Fusing Eigenvalues

Shahab Basiri, Esa Ollila, Gordana Draskovic, Frédéric Pascal

HAL Id: hal-02186196

https://hal-centralesupelec.archives-ouvertes.fr/hal-02186196

Submitted on 26 Feb 2020

FUSING EIGENVALUES

Shahab Basiri, Esa Ollila

Aalto University
Dept. of Signal Processing and Acoustics
P.O.Box 13000, FI-00076 Aalto, Finland

Gordana Drašković, Frédéric Pascal

L2S/CentraleSupélec
3 rue Joliot-Curie,
91192 Gif-sur-Yvette, France

ABSTRACT

In this paper, we propose a new regularized (penalized) covariance matrix estimator which encourages grouping of the eigenvalues by penalizing large differences (gaps) between successive eigenvalues. This is referred to as fusing eigenvalues (eFusion), The proposed penalty function utilizes Tukey's biweight function that is widely used in robust statistics. The main advantage of the proposed method is that it has very small bias for sufficiently large values of penalty parameter. Hence, the method provides accurate grouping of eigenvalues. Such benefits of the proposed method are illustrated with a numerical example, where the method is shown to perform favorably compared to a state-of-art method.

Index Terms— eFusion, Penalized sample covariance matrix, Tuckey's biweight function, Iteratively reweighted algorithm

1. INTRODUCTION

Most data mining, machine learning and statistical multivariate analysis techniques require either an estimate of the covariance matrix (CM), $\mathbf{\Sigma} = \mathbb{E}[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{x} - \mathbb{E}[\mathbf{x}])^\top]$, or some nonlinear function of it, e.g., the inverse CM or its eigenvalues/eigenvectors. However, insufficient number of samples causes significant estimation error in many conventional estimators of the CM including the sample covariance matrix (SCM). When the sample size n is not orders of magnitude larger than the dimensionality, p, it has long been recognized that larger eigenvalues of the SCM tend to overestimate, whereas the smaller eigenvalues tend to underestimate the true eigenvalues. Consequently, regularized or penalized estimators of CM have been introduced in a series of papers [1–10].

A regularized estimator of CM may be an optimally weighted average of the SCM and a well-structured target estimator, which determines what type of structure is imposed on the estimator. The weight parameter controls how much structure is required [2,9–11]. Another approach in regularizing the SCM is to shrink the eigenvalues towards each other, and not towards a predefined target value. Such an approach,

called elasso, was developed in [12], where the authors developed a family of non-smooth penalty functions that not only shrink the eigenvalues towards each other, but they may result in partitioning the eigenvalues into sub-groups.

Given a sample $\mathbf{x}_1, \ldots, \mathbf{x}_n$ of independent and identically distributed (i.i.d.) p-variate observations, the sample covariance matrix, $\mathbf{S}_n = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top$, uniquely minimizes the loss function

$$l(\mathbf{\Sigma}; \mathbf{S}_n) = \text{Tr}(\mathbf{\Sigma}^{-1}\mathbf{S}_n) + \log\{\det(\mathbf{\Sigma})\} \qquad (1)$$

over $\mathbf{\Sigma} \in \mathbb{S}_{++}^{p \times p}$ given that $p < n$, where $\mathbb{S}_{++}^{p \times p}$ denotes the set of $p \times p$ positive definite symmetric matrices. The loss function (1) corresponds to two times the negative log-likelihood function when sampling from p-variate normal distribution. A *regularized sample covariance matrix (RSCM)* $\hat{\mathbf{\Sigma}}$ is then defined as the minimizer of

$$L(\mathbf{\Sigma}; \mathbf{S}_n; \eta) = l(\mathbf{\Sigma}; \mathbf{S}_n) + \eta\Pi(\mathbf{\Sigma}), \qquad (2)$$

where $\Pi(\mathbf{\Sigma})$ denotes a nonnegative penalty function, with $\eta \geq 0$ being the *regularization parameter*.

In this paper, we exploit the assumption that $\mathbf{\Sigma}$ has a structure with only a few distinct eigenvalues, i.e., there are groups of identical eigenvalues. We propose a RSCM estimator that groups the eigenvalues by penalizing large differences between successive eigenvalues. For this purpose, let $d_1 \geq \cdots \geq d_p > 0$ and $\lambda_1 \geq \cdots \geq \lambda_p > 0$ denote the ordered eigenvalues of $\mathbf{S}_n$ and $\mathbf{\Sigma}$, respectively. Furthermore, let $r_j = \log(\lambda_j) - \log(\lambda_{j+1})$ denote the differences, referred to as *gaps*, between successive log-eigenvalues of $\mathbf{\Sigma}$ and let $r_j^{[0]} = \log(d_j) - \log(d_{j+1})$ denote the gaps between log-eigenvalues of $\mathbf{S}_n$ for $j = 1, \ldots, p-1$. We propose to find $\hat{\mathbf{\Sigma}}$ as the minimizer of (2) based on the following non-convex penalty

$$\Pi(\mathbf{\Sigma}) = \sum_{j=1}^{p-1} \rho_c\left(\frac{r_j}{s}\right), \qquad (3)$$

where $\rho_c(\cdot) : \mathbb{R} \to \mathbb{R}$ denotes Tukey's biweight function [13]:

$$\rho_c(r) = \frac{1}{6} \cdot \min\left\{1, 1 - \left(1 - \frac{r^2}{c^2}\right)^3\right\}, \quad r \in \mathbb{R}, \qquad (4)$$

where c is a user-defined tuning parameter and s is the sample standard deviation (SD) of $r_j^{[0]}$ for $j = 1, \ldots, p - 1$.

In the minimization of (2), our penalty function (3) assigns relatively large weights to smaller gaps r_j's, whereas very large gaps attain relatively smaller weights due to the boundedness of Tukey's loss function. The latter property is required in oder to achieve the grouping effect of eigenvalues. We refer to (3) as *eFusion penalty* and the corresponding estimator $\hat{\mathbf{\Sigma}}$ as the eFusion RSCM estimator. We devise an iteratively reweighting (IR) algorithm for computing the proposed RSCM estimator. The IR algorithms are commonly used in finding approximate solutions to such non-convex optimization problems [14, 15].

Despite numerical stability, most RSCM estimators in the literature are severely biased, i.e., the penalized eigenvalues significantly deviate from the true values [12, 16]. Except in specific scenarios [10, 17, 18], optimum tuning parameters η may not be analytically derived in general, without making prior assumptions on the distribution of the data or model parameters. Our numerical example illustrates that the proposed eFusion estimator, has significantly smaller bias than the benchmark method, the elasso estimator [12]. At the same time, poor choice of η appears to be less detrimental compared to elasso; namely, when $\mathbf{\Sigma}$ has groups of identical eigenvalues, larger values of η tend to provide better separation of the groups as eigenvalues within each group are being more shrunk towards each other.

2. ESTIMATING EQUATIONS AND THE ITERATIVE REWEIGHTING ALGORITHM

According to [12, Lemma 2.2], for an orthogonally invariant penaty $\Pi(\mathbf{\Sigma})$, the RSCM estimator $\hat{\mathbf{\Sigma}}$ and the SCM $\mathbf{S}_n$ possess the same set of eigenvectors, with the associated eigenvalues following the same ordering. Note that the eFusion penalty (3) is orthogonally invariant and hence we only need to solve the eigenvalues $\hat{\lambda}_1, \ldots, \hat{\lambda}_p$. Thus, due to [12], the eigenvalues of the eFusion estimator $\hat{\mathbf{\Sigma}}$ can be found as minimizers of

$$\mathcal{L}(\boldsymbol{\lambda}; \boldsymbol{d}, \eta) = \sum_{j=1}^{p} \left(\frac{d_j}{\lambda_j} + \log(\lambda_j) \right) + \eta \sum_{j=1}^{p-1} \rho_c\left(\frac{r_j}{s}\right)$$

$$= \mathbf{d}^\top \boldsymbol{\lambda}^{-1} + \log(\boldsymbol{\lambda})^\top \mathbf{1} + \eta \sum_{j=1}^{p-1} \rho_c\left(\frac{r_j}{s}\right), \quad (5)$$

over $\lambda_1 \geq \cdots \geq \lambda_p > 0$, i.e., over the ordered eigenvalues. Above $\mathbf{d} = (d_1, \ldots, d_p)^\top$, $\boldsymbol{\lambda}^{-1} = (1/\lambda_1, \ldots, 1/\lambda_p)^\top$ and $\mathbf{1}$ is a vector of size $p \times 1$ with all elements equal to one. By setting the gradient of (5) w.r.t. $\boldsymbol{\lambda}$ to zero $\nabla_{\boldsymbol{\lambda}} \mathcal{L} = \mathbf{0}$ we get

$$-\text{diag}(\boldsymbol{\lambda})^{-2} \mathbf{d} + \text{diag}(\boldsymbol{\lambda})^{-1} \left(\mathbf{1} + \frac{\eta}{s} \sum_{j=1}^{p-1} \rho_c'\left(\frac{r_j}{s}\right) \boldsymbol{\kappa}_j \right) = \mathbf{0},$$

Algorithm 1: Iteratively reweighted eFusion algorithm

Input : $\mathbf{d}$: Eigenvalues of the SCM $\mathbf{S}_n$;
$\quad\quad\quad$ η: Penalty parameter;
$\quad\quad\quad$ c: Tukey tuning constant.
Output : $\hat{\boldsymbol{\lambda}}$: Penalized eigenvalues verifying (9)

Initialize: $k \leftarrow 0$; $\boldsymbol{\lambda}^{[0]} \leftarrow \mathbf{d}$

1 Compute $s = \text{SD}(\mathbf{r}^{[0]})$,
  Repeat

2 $\quad$ Update the gaps:

$\quad$ $r_j^{[k]} \leftarrow \log(\lambda_j^{[k]}) - \log(\lambda_{j+1}^{[k]}), \ j = 1, \ldots, p-1,$

3 $\quad$ Update the weights:

$\quad$ $w_j^{[k]} \leftarrow \rho_c'(r_j^{[k]}/s)/(r_j^{[k]}/s), \ j = 1, \ldots, p-1,$

4 $\quad$ Update the eigenvalue estimates:

$\quad$ $\log \lambda_j^{[k+1]} \leftarrow \frac{1}{w_j^{[k]} + w_{j-1}^{[k]}} \left(\frac{s^2}{\eta}(d_j/\lambda_j^{[k]} - 1) + \right.$

$\quad\quad$ $\left. w_j^{[k]} \log \lambda_{j+1}^{[k]} + w_{j-1}^{[k]} \log \lambda_{j-1}^{[k+1]} \right),$

$\quad$ for $j = 1, \ldots, p$.

5 $\quad$ $k \leftarrow k + 1$

  until *convergence*

6 $\hat{\boldsymbol{\lambda}} \leftarrow \left(\exp(\log \lambda_1^{[k+1]}), \ldots, \exp(\log \lambda_p^{[k+1]}) \right)^\top$

where

$$\kappa_j = \begin{bmatrix} \underbrace{0 \quad \cdots \quad 0}_{j-1} & 1 & -1 & 0 & \cdots & 0 \end{bmatrix}_{1 \times p}^\top.$$

After some straightforward math, we obtain the following estimating equation.

$$\mathbf{f}(\boldsymbol{\lambda}) = \text{diag}\left(\mathbf{1} + \frac{\eta}{s}\mathbf{v}\right)\boldsymbol{\lambda} - \mathbf{d} = \mathbf{0}, \quad (6)$$

where $\mathbf{v} = (v_1, \ldots, v_p)^\top$ with $v_j = \rho_c'(r_j/s) - \rho_c'(r_{j-1}/s)$ for $j \in \{1, \ldots, p\}$. Note that, $v_1 = \rho_c'(r_1/s)$ and $v_p = -\rho_c'(r_{p-1}/s)$.

The solution to (6) can be obtained by solving the following system of equations,

$$1 + \frac{\eta}{s}\left(\rho_c'\left(\frac{r_j}{s}\right) - \rho_c'\left(\frac{r_{j-1}}{s}\right)\right) - \frac{d_j}{\lambda_j} = 0, \quad (7)$$

for $j = 1, \ldots, p$. This can be reformulated as

$$1 + \eta(r_j w_j - r_{j-1} w_{j-1})/s^2 - d_j/\lambda_j = 0, \quad (8)$$

where $w_j = \rho_c'(r_j/s)/(r_j/s)$ are referred to as weights. By substituting $r_j = \log(\lambda_j) - \log(\lambda_{j+1})$ and $r_{j-1} = $

$\log(\lambda_{j-1}) - \log(\lambda_j)$ to (8), we obtain the following system of fixed-point equations.

$$\log(\lambda_j) = \frac{\frac{s^2}{\eta}(d_j/\lambda_j - 1) + w_j \log \lambda_{j+1} + w_{j-1} \log \lambda_{j-1}}{w_j + w_{j-1}}, \tag{9}$$

for $j = 1, \ldots, p$. Not that for $j = 1$ and $j = p$, (9) reduces to

$$\log(\lambda_1) = \frac{s^2}{w_1 \eta}\left(\frac{d_1}{\lambda_1} - 1\right) + \log \lambda_2,$$

$$\log(\lambda_p) = \frac{s^2}{w_{p-1} \eta}\left(\frac{d_p}{\lambda_p} - 1\right) + \log \lambda_{p-1}.$$

In the spirit of Iteratively Reweighted Least Squares (IRLS), we devise an IR-eFusion algorithm to find the solution $\hat{\lambda}_1, \ldots, \hat{\lambda}_p$ that verify (9).

Using $\mathbf{d}$ as the initial value for $\boldsymbol{\lambda}$, our approach, detailed in Algorithm 1, iterates the following steps until convergence. First, the gaps r_j and the weights w_j are computed for all the eigenvalues. Then, (9) is used to update each eigenvalue estimate in a coordinate-wise fashion, i.e., in updating $\log(\lambda_j)$ we use the already updated $\log(\lambda_{j-1})$. Note that, In Step 4, we avoid updating $\log \lambda_j^{[k+1]}$ if $w_j^{[k]} + w_{j-1}^{[k]} = 0$, i.e., $\log \lambda_j^{[k+1]} \leftarrow \log \lambda_j^{[k]}$.

3. ON CHOOSING THE TUNING PARAMETER

In order to find an optimal value of the tuning parameter c for Tukey's biweight function we will analyze the distribution of $r_j^{[0]} = \log(d_j) - \log(d_{j+1})$. To that end, let us consider the following binary hypothesis test

$$\begin{cases} \mathcal{H}_0 : \lambda_j = \lambda_{j+1}, \\ \mathcal{H}_1 : \lambda_j > \lambda_{j+1}, \quad j = 1, \ldots, p-1. \end{cases}$$

Our goal is to detect when two consecutive eigenvalues are equal and thus, we want to derive the distribution of $r_j^{[0]} = \log(d_j) - \log(d_{j+1})$ under the null hypothesis $\mathcal{H}_0$. Once the distribution is derived, the tuning parameter can be obtained as a threshold that assures a given probability of false alarm (P_{fa}). The distribution of $r_j^{[0]}$ can be derived using the result for the joint distribution of d_j and d_{j+1} derived in [19]. More precisely, under the assumption that the data is uncorrelated ($\boldsymbol{\Sigma} = \mathbf{I}$), one has the following joint distribution of two ordered consecutive eigenvalues of a Wishart-distributed matrix

$$f_{d_j, d_{j+1}}(x_j, x_{j+1}) = \frac{K}{(j-1)!} \sum_{\mathbf{n}} \sum_{\mathbf{m}} s(\mathbf{n}, \mathbf{m})$$
$$|\mathbf{D}(x_{j+1})| \prod_{k=j}^{j+1} \varphi(n_k, m_k, x_k) \prod_{k=1}^{j-1} g(k)$$

where $g(k) = \Gamma(n - p + n_k + m_k - 1, x_j)$, the $(s, t)^{\text{th}}$ element of $\mathbf{D}(y)$ is given by $\gamma(n - p + l_{s,\mathbf{n}} + l_{t,\mathbf{m}} - 1, y)$,

Γ and γ are respectively upper and lower incomplete Gamma functions, K is a normalizing constant, $\varphi(n_k, m_k, x_k) = x_k^{n-p+n_k+m_k-2} \exp(-x_k)$ and

$$\sum_{\mathbf{n}} = \sum_{n_1=1}^{p} \sum_{n_2=1, n_2 \neq n_1}^{p} \cdots \sum_{n_{j+1}=1, n_{j+1} \neq \{n_1, \ldots, n_j\}}^{p}$$

(analogous for $\sum_{\mathbf{m}}$). The definitions for $s(\mathbf{n}, \mathbf{m})$, $l_{s,\mathbf{n}}$ and $l_{t,\mathbf{m}}$ can be found in [19]. In order to derive the distribution of r_j we perform a change of variables by introducing $x = \log(x_j/x_{j+1})$ and $y = x_{j+1}$. Then, computing the Jacobian $J(x, y) = \frac{1}{x_j} = \exp(-x)/y$ and using $f_{r_j}(x) = \int_0^{+\infty} f_{d_j, d_{j+1}}(x, y) |J(x, y)|^{-1} dy$ one obtains the result

$$f_{r_j^{[0]}}(x) = \int_0^{+\infty} \left(\frac{K}{(j-1)!} \sum_{\mathbf{n}} \sum_{\mathbf{m}} s(\mathbf{n}, \mathbf{m}) |\mathbf{D}(y)| \right.$$
$$\varphi(n_j, m_j, \exp(x)y)\varphi(n_{j+1}, m_{j+1}, y)$$
$$\left. \prod_{k=1}^{j-1} \Gamma(n - p + n_j + m_j - 1, \exp(x)y) \exp(x)y \right) dy. \tag{10}$$
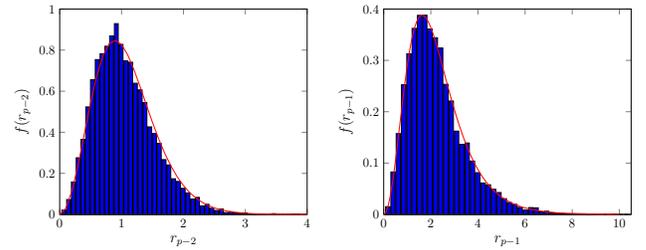

Fig. 1. Empirical distribution of $r_{p-2}^{[0]}$ (left panel) and resp. $r_{p-1}^{[0]}$ (right panel) with $p = 100$ compared to the corresponding theoretical distribution for $p = 5$; $n = p$.

One can note that the distribution of $r_j^{[0]}$ depends on the position j, on the data dimension p and on the number of observations n. Indeed, for a small sample size n the SCM's eigenvalues are very distinct, $r_j^{[0]}$'s have high fluctuations, especially for large j (small eigenvalues). On the other hand, for a sufficiently large n the eigenvalues of SCM are closer to their true values and the gaps between them are smaller. This implies that for larger n the interval of acceptable parameter c is wider. For instance, with our settings and $n = 700$ (instead of 3000) the "optimal" parameter c is approximately between 1.13 and 1.5. When n increases the gaps that correspond to true identical eigenvalues decrease, while the gaps corresponding to distinct eigenvalues rapidly increase. Therefore the choice of c is much more flexible, e.g. for $n = 3000$ all values between 0.42 and 2.96 give good results. Small values of c can result in more groups than expected, while large values have tendency to fuse even very different eigenvalues.

For high dimensional p and for small eigenvalues (so large j), Eq. (10) is computationally demanding, but we can empirically show that some cases can be well approximated with distributions obtained with significantly smaller parameters. Figure 1 illustrates this point. As it can be seen from Figure 1, the variations of $r_j^{[0]}$ are much higher for smaller eigenvalues. The range becomes narrower when j decreases. Here, we have plotted the limited case when $n = p$. Obviously, in order to better estimate the covariance matrix one needs c that depends on the position j. In this paper, we assume that n is big enough that the variation of $r_j^{[0]}$ are smaller under $\mathcal{H}_0$ and thus a unique c can be used for all j. More general case will be analyzed in the extended version of this paper. In order to compute that value, we propose then to look, for instance, at the distribution of $r_j^{[0]}$ for $p = 4$, $n = 16$ and $P_{fa} = 0.01$. For j equal to $1, 2, 3$ one obtains c equal to $0.794, 0.86, 1.159$, respectively. These are used in Section 4 as candidate values.

4. NUMERICAL EXAMPLE

We compare the performance of the proposed estimator with the elasso [12]. In elasso, $\Pi(\mathbf{\Sigma}) = \sum_{j=1}^{p} a_j \log(\lambda_j)$, is used as the penalty function, where the weights a_j are obtained by centering decreasing quantiles from the Marčenko-Pastur law. We generate a random sample of size $n = 3000$ from a $p = 100$ dimensional multivariate normal distribution. Similar to [12], the covariance matrix $\mathbf{\Sigma}$ has 40 eigenvalues equal to 20, 30 equal to 10 and 30 equal to 2.

Figure 2 display the process of grouping eigenvalues with elasso (top panel) and eFusion (bottom panel). The results showcase a significant improvement that our estimator can offer. First, it gives an unbiased estimation as the three groups of eigenvalues are well separated and close to their true value. Second, one does not need to search for optimal penalty parameter since for large enough η, the method provides good results. The eFusion estimator offers similar results for different choices of c and in order to compare these results we can measure an affine-invariant (Riemannian) distance between positive definite matrices $\mathbf{\Sigma}$ and $\widehat{\mathbf{\Sigma}}$, defined as $d(\mathbf{\Sigma}, \widehat{\mathbf{\Sigma}}) = \| \log(\mathbf{\Sigma}^{-1/2}\widehat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1/2})\|_F$. Table below lists the values of the distance for different values of c for the eFusion estimator:

c	0.794	0.86	1.159
$d(\mathbf{\Sigma}, \widehat{\mathbf{\Sigma}})$	1.548	1.5066	1.5050

One can note that the values of the distance are very similar for these choices of c. In general, to find a satisfactory value of c one should adjust p/n to be larger than the one of high dimensional data (given sample). Otherwise, we risk to get too small value for c and finally a biased estimation. One should note that we did not take into account the parameter s (SD of $r_j^{[0]}$) assuming that it is near 1 which can be a strong approximation even for high n since the eigenvalues are cor-
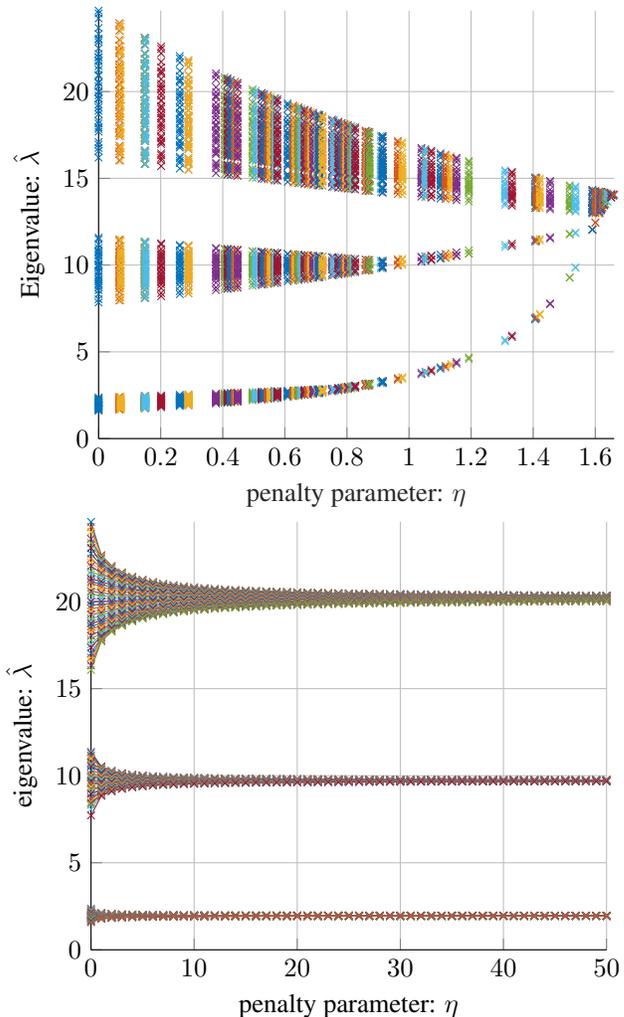

Fig. 2. An example of grouping of eigenvalues of elasso (top panel) and eFusion (bottom panel) for $c = 0.794$.

related. From table above one could conclude that higher values of c give better estimation. This is still under study as well as the distribution of gaps in the general case ($\mathbf{\Sigma} \neq \mathbf{I}$).

5. CONCLUSION

We introduced a new regularized covariance matrix estimator based on a novel eFusion penalty (3) that promotes similarity and grouping of eigenvalues. The grouping effect is achieved as large gaps between successive eigenvalues are not penalized excessively. This feature is obtained by utilizing Tukey's function as the penalty function for the gaps. The important topic of how to choose the tuning parameter c of Tukey's function was addressed along with some simulation results. Our numerical example revealed that the main benefits of the eFusion are unbiasedness (accurate grouping) and robustness to the choice of the penalty parameter.

6. REFERENCES

[1] Olivier Ledoit and Michael Wolf, "Honey, I shrunk the sample covariance matrix," 2003.

[2] Olivier Ledoit and Michael Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *J. Multivar. Anal.*, vol. 88, no. 2, pp. 365–411, Feb. 2004.

[3] Yilun Chen, Ami Wiesel, Yonina C Eldar, and Alfred O Hero, "Shrinkage algorithms for mmse covariance estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5016–5029, 2010.

[4] Olivier Besson and Yuri I Abramovich, "Regularized covariance matrix estimation in complex elliptically symmetric distributions using the expected likelihood approach—part 2: The under-sampled case," *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5819–5829, 2013.

[5] Yuri I Abramovich and Olivier Besson, "Regularized covariance matrix estimation in complex elliptically symmetric distributions using the expected likelihood approach—part 1: The over-sampled case," *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5807–5818, 2013.

[6] F. Pascal, Y. Chitour, and Yihui Quek, "Generalized Robust Shrinkage Estimator and Its Application to STAP Detection Problem," *IEEE Transactions on Signal Processing*, vol. 62, no. 21, pp. 5640–5651, Nov 2014.

[7] Ying Sun, Prabhu Babu, and Daniel P Palomar, "Regularized tyler's scatter estimator: Existence, uniqueness, and algorithms," *IEEE Transactions on Signal Processing*, vol. 62, no. 19, pp. 5143–5156, 2014.

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

[9] Jacob Bien and Robert J. Tibshirani, "Sparse estimation of a covariance matrix," *Biometrika*, vol. 98, no. 4, pp. 807–820, 2011.

[10] Esa Ollila and Elias Raninen, "Optimal shrinkage covariance matrix estimation under random sampling from elliptical distributions," *arXiv preprint arXiv:1808.10188*, 2018.

[11] Olivier Ledoit and Michael Wolf, "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection," *Journal of Empirical Finance*, vol. 10, no. 5, pp. 603–621, 2003.

[12] David E. Tyler and Mengxi Yi, "Lassoing Eigenvalues," *arXiv:1805.08300v1*, 2018.

[13] Peter J. Rousseeuw and Annick M. Leroy, *Robust regression and outlier detection*, John Wiley, New York, 1987.

[14] P. J. Green, "Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 46, no. 2, pp. 149–192, 1984.

[15] R. Wolke and H. Schwetlick, "Iteratively reweighted least squares: Algorithms, convergence analysis, and numerical comparisons," *SIAM Journal on Scientific and Statistical Computing*, vol. 9, no. 5, pp. 907–921, 1988.

[16] Abla Kammoun, Romain Couillet, Frérdéric Pascal, and Mohamed-Slim Alouini, "Convergence and fluctuations of Regularized Tyler estimatorsyler estimators," *IEEE Transactions on Signal Processing*, vol. 64, no. 4, pp. 1048–1060, 2016.

[17] Romain Couillet and Matthew McKay, "Large dimensional analysis and optimization of robust shrinkage covariance matrix estimators," *Journal of Multivariate Analysis*, vol. 131, pp. 99–120, 2014.

[18] Abla Kammoun, Romain Couillet, Frédéric Pascal, and Mohamed-Slim Alouini, "Optimal design of the adaptive normalized matched filter detector using regularized tyler estimators," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 2, pp. 755–769, 2018.

[19] A. Zanella, M. Chiani, and M. Z. Win, "On the marginal distribution of the eigenvalues of wishart matrices," *IEEE Transactions on Communications*, vol. 57, no. 4, pp. 1050–1060, April 2009.