Fusing eigenvalues

Shahab Basiri [1] Esa Ollila [1] Gordana Drašković [2] Frédéric Pascal [2]

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

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

May 16, 2019

Contribution

- A new regularized covariance matrix estimator is proposed.

  Properties:
  - promotes grouping of eigenvalues (fusing eigenvalues),
  - has significantly smaller bias, compared to state-of-the-art methods,
  - less sensitive to the choice of regularization parameter.

Outline

- Preliminaries:
  - Sample Covariance Matrix (SCM),
  - Regularized Sample Covariance Matrix (RSCM).

Outline

- Preliminaries:
  - Sample Covariance Matrix (SCM),
  - Regularized Sample Covariance Matrix (RSCM).
- Fusing eigenvalues of the SCM (eFusion):
  - Properties of the eFusion penalty function,
  - Estimating equations,
  - Choosing the tuning parameters,
  - Iteratively reweighted algorithm for eFusion.

Outline

- Preliminaries:
  - Sample Covariance Matrix (SCM),
  - Regularized Sample Covariance Matrix (RSCM).
- Fusing eigenvalues of the SCM (eFusion):
  - Properties of the eFusion penalty function,
  - Estimating equations,
  - Choosing the tuning parameters,
  - Iteratively reweighted algorithm for eFusion.
- Numerical example.
- Conclusion.

Preliminaries

The Sample Covariance Matrix (SCM)

Given a sample $\mathbf{x}_1, \ldots, \mathbf{x}_n$ of i.i.d. p-variate observations, the Sample Covariance Matrix (SCM) is defined to be

$$\mathbf{S}_n = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top,$$

where $\bar{\mathbf{x}}$ denotes the sample mean.

- Well-known problem: When $n \not\gg p$, the SCM, tends to
  - overestimate the larger eigenvalues, of the true CM,
  - underestimate the smaller eigenvalues of the true CM.
- Possible solution: Regularized or penalized CM estimators have been introduced in a series of papers.

Preliminaries

Regularized Sample Covariance Matrix (RSCM)

- The SCM uniquely minimizes,

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

  over $\boldsymbol{\Sigma} \in \{p \times p$ positive definite symmetric matrices$\}$.
- Regularized Sample Covariance Matrix (RSCM) is then defined as

$$\hat{\boldsymbol{\Sigma}} = \min_{\boldsymbol{\Sigma}}\{l(\boldsymbol{\Sigma}; \mathbf{S}_n) + \eta\Pi(\boldsymbol{\Sigma})\}, \tag{2}$$

  where
  - $\Pi(\boldsymbol{\Sigma})$ denotes a non-negative penalty function,
  - $\eta \geq 0$ being the *regularization parameter*.

Note: The properties of RSCMs depend on the choice of the penalty function.

Preliminaries

Two paradigms in regularizing the SCM:

- Optimally weighted average of the SCM and a well-structured target estimator,
  - e.g.,[Ledoit and Wolf, 2003, Ledoit and Wolf, 2004, Bien and Tibshirani, 2011, Ollila and Raninen, 2018].
- Shrinking the SCM eigenvalues towards each other, and not towards a predefined target estimator,
  - e.g., eLasso [Tyler and Yi, 2018], our proposed eFusion.

Preliminaries

An example of RSCM estimator

- eLasso [Tyler and Yi, 2018], shrinks the eigenvalues towards each other using a penalty function

$$\Pi(\boldsymbol{\Sigma}) = \sum_{j=1}^{p} a_j \log(\lambda_j),$$

  - λ_j: the jth eigenvalue of $\boldsymbol{\Sigma}$,
  - a_j: weights obtained from decreasing quantiles of the Marčenko-Pastur distribution.
- Depending on the choice the regularization parameter, eLasso may result in partitioning the eigenvalues into sub-groups.

Preliminaries

Despite numerical stability, in most RSCM estimators,

- ✗ penalized eigenvalues significantly deviate from the true values,
- ✗ optimum η may not be analytically derived without making prior assumptions on the distribution of the data or model parameters.

The proposed eFusion estimator,

- ✓ has significantly smaller bias than the state-of-the-art methods,
- ✓ poor choices of η appear to be less detrimental, when $\boldsymbol{\Sigma}$ possesses groups of identical eigenvalues.

eFusion

Definition

$$\hat{\boldsymbol{\Sigma}} = \min_{\boldsymbol{\Sigma}} \{ l(\boldsymbol{\Sigma}; \mathbf{S}_n) + \eta \Pi(\boldsymbol{\Sigma}) \},$$

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

- $\rho_c(\cdot) : \mathbb{R} \to \mathbb{R}$ denotes Tukey' s biweight function:

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

eFusion

Definition

$$\hat{\boldsymbol{\Sigma}} = \min_{\boldsymbol{\Sigma}} \{ l(\boldsymbol{\Sigma}; \mathbf{S}_n) + \eta \Pi(\boldsymbol{\Sigma}) \},$$

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

- $r_j = \log(\lambda_j) - \log(\lambda_{j+1})$: the *gaps* between successive log-eigenvalues of $\boldsymbol{\Sigma}$, i.e., $\lambda_1 \geq \cdots \geq \lambda_p > 0$.
- s: the sample *standard deviation* of the gaps between successive log-eigenvalues of $\mathbf{S}_n$, denoted by $r_j^{[0]} = \log(d_j) - \log(d_{j+1})$.

eFusion
Orthogonally invariant

The eFusion penalty function is orthogonally invariant

$$\Pi(\mathbf{Q}\mathbf{\Sigma}\mathbf{Q}^\top) = \Pi(\mathbf{\Sigma}) = \sum_{j=1}^{p-1} \rho_c\Big(\frac{r_j}{s}\Big),$$

for any **Q** in the set of orthogonal matrices of order p.

[Tyler and Yi, 2018, Lemma 2.2]:
If $\Pi(\mathbf{\Sigma})$ is orthogonally invariant, then the RSCM and SCM possess the same set of eigenvectors, with the associated eigenvalues following the same ordering.

eFusion

Orthogonally invariant

$$\hat{\boldsymbol{\Sigma}} = \min_{\boldsymbol{\Sigma}} \left\{ \operatorname{Tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}_n) + \log\{\det(\boldsymbol{\Sigma})\} + \eta \sum_{j=1}^{p-1} \rho_c\left(\frac{r_j}{s}\right) \right\},$$

reduces to

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

where

- $\mathbf{d} = (d_1, \ldots, d_p)^\top$,
- $\boldsymbol{\lambda}^{-1} = (1/\lambda_1, \ldots, 1/\lambda_p)^\top$, with $\lambda_1 \geq \cdots \geq \lambda_p > 0$
- $\mathbf{1}$ is a vector of size $p \times 1$ with all elements equal to one.

eFusion

Estimating equations

$$\hat{\boldsymbol{\lambda}} = \min_{\boldsymbol{\lambda}} \Big\{ \overbrace{\mathbf{d}^\top \boldsymbol{\lambda}^{-1} + \log(\boldsymbol{\lambda})^\top \mathbf{1} + \eta \sum_{j=1}^{p-1} \rho_c\left(\frac{r_j}{s}\right)}^{\mathcal{L}(\boldsymbol{\lambda};\boldsymbol{d},\eta)} \Big\}, \tag{4}$$

$$\mathbf{f}(\boldsymbol{\lambda}) = \mathrm{diag}\left(\mathbf{1} + \frac{\eta}{s}\mathbf{v}\right)\boldsymbol{\lambda} - \mathbf{d} = \mathbf{0}, \tag{5}$$

where

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

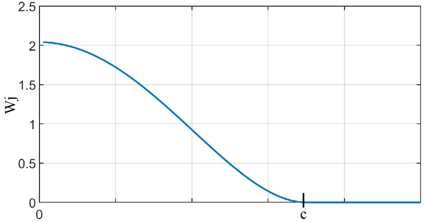
eFusion

Fixed-point estimating equations

By rearranging the terms in (5), we obtain the following system of fixed-point equations.

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

where $w_j = \rho'_c(r_j/s)/(r_j/s)$ are referred to as weights, for $j = 1, \ldots, p$.



eFusion

Finding an optimal tuning parameter for Tukey's biweight function

- We formulate the following binary hypothesis to detect if two successive eigenvalues are equal:

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

- In order to test such a hypothesis:
  - The distribution of $r_j^{[0]} = \log(d_j) - \log(d_{j+1})$ is derived under the null hypothesis $\mathcal{H}_0$.
- The tuning parameter c is obtained as a threshold that assures a given probability of false alarm (P_{fa}).

eFusion

On the choice of tuning parameter c

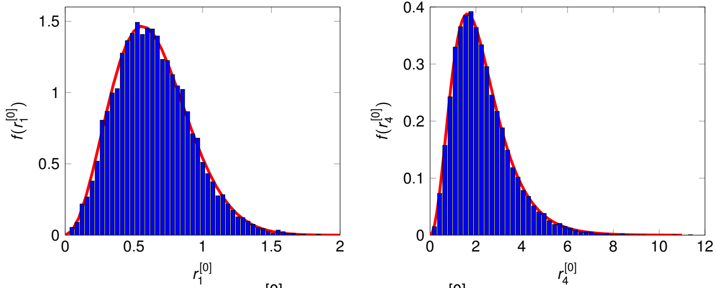


Figure: Empirical distribution of $r_1^{[0]}$ (left panel) and $r_4^{[0]}$(right panel) compared to the corresponding theoretical distribution for $\boldsymbol{\Sigma} = \mathbf{I}$.

- The distribution of $r_j^{[0]}$ has higher variation for larger j (smaller eigenvalues).
- The choice of c is more flexible for small p/n, e.g.,
  for $p = 100$, $n = 700$, $c \in [1.13, 1.50]$
  for $p = 100$, $n = 3000$, $c \in [0.42, 2.96]$.

Algorithm 1: Iteratively reweighted eFusion algorithm

Input : $\mathbf{d}$: Eigenvalues of the SCM $\mathbf{S}_n$;
η: Penalty parameter; c: Tukey tuning constant.

Output : $\hat{\boldsymbol{\lambda}}$: eFusion eigenvalues verifying (13)

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

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

Repeat

Update the gaps: $r_j^{[k]} \leftarrow \log(\lambda_j^{[k]}) - \log(\lambda_{j+1}^{[k]})$,

Update the weights: $w_j^{[k]} \leftarrow \rho_c'(r_j^{[k]}/s)/(r_j^{[k]}/s)$,

Update the eigenvalue estimates for $j = 1, \ldots, p$:

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

$k \leftarrow k + 1$

until *convergence*

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

Numerical example

We compare the performance of the proposed eFusion with eLasso.

Problem setting:

- Random sample of size $n = 3000$ from a $p = 100$ dimensional multivariate normal distribution,
- The covariance matrix $\boldsymbol{\Sigma}$ has
  - 40 eigenvalues equal to 20,
  - 30 equal to 10,
  - 30 equal to 2.

eFusion VS eLasso

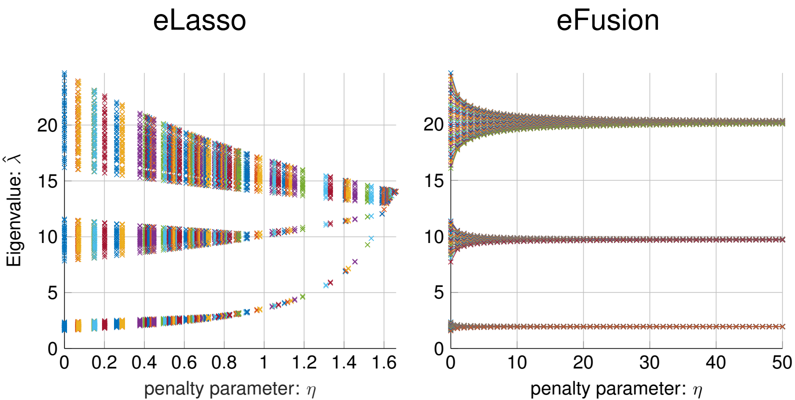


Figure: Grouping of eigenvalues using eLasso and eFusion.

Conclusion

- We introduced a new regularized covariance matrix estimator based on the novel eFusion penalty function that promotes grouping of eigenvalues.
- The topic of choosing the tuning parameter c was addressed along with simulation studies.
- An efficient iteratively reweighted algorithm was proposed for computing the estimator.
- The main benefits of the eFusion:
  - Unbiasedness (accurate grouping),
  - Robustness to the choice of the penalty parameters.

References

Bien, J. and Tibshirani, R. J. (2011).
Sparse estimation of a covariance matrix.
Biometrika, 98(4):807–820.

Ledoit, O. and Wolf, M. (2003).
Improved estimation of the covariance matrix of stock returns with an application to portfolio selection.
Journal of Empirical Finance, 10(5):603–621.

Ledoit, O. and Wolf, M. (2004).
A well-conditioned estimator for large-dimensional covariance matrices.
J. Multivar. Anal., 88(2):365–411.

Ollila, E. and Raninen, E. (2018).
Optimal shrinkage covariance matrix estimation under random sampling from elliptical distributions.
arXiv preprint arXiv:1808.10188.

Tyler, D. E. and Yi, M. (2018).
Lassoing Eigenvalues.
arXiv:1805.08300v1.