Engineering Risk Analysis Group
Prof. Dr Daniel Straub

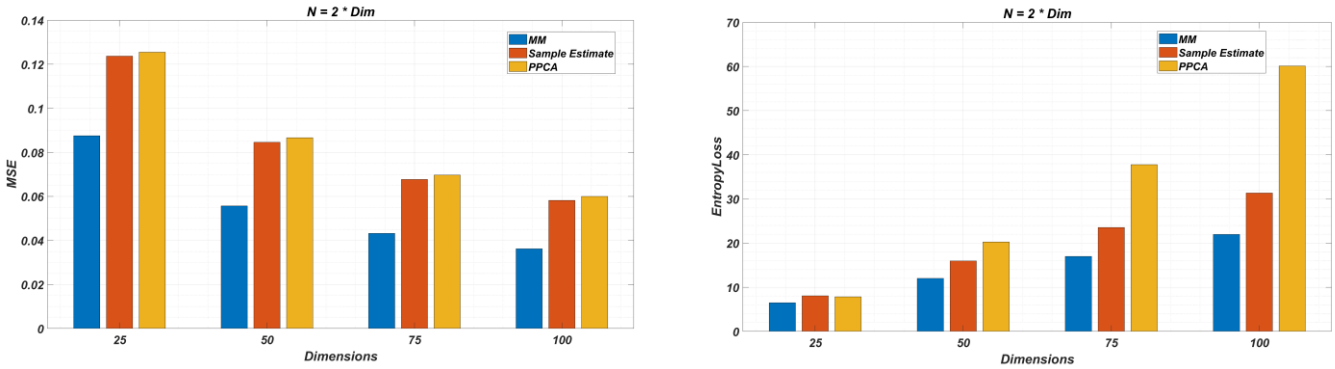
MSc Thesis/ Computational Mechanics

Sparse Estimation of Covariance Matrix

Mohamed Kerish, December 2021

Background

Covariance matrices mainly shows the correlation between random variables; however, in some fields related to Economics or Image processing, for example, sparse covariance matrices where most random variables are marginally independent are often used. In general, estimation of a covariance matrix requires the estimation of free parameters, and in the case of a high dimensional covariance structure, the number of free parameters to estimated would be in the order of $\boldsymbol{O}\ (Dim^2)$**;** therefore, obtaining the sample estimate covariance matrix would be computationally infeasible if the number of samples is small. The main essence of this paper is to explore two tailored methods for estimating sparse covariance structures with a low number of free parameters. The first method is to obtain an estimate using the Probabilistic Principal Component Analysis $\boldsymbol{\Sigma_{PPCA}}$ while the second method is to obtain an estimate using the Majorization-Minimization Algorithm $\boldsymbol{\Sigma_{MM}}$



Left: MSE Comparisons between different estimation methods for number of samples twice the dimension size. Right: Entropy Loss Comparisons between different estimation methods for number of samples twice the dimension size

Methodology

Several covariance structures were used in synthetic data generation and the second step was to perform simulations of 50 runs to measure the discrepancy between the population matrix $\boldsymbol{\Sigma}$ and estimates of the covariance matrices obtained using $\boldsymbol{S}$, $\boldsymbol{\Sigma_{PPCA}}$ and $\boldsymbol{\Sigma_{MM}}$ using two error measures mostly used in the field of comparing covariance matrices, Entropy Loss and Root Mean Square Error

Conclusion

The Majorization-Minimization Algorithm, $\boldsymbol{\Sigma_{MM}}$, outperformed the other estimated covariance matrices $\boldsymbol{S}$ and $\boldsymbol{\Sigma_{PPCA}}$ in terms of obtaining the smallest error values when compared to the population matrix $\boldsymbol{\Sigma}$**.** The integration of the explored methods is highly recommended in the framework of Cross-Entropy based approaches for Reliability Assessment and Bayesian Inference.

Supervised by Dr. Ing. Iason Papaioannou

Dr. Oindrila Kanjilal