Supplement to "Shrinking the eigenvalues of M-estimators of covariance matrix"

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

***Abstract*—This report provides some supplementary studies and examples for the paper "Shrinking the eigenvalues of M-estimators of covariance matrix" [1]. Specifically, we provide simulation studies of the proposed shrinkage covariance matrix estimators for complex-valued data as well as an additional example of portfolio optimization using real historical stock returns data on the same set-up as in [2].**

***Index Terms*—M-estimators, sample covariance matrix, shrinkage, regularization, elliptical distributions**

I. Introduction

Throughout, we drop the suffix -Ell1 from the proposed shrinkage M-estimators for simplicity. Thus the proposed estimators, described in detail in our paper [1] are referred to as **RSCM**, **RMVT**, **RHub**, **RTyl** and **CV** as in the paper. We remind the reader that CV uses regularized sample covariance matrix (RSCM) estimator $\mathbf{S}_\beta$ for which the shrinkage parameter β is chosen by 5-fold cross-validation (CV) over a grid of β-values. See [1] for more details of all the approaches.

II. Simulation study: complex-valued data

We replicate the simulation study of [1, Section V] but using complex-valued data. We generate samples from centered complex elliptically symmetric [3] distributions with a scatter matrix parameter $\boldsymbol{\Sigma}$ following an AR(1) structure, $(\boldsymbol{\Sigma})_{ij} = \tau \varrho^{|i-j|}$, where $\varrho = |\varrho| e^{\jmath\theta}$ with $|\varrho| \in (0,1)$ and scale parameter $\tau = \mathrm{tr}(\boldsymbol{\Sigma})/p = 10$. When $|\varrho| \downarrow 0$, then $\boldsymbol{\Sigma}$ is close to an identity matrix scaled by τ, and when $|\varrho| \uparrow 1$, $\boldsymbol{\Sigma}$ tends to a singular matrix of rank 1. All simulation results in this section are averages over 2000 Monte-Carlo (MC) trials and $\theta = \mathrm{Arg}(\varrho)$ is generated from $Unif(0, 2\pi)$ distribution for each MC trial.

We compare the proposed estimators to the ones proposed by Chen, Wiesel and Hero [4] (see also [1, Section IV-C]) and Coluccia [5], respectively, referred to as **CWH** and **EB**, respectively. EB stands for empirical Bayes shrinkage estimator. EB assumes that the data follows complex circular Gaussian distribution and hence the method is not expected to perform very well under heavy-tailed non-Gaussian distributions. For tuning parameter ν of the prior distribution of EB, we used the recommended values $\nu = p+1$ and $\nu = n+p+1$, but we report results only for the former which provided much better performance.

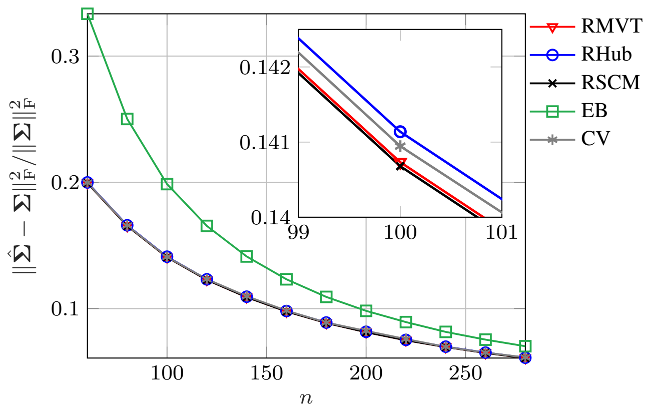

Fig. 1. NMSE as a function of n when samples are drawn from a complex circular Gaussian distribution $\mathbb{C}\mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$ with an AR(1) covariance structure; $|\varrho| = 0.6$ and $p = 40$.

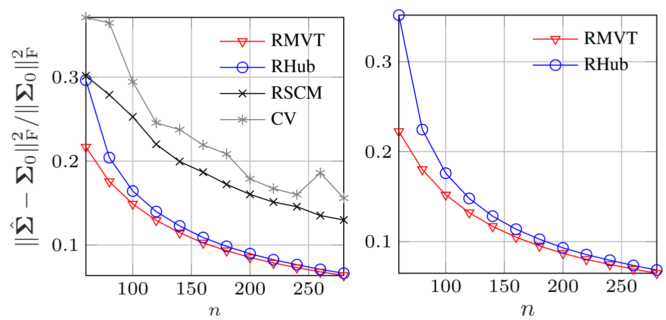

Fig. 2. NMSE as a function of n when samples are draw from a complex circular p-variate $t-$distribution with $\nu = 5$ (left panel) and $\nu = 3$ (right panel) degrees of freedom. The scatter matrix has an AR(1) covariance structure; $|\varrho| = 0.6$ and $p = 40$.

A. Gaussian data

We generate the data from the complex multivariate normal distribution $\mathbb{C}\mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$. The correlation parameter of AR(1) covariance matrix verifies $|\varrho| = 0.6$. The dimension is $p = 40$ and n varies from 60 to 280. As in [1], value $q = 0.7$ is used for RHub estimator. Since Huber's M-estimator is scaled to be consistent to the covariance matrix for Gaussian samples, the underlying population parameter $\boldsymbol{\Sigma}_0$ coincides with the covariance matrix $\boldsymbol{\Sigma}$ in this case. We also scaled the MVT-weight $u_{\mathrm{T}}(t;\nu)$ so that it is consistent to $\boldsymbol{\Sigma}$ for Gaussian data. Figure 1 compares the normalized MSE (NMSE) $\|\hat{\boldsymbol{\Sigma}}_\beta - \boldsymbol{\Sigma}\|_{\mathrm{F}}^2/\|\boldsymbol{\Sigma}\|_{\mathrm{F}}^2$ of different estimators w.r.t. increasing

E. Ollila is with the Department of Signal Processing and Acoustics, Aalto University, P.O. Box 15400, FI-00076 Aalto, Finland. D. Palomar is with the Hong Kong University of Science and Technology, Hong Kong. F. Pascal is with Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes, 91190, Gif-sur-Yvette, France.

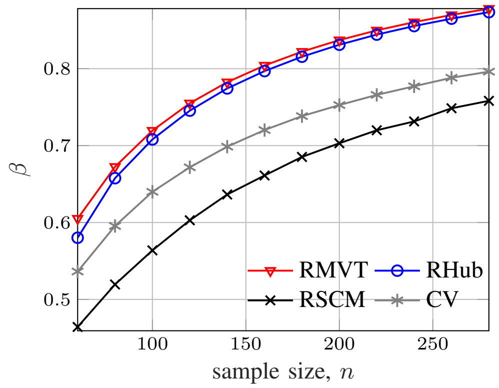

Fig. 3. Shrinkage parameter β as a function of n when samples are drawn from a complex circular p-variate t-distribution with an AR(1) covariance structure; $\nu = 5$, $|\varrho| = 0.6$ and $p = 40$.

sample length n. It can be noted that the proposed estimators (RSCM, RHub, and RMVT) provide essentially equally good estimator of the covariance matrix $\boldsymbol{\Sigma}$ for Gaussian data; RSCM and RMVT are performing equally well, largely due to data-adaptive estimation of d.o.f. parameter ν. Yet, differences are marginal and can be spotted only by zooming in as in the sub-plot of Figure 1. EB estimator is is performing much worse than the proposed shrinkage estimators. This is probably due to the difficulty of selecting the parameter ν of the prior distribution. As noted earlier, we used the value $\nu = p + 1$ which was recommended in [5]. All in all, above results for complex-valued Gaussian data confirm the results found for real-valued Gaussian data in [1].

B. Heavy-tailed data

Next we computed the NMSE curves when the data is generated from a complex circular t-distribution [3] with $\nu = 5$ and $\nu = 3$ degrees of freedom (d.o.f). EB was performing poorly here due to its strict assumption of Gaussianity and hence is not shown in the plots. In the latter case ($\nu = 3$), also the non-robust RSCM provided large NMSE and is left out from the plot. This was expected since t-distribution with $\nu = 3$ d.o.f. is very heavy-tailed with non-finite kurtosis. Figure 2 displays the results. Note that NMSE of each estimator is now compared against the underlying scatter matrix parameter $\boldsymbol{\Sigma}_0$ that they are estimating. As can be seen, the proposed robust RHub and RMVT estimators provide significantly improved performance compared to RSCM. We can also notice that RMVT estimator that adaptively estimates the d.o.f. ν from the data is able to outperform the regularized Huber's estimator (RHub).

Figure 3 depicts the (average) shrinkage parameter β as a function of n in the case that samples are drawn from a compex circular t-distribution with $\nu = 5$ degrees of freedom. As can be seen the robust shrinkage estimators (RHub and RMVT) use a larger shrinkage parameter value β than the non-robust RSCM estimator.

Figure 4 displays the normalized MSE of different shrinkage shape matrix estimators, $\|\hat{\mathbf{V}} - \mathbf{V}\|_{\mathrm{F}}^2 / \|\mathbf{V}\|_{\mathrm{F}}^2$, as a function of sample length n, where $\hat{\mathbf{V}}$ denotes either CWH or RTyl estimator or the normalized RHub or RMVT estimator defined

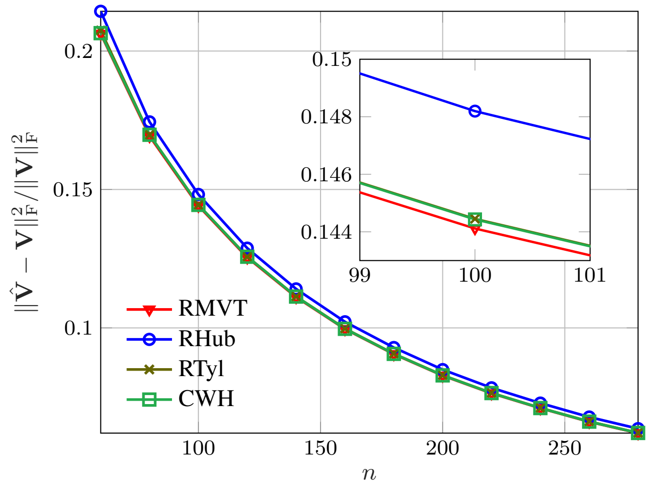

Fig. 4. NMSE of different shrinkage estimators of shape matrix $\mathbf{V}$ as a function of n when samples are drawn from a complex circular t-distribution with an AR(1) scatter matrix; $\nu = 5$, $|\varrho| = 0.6$ and $p = 40$.

as $\hat{\mathbf{V}} = p\hat{\boldsymbol{\Sigma}}_\beta / \operatorname{tr}(\hat{\boldsymbol{\Sigma}}_\beta)$. Note that such normalization is not necessary for CWH or RTyl since they verify $\operatorname{tr}(\hat{\mathbf{V}}) = p$. As can be seen all estimators are performing well, and RMVT and CWH are performing equally well. We can also notice that the two very different approaches for shrinking Tyler's M-estimator, so RTyl and CWH, provide essentially the same NMSE performance.

III. Additional portfolio optimization example

In this Section, we investigate the out-of-sample portfolio performance of the proposed shrinkage M-estimators. In particular, we use the dividend adjusted daily closing prices downloaded from the Yahoo! Finance (http://finance.yahoo.com) database to obtain the net returns for 50 stocks that are currently included in the Hang Seng Index (HSI) for two different time periods, from Jan. 4, 2010 to Dec. 24, 2011, and from Jan. 1, 2016 to Dec. 27, 2017 (excluding the weekends and public holidays). In both cases, the time series contains $T = 491$ trading days. For the first period (2010-2011), we had full length time series for only $p = 45$ stocks, whereas in the latter case we had full length time series for all stocks, so $p = 50$. The data sets and the study is the same as described in [2].

At a particular day t, we used the previous n days (i.e., from $t - n$ to $t - 1$) as the training window to estimate the covariance matrix, and the portfolio weight vector. The obtained portfolio weight vector was then used to compute the portfolio returns for the following 20 days. Next, the window was shifted 20 trading days forward, a new weight vector was computed, and the portfolio returns for another 20 days were computed. Hence, this scenario corresponds to the case that the portfolio manager holds the assets for approximately a month (20 trading days), after which they are liquidated and new weights are computed. In this manner, we obtained $T - n$ daily returns from which the realized risk was computed as the sample standard deviation of the obtained portfolio returns. To obtain the annualized realized risk, the sample standard deviations of the daily returns were multiplied by $\sqrt{250}$. In our tests, different training window lengths n were considered.

Let $\mathbf{r}_t \in \mathbb{R}^p$ denote the net returns of p assets at time t. The *global mean variance portfolio* (GMVP) optimization strategy aims to solve the following optimization problem

$$\underset{\mathbf{w}\in\mathbb{R}^p}{\text{minimize}}\ \mathbf{w}^\top \boldsymbol{\Sigma}\mathbf{w} \quad \text{subject to} \quad \mathbf{1}^\top \mathbf{w} = 1,$$

where $\mathbf{1}$ denotes is a p-vector of ones and $\boldsymbol{\Sigma}$ denotes the covariance matrix of the vector $\mathbf{r}_t$ of returns. The solution is

$$\mathbf{w}_o = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}}. \tag{1}$$

The portfolio allocations are estimated by GMVP using different covariance matrix estimators and different training window lengths n.

Figure 5 depicts the annualized realized risks for of HSI data for time period 2010-2011. We also included in our study the robust GMVP weight estimator proposed in [6] that uses a robust regularized Tyler's M-estimator with a tuning parameter selection that is optimized for the GMVP problem using random matrix theory (RMT). In [6], it was shown hat their estimator outperforms a large selection of regularized covariance matrix estimators both for simulated and real financial data. The method is referred to as **RMT** in the sequel. Furthermore, since the stock data is not very heavy-tailed, we use $q = 0.95$ as the threshold constant of Huber's weight.

As can be seen from Figure 5, for period 2010-2011, the proposed RSCM, CV and RHub estimators achieved the smallest realized risk, outperforming all the other estimators for almost all window lengths. One can state that the proposed RSCM method was the best performing method with lowest realized risk except at window length $n = 110$, when the proposed RHub method obtained the lowest realized risk. The robust RMT of [6] proposed to minimize the risk, however, did not perform better than the proposed shrinkage M-estimators RSCM and RHub for any window length. Furthermore, RMT was also the worst method for a very small window length ($n = 50$). Figure 5 also illustrates that Ledoit-Wolf estimator [7] (denoted LW) had the largest realized minimum risk. In this sense, LW method had the least favourable performance.

For period 2016-2017, the differences between the estimators were not as large as in the period 2010-2011 as can be seen from Figure 6. Here we observed that for many window lengths, the robust RMT method [6] and the proposed RHub method had almost identical behaviour (e.g., for $n \geq 210$). Overall, however, the proposed RSCM and CV methods were the best performing methods.

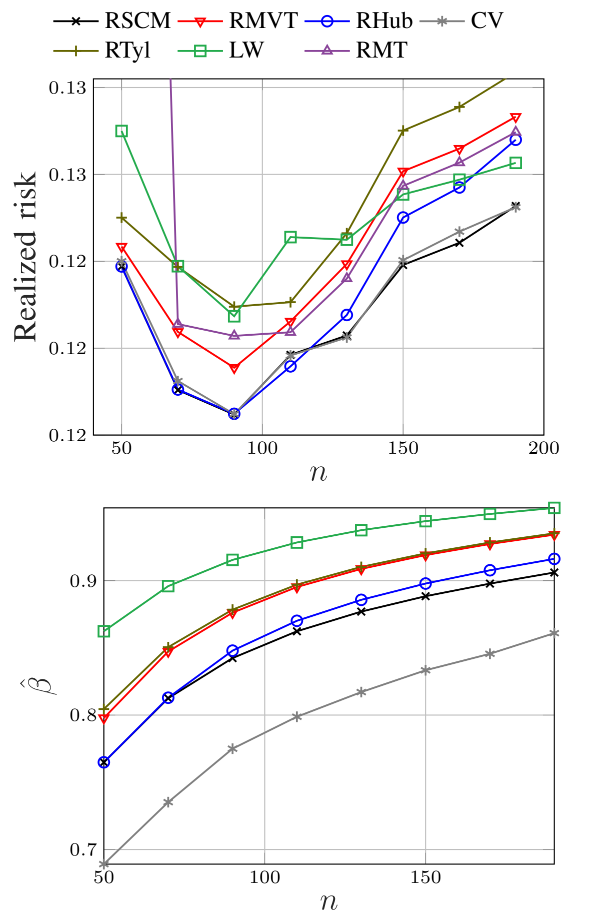

Fig. 5. Annualized realized portfolio risk (top panel) and average $\hat{\beta}$ (bottom panel) achieved out-of-sample for a portfolio consisting of $p = 45$ stocks in HSI for Jan. 4, 2010 to Dec. 24, 2011 containing 491 trading days.

References

[1] E. Ollila, D. P. Palomar, and F. Pascal, "Shrinking the eigenvalues of M-estimators of covariance matrix," *IEEE Transactions on Signal Processing (submitted)*, June 2020.

[2] E. Ollila and E. Raninen, "Optimal shrinkage covariance matrix estimation under random sampling from elliptical distributions," *IEEE Transactions on Signal Processing*, vol. 67, no. 10, pp. 2707–2719, May 2019.

[3] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, "Complex elliptically symmetric distributions: survey, new results and applications," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5597–5625, 2012.

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

[5] A. Coluccia, "Regularized covariance matrix estimation via empirical bayes," *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 2127–2131, 2015.

[6] L. Yang, R. Couillet, and M. R. McKay, "A robust statistics approach to minimum variance portfolio optimization," *IEEE Transactions on Signal Processing*, vol. 63, no. 24, pp. 6684–6697, 2015.

[7] 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.

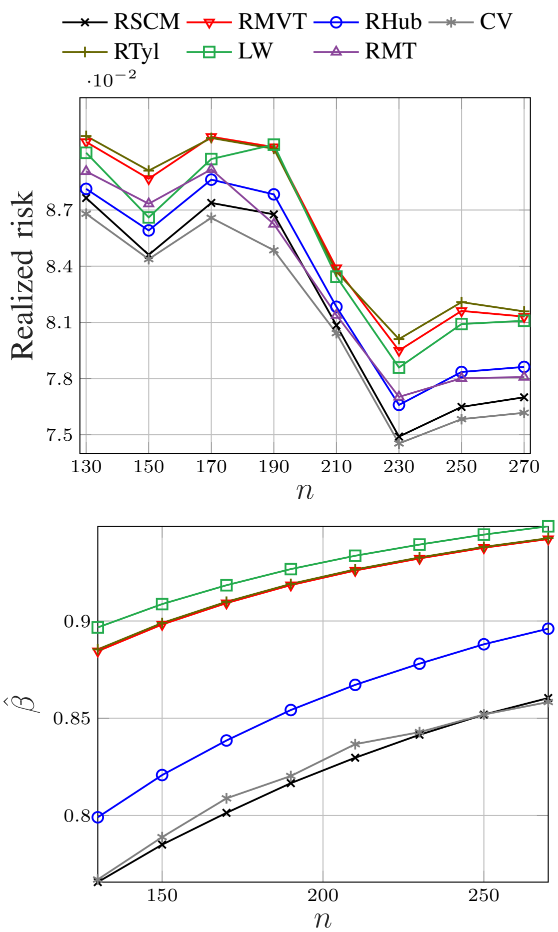

Fig. 6. Annualized realized portfolio risk and average $\hat{\beta}$ achieved out-of-sample for a portfolio consisting of $p = 50$ stocks in HSI for Jan. 1, 2016 to Dec. 27, 2017 containing 491 trading days.