M-estimators of scatter with eigenvalue shrinkage

Esa Ollila

Department of Signal Processing and Acoustics
Aalto University, Finland

ICASSP SPTM-L1.3

Tuesday, May 5th 2020, 12:10 - 12:30

Aalto University

Joint work with

Daniel P. Palomar

Frédéric Pascal

Menu

Covariance estimation problem

- $\mathbf{x}$: p-variate (centered) random vector.
- $\mathbf{x}_1, \ldots, \mathbf{x}_n$ i.i.d. realizations of $\mathbf{x}$ and assume $n > p$.
- The unknown covariance matrix $\mathbb{E}\big[\mathbf{x}\mathbf{x}^\top\big] \in \mathbb{S}_{++}^{p \times p}$ is commonly esimated using the sample covariance matrix (SCM)

$$\mathbf{S} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^\top.$$

- When $p = O(n)$, then $\mathbf{S}$ is very inaccurate estimator.
- Regularization: **regularized (shrinkage) SCM (RSCM)**

$$\mathbf{S}_\beta = \beta \mathbf{S} + (1 - \beta) \frac{\mathrm{tr}(\mathbf{S})}{p} \mathbf{I}, \quad \beta \in (0, 1]$$

shrinks the eigenvalues towards the grand mean of the eigenvalues.

Non-gaussian data and outliers

- (R)SCM is sensitive to outliers and non-Gaussianity of the data.
- **M-estimators of scatter** [Mar76] provides a robust alternative:

$$\hat{\mathbf{\Sigma}} = \frac{1}{n} \sum_{i=1}^{n} u(\mathbf{x}_i^\top \hat{\mathbf{\Sigma}}^{-1} \mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i^\top ,$$

  where $u : [0, \infty) \to [0, \infty)$ is a non-increasing **weight function**.
- We study the natural alternative to the RSCM

$$\mathbf{S}_{\beta} = \beta \mathbf{S} + (1 - \beta) \frac{\mathrm{tr}(\mathbf{S})}{p} \mathbf{I}, \quad \beta \in (0, 1].$$

$\Rightarrow$ we propose simple and data-adaptive computation of the optimal MMSE parameter β for $\hat{\mathbf{\Sigma}}_{\beta}$ for *any* weight function u.

Non-gaussian data and outliers

- (R)SCM is sensitive to outliers and non-Gaussianity of the data.
- **M-estimators of scatter** [Mar76] provides a robust alternative:

$$\hat{\mathbf{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n} u(\mathbf{x}_i^\top \hat{\mathbf{\Sigma}}^{-1}\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^\top,$$

where $u : [0,\infty) \to [0,\infty)$ is a non-increasing **weight function**.

- We study the natural alternative to the RSCM

$$\hat{\mathbf{\Sigma}}_\beta = \beta\hat{\mathbf{\Sigma}} + (1-\beta)\frac{\mathrm{tr}(\hat{\mathbf{\Sigma}})}{p}\mathbf{I}, \quad \beta \in (0,1].$$

⇒ we propose simple and data-adaptive computation of the optimal MMSE parameter β for $\hat{\Sigma}_\beta$ for *any* weight function u.

Non-gaussian data and outliers

- (R)SCM is sensitive to outliers and non-Gaussianity of the data.
- **M-estimators of scatter** [Mar76] provides a robust alternative:

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n} u(\mathbf{x}_i^\top \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^\top,$$

  where $u : [0, \infty) \to [0, \infty)$ is a non-increasing **weight function**.
- We study the natural alternative to the RSCM

$$\hat{\boldsymbol{\Sigma}}_\beta = \beta\hat{\boldsymbol{\Sigma}} + (1-\beta)\frac{\mathrm{tr}(\hat{\boldsymbol{\Sigma}})}{p}\mathbf{I}, \quad \beta \in (0, 1].$$

$\Rightarrow$ we propose simple and data-adaptive computation of the optimal MMSE parameter β for $\hat{\boldsymbol{\Sigma}}_\beta$ for *any* weight function u.

Menu

Shrinkage M-estimators of scatter

$$\hat{\mathbf{\Sigma}}_\beta = \beta \hat{\mathbf{\Sigma}} + (1-\beta)\frac{\mathrm{tr}(\hat{\mathbf{\Sigma}})}{p}\mathbf{I}, \quad \beta \in (0,1].$$

- An M-estimator $\hat{\mathbf{\Sigma}}$ is consistent to underlying population parameter, defined as a solution to

$$\mathbf{\Sigma}_0 = \mathbb{E}\big[u(\mathbf{x}^\top \mathbf{\Sigma}_0^{-1}\mathbf{x})\mathbf{x}\mathbf{x}^\top\big].$$

- Ideally, we would like to find

$$\beta_o = \arg\min_\beta \Big\{ \mathrm{MSE}(\hat{\mathbf{\Sigma}}_\beta) = \mathbb{E}\Big[\big\|\hat{\mathbf{\Sigma}}_\beta - \mathbf{\Sigma}_0\|_\mathrm{F}^2\Big]\Big\},$$

but the problem is not tractable due to implicit form of M-estimators.

Surrogate for the M-estimator

- An M-estimator $\hat{\mathbf{\Sigma}}$ can be computed by iterating

$$\hat{\mathbf{\Sigma}}_{k+1} = \frac{1}{n}\sum_{i=1}^{n} u(\mathbf{x}_i^\top \hat{\mathbf{\Sigma}}_k^{-1}\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^\top, \quad k = 0, 1, \ldots$$

- Consider a 1-step M-estimator that starts from true parameter $\mathbf{\Sigma}_0$:

$$\mathbf{C} = \frac{1}{n}\sum_{i=1}^{n} u(\mathbf{x}_i^\top \mathbf{\Sigma}_0^{-1}\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^\top.$$

- Then use a 1-step M-estimator

$$\mathbf{C}_\beta = \beta\mathbf{C} + (1-\beta)[\mathrm{tr}(\mathbf{C})/p]\mathbf{I}$$

  as a proxy for $\hat{\mathbf{\Sigma}}_\beta$.
- Naturally, $\mathbf{C}_\beta$ is fictional, as the initial value $\mathbf{\Sigma}_0$ is unknown.

Approximation of the optimum shrinkage

- We use

$$\beta_o^{\mathrm{app}} = \arg\min_{\beta} \left\{ \mathrm{MSE}(\mathbf{C}_\beta) = \mathbb{E}\Big[\big\| \mathbf{C}_\beta - \mathbf{\Sigma}_0 \big\|_{\mathrm{F}}^2 \Big] \right\}.$$

as approximation for β_o. Similar approach was used in [CWH11].

- **Theorem 1.** Given $\mathbb{E}[\mathrm{tr}(\mathbf{C}^2)] < \infty$, one has that

$$\beta_o^{\mathrm{app}} = \frac{p(\gamma - 1)\eta_o^2}{\mathbb{E}[\mathrm{tr}(\mathbf{C}^2)] - p^{-1}\mathbb{E}[\mathrm{tr}(\mathbf{C})^2]}$$

where $\eta_o = \frac{\mathrm{tr}(\mathbf{\Sigma}_0)}{p}$ is a scale and γ is a **sphericity measure**,

$$\gamma = \frac{p\,\mathrm{tr}(\mathbf{\Sigma}_0^2)}{\mathrm{tr}(\mathbf{\Sigma}_0)^2}.$$

- The expression for β_o^{app} can be simplified assuming that samples are from elliptically symmetric distribution.

Approximation of the optimum shrinkage

- We use

$$\beta_o^{\text{app}} = \arg\min_{\beta} \Big\{ \text{MSE}(\mathbf{C}_\beta) = \mathbb{E}\Big[\big\| \mathbf{C}_\beta - \boldsymbol{\Sigma}_0 \big\|_{\text{F}}^2 \Big] \Big\}.$$

  as approximation for β_o. Similar approach was used in [CWH11].

- **Theorem 1.** Given $\mathbb{E}[\text{tr}(\mathbf{C}^2)] < \infty$, one has that

$$\beta_o^{\text{app}} = \frac{p(\gamma - 1)\eta_o^2}{\mathbb{E}[\text{tr}(\mathbf{C}^2)] - p^{-1}\mathbb{E}[\text{tr}(\mathbf{C})^2]}$$

  where $\eta_o = \frac{\text{tr}(\boldsymbol{\Sigma}_0)}{p}$ is a scale and γ is a **sphericity measure**,

$$\gamma = \frac{p\,\text{tr}(\boldsymbol{\Sigma}_0^2)}{\text{tr}(\boldsymbol{\Sigma}_0)^2}.$$

- The expression for β_o^{app} can be simplified assuming that samples are from elliptically symmetric distribution.

Approximation of the optimum shrinkage

- We use

$$\beta_o^{\mathrm{app}} = \arg\min_{\beta} \left\{ \mathrm{MSE}(\mathbf{C}_\beta) = \mathbb{E}\Big[\big\| \mathbf{C}_\beta - \mathbf{\Sigma}_0 \big\|_{\mathrm{F}}^2 \Big] \right\}.$$

  as approximation for β_o. Similar approach was used in [CWH11].

- **Theorem 1.** Given $\mathbb{E}[\mathrm{tr}(\mathbf{C}^2)] < \infty$, one has that

$$\beta_o^{\mathrm{app}} = \frac{p(\gamma - 1)\eta_o^2}{\mathbb{E}[\mathrm{tr}(\mathbf{C}^2)] - p^{-1}\mathbb{E}[\mathrm{tr}(\mathbf{C})^2]}$$

  where $\eta_o = \frac{\mathrm{tr}(\mathbf{\Sigma}_0)}{p}$ is a scale and γ is a **sphericity measure**,

$$\gamma = \frac{p\,\mathrm{tr}(\mathbf{\Sigma}_0^2)}{\mathrm{tr}(\mathbf{\Sigma}_0)^2}.$$

- The expression for β_o^{app} can be simplified assuming that samples are from elliptically symmetric distribution.

Menu

1. Introduction
2. Shrinkage M-estimators of scatter
3. Shrinkage parameter computation
4. Simulation studies

Elliptically symmetric (ES) distributions

$\mathbf{x} \sim \mathcal{E}_p(\mathbf{0}, \mathbf{\Sigma}, g)$ when its pdf is [FKN90, OTKP12]

$$f(\mathbf{x}) \propto |\mathbf{\Sigma}|^{-1/2} g\big(\mathbf{x}^\top \mathbf{\Sigma}^{-1} \mathbf{x}\big),$$

where

- $\mathbf{\Sigma} \in \mathbb{S}_{++}^{p \times p}$ is the unknown **scatter matrix** parameter
- $g : [0, \infty) \to [0, \infty)$ is **density generator**

- Multivariate normal (MVN) : $g(t) = \exp(-t/2)$
- Multivariate t (MVT) with ν d.o.f : $g(t) = (1 + t/\nu)^{-\frac{p+\nu}{2}}$.
- $\mathbb{E}[\mathbf{x}\mathbf{x}^\top] \propto \mathbf{\Sigma}$

Shrinkage parameter

- The relationship between $\mathbf{\Sigma}$ and the population M-estimator $\mathbf{\Sigma}_0$:

$$\mathbf{\Sigma}_0 = \sigma \mathbf{\Sigma},$$

  where $\sigma > 0$ is a solution to an equation

$$\mathbb{E}\left[\psi\left(\frac{\mathbf{x}^\top \mathbf{\Sigma}^{-1}\mathbf{x}}{\sigma}\right)\right] = p,$$

  where $\psi(t) = u(t)t$ and $u(t)$ is the weight fnc of the M-estimator.
- Define a constant

$$\psi_1 = \frac{1}{p(p+2)}\mathbb{E}\left[\psi\left(\frac{\mathbf{x}^\top \mathbf{\Sigma}^{-1}\mathbf{x}}{\sigma}\right)^2\right]$$

- **Theorem 2.** For $\{\mathbf{x}_i\} \overset{iid}{\sim} \mathcal{E}_p(\mathbf{0}, \mathbf{\Sigma}, g)$

$$\beta_o^{\text{app}} = \frac{\gamma - 1}{(\gamma - 1)(1 - 1/n) + \psi_1(1 - 1/p)(2\gamma + p)/n}$$

  where γ is the sphericity measure.

Shrinkage parameter estimation

- β_o^{app} depends on
  - sphericity measure $\gamma = \frac{p\,\mathrm{tr}(\boldsymbol{\Sigma}_0^2)}{\mathrm{tr}(\boldsymbol{\Sigma}_0)^2}$

  $\Rightarrow$ we use the same estimator as in [ZW16, OR19]:

$$\hat{\gamma}^{\textsf{Ell1}} = \frac{n}{n-1}\left(\frac{p}{n}\sum_{i=1}^{n}\frac{\mathbf{x}_i\mathbf{x}_i^\top}{\|\mathbf{x}_i\|^2} - \frac{p}{n}\right)$$

  - Constant depending on the weight function u:

$$\psi_1 = \frac{1}{p(p+2)}\mathbb{E}\left[\psi\left(\frac{\mathbf{x}^\top\boldsymbol{\Sigma}^{-1}\mathbf{x}}{\sigma}\right)^2\right]$$

  where $\psi(t) = u(t)t$.

  $\Rightarrow$ $\hat{\psi}_1$ is discussed next for different M-estimators.
- Compute $\hat{\boldsymbol{\Sigma}}_\beta$, where $\beta = \beta_o^{\mathrm{app}}(\hat{\gamma}^{\textsf{Ell1}}, \hat{\psi}_1)$.

Regularized SCM (RSCM) M-estimator

- Choose $u(t) \equiv 1$ for all t

$\Rightarrow$ $\hat{\boldsymbol{\Sigma}} = \mathbf{S}$ and $\mathbf{C}_\beta = \mathbf{S}_\beta$ and hence $\beta_o = \beta_o^{\text{app}}$ (approximation is exact)

- The constant ψ_1 is

$$\psi_1 = \frac{\mathbb{E}[(\mathbf{x}^\top \boldsymbol{\Sigma}^{-1} \mathbf{x})^2]}{p(p+2)} = 1 + \kappa$$

where $\kappa = \frac{1}{3}\text{kurt}(x_i)$

$\Rightarrow$ $\hat{\psi}_1 = 1 + \hat{\kappa}$.

- $\mathbf{S}_\beta = \beta \mathbf{S} + (1-\beta)\frac{\text{tr}(\mathbf{S})}{p}\mathbf{I}$, where $\beta = \beta_o^{\text{app}}(\hat{\gamma}^{\text{Ell1}}, \hat{\psi}_1)$.

Regularized Huber's (RHub) M-estimator

- Huber's weight function

$$u_{\mathsf{H}}(t;c) = \begin{cases} 1/b, & \text{for } t \leqslant c^2 \\ c^2/(tb), & \text{for } t > c^2 \end{cases}$$

where $c > 0$ is a user defined *tuning constant* and b is a scaling factor.

- Define a winsorized observation $\mathbf{w}$:

$$\mathbf{w} = \text{wins}(\mathbf{x}) = \frac{1}{\sqrt{b}} \times \begin{cases} \mathbf{x}, & \|\boldsymbol{\Sigma}^{-1/2}\mathbf{x}\|^2 \leqslant c^2 \\ c\dfrac{\mathbf{x}}{\|\boldsymbol{\Sigma}^{-1/2}\mathbf{x}\|}, & \|\boldsymbol{\Sigma}^{-1/2}\mathbf{x}\|^2 > c^2 \end{cases}$$

- Constant ψ_1 is then

$$\psi_1 = 1 + \kappa_{\mathbf{w}}$$

where $\kappa_{\mathbf{w}} = (1/3)\text{kurt}(w_i) \Rightarrow \hat{\psi}_1 = 1 + \hat{\kappa}_{\mathbf{w}}$.

Regularized MVT (RMVT) estimator

- Suppose the data follows MVT distribution with ν d.o.f..
- The ML-weight function is

$$u_{\mathsf{T}}(t;\nu) = \frac{p+\nu}{\nu+t}$$

- The constant ψ_1 is then

$$\psi_1 = \frac{p+\nu}{2+p+\nu}.$$

RMVT estimator:

- Compute estimate $\hat{\nu}$ of ν based on the data.
- Compute the M-estimator $\hat{\mathbf{\Sigma}}$ using $u_{\mathsf{T}}(t;\hat{\nu})$.
- Compute $\hat{\psi}_1$ using $\hat{\nu}$.
- Compute $\hat{\mathbf{\Sigma}}_\beta$, where $\beta = \beta_o^{\mathrm{app}}(\hat{\gamma}^{\mathsf{Ell1}}, \hat{\psi}_1)$.

Menu

Simulation studies

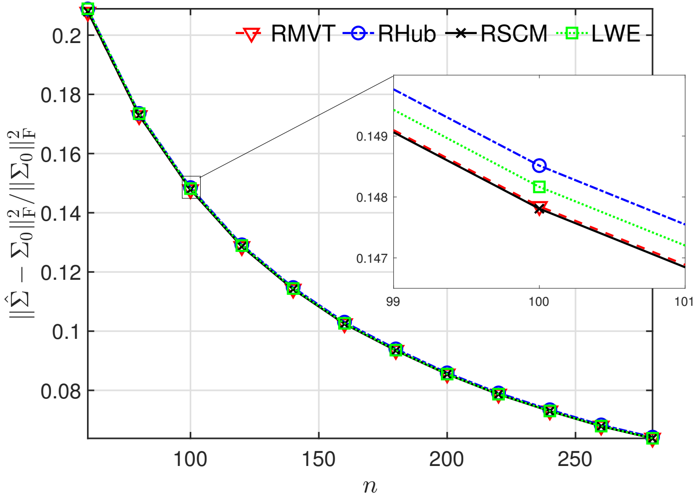
Set-up:

- i.i.d. data from $\mathcal{E}_p(\mathbf{0}, \mathbf{\Sigma}, g)$
- $\mathbf{\Sigma}$ has an AR(1) structure, $(\mathbf{\Sigma})_{ij} = 10\varrho^{|i-j|}$, where $\varrho = 0.6$.
- $p = 40$ and n varies from 60 to 280.
- We compute the normalized MSE (NMSE)

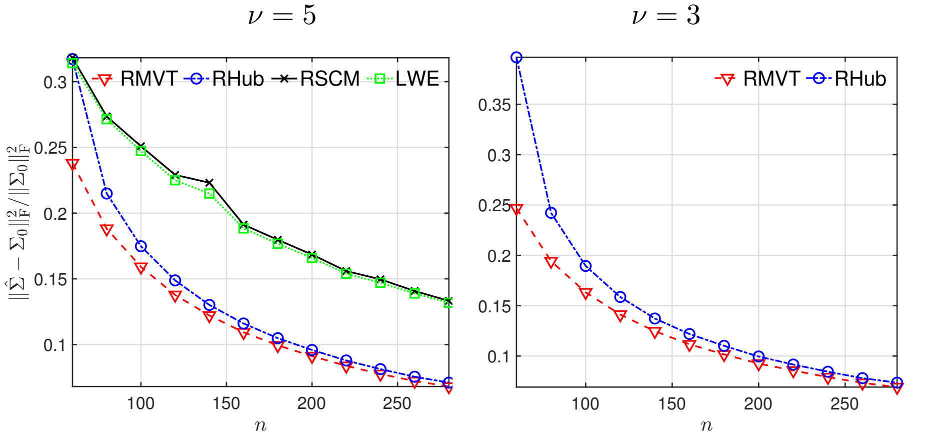
$$\|\hat{\mathbf{\Sigma}}_\beta - \mathbf{\Sigma}_0\|_{\mathrm{F}}^2 / \|\mathbf{\Sigma}_0\|_{\mathrm{F}}^2$$

  averaged over 2000 MC trials.
- We compare with the Ledoit-Wolf estimator [LW04].

MVN (Gaussian) data

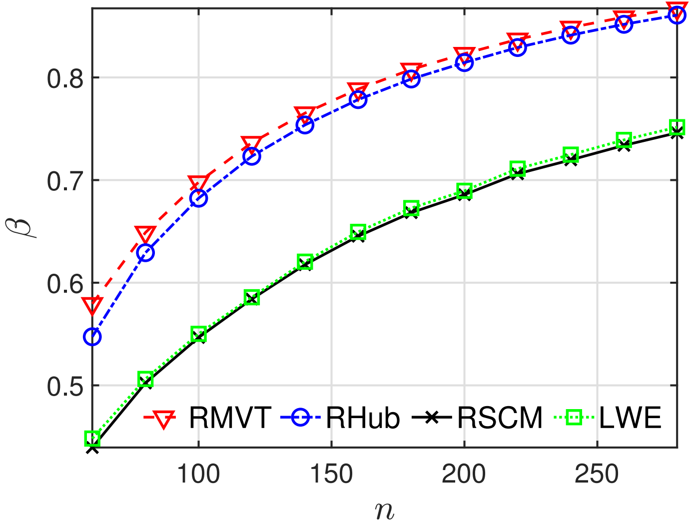


MVT (t-distributed) data



Estimate of shrinkage parameter

MVT data with $\nu = 5$ d.o.f.



What's cooking

A journal extension is currently being finalized ... It includes

- Extension to complex-valued data
- Principled approaches for estimating parameter ν of the MVT distribution.
- Tyler's M-estimator is also considered.
- Advanced approaches for estimating the shrinkage parameter for each specific M-estimator (Huber, Gaussian, Tyler, MVT).
- Application to portfolio optimization: investigation using both synthetic and real stock returns data.
- Journal extension will be sent to ArXiv with Matlab and R codes made publicly available.

References

Yilun Chen, Ami Wiesel, and Alfred O. Hero, *Robust shrinkage estimation of high-dimensional covariance matrices*, IEEE Trans. Signal Process. **59** (2011), no. 9, 4097 – 4107.

Kai-Tai Fang, Samuel Kotz, and Kai-Wang Ng, *Symmetric multivariate and related distributions*, Chapman and hall, London, 1990.

Olivier Ledoit and Michael Wolf, *A well-conditioned estimator for large-dimensional covariance matrices*, Journal of multivariate analysis **88** (2004), no. 2, 365–411.

Ricardo A. Maronna, *Robust M-estimators of multivariate location and scatter*, Ann. Stat. **5** (1976), no. 1, 51–67.

Esa Ollila and Elias Raninen, *Optimal shrinkage covariance matrix estimation under random sampling from elliptical distributions*, IEEE Transactions on Signal Processing **67** (2019), no. 10, 2707–2719.

Esa Ollila, David E. Tyler, Visa Koivunen, and H. Vincent Poor, *Complex elliptically symmetric distributions: survey, new results and applications*, IEEE Trans. Signal Process. **60** (2012), no. 11, 5597–5625.

Teng Zhang and Ami Wiesel, *Automatic diagonal loading for Tyler's robust covariance estimator*, IEEE Statistical Signal Processing Workshop (SSP'16), 2016, pp. 1–5.