AUTOMATIC DIAGONAL LOADING FOR TYLER'S ROBUST COVARIANCE ESTIMATOR

Teng Zhang

University of Central Florida
Department of Mathematics

*Ami Wiesel**

The Hebrew University of Jerusalem
School of Computer Science and Engineering

ABSTRACT

An approach of regularizing Tyler's robust M-estimator of the covariance matrix is proposed. We also provide an automatic choice of the regularization parameter in the high-dimensional regime. Simulations show its advantage over the sample covariance estimator and Tyler's M-estimator when data is heavy-tailed and the number of samples is small. Compared with the previous approaches of regularizing Tyler's M-estimator, our approach has a similar performance and a much simpler way of choosing the regularization parameter automatically.

Index Terms— high-dimensional statistics, robust estimation.

1. INTRODUCTION

In this paper, we consider the problem of regularizing Tyler's covariance estimator. Tyler's covariance estimator had been popular in signal processing fields since it is robust to outliers and has theoretical guarantees on its robustness. We refer the reader to [1, 2] for more detailed discussions. However, the traditional analysis is performed under the setting when p is fixed and n goes to infinity. When n is not much greater than p, most covariance estimators require regularization to ensure good performance and stability [3–7]. This is especially true in modern problems where n is of the same order as p or even smaller.

To our knowledge, the first paper on regularizing Tyler's estimator is [8], and numerous follow up works analyzed its properties and extended it [9–19]. While these estimators perform well, they all depend on a regularization parameter which needs to be chosen in advance. Some automatic choices of the regularization parameters are proposed, but they are rather sophisticated [18]. The main contribution of this paper is a new approach of regularizing Tyler's M-estimator, which has a similar performance and a much simpler way of choosing the regularization parameter.

Our paper is organized as follows. Section 2 introduces the regularized sample covariance estimator, and Section 3 reviews two approaches of regularizing Tyler's M-estimator. Our main contributions, a new approach of regularizing Tyler's M-estimator and an automatic choice of its parameter, are presented in Section 4. Section 5 compares the proposed approach with other estimators by simulated data sets.

In this paper, we denote vectors by boldface lowercase letters, e.g., $\mathbf{x} \in \mathbf{R}^n$, and matrices by boldface uppercase letters, e.g., $\mathbf{A} \in \mathbf{R}^{n\times m}$. The identity matrix of appropriate dimension is written as $\mathbf{I}$. For a square matrix $\mathbf{A}$, $\mathrm{Tr}\left\{\mathbf{A}\right\}$ is the trace, $|\mathbf{A}|$ is the determinant. The standard Euclidean norm of $\mathbf{x}$ is denoted by $\|\mathbf{x}\|$.

*This work was partially supported by the Israel Science Foundation Grant 1339/15.

2. REGULARIZED SAMPLE COVARIANCE ESTIMATOR

We begin with a brief overview of regularization of the sample covariance in the Gaussian case. Given $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n \in \mathbb{R}^p$ sampled from a distribution with zero mean and covariance $\mathbf{Q}$, the sample covariance estimator is a $p \times p$ matrix defined by $\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T$. Due to its popularity and wide use, we focus on the simplest type of regularization: shrinkage towards the identity matrix, also known as diagonal loading [7, 20, 21]. Basically, the main idea is to replace the classical sample covariance $\mathbf{S} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T$ with a weighted version of $\mathbf{S}$ and the identity matrix

$$\hat{\mathbf{Q}}\left(\alpha\right) = \left(1-\alpha\right)\mathbf{S} + \alpha\frac{\mathrm{Tr}\left\{\mathbf{S}\right\}}{p}\mathbf{I} \tag{1}$$

where $0 \le \alpha \le 1$ is a regularization parameter.

Diagonal loading is a simple method which is computationally cheap yet statistically powerful. The main question is how to tune the regularization parameter α. This choice depends on the goal and different procedures are optimal with respect to different objectives. It is possible to use simple data-independent tuning which chooses α as a function of n and p. For example, a simple heuristic is to choose $\alpha = \frac{p}{p+n}$ which is optimal in the two extreme cases in which $p \ll n$ or $p \gg n$. Better performance may be obtained by modern methods that also take into account the data. We now briefly review two common approaches. A popular general-purpose approach to parameter tuning is K-fold cross-validation. Its main drawbacks are that we "waste" $1/K$-th of the samples for approximating the expectation, and that it is computationally intensive since the estimates must be computed over a grid of possible parameters for each group division.

An alternative approach to regularization tuning in covariance estimation is due to Ledoit and Wolf [4]. The idea is to find the optimal α in a closed form which depends on $\mathbf{Q}$. Then, instead of estimating the full matrix $\mathbf{Q}$, we consistently estimate the scalar α directly. Following this idea, we begin by deriving the optimal clairvoyant regularization for minimum mean squared error, where the first step is identical to [4] and the rest steps are similar:

$$\begin{aligned}\alpha\left(\mathbf{Q}\right) &= \arg\min_{\alpha} \mathrm{E}\left\{\left\|\left(\left(1-\alpha\right)\mathbf{S} + \alpha\frac{\mathrm{Tr}\left\{\mathbf{S}\right\}}{\mathrm{p}}\mathbf{I}\right) - \mathbf{Q}\right\|_{\mathrm{Fro}}^2\right\} \\ &= \frac{1}{n}\frac{\left[\left(\frac{1}{p}-\frac{2}{p^2}\right)\frac{p\mathrm{Tr}\left\{\mathbf{Q}^2\right\}}{\mathrm{Tr}^2(\mathbf{Q})}+1\right]}{\left(\frac{1}{p}+\frac{1}{pn}-\frac{2}{p^2 n}\right)\frac{p\mathrm{Tr}\left\{\mathbf{Q}^2\right\}}{\mathrm{Tr}^2(\mathbf{Q})}+\left(\frac{1}{n}-\frac{1}{p}\right)} \\ &\approx \frac{1}{n}\frac{\left[\zeta\left(\mathbf{Q}\right)+1+p\right]}{\zeta\left(\mathbf{Q}\right)+\frac{p}{n}}\end{aligned} \tag{2}$$

where $\zeta\left(\mathbf{Q}\right) = \frac{p\mathrm{Tr}\left\{\mathbf{Q}^2\right\}}{\mathrm{Tr}^2(\mathbf{Q})} - 1$ and the last step is obtained by dropping $(\frac{1}{pn} - \frac{2}{p^2 n})(\zeta\left(\mathbf{Q}\right)+1)$ and $-\frac{2}{p^2}(\zeta\left(\mathbf{Q}\right)+1)$, which are small when p and n are large. The value of $\zeta(\mathbf{Q})$ measures sphericity, i.e., how close $\mathbf{Q}$ is to a scaled identity matrix. It satisfies

$$0 \le \zeta(\mathbf{Q}) \le p-1 \quad (3)$$

where the upper bound is achieved for rank one matrices and the lower bound is achieved for any scaled identity matrix. To gain intuition about this choice of α, it is instructive to consider a few special cases. If $n \gg p$ then clearly $\alpha \to 0$ since there is no need for regularization. If $\zeta \to 0$ then $\alpha \to 1$ and we choose the scaled identity matrix as the estimator. If $\zeta \to p-1$ then $\alpha \to \frac{1}{n}$ which is a reasonable regularization approach when no prior knowledge exists. Clearly, it is impossible to implement this clairvoyant tuning which depends on the unknown covariance via $\zeta(\mathbf{Q})$. Ledoit and Wolf proposed a simple approximation using the available data [4]. The following result based on [22] suggests a similar consistent estimate for the RMT regime in which $\{\mathbf{x}_i\}_{i=1}^n \subset \mathbb{R}^p$ are sampled from $N(0,\mathbf{Q})$, where $p,n\to\infty$, $\lim_{p,n\to\infty}\frac{p}{n}=c$, and the set of eigenvalues of $\mathbf{Q}$ converges to a fixed spectrum. In addition, $\mathrm{Tr}\left\{\mathbf{Q}^i\right\}/p$ remain positive and finite for $i=1,2,3,4$ as $p\to\infty$. Similar closed form estimates are available in [18].

Lemma 2.1. *Consider the RMT regime. Then,*

$$\zeta(\mathbf{S})-c=\frac{p\mathrm{Tr}\left\{\mathbf{S}^2\right\}}{\mathrm{Tr}^2(\mathbf{S})}-c-1=\frac{p\mathrm{Tr}\left\{\mathbf{Q}^2\right\}}{\mathrm{Tr}^2(\mathbf{Q})}-1+o(p)=\zeta(\mathbf{Q})+o(p). \quad (4)$$

Plugging the approximation in (4) into the RHS of (2) yields

$$\frac{1}{n}\frac{[\zeta(\mathbf{Q})+1+p]}{\zeta(\mathbf{Q})+\frac{p}{n}}=\frac{1}{n}\frac{[\zeta(\mathbf{S})-c+1+p]}{\zeta(\mathbf{S})}+o(\frac{1}{n}).$$

and the following closed form regularization parameter

$$\hat{\alpha}=\frac{1}{n}\frac{[\zeta(\mathbf{S})-c+1+p]}{\zeta(\mathbf{S})}. \quad (5)$$

The result is a simple, yet consistent, data-dependent closed form expression that can be easily used. It can be extended and improved via additional assumptions as Gaussianity [9]. A similar approach can be considered with respect to other performance measures [23]. Its main drawback is that it is only asymptotically optimal and may perform poorly when n is small.

3. REGULARIZING TYLER'S ESTIMATOR

A well known robust covariance estimator is Tyler's estimator introduced in [1]. The estimator is a distribution-free variation on the class of M-estimators by Maronna [24], which is defined as the positive definite solution to the fixed point equation $\hat{\mathbf{Q}}=\frac{p}{n}\sum_{i=1}^n\frac{\mathbf{x}_i\mathbf{x}_i^T}{\mathbf{x}_i^T\hat{\mathbf{Q}}^{-1}\mathbf{x}_i}$. This implicit definition is accompanied by a simple fixed-point iteration:

$$\mathbf{Q}_{k+1}=p\sum_{i=1}^n\frac{\mathbf{x}_i\mathbf{x}_i^T}{\mathbf{x}_i^T\hat{\mathbf{Q}}_k^{-1}\mathbf{x}_i}\Big/\mathrm{Tr}\left\{\sum_{i=1}^n\frac{\mathbf{x}_i\mathbf{x}_i^T}{\mathbf{x}_i^T\mathbf{Q}_k^{-1}\mathbf{x}_i}\right\}.$$

Tyler's estimator requires regularization when the number of samples is not much larger than the dimension. In the case of the sample covariance, shrinkage is important for better accuracy and to ensure stable inversion. In Tyler's case, regularization is even more critical since the algorithm itself uses inversions. For example, consider the extreme case in which $n<p$. Here, the sample covariance is inaccurate but can be computed. In contrast, Tyler's estimator cannot even be computed since the inverse in the denominator of the iteration does not exist. In this sense, it must be regularized. In this section we review two modern approaches for shrinking Tyler's estimate towards the identity matrix.

The first approach to regularizing Tyler's estimator is based on modifying the fixed-point iteration while ignoring its MLE interpretation [8]:

$$\mathbf{Q}_{k+1}=p\frac{(1-\alpha)\frac{p}{n}\sum_{i=1}^n\frac{\mathbf{x}_i\mathbf{x}_i^T}{\mathbf{x}_i^T\mathbf{Q}_k^{-1}\mathbf{x}_i}+\alpha\mathbf{I}}{\mathrm{Tr}\left\{(1-\alpha)\frac{p}{n}\sum_{i=1}^n\frac{\mathbf{x}_i\mathbf{x}_i^T}{\mathbf{x}_i^T\mathbf{Q}_k^{-1}\mathbf{x}_i}+\alpha\mathbf{I}\right\}}. \quad (6)$$

The method is guaranteed to converge to a unique limit for any positive definite initial matrix, and the proof follows directly from concave Perron-Frobenius theory [25]. Unlike Tyler's original method, it is not clear whether the output is the optimal solution of any optimization problem, and whether it can be considered as a regularized MLE.

The second approach to regularizing Tyler's estimator is via a regularized MLE [11]. In particular, it proposes to solve

$$\min_{\mathbf{Q}}\frac{p}{n}\sum_{i=1}^n\log\left(\mathbf{x}_i^T\mathbf{Q}^{-1}\mathbf{x}_i\right)+\log|\mathbf{Q}|+\gamma h(\mathbf{Q}) \quad (7)$$

where $h(\mathbf{Q})=\mathrm{Tr}\left\{\mathbf{Q}^{-1}\right\}+\log|\mathbf{Q}|$ and $\gamma=\alpha/(1-\alpha)$.

The unique solution can be calculated by the iterative procedure

$$\mathbf{Q}_{k+1}=(1-\alpha)\frac{p}{n}\sum_{i=1}^n\frac{\mathbf{x}_i\mathbf{x}_i^T}{\mathbf{x}_i^T\mathbf{Q}_k^{-1}\mathbf{x}_i}+\alpha\mathbf{I}. \quad (8)$$

However, choosing the right regularization parameter $0<\alpha\le 1$ is a highly non-trivial problem for both methods. Just like the sample covariance estimator, there are different options for this choice, including: data-independent heuristics, computationally expensive cross-validation, and sophisticated RMT based approaches, e.g. [18, Proposition 2], which requires calculating the regularized estimator for various regularization parameters.

4. NORMALIZED REGULARIZATION

The two regularized versions of Tyler's estimator discussed above are quite intuitive, but do not lend themselves to simple parameter tuning. In this section, we propose an alternative approach with a slightly more complicated iteration, but with a much simpler choice of regularization parameter.

We already know that the spectral norm of a properly scaled tyler's M-estimator and the sample covariance converges to zero in the RMT regime [26]. The latter has a simple regularized version (1) with a natural choice for α in (2). This raises the hope that we could re-use the same tuning approach. Unfortunately, this is not the case for the regularized estimators derived above. The main difficulty is that these estimators do not converge to (1) but rather to $\hat{\mathbf{Q}}(\alpha)=(1-\alpha)\beta(\alpha)\mathbf{S}+\alpha\frac{\mathrm{Tr}\{\mathbf{S}\}}{p}\mathbf{I}$, where $\beta(\alpha)$ is a nonlinear function of α which complicates the derivations (see [18]). In what follows, we present a simple tweak that eliminates $\beta(\alpha)$. In particular, we propose a new regularized Tyler's estimator:

$$\hat{\mathbf{Q}} = \frac{p}{n}\sum_{i=1}^n\frac{(1-\alpha)\mathbf{x}_i\mathbf{x}_i^T+\alpha\frac{\|\mathbf{x}_i\|^2}{p}\mathbf{I}}{(1-\alpha)\mathbf{x}_i^T\hat{\mathbf{Q}}^{-1}\mathbf{x}_i+\alpha\frac{\|\mathbf{x}_i\|^2}{p}\mathrm{Tr}\left\{\hat{\mathbf{Q}}^{-1}\right\}}. \quad (9)$$

The motivation for this definition is that, unlike for the previous regularized estimators, the denominator of (9) normalizes both the data term and the regularization term. This ensures equal scalings and convergence to (1) under the RMT regime, which can be proved by following the proof in [26].

It is easy to check that the normalized regularized Tyler's estimator in (9) is the minimizer (up to a scaling) of

$$g(\mathbf{Q})=\frac{p}{n}\sum_{i=1}^{n}\log\left[(1-\alpha)\mathbf{x}_i^T\mathbf{Q}^{-1}\mathbf{x}_i+\alpha\frac{\|\mathbf{x}_i\|^2}{p}\mathrm{Tr}\left\{\mathbf{Q}^{-1}\right\}\right]+\log|\mathbf{Q}|. \tag{10}$$

The solution can be obtained by Algorithm 1 as stated below, and its property is summarized in the following Theorem. Its proof is based on the geodesic convexity of the objective function in (10) and is available in [27].

Algorithm 1: Normalized regularized Tyler's estimator

input : $\{\mathbf{x}_i\}_{i=1}^n$ of dimension p, $0<\alpha\leq 1$
output: $\hat{\mathbf{Q}}$
$\hat{\mathbf{Q}}\leftarrow\mathbf{I}$
while *not converged* **do**
  $\hat{\mathbf{Q}}\leftarrow\frac{p}{n}\sum_{i=1}^{n}\frac{(1-\alpha)\mathbf{x}_i\mathbf{x}_i^T+\alpha\|\mathbf{x}_i\|^2\mathbf{I}/p}{(1-\alpha)\mathbf{x}_i^T\mathbf{Q}^{-1}\mathbf{x}_i+\frac{1}{p}\alpha\|\mathbf{x}_i\|^2\mathrm{Tr}\left\{\mathbf{Q}^{-1}\right\}}$
  $\hat{\mathbf{Q}}\leftarrow p\frac{\hat{\mathbf{Q}}}{\mathrm{Tr}\{\hat{\mathbf{Q}}\}}$

Theorem 4.1. *[27, Theorem 4.3] Assume that $\mathbf{x}_i\neq 0$ for $i=1,\cdots,n$. The normalized regularized Tyler's estimator $\hat{\mathbf{Q}}$ defined in (10) exists and is unique up to a scaling factor. Algorithm 1 outputs the solution that satisfies $\mathrm{Tr}\left\{\hat{\mathbf{Q}}\right\}=p$.*

The normalized regularized estimator in Algorithm 1 has two main advantages over its competitors. First, it is well-defined for any $0<\alpha\leq 1$. Second, tuning this parameter is quite easy. As a result, we can use the regularization parameter of (1) defined in (2). This requires knowledge of $\zeta(\mathbf{Q})$ which measures how close $\mathbf{Q}$ is to a scaled identity matrix. In the Gaussian case, $\zeta(\mathbf{Q})$ can be approximated via its sample version $\zeta(\mathbf{S})$. This may not work in heavy-tailed data due to sensitivity of $\mathbf{S}$. Instead, a more robust approach is to define the normalized sample covariance $\widetilde{\mathbf{S}}=\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{x}_i\mathbf{x}_i^T}{\|\mathbf{x}_i\|^2}$. Like Tyler's estimator, the normalized sample covariance uses $\frac{\mathbf{x}_i}{\|\mathbf{x}_i\|}$ which are less sensitive to heavy tails than the original samples $\mathbf{x}_i$. It is not a consistent estimator of the covariance, but the following lemma shows that it is useful for approximating $\zeta(\mathbf{Q})$.

Lemma 4.1. *Consider the RMT regime. Then,*

$$p\mathrm{Tr}\left\{\widetilde{\mathbf{S}}^2\right\}-c-1=p\frac{\mathrm{Tr}\left\{\mathbf{Q}^2\right\}}{\mathrm{Tr}^2(\mathbf{Q})}-1+o(p)=\zeta(\mathbf{Q})+o(p). \tag{11}$$

Proof. Following [18] and similarly to (4), we have $\frac{1}{p}\mathrm{Tr}\left\{\mathbf{S}^2\right\}\to\frac{\mathrm{Tr}\left\{\mathbf{Q}^2\right\}}{p}+\frac{\mathrm{Tr}^2(\mathbf{Q})}{pn}$ and $\frac{1}{p}\|\mathbf{x}_i\|^2\to\frac{\mathrm{Tr}\{\mathbf{Q}\}}{p}$ as $p,n\to\infty$. Dividing these limits leads to the stated result. □

Thus, we can consistently estimate $\zeta(\mathbf{Q})$ and define α in Algorithm 1 using the available $\widetilde{\mathbf{S}}$ as follows:

$$\begin{aligned}\zeta &= p\mathrm{Tr}\left\{\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{x}_i\mathbf{x}_i^T}{\|\mathbf{x}_i\|^2}\right)^2\right\}-\frac{p}{n}-1,\\ \alpha &= \frac{1}{n}\frac{[\zeta+1+p]}{\zeta+\frac{p}{n}}.\end{aligned} \tag{12}$$

5. NUMERICAL RESULTS

In this section, we demonstrate the advantages of Algorithm 1 with automatic parameter (12) via numerical simulations. Similar results can be obtained for the other regularized versions, but these involve more complicated tuning. In particular, we compare our regularized Tyler's M-estimator with the two procedures in Section 3 with the automatic parameter choices in [18, Proposition 2], denoted as the AbramovichPascal estimate and the Chen estimate. In these simulations, we consider data originating from multivariate T-distributions, where the probability density function is defined as follows, with $\nu=3$ and various choices of $\mathbf{Q}$

$$p(\mathbf{x})=\frac{\Gamma\left[\frac{\nu+p}{2}\right]}{\Gamma\left(\frac{\nu}{2}\right)\nu^{p/2}\pi^{p/2}|\mathbf{Q}|^{1/2}\left[1+\frac{1}{\nu}\mathbf{x}^T\mathbf{Q}^{-1}\mathbf{x}\right]^{\frac{\nu+p}{2}}}. \tag{13}$$

In the first simulations, we follow the previous works [11, 14, 18] and use the covariance matrix $\mathbf{Q}\in\mathbb{R}^{50\times 50}$ such that $\mathbf{Q}_{i,j}=\rho^{|i-j|}$, and $\rho=0.8$ or 0.5. To avoid the scaling ambiguity, we assume that the trace of the matrix is known to be p, and scale all the estimates so that it has the correct trace (for example, from $\hat{\mathbf{Q}}$ to $p\hat{Q}/\mathrm{Tr}\left\{\hat{\mathbf{Q}}\right\}$). We compare six estimators: the classical sample covariance $\mathbf{S}$, its regularized version in (1) with parameter (2), Tyler's M-estimator, its regularized version in Algorithm 1 with parameter (12), the AbramovichPascal estimate and the Chen estimate. We present the normalized mean squared error (in Frobenius norm) of the covariance estimate averaged over 100 Monte Carlo runs in Figure 1. We remark that the performance of Abramovich-Pascal estimate and the Chen estimate are very close consistently across all our simulations, and the lines representing their performance overlap. From Figure 1 it is easy to see the superiority of the robust versions and the advantage of regularization, and and our estimator has a very similar MSE error to the AbramovichPascal estimate and the Chen estimate. This is achieved even when our choice of the regularization parameter is simpler and less computationally expensive: To find the optimal parameters of the Abramovich-Pascal estimate and the Chen estimate, we compute the estimators in Section 3 with parameters $\alpha=0,0.01,0.02,\ldots,1$ and choose the optimal one by [18, Proposition 2]. In comparison, our choice of the regularization does not require such computations.

While the performance of our estimator is close to the Abramovich-Pascal estimate and the Chen estimate in Figure 1, there exists scenarios where our estimator does not perform as well, or outperforms these estimators. A particular scenario is that, when $\mathbf{Q}$ has one (or a few) large eigenvalue. In particular, we run the following simulation: Let $\mathbf{Q}$ be a diagonal matrix of size 50×50, where m eigenvalues are 1 and the remaining $50-m$ eigenvalues are 0.01. For the case $n=80$, we present the normalized mean squared error (in Frobenius norm) of the covariance estimate averaged over 100 Monte Carlo runs as a function of m in Figure 2. From it we can see that when the number of large eigenvalues is small, the Abramovich-Pascal estimate and the Chen estimate work better. The reason is that, our regularized Tyler's M-estimator is not well-approximated

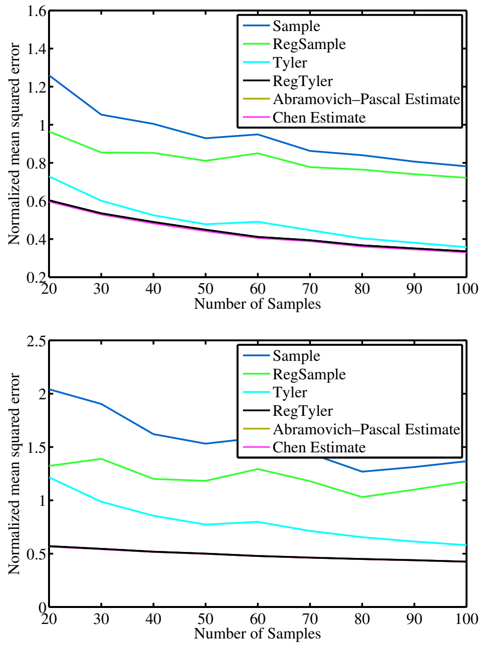

Fig. 1. Regularized Tyler's estimator in heavy-tailed distributions with $\mathbf{Q}_{i,j} = \rho^{|i-j|}$. The first figure corresponds to $\rho = 0.8$ and the second figure corresponds to $\rho = 0.5$.

by a scaled version of $(1-\alpha)\,\mathbf{S} + \alpha\frac{\mathrm{Tr}\{\mathbf{S}\}}{p}\mathbf{I}$ under this setting (although it should be true asymptotically). However, Figure 2 also shows a slight advantage of our estimator when the number of large eigenvalues are between 5 and 20.

In simulations, we also observe that when the spectrum of $\mathbf{Q}$ converges to several different eigenvalues, our algorithm generally outperforms the Abramovich-Pascal estimate and the Chen estimate. In particular, if $\mathbf{Q}$ has 15 eigenvalues of 100, 20 eigenvalues of 1 and 15 eigenvalues of 0.01, the performance of these algorithms are shown in Figure 3, which shows a slight but consistent advantage of our approach.

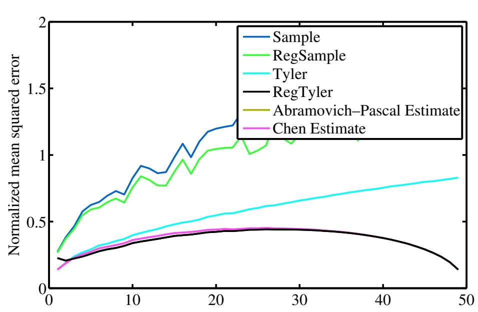

Fig. 2. Regularized Tyler's estimator in heavy-tailed distributions with $\mathbf{Q}$ that has m eigenvalues of 1 and $50-m$ eigenvalues of 0.01.

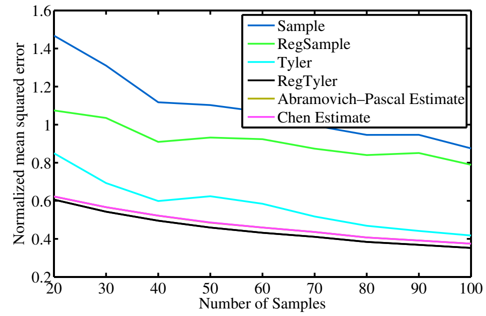

Fig. 3. Regularized Tyler's estimator in heavy-tailed distributions with $\mathbf{Q}$ that has 15 eigenvalues of 100, 20 eigenvalues of 1 and 15 eigenvalues of 0.01.

6. SUMMARY

This work proposes an approach of regularizing Tyler's M-estimator. Based on the high-dimensional random matrix model, it also provides an estimation of the optimal tuning parameter, and simulations show the advantage of the proposed estimator: It achieves a similar performance to other regularizations with a much simpler regularization parameter.

7. REFERENCES

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

[2] R. Maronna, D. Martin, and V. Yohai, *Robust statistics*, John Wiley & Sons, Chichester. ISBN, 2006.

[3] C. Stein, "Estimation of a covariance matrix," *Rietz Lecture*, 1975.

[4] 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.

[5] J. Schafer and K. Strimmer, "A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics," *Statistical applications in genetics and molecular biology*, vol. 4, no. 1, pp. 1175, 2005.

[6] P. Stoica, L. Jian, Z. Xumin, and J.R. Guerci, "On using a priori knowledge in space-time adaptive processing," *IEEE transactions on signal processing*, vol. 56, no. 6, pp. 2598–2602, 2008.

[7] B. D. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 24, no. 4, pp. 397–401, 1988.

[8] Y. I. Abramovich and N. K. Spencer, "Diagonally loaded normalised sample matrix inversion (LNSMI) for outlier-resistant adaptive filtering," in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2007, ICASSP 2007*. IEEE, 2007, vol. 3.

[9] Y. Chen, A. Wiesel, Y.C. Eldar, and A.O. Hero, "Shrinkage algorithms for MMSE covariance estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5016–5029, 2010.

[10] F. Bandiera, O. Besson, and G. Ricci, "Knowledge-aided covariance matrix estimation and adaptive detection in compound-Gaussian noise," *IEEE Trans. on Signal Processing*, vol. 58, no. 10, pp. 5391 –5396, oct. 2010.

[11] Y. Chen, A. Wiesel, and A.O. Hero, "Robust shrinkage estimation of high-dimensional covariance matrices," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4097–4107, 2011.

[12] A. Wiesel, "Unified framework to regularized covariance estimation in scaled Gaussian models," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 29 –38, 2012.

[13] A. Aubry, A. De Maio, L. Pallotta, and A. Farina, "Maximum likelihood estimation of a structured covariance matrix with a condition number constraint," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 3004–3021, 2012.

[14] E. Ollila and D. E. Tyler, "Regularized M-estimators of scatter matrix," *IEEE Transactions on Signal Processing*, vol. 62, no. 22, pp. 6059–6070, 2014.

[15] Y.I Abramovich and O. Besson, "Regularized covariance matrix estimation in complex elliptically symmetric distributions using the expected likelihood approach; part 1: The oversampled case," *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5807–5818, Dec 2013.

[16] O. Besson and Y.I Abramovich, "Regularized covariance matrix estimation in complex elliptically symmetric distributions using the expected likelihood approach; part 2: The undersampled case," *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5819–5829, Dec 2013.

[17] Y. Sun, P. Babu, and D. P. Palomar, "Regularized Tyler's scatter estimator: Existence, uniqueness, and algorithms," *IEEE Transactions on Signal Processing*, vol. 62, no. 19, pp. 5143–5156, 2014.

[18] R. Couillet and M. R. McKay, "Large dimensional analysis and optimization of robust shrinkage covariance matrix estimators," *Journal of Multivariate Analysis*, vol. 131, no. 0, pp. 99 – 120, 2014.

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

[20] S. Vorobyov, A. B. Gershman, and Z. Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem," *IEEE Transactions on Signal Processing*, vol. 51, no. 2, pp. 313–324, 2003.

[21] J. Li, P. Stoica, and Z. Wang, "On robust Capon beamforming and diagonal loading," *IEEE Transactions on Signal Processing*, vol. 51, no. 7, pp. 1702–1715, 2003.

[22] O. Ledoit and M. Wolf, "Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size," *Annals of Statistics*, pp. 1081–1102, 2002.

[23] X. Mestre and M. A. Lagunas, "Finite sample size effect on minimum variance beamformers: Optimum diagonal loading factor for large arrays," *IEEE Transactions on Signal Processing*, vol. 54, no. 1, pp. 69–82, 2006.

[24] R. A. Maronna and V. J. Yohai, "Robust estimation of multivariate location and scatter," *Encyclopedia of Statistical Sciences*, 1998.

[25] U. Krause, "Concave Perron–Frobenius theory and applications," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 47, no. 3, pp. 1457–1466, 2001.

[26] T. Zhang, X. Cheng, and A. Singer, "Marchenko-Pastur law for Tyler's and Maronna's M-estimators," *arXiv preprint arXiv:1401.3424*, 2014.

[27] A. Wiesel and T. Zhang, "Structured robust covariance estimation," *Foundations and Trends in Signal Processing*, vol. 8, no. 3, pp. 127–216, 2014.