf{V}^{\frac{1}{2}}\mathbf{u}_i} = \frac{p}{n}\sum_{i=1}^{n}\frac{\mathbf{V}^{\frac{1}{2}}\mathbf{u}_i\mathbf{u}_i^\top(\mathbf{V}^{\frac{1}{2}})^\top}{\mathbf{u}_i^\top(\mathbf{V}^{\frac{1}{2}})^\top\hat{\mathbf{V}}^{-1}\mathbf{V}^{\frac{1}{2}}\mathbf{u}_i},$$

which means that the sample observations $r_1, r_2, \ldots, r_n$ of the modulus r have no impact on Tyler's shape matrix at all. This holds true even if some r_i becomes negative, since r_i^2 does not depend on the sign of r_i, and also if r_i depends on $\mathbf{u}_i$. Hence, Tyler's shape matrix is distribution free if the data are generalized elliptically distributed—provided that r has no atom at $\mathbf{0}$ and we know the location vector $\boldsymbol{\mu}$ (Frahm & Jaekel, 2015). Here, we have chosen $\mathbf{V}^{\frac{1}{2}}$ as a transformation matrix. Indeed, the decomposition of $\mathbf{V}$, i.e., the precise meaning of the $p \times p$ matrix $\mathbf{U}$ in $\mathbf{V} = \mathbf{U}\mathbf{U}^\top$, is *not* arbitrary if r depends on $\mathbf{u}$, but our arguments still remain valid if we choose any other decomposition of $\mathbf{V}$.

Originally, conditions for the existence of Tyler's shape matrix were listed in Tyler (1987a) and it was shown that the matrix is unique up to a positive scaling constant. In Tyler (1987a) the shape matrix was chosen so that $\mathrm{tr}(\hat{\mathbf{V}}) = p$ and in Tatsuoka &

Tyler (2000) $\det(\hat{\mathbf{V}}) = 1$ was used. We use here the first option. Tyler's shape matrix can be computed simply by starting with an initial value, e.g., $\hat{\mathbf{V}}_0 = \mathbf{I}_p$, and then iterating

$$\mathbf{z}_i = \hat{\mathbf{V}}_{k-1}^{-\frac{1}{2}} \mathbf{x}_i,$$
$$\hat{\mathbf{V}}_k \leftarrow \hat{\mathbf{V}}_{k-1}^{\frac{1}{2}} \frac{p}{n} \sum_{i=1}^{n} \mathbf{U}(\mathbf{z}_i)\mathbf{U}(\mathbf{z}_i)^\top \hat{\mathbf{V}}_{k-1}^{\frac{1}{2}},$$

until convergence. The scale can be fixed either at each iteration step or in the end so that $\mathrm{tr}(\hat{\mathbf{V}}) = p$. In Tyler (1987a) weak conditions for the convergence are given. See also Kent & Tyler (1988) for the existence of the solution under general distributions.

Recently, in Wiesel (2012) a new viewpoint for the investigation of covariance matrices was developed. In this framework covariance matrices can be seen as elements of the Riemannian manifold of symmetric positive definite matrices which can also be used to study Tyler's shape matrix. The use of the concept of geodesic convexity provides then a new set of tools to prove existence and uniqueness of Tyler's shape matrix. Dümbgen & Tyler (2016) give a very detailed treatment of the geodesic approach to M-estimation of scatter in general and to Tyler's shape matrix in particular. Another advantage of this framework is the development of fast Newton-Raphson type algorithms for Tyler's shape matrix (Dümbgen et al., 2016; Dümbgen & Tyler, 2016) which are from a computational point of view more efficient than the fixed point algorithm mentioned above. Franks & Moitra (2020) show the connection between Tyler's shape matrix and operator scaling. This connection is then used to derive non-asymptotic bounds and to show that the iterative algorithm from above converges in polynomially many steps. Other results concerning non-asymptotic performance are given in Soloveychik & Wiesel (2015).

Now, consider the limiting behavior of Tyler's shape matrix, more precisely, the consistency of $\hat{\mathbf{V}}$ and the asymptotic distribution of $\sqrt{n}\big(\hat{\mathbf{V}} - \mathbf{V}\big)$, where $\hat{\mathbf{V}}$ is based on a random sample $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \sim \mathbf{x}$. We need not require that $\mathbf{x}$ is elliptically distributed. The matrix $\mathbf{V}$ just represents a solution of

$$\mathbf{V} = p\mathbf{E}\left(\frac{\mathbf{x}\mathbf{x}^\top}{\mathbf{x}^\top\mathbf{V}^{-1}\mathbf{x}}\right).$$

This solution exists and is unique—up to scale—if the distribution of $\mathbf{x}$ is continuous (Tyler, 1987a). Further, in this case, $\hat{\mathbf{V}}$ is strongly consistent, i.e., it converges almost surely to $\mathbf{V}$. In order to prove that $\sqrt{n}\big(\hat{\mathbf{V}} - \mathbf{V}\big) \to N_{p\times p}(\mathbf{0}, \mathbf{C})$ as $n \to \infty$, Tyler (1987a) applies the normalization $\hat{\mathbf{V}}_0 = p\hat{\mathbf{V}}/\mathrm{tr}(\mathbf{V}^{-1}\hat{\mathbf{V}})$. The asymptotic covariance matrix of $\sqrt{n}\big(\hat{\mathbf{V}}_0 - \mathbf{V}\big)$ is quite complicated (Tyler, 1987a, Theorem 3.2).

If $\mathbf{x}$ is elliptically distributed, the asymptotic covariance matrix of $\sqrt{n}\big(\hat{\mathbf{V}}_0 - \mathbf{V}\big)$ simplifies, essentially. More precisely, it holds that

$$\mathbf{C} = \frac{p+2}{p}\big(\mathbf{I}_{p^2} + \mathbf{K}_{p^2}\big)\big(\mathbf{V}\otimes\mathbf{V}\big) + \frac{2(p+2)}{p^2}\,\mathrm{vec}(\mathbf{V})\,\mathrm{vec}(\mathbf{V})^\top,$$

where $\mathbf{I}_{p^2}$ is the $p^2 \times p^2$ identity matrix, $\mathbf{K}_{p^2}$ is the $p^2 \times p^2$ commutation matrix, and $\mathrm{vec}(\mathbf{V})$ is the p^2-variate vector obtained by stacking the columns of $\mathbf{V}$ on top of each other. Frahm (2009, Corollary 1) shows that we obtain the same asymptotic covariance matrix for $\sqrt{n}(\hat{\mathbf{V}} - \mathbf{V})$ when choosing the canonical scale function, i.e., requiring that $\det(\hat{\mathbf{V}}) = 1$.

It can be seen that $\hat{\mathbf{V}}_0$ is affine equivariant. However, in general, this is no longer true if we choose another normalization of $\hat{\mathbf{V}}$. The chosen scale function has an essential impact on the asymptotic covariance matrix. More precisely, we have that

$$\mathbf{C} = \frac{p+2}{p}\psi\left(\mathbf{I}_{p^2} + \mathbf{K}_{p^2}\right)\left(\mathbf{V} \otimes \mathbf{V}\right)\psi^\top$$

with $\psi := \mathbf{I}_{p^2} - \mathrm{vec}(\mathbf{V})\mathcal{J}_{\sigma^2}$, where $\mathcal{J}_{\sigma^2}$ is the Jabobian of the scale function σ^2 (Frahm, 2009; Frahm & Jaekel, 2010; Frahm et al., 2020). See also Sirkiä et al. (2009) and Taskinen & Oja (2016), among others, for the limiting distributions of $\sqrt{n}(\hat{\mathbf{V}} - \mathbf{V})$ under different choices of σ^2. In any case, since Tyler's shape matrix is distribution free within the class of generalized elliptical distributions, the asymptotic covariance matrix never depends on the distribution of the generating variate r. Further, the breakdown point of Tyler's shape matrix is between $1/(p+1)$ and $1/p$ (Yohai & Maronna, 1990; Dümbgen & Tyler, 2005). In Adrover (1998) the Tyler's shape matrix is shown to be minimax bias-robust.

Tyler (1987a) points out that his shape matrix is the "most robust" estimator for the shape matrix of an absolutely continuous elliptical population. More precisely, let h be a real-valued, differentiable, and scale invariant function of $\mathbf{\Sigma} > 0$. That is, we have that $h(\alpha\mathbf{\Sigma}) = h(\mathbf{\Sigma})$ for all $\alpha > 0$ and $\mathbf{\Sigma} > 0$. Consider some parameter $\theta = h(\mathbf{\Sigma})$ and some estimator $\hat{\theta} = h(\hat{\mathbf{\Sigma}})$. It is clear that we can substitute $\mathbf{\Sigma}$ with $\mathbf{V}$ and $\hat{\mathbf{\Sigma}}$ with $\hat{\mathbf{V}}$. Now, Tyler's shape matrix minimizes the maximum asymptotic variance of $\hat{\theta} = h(\hat{\mathbf{V}})$ among all consistent estimators $\hat{\mathbf{V}}$ such that $\sqrt{n}(\hat{\mathbf{V}} - \mathbf{V}) \to N_{p\times p}(\mathbf{0}, \mathbf{C})$.

Tyler's shape matrix is usually introduced as a general M-estimator of shape, however it can also be derived as the ML-estimator for $\mathbf{\Sigma}$ under the angular central Gaussian distribution, as shown in Tyler (1987b). Later Ollila & Tyler (2012) showed the similar result under more general model of elliptical distributions of proportional covariance matrices. See also Gini & Greco (2002); Conte et al. (2002).

Above, we assumed that the location vector of the elliptical distribution is known. However, Tyler (1987a) considers also the case in which the location is unknown. One can, for example, center the observations using any $\sqrt{n}$-consistent location estimate before computing the shape matrix. The asymptotic properties of the resulting shape matrix estimate will remain the same. Tyler (1987a) also mentions a possibility of estimating the location vector and shape matrix simultaneously in a similar fashion as in Maronna (1976); Huber (1981) and recognizes the limitations of such an approach. We will discuss the simultaneous estimation in Section 3 along with other extensions of Tyler's shape matrix.

To conclude this section, note that Paindaveine & Van Bever (2019) introduce the concept of Tyler shape depth which can be used to order shape matrices. The deepest shape matrix is then related to the definition of Tyler's shape matrix.

3 Extensions

In the exposition above Tyler's shape matrix was considered for real data observations with known location and for data without missing values. It was also assumed that the shape matrix does not follow any special structure and that the sample size n is larger than the dimension p. All the issues listed above have been recently addressed in the literature and in the following we will give an overview of the approaches that tackle these issues.

As custom in statistics we will continue focusing on real valued data. Especially in the signal processing community the theory is however often developed considering complex-valued data and most of the methods described below can also be applied in such a context. The interested reader is referred for example to Kent (1997); Gini & Greco (2002); Conte et al. (2002); Pascal et al. (2008); Ollila & Tyler (2012); Ollila et al. (2012), and references therein.

3.1 Joint estimation of location and Tyler's shape matrix - the Hettmansperger-Randles estimators

Hettmansperger & Randles (2002) tackle the problem of simultaneous estimation of location vector and shape matrix utilizing spatial sign scores. Write now $\mathbf{z}_i = \hat{\mathbf{V}}^{-\frac{1}{2}}(\mathbf{x}_i - \hat{\boldsymbol{\mu}})$, $i = 1, \ldots, n$, for transformed observations. Then the Hettmansperger-Randles (HR) estimators of location vector and shape matrix, $\hat{\boldsymbol{\mu}}$ and $\hat{\mathbf{V}}$, solve

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{U}(\mathbf{z}_i) = \mathbf{0} \quad \text{and} \quad \frac{p}{n}\sum_{i=1}^{n}\mathbf{U}(\mathbf{z}_i)\mathbf{U}(\mathbf{z}_i)^\top = \mathbf{I}_p, \tag{3}$$

and $\hat{\mathbf{V}}$ is standardised so that $\mathrm{tr}(\hat{\mathbf{V}}) = p$ (for example). The resulting location vector estimate is known as the transformation-retransformation (TR) spatial median (Chakraborty et al., 1998) and the shape matrix is the Tyler's shape matrix with respect to the TR spatial median. Notice that the classical spatial median that solves $n^{-1}\sum_{i=1}^{n}\mathbf{U}(\mathbf{x}_i - \hat{\boldsymbol{\mu}})$ is only rotation equivariant whereas the TR spatial median is affine equivariant. For the robustness properties and limiting distributions of HR estimates, see Hettmansperger & Randles (2002); Möttönen et al. (2010); Oja (2010).

HR estimates are easy to compute as estimating equations in (3) yield to following iteration steps

$$\begin{aligned}
\mathbf{z}_i &= \hat{\mathbf{V}}_{k-1}^{-\frac{1}{2}}(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{k-1}),\\
\hat{\boldsymbol{\mu}}_k &\leftarrow \hat{\boldsymbol{\mu}}_{k-1} + \frac{\hat{\mathbf{V}}_{k-1}^{\frac{1}{2}}\sum_{i=1}^{n}\mathbf{U}(\mathbf{z}_i)}{\sum_{i=1}^{n}||\mathbf{z}_i||^{-1}},\\
\hat{\mathbf{V}}_k &\leftarrow \hat{\mathbf{V}}_{k-1}^{\frac{1}{2}}\,\frac{p}{n}\sum_{i=1}^{n}\mathbf{U}(\mathbf{z}_i)\mathbf{U}(\mathbf{z}_i)^\top\,\hat{\mathbf{V}}_{k-1}^{\frac{1}{2}}.
\end{aligned}$$

See also Hettmansperger & Randles (2002) for computation of HR estimates. Unfortunately, as far as we know, there is no proof for the convergence of the above algorithm. Also, as Tyler (1987a) already pointed out, the existence and uniqueness of the HR estimates remains an open question as the estimates do not satisfy the conditions that guarantee the existence and uniqueness of simultaneous M-estimates (Maronna, 1976; Huber, 1981). Motivated by this, Taskinen & Oja (2016) proposed k-step HR estimators for location and shape, that is, one starts with initial $\sqrt{n}$-consistent estimates $\hat{\boldsymbol{\mu}}_0$ and $\hat{\mathbf{V}}_0$ and repeats the above iteration steps k times. Resulting estimates are affine equivariant if the initial estimates are affine equivariant. The limiting distributions depend on the limiting distributions of the initial pair of estimates and those of HR estimates. The larger the k, the more similar are the distributions to those of the HR estimates (Taskinen & Oja, 2016). For the robustness properties of k-step estimates, see Croux et al. (2010); Taskinen & Oja (2016).

3.2 The symmetrized variant of Tyler's shape matrix - Dümbgen's estimator

Tyler (1987a) assumes that the location center is known or given. Dümbgen (1998) avoids this assumption and proposes a symmetrised version of the Tyler's shape matrix. Write now $\mathbf{z}_i = \hat{\mathbf{V}}^{-\frac{1}{2}}\mathbf{x}_i$, $i = 1, \ldots, n$. Dümbgen's shape matrix $\hat{\mathbf{V}}$ is then chosen to solve

$$\begin{aligned}\frac{1}{p}\mathbf{I}_p &= \binom{n}{2}^{-1}\sum\sum_{i<j}\frac{\hat{\mathbf{V}}^{-\frac{1}{2}}(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^\top\hat{\mathbf{V}}^{-\frac{1}{2}}}{||\hat{\mathbf{V}}^{-\frac{1}{2}}(\mathbf{x}_i - \mathbf{x}_j)||^2} \\ &= \binom{n}{2}^{-1}\sum\sum_{i<j}\mathbf{U}(\mathbf{z}_i - \mathbf{z}_j)\mathbf{U}(\mathbf{z}_i - \mathbf{z}_j)^\top\end{aligned}$$

and standardised so that $\mathrm{tr}(\hat{\mathbf{V}}) = p$ (for example). This shape matrix is thus Tyler's shape matrix computed on pairwise differences.

Statistical properties of Dümbgen's shape matrix are studied in detail in Dümbgen (1998); Dümbgen & Tyler (2005); Rublik (2021), and later in Sirkiä et al. (2007); Dümbgen et al. (2015) under a framework of symmetrized M-estimators of scatter. The shape matrix obtained using pairwise differences is highly efficient (under elliptical model). It also possesses the so-called joint (and block) independence properties which means that the matrix is a (block) diagonal matrix if the components of $\mathbf{x}$ are mutually (block) independent (Nordhausen & Tyler, 2015). The joint independence property is rare among scatter and shape functionals and needed for example in the independent component model. The use of symmetrized scatter functionals for independent component analysis is discussed in Oja et al. (2006). Other multivariate methods which require the joint or block independence property are discussed in Nordhausen & Tyler (2015).

Dümbgen's shape matrix can be computed using the algorithm proposed in Tyler (1987a), that is, one can simply start with an initial value, e.g., $\hat{\mathbf{V}}_0 = \mathbf{I}_p$, and then iterate

$$\mathbf{z}_i = \hat{\mathbf{V}}_{k-1}^{-\frac{1}{2}}\mathbf{x}_i,$$
$$\hat{\mathbf{V}}_k \leftarrow \hat{\mathbf{V}}_{k-1}^{\frac{1}{2}} \binom{n}{2}^{-1} \sum\sum_{i<j} \mathbf{U}(\mathbf{z}_i - \mathbf{z}_j)\mathbf{U}(\mathbf{z}_i - \mathbf{z}_j)^\top \hat{\mathbf{V}}_{k-1}^{\frac{1}{2}}.$$

The standardisation can be done after the algorithm has converged. Although the above algorithm is easy to apply in practise, it has a drawback of being highly intensive when the sample size is large. Due to this issue, several new computational approaches and variants of the Dümbgen's shape matrix are introduced in the literature. For alternative algorithms, see Miettinen et al. (2016); Dümbgen et al. (2016). In Taskinen et al. (2010) k-step estimator of the Dümbgen's shape matrix was considered. Finally notice that iteration steps

$$\hat{\mathbf{V}}_k \leftarrow \hat{\mathbf{V}}_{k-1}^{\frac{1}{2}} \frac{1}{n(n-1)^2} \sum\sum\sum_{i\neq j, i\neq k} \mathbf{U}(\mathbf{z}_i - \mathbf{z}_j)\mathbf{U}(\mathbf{z}_i - \mathbf{z}_k)^\top \hat{\mathbf{V}}_{k-1}^{\frac{1}{2}}$$

yield a related shape matrix estimator based on spatial rank vectors (Möttönen & Oja, 1995). We refer interested readers to Sirkiä et al. (2009), for more details.

3.3 Estimation under missing data

In real-life applications, practitioners often face the problem that some data are missing. Nevertheless, it may be of interest to estimate the scatter matrix by using *all* available observations—not only the observations that are complete. Under the assumption that the data are missing at random, maximum-likelihood methods based on the so-called observed-data likelihood function are well-developed (Schafer, 1997). In order to generalize Tyler's shape matrix to the case of incomplete data Frahm & Jaekel (2010) use the fact that Tyler's shape matrix $\hat{\mathbf{V}}$ is a ML-estimator under the angular central Gaussian distribution. More precisely, they show that $\hat{\mathbf{V}}$ represents an observed-data ML-estimator under the assumption that the data stem from a generalized elliptical distribution. They also point out that the incomplete data must be missing *completely* at random to guarantee the consistency of $\hat{\mathbf{V}}$.

Frahm & Jaekel (2010) provide a fixed-point algorithm for the computation of Tyler's shape matrix in the case of incomplete data. An extension to the case of the Hettmansperger-Randles estimator is also given. Since the notation convention in the missing-data framework is nonstandard, we omit details here and refer to Frahm & Jaekel (2010). Theoretical properties of M-estimators, in particular for Tyler's shape matrix, in the case of independent and dependent observations are derived by Frahm et al. (2020). The aforementioned authors show that, when applying M-weight functions to incomplete data, the critical scaling condition expressed by (1) must be satisfied, correspondingly, for each incomplete observation, in order to guarantee that the given M-estimator for scatter is consistent. They resolve the scaling problem by introducing the class of power M-estimators for location and scatter. Both the Gauss-type weight function, $r^2 \mapsto 1$, and Tyler's weight function, $r^2 \mapsto d/r^2$, represent two distinguished power M-weight functions, which implicitly satisfy the critical scaling condition for incomplete data.

3.4 Structured Tyler's shape estimation

In many applications there is some prior knowledge about the structure of the scatter/shape matrix available. Such structures include for example Toeplitz structure, spiked covariance structure, group symmetry and Kronecker structure, among many others. Originally, structured estimation was considered in the context of the covariance matrix estimation for iid Gaussian data and it was shown that enforcing the structure improves the performance of the estimator. Recently, especially in the signal processing community, there has been an increasing interest in estimating robust structured scatter matrices in the context of elliptical distributions, and the research has focused especially on Tyler's shape matrix (see for example Soloveychik & Wiesel, 2014; Soloveychik et al., 2016; Sun et al., 2016; Soloveychik & Trushin, 2016; Mériaux et al., 2021, and references therein). In general, algorithms to estimate the structured shape matrix are usually tailored for the specific structure. A lot depends on the convexity of the assumed structure. As the unstructured Tyler's shape matrix is geodesic convex, it can be concluded that the minimizer of the cost function under a constraint that is also geodesic convex leads to a global maximizer, which is for example the case under a group symmetric constraint (Soloveychik et al., 2016).

In the following assume a centered sample $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ with an unstructured estimate $\hat{\mathbf{V}}$ of Tyler's shape matrix and denote $\mathcal{S}$ as a known closed convex subset of all positive semi-definite $p \times p$ matrices $\mathcal{V}$ under an appropriate scale constraint, i.e., $\mathcal{S} \subset \mathcal{P}$. The subset $\mathcal{S}'$ of $\mathcal{P}$ is closed but not necessarily convex. In the following we only outline some general approaches for structured estimation and provide some references for more details.

Convex projection projects an unconstrained estimate onto the closest element of the constrained set, that is, structured shape matrix $\mathbf{V}^s$ is found as a solution to

$$\min_{\mathbf{V}^s \in \mathcal{S}} ||\mathbf{V}^s - \hat{\mathbf{V}}||,$$

where $||\cdot||$ denotes some norm. This is a convex optimization problem but it consists of a two-step procedure and therefore does not couple structural and distributional properties in the estimation process (Soloveychik & Wiesel, 2014).

Convex constrained covariance estimation (COCA, Soloveychik & Wiesel, 2014) is based on the general methods of moments (GMM) approach and it seeks an approximate solution to

$$\min_{\mathbf{V}^s \in \mathcal{S}} \left\| \mathbf{V}^s - \frac{p}{n} \sum_{i=1}^{n} \frac{\mathbf{x}_i \mathbf{x}_i^\top}{\mathbf{x}_i^\top \mathbf{V}^{s-1} \mathbf{x}_i} \right\|.$$

This problem is however not convex and for general practical computation a convex relaxation of the above equation is suggested which then allows the use of general optimizers for solving the problem. It is then shown that in the unconstrained case COCA is equivalent to Tyler's shape matrix and in the constrained case the two matrices are asymptotically equivalent.

The most general form is the majorization-minimization (MM) approach of Sun et al. (2015, 2016) which starts from a log-likelihood point of view and aims at solving

$$\min_{\mathbf{V}^s \in \mathcal{S}'} \log \det(\mathbf{V}^s) + \frac{p}{n} \sum_{i=1}^{n} \log(\mathbf{x}_i^\top \mathbf{V}^{s-1} \mathbf{x}_i). \tag{4}$$

Due to the complexity of the problem, the MM approach searches for stationary points of (4) and therefore does not necessarily provide the global optimum. Sun et al. (2015, 2016) then provide many tailored MM algorithms for specific structures whose properties depend on the structure at hand. These include convex (e.g. Toeplitz structure, sum of rank one matrices structure) and non-convex structures (e.g. spiked covariance model structure, Kronecker structure).

3.5 Regularized estimators

A topic of increasing interest in multivariate statistics is high-dimensionality, as the dimension p of modern data can be very large and increasingly often even larger than the available sample size n. Therefore, the behaviour of multivariate methods is nowadays often investigated in settings where n and/or p grow.

A key result regarding scatter matrix estimation is given in Tyler (2010) which states that for finite data, if $n \leq p + 1$ and the data is in general position, then any affine equivariant scatter matrix is proportional to the covariance matrix, where the proportionality factor does not depend on the data. The question is then, what is the behaviour of Tyler's shape matrix if n and p grow, i.e., if $p/n \to c$ when $n \to \infty$ and $p \to \infty$. Dümbgen (1998); Frahm & Glombek (2012) consider the case $c = 0$ and show that the condition number of Tyler's shape matrix is $1 + 4\sqrt{p/n} + o(\sqrt{p/n})$ and that the spectral distribution of $\sqrt{n/p}(\hat{\mathbf{V}} - \mathbf{I}_p)$ converges weakly to a semicircle distribution. Further, Zhang et al. (2016) show that in the case $0 < c < 1$ the spectral distribution of Tyler's shape matrix converges weakly to the Marčenko–Pastur distribution. Notice that these results are derived in the context of iid samples from elliptical distributions while similar results for the covariance matrix require usually iid samples from the Gaussian distribution or are less useful in case of elliptical distributions (Karoui, 2009; Zhang et al., 2016).

As the estimation in high-dimensional setting is quite challenging, often estimators are regularized in such a setup and include shrinkage. For Tyler's shape matrix basically three different options for shrinkage are considered: (i) shrinking the eigenvalues of an already computed shape matrix, (ii) adding an penalty term to the M-estimation objective function or (iii) modifying the M-estimation equation.

In the following we outline some recent suggestions to regularize Tyler's shape matrix and refer for further details to the provided references. We first consider shrinking the eigenvalues which assumes a framework with $n > p$ and that we are able to compute Tyler's shape matrix $\hat{\mathbf{V}}$ with $\mathrm{tr}(\hat{\mathbf{V}}) = p$ for the centered sample $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$. The shrinkage regularized Tyler's shape matrix is then defined as

$$\hat{\mathbf{V}}^r_\alpha = \alpha\hat{\mathbf{V}} + (1-\alpha)\mathbf{I}_p,$$

where $\alpha \in [0,1]$ is a regularization parameter. Thus $\hat{\mathbf{V}}^R_\alpha$ shares the same eigenvectors as $\hat{\mathbf{V}}$ but the eigenvalues of it are shrank towards the mean of the eigenvalues of $\hat{\mathbf{V}}$. This type of estimator is considered for example in Chen et al. (2011); Ollila et al. (2021). Ollila et al. (2021) suggest to choose α as the minimizer of

$$\alpha_o = \min_\alpha \mathrm{MSE}(\hat{\mathbf{V}}^r_\alpha),$$

where $\mathrm{MSE}(\hat{\mathbf{V}}^r_\alpha) = \mathbf{E}[||\hat{\mathbf{V}}^r_\alpha - \mathbf{V}||^2]$, for which a closed-form expression can be obtained in case of elliptical distributions or using cross-validation (CV).

To allow $p > n$ Abramovich & Spencer (2007) suggest to load the diagonals in the fixed point algorithm of Tyler's shape matrix by modifying the updating steps as follows

$$\tilde{\mathbf{V}}^r_{k,\beta} \leftarrow \beta\,\frac{p}{n}\sum_{i=1}^{n}\frac{\mathbf{x}_i\mathbf{x}_i^\top}{\mathbf{x}_i^\top\hat{\mathbf{V}}^R_{k,\beta}\mathbf{x}_i} + (1-\beta)\mathbf{I}_p,$$
$$\hat{\mathbf{V}}^r_{k+1,\beta} \leftarrow \frac{\tilde{\mathbf{V}}^r_{k+1,\beta}}{\mathrm{tr}(\tilde{\mathbf{V}}^r_{k+1,\beta})},$$

and iterate until convergence. Here $\beta \in [0,1]$ is a shrinkage coefficient. Chen et al. (2011) establish uniqueness of the estimator and suggest a way to choose β. The above estimator has however been criticized for being heuristic as it is not related to any cost function which it would minimize. Therefore, Wiesel (2012) start again from a log-likelihood point of view and suggest to minimize the following penalized log-likelihood function, that is, to solve

$$\min_{\mathbf{V}\in\mathcal{S}} \log\det(\mathbf{V}) + \frac{p}{n}\sum_{i=1}^{n}\log(\mathbf{x}_i^\top\mathbf{V}^{-1}\mathbf{x}_i) + \gamma P(\mathbf{V}), \tag{5}$$

where $P(\mathbf{V})$ is the penalty function and $\gamma \geq 0$ is a regularization parameter. Wiesel (2012) uses $P(\mathbf{V}) = p\log(\mathrm{tr}(\mathbf{V}^{-1}\mathbf{T})) + \log(|\mathbf{V}|)$ which has its minimum at $\mathbf{T}$ which is the desired target matrix towards which $\mathbf{V}$ should be shrunk to. The minimizer of (5) is denoted as $\hat{\mathbf{V}}^r_\gamma$. Wiesel (2012) and Dümbgen & Tyler (2016) also list several other penalty functions and discuss their appropriateness for different data settings.

The regularization parameter can either be fixed or chosen data dependent using so-called oracle type estimator as discussed in Chen et al. (2011); Ollila & Tyler (2014). The use of cross validation was suggested in Dümbgen & Tyler (2016). It is shown that statistical properties, such as existence and uniqueness, depend on the used penalty function and special attention is given to penalty functions which are geodesic convex in **V**. For example Sun et al. (2014) show that if one uses as penalty the Kullback-Leibler distance between two zero mean Gaussian distributions with covariance matrices **V** and **T**, an estimate very similar to the diagonal loading method mentioned above is obtained.

For further discussions on regularized Tyler's shape matrices we refer to Pascal et al. (2014); Couillet & McKay (2014); Sun et al. (2014); Ollila & Tyler (2014); Dümbgen & Tyler (2016), where maybe Dümbgen & Tyler (2016) provide the most general treatment of regularized Tyler's shape matrices and suggest also a cross validation procedure.Corresponding algorithms are discussed for example in Sun et al. (2014); Dümbgen & Tyler (2016). Robustness properties of previous regularized estimator are studied in Tyler & Yi (2020) showing that, under certain conditions on the tuning parameter, the breakdown point of regularized Tyler's shape matrix could be 1, if not estimating the center μ.

None of the above methods guarantee a sparse solution. To obtain a sparse estimate based on a (regularized) Tyler's shape matrix, Goes et al. (2020) discuss thresholding. Entry-wise thresholding of a matrix $\mathbf{A} = (a_{ij})$ and a threshold $t > 0$ is defined as

$$\tau_t(\mathbf{A}) = (I(|a_{ij}| > t)a_{ij}).$$

Applying such a entry-wise tresholding to an estimate of Tyler's shape matrix, which can also be regularized, yields the thresholded estimate

$$\hat{\mathbf{V}}^t = \tau_t(\hat{\mathbf{V}}),$$

where it is assumed that $\hat{\mathbf{V}}$ has unit trace. Under the assumption of elliptical data with approximately sparse scatter matrix, Goes et al. (2020) provide many properties of $\hat{\mathbf{V}}^t$, especially that these estimators are rate optimal, meaning that the rate coincides with the minimax rate for sparse covariance estimation for sub-Gaussian ellipitical data but in addition holds also for heavy tailed ellipitical data. There seems, however, to be no suggestion yet for choosing the threshold t in a data-driven fashion.

4 Discussion

The seminal paper introducing Tyler's shape matrix (Tyler, 1987a) has been cited according to the Web of Science up to date 378 times[3]. Since its appearance, Tyler's shape matrix has been used in many application areas such as antenna array processing (Ollila & Koivunen, 2003), radar detection (Ollila & Tyler, 2012) or image

[3] Access date: 09.05.2022.

analysis based subspace recovery (Zhang, 2015). Applications in the field of finance are discussed for example in Frahm & Jaekel (2015) and Yang et al. (2015).

This paper is a short and restricted review which shows that due to its nonparametric nature with many excellent statistical properties and computational simplicity, Tyler's shape matrix is still, 35 years after its introduction, an active research area. Tyler's shape matrix continues to exhibit great promise and can be extended in different directions driven by the complex nature of modern data sets.

Acknowledgements

The authors would like to thank the editor and referees for their insightful comments and suggestions.

References

Abramovich, Y. I., & Spencer, N. K. (2007). Diagonally loaded normalised sample matrix inversion (LNSMI) for outlier-resistant adaptive filtering. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, vol. 3, (pp. III–1105–III–1108).

Adrover, J. G. (1998). Minimax bias-robust estimation of the dispersion matrix of a multivariate distribution. *The Annals of Statistics*, *26*, 2301–2320.

Chakraborty, B., Chaudhuri, P., & Oja, H. (1998). Operating transformation retransformation on spatial median and angle test. *Statistica Sinica*, *8*, 767–784.

Chen, Y., Wiesel, A., & Hero, A. O. (2011). Robust shrinkage estimation of high-dimensional covariance matrices. *IEEE Transactions on Signal Processing*, *59*, 4097–4107.

Conte, E., De Maio, A., & Ricci, G. (2002). Recursive estimation of the covariance matrix of a compound-gaussian process and its application to adaptive CFAR detection. *IEEE Transactions on Signal Processing*, *50*(8), 1908–1915.

Couillet, R., & McKay, M. (2014). Large dimensional analysis and optimization of robust shrinkage covariance matrix estimators. *Journal of Multivariate Analysis*, *131*, 99–120.

Croux, C., Dehon, C., & Yadine, A. (2010). The k-step spatial sign covariance matrix. *Advances in Data Analysis and Classification*, *4*, 137–150.

Croux, C., & Haesbroeck, G. (2000). Principal component analysis based on robust estimators of the covariance or correlation matrix: Influence functions and efficiencies. *Biometrika*, *87*, 603–618.

Dümbgen, L. (1998). On Tyler's M-functional of scatter in high dimension. *Annals of the Institute of Statistical Mathematics*, *50*(3), 471–491.

Dümbgen, L., Nordhausen, K., & Schuhmacher, H. (2016). New algorithms for M-estimation of multivariate scatter and location. *Journal of Multivariate Analysis*, *144*, 200–217.

Dümbgen, L., Pauly, M., & Schweizer, T. (2015). M-functionals of multivariate scatter. *Statistics Surveys*, *9*, 32–105.

Dümbgen, L., & Tyler, D. E. (2005). On the breakdown properties of some multivariate M-functionals. *Scandinavian Journal of Statistics*, *32*, 247–264.

Dümbgen, L., & Tyler, D. E. (2016). Geodesic convexity and regularized scatter estimators. *arXiv*, *1607.05455*.

Fang, K., Kotz, S., & Ng, K. (1990). *Symmetric Multivariate and Related Distributions*. London: Chapman & Hall.

Frahm, G. (2009). Asymptotic distributions of robust shape matrices and scales. *Journal of Multivariate Analysis*, *100*, 1329–1337.

Frahm, G. (2022). Power M-estimators for location and scatter. In M. Yi, & K. Nordhausen (Eds.) *Robust and Multivariate Statistical Methods – Festschrift in Honor of David E. Tyler*. Springer.

Frahm, G., & Glombek, K. (2012). Semicircle law of Tyler's M-estimator for scatter. *Statistics & Probability Letters*, *82*, 959–964.

Frahm, G., & Jaekel, U. (2010). A generalization of Tyler's M-estimators to the case of incomplete data. *Computational Statistics & Data Analysis*, *54*, 374–393.

Frahm, G., & Jaekel, U. (2015). Tyler's M-estimator in high-dimensional financial-data analysis. In K. Nordhausen, & S. Taskinen (Eds.) *Modern Nonparametric, Robust and Multivariate Methods: Festschrift in Honour of Hannu Oja*, (pp. 289–305). Cham: Springer International Publishing.

Frahm, G., Nordhausen, K., & Oja, H. (2020). M-estimation with incomplete and dependent multivariate data. *Journal of Multivariate Analysis*, *176*, 104569.

Franks, W. C., & Moitra, A. (2020). Rigorous guarantees for Tyler's M-estimator via quantum expansion. In J. Abernethy, & S. Agarwal (Eds.) *Proceedings of Thirty Third Conference on Learning Theory*, vol. 125 of *Proceedings of Machine Learning Research*, (pp. 1601–1632).

Gini, F., & Greco, M. (2002). Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter. *Signal Processing*, *82*, 1847–1859.

Goes, J., Lerman, G., & Nadler, B. (2020). Robust sparse covariance estimation by thresholding Tyler's M-estimator. *The Annals of Statistics*, *48*, 86–110.

Hallin, M., Oja, H., & Paindaveine, D. (2006). Semiparametrically efficient rank-based inference for shape. II. Optimal R-estimation of shape. *The Annals of Statistics*, *34*, 2757–2789.

Hallin, M., & Paindaveine, D. (2006). Semiparametrically efficient rank-based inference for shape. I. optimal rank-based tests for sphericity. *The Annals of Statistics*, *34*, 2707–2756.

Hettmansperger, T. P., & Randles, R. H. (2002). A practical affine equivariant multivariate median. *Biometrika*, *89*, 851–860.

Huber, P. J. (1981). *Robust Statistics*. New York: Wiley.

Karoui, N. E. (2009). Concentration of measure and spectra of random matrices: Applications to correlation matrices, elliptical distributions and beyond. *The Annals of Applied Probability*, *19*, 2362–2405.

Kent, J. T. (1997). Data analysis for shapes and images. *Journal of Statistical Planning and Inference*, *57*, 181–193. Robust Statistics and Data Analysis, Part II.

Kent, J. T., & Tyler, D. E. (1988). Maximum likelihood estimation for the wrapped cauchy distribution. *Journal of Applied Statistics*, *15*, 247–254.

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

Maronna, R. A., Martin, R. D., Yohai, V. J., & Salibian-Barrera, M. (2018). *Robust Statistics: Theory and Methods (with R)*. Hoboken: Wiley, 2nd edition ed.

Miettinen, J., Nordhausen, K., Taskinen, S., & Tyler, D. E. (2016). On the computation of symmetrized M-estimators of scatter. In C. Agostinelli, A. Basu, P. Filzmoser, & D. Mukherjee (Eds.) *Recent Advances in Robust Statistics: Theory and Applications*, (pp. 151–167). New Delhi: Springer India.

Möttönen, J., Nordhausen, K., & Oja, H. (2010). Asymptotic theory of the spatial median. In J. Antoch, M. Huskova, & P. Sen (Eds.) *Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in honor of Professor Jana Jureckova*, (pp. 182–193). Beachwood, Ohio, USA: Institute of Mathematical Statistics.

Möttönen, J., & Oja, H. (1995). Multivariate spatial sign and rank methods. *Journal of Nonparametric Statistics*, *5*, 201–213.

Mériaux, B., Ren, C., Breloy, A., Korso, M. N. E., & Forster, P. (2021). Matched and mismatched estimation of Kronecker product of linearly structured scatter matrices under elliptical distributions. *IEEE Transactions on Signal Processing*, *69*, 603–616.

Nordhausen, K., & Oja, H. (2018a). Independent component analysis: A statistical perspective. *WIREs: Computational Statistics*, *10*, e1440.

Nordhausen, K., & Oja, H. (2018b). Robust nonparametric inference. *Annual Review of Statistics and Its Application*, *5*, 473–500.

Nordhausen, K., & Tyler, D. E. (2015). A cautionary note on robust covariance plug-in methods. *Biometrika*, *102*, 573–588.

Oja, H. (2010). *Multivariate Nonparametric Methods with R: An Approach Based on Spatial Signs and Ranks*. New York: Springer.

Oja, H., Sirkiä, S., & Eriksson, J. (2006). Scatter matrices and independent component analysis. *Australian Journal of Statistics*, *35*, 175–189.

Ollila, E., & Koivunen, V. (2003). Robust antenna array processing using M-estimators of pseudo-covariance. In *14th IEEE Proceedings on Personal, Indoor and Mobile Radio Communications, 2003. PIMRC 2003.*, vol. 3, (pp. 2659–2663).

Ollila, E., Palomar, D. P., & Pascal, F. (2021). Shrinking the eigenvalues of M-estimators of covariance matrix. *IEEE Transactions on Signal Processing*, *69*, 256–269.

Ollila, E., & Tyler, D. E. (2012). Distribution-free detection under complex elliptically symmetric clutter distribution. In *2012 IEEE 7th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, (pp. 413–416).

Ollila, E., & Tyler, D. E. (2014). Regularized M-estimators of scatter matrix. *IEEE Transactions on Signal Processing*, *62*, 6059–6070.

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

Paindaveine, D. (2008). A canonical definition of shape. *Statistics & Probability Letters*, *78*, 2240–2247.

Paindaveine, D., & Van Bever, G. (2019). Tyler shape depth. *Biometrika*, *106*, 913–927.

Pascal, F., Chitour, Y., Ovarlez, J.-P., Forster, P., & Larzabal, P. (2008). Covariance structure maximum-likelihood estimates in compound gaussian noise: Existence and algorithm analysis. *IEEE Transactions on Signal Processing*, *56*, 34–48.

Pascal, F., Chitour, Y., & Quek, Y. (2014). Generalized robust shrinkage estimator and its application to STAP detection problem. *IEEE Transactions on Signal Processing*, *62*, 5640–5651.

Rublik, F. (2021). On jackknifing the symmetrized Tyler matrix. *Statistics*, *55*, 195–230.

Salibián-Barrera, M., Aelst, S. V., & Willems, G. (2006). Principal components analysis based on multivariate MM estimators with fast and robust bootstrap. *Journal of the American Statistical Association*, *101*, 1198–1211.

Schafer, J. (1997). *Analysis of Incomplete Multivariate Data*. New York: Chapman & Hall.

Sirkiä, S., Taskinen, S., & Oja, H. (2007). Symmetrised M-estimators of multivariate scatter. *Journal of Multivariate Analysis*, *98*, 1611–1629.

Sirkiä, S., Taskinen, S., Oja, H., & Tyler, D. E. (2009). Tests and estimates of shape based on spatial signs and ranks. *Journal of Nonparametric Statistics*, *21*, 155–176.

Soloveychik, I., & Trushin, D. (2016). Gaussian and robust Kronecker product covariance estimation: Existence and uniqueness. *Journal of Multivariate Analysis*, *149*, 92–113.

Soloveychik, I., Trushin, D., & Wiesel, A. (2016). Group symmetric robust covariance estimation. *IEEE Transactions on Signal Processing*, *64*, 244–257.

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

Soloveychik, I., & Wiesel, A. (2015). Performance analysis of Tyler's covariance estimator. *IEEE Transactions on Signal Processing*, *63*, 418–426.

Sun, Y., Babu, P., & Palomar, D. P. (2014). Regularized Tyler's scatter estimator: Existence, uniqueness, and algorithms. *IEEE Transactions on Signal Processing*, *62*, 5143–5156.

Sun, Y., Babu, P., & Palomar, D. P. (2015). Robust estimation of structured covariance matrix for heavy-tailed distributions. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (pp. 5693–5697).

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

Taskinen, S., Croux, C., Kankainen, A., Ollila, E., & Oja, H. (2006). Influence functions and efficiencies of the canonical correlation and vector estimates based on scatter and shape matrices. *Journal of Multivariate Analysis*, *97*, 359–384.

Taskinen, S., & Oja, H. (2016). Influence functions and efficiencies of k-step Hettmansperger-Randles estimators for multivariate location and regression. In R. Y. Liu, & J. W. McKean (Eds.) *Robust Rank-Based and Nonparametric Methods*, (pp. 189–207). Cham: Springer International Publishing.

Taskinen, S., Sirkiä, S., & Oja, H. (2010). k-Step shape estimators based on spatial signs and ranks. *Journal of Statistical Planning and Inference*, *140*, 3376–3388.

Tatsuoka, K. S., & Tyler, D. E. (2000). On the uniqueness of S-functionals and M-functionals under nonelliptical distributions. *The Annals of Statistics*, *28*, 1219–1243.

Tyler, D. E. (1982). Radial estimates and the test for sphericity. *Biometrika*, *69*, 429–436.

Tyler, D. E. (1983). Robustness and efficiency properties of scatter matrices. *Biometrika*, *70*, 411–420.

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

Tyler, D. E. (1987b). Statistical analysis for the angular central Gaussian distribution on the sphere. *Biometrika*, *74*, 579–589.

Tyler, D. E. (2010). A note on multivariate location and scatter statistics for sparse data sets. *Statistics & Probability Letters*, *80*, 1409–1413.

Tyler, D. E., & Yi, M. (2020). Breakdown points of penalized and hybrid M-estimators of covariance. *arXiv*, (p. 2003.00078).

Wiesel, A. (2012). Geodesic convexity and covariance estimation. *IEEE Transactions on Signal Processing*, *60*, 6182–6189.

Yang, L., Couillet, R., & McKay, M. (2015). A robust statistics approach to minimum variance portfolio optimization. *IEEE Transactions on Signal Processing*, *63*, 6684–6697.

Yohai, V. J., & Maronna, R. A. (1990). The maximum bias of robust covariances. *Communications in Statistics - Theory and Methods*, *19*, 3925–3933.

Zhang, T. (2015). Robust subspace recovery by Tyler's M-estimator. *Information and Inference*, *5*, 1–21.

Zhang, T., Cheng, X., & Singer, A. (2016). Marcenko–Pastur law for Tyler's M-estimator. *Journal of Multivariate Analysis*, *149*, 114–123.