1 Results

1.1 Simulation Results

We begin by illustrating the performance of the *CorShrink* method on simulated data. We randomly generated covariance matrices of dimensions p varying in the range 100, 500 and 1000 and then generated n samples from a Multivariate Normal distribution centered at 0 and with the above generated covariance matrix. We considered four choices of $n = 5, 10, 50, 200$ that spans all three scenarios - $n << p$, $n < p$ and $n > p$. We then performed shrinkage on the sample covariance matrix using the three versions of *CorShrink*, GLASSO at 4 different regularization parameter values ranging from low to high shrinkage and the Schäfer-Strimmer method [5] [6] [2].

In Figure 1, we compare the eigenvalues of the shrunk covariance matrices under different shrinkage schemes and different choices of n for $p = 100$. Results corresponding to other choices of p can be checked here [?Link]. Figure 1 shows that the trends of the eigenvalues from the *CorShrink-ML*, *CorShrink-VEM2* and Schäfer-Strimmer shrinkage methods are consistently close to the population eigenvalues across all four choices of n. The *CorShrink-VEM* version performs well for the $n << p$ scenarios, however its performance is not so good for moderate to large values of n. This is probably due to the fact that for larger data, the strong weight of the Dirichlet hyperprior on the null component and the fixed grid of component variances of underlying mixture model on Fisher Z-scores makes the model inflexible to adapt itself to data, a problem that is solved in VEM2 when the component variances are more flexibly chosen by the model. GLASSO for low shrinkage (regularization parameter $\rho = 0.05$) is very close to the sample covariance matrix and for high shrinkage ($\rho = 10$) provides a matrix close to diagonal and therefore is a bad fit to the population covariance matrix. The $\rho = 0.5$ or $\rho = 1$ provide slightly better fit to the population covariance matrix in terms of eigenvalue patterns, but noticeably, for $n << p$ cases, the top eigenvalues (with the highest magnitude) remain very close to that of the sample covariance matrix despite increasing the level of shrinkage and it is the lower order eigenvalues that adapt more rapidly with increasing ρ. Here we must emphasize that usually, the top few eigenvalues are of principal interest to researchers interested in lower dimensional representation, and under $n << p$ scenario, the three versions of the *CorShrink* approach and the Schäfer-Strimmer shrinkage method are more effective than GLASSO in mapping the top eigenvalues close to the ones from the population covariance matrix.

Besides the eigenvalue trends and the top few eigenvalues, another important consideration in comparing these shrinkage methods is how close the eigenvectors from the shrunk covariance matrices are with respect to the population covariance matrix. Table **??** present the average distance between the top 5 eigenvectors of the each shrinkage method with respect to the population covariance. Again we find that the Schäfer-Strimmer, *CorShrink-ML* and *CorShrink-VEM2* produce shrunk covariance matrices closest to the population covariance matrix in terms of the eigen-spaces corresponding to the top eigenvalues. In Figure 2, we plot the distribution of the correlations from the shrunk matrices obtained using different shrinkage methods . We observed that the distribution is more concentrated around 0 for the *CorShrink* models when compared to the GLASSO and Schäfer-Strimmer methods. Additionally, *CorShrink* retains some correlation values with large magnitudes. This characteristic of the *CorShrink* approach would ensure a sparse representation when used in building correlation networks.

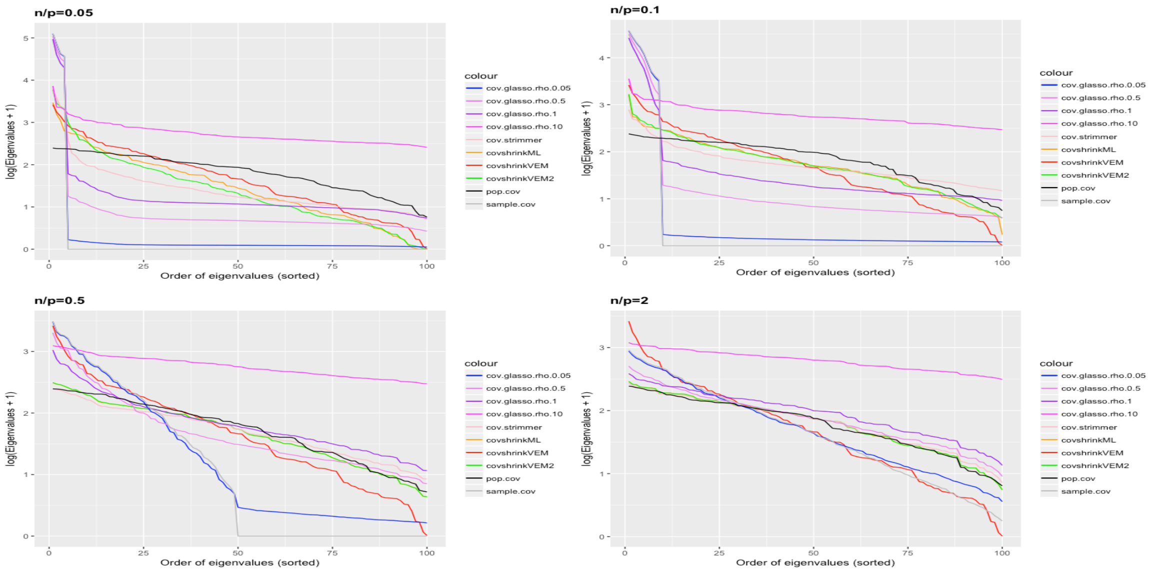

Figure 1: Distribution of the sorted eigenvalues for the shrunk covariance matrices due to the three versions of *CorShrink*- namely *CorShrinkML*, *CorShrinkVEM* and *CorShrinkVEM2*, three versions of GLASSO for three regularization parameters ρ, varying from 0.05, 0.5 and 1 [low to high shrinkage] and the Schäfer-Strimmer shrinkage method, along the distributions of the eigenvalues for the sample covariance and the population covariance matrices.

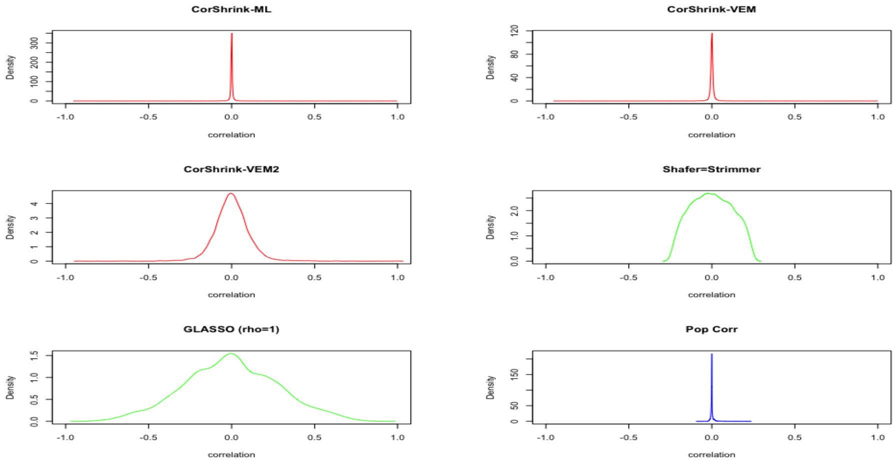

Figure 2: Overall distribution of the correlations after shrinking the covariance matrix using the three models of *CorShrink*, Schäfer-Strimmer method and the GLASSO method for regularization parameter of 1 (which was found to be the best fit in terms of eigenvalue patterns out of the four choices considered). The correlation distributions are compared with the actual correlation distribution from the population covariance matrix.

Table 1: Average Distance of the first 5 eigenvectors of the shrunk covariance matrices from different shrinkage methods versus population covariance matrices.

Methods	$n/p = 0.05$	$n/p = 0.1$	$n/p = 0.5$	$n/p = 2$
CorShrink-ML	0.030	0.034	0.033	0.036
CorShrink-VEM	0.037	0.039	0.057	0.062
CorShrink-VEM2	0.031	0.046	0.043	0.037
Schäfer-Strimmer	0.037	0.054	0.051	0.044
GLASSO ($\rho = 0.05$)	0.083	0.085	0.085	0.069
GLASSO ($\rho = 0.1$)	0.078	0.083	0.084	0.061
GLASSO ($\rho = 0.5$)	0.064	0.081	0.081	0.057
GLASSO ($\rho = 1$)	0.063	0.079	0.080	0.057
Sample cov	0.083	0.084	0.085	0.076

2 Discussion

Our goal here is to highlight the potential of the *CorShrink* models in performing adaptive correlation and covariance shrinkage that has comparable to better performance over the Schäfer Strimmer shrinkage approach in terms of eigen-space and eigenvalue patterns comparisons and that outperforms GLASSO as a correlation shrinkage method irrespective of the choice of regularization parameter used for the latter (see Figure 1 and Table **??**). In terms of computational time, the computation time for the *CorShrink-ML* method is comparable with GLASSO under medium to high shrinkage and Schäfer Strimmer method, while *CorShrink-VEM* and *CorShrink-VEM2* are slower in comparison. For instance, the time taken to run *CorShrink-ML*, *CorShrink-VEM* and *CorShrink-VEM2* on the Deng et al samples data were ?? , ?? and ?? seconds whereas that for GLASSO ($\rho = 1$) and Schäfer-Strimmer methods were ?? .

The *CorShrink* can be easily extended to partial correlation and partial covariance matrices and also leads to efficient computation of the inverse correlation and covariance matrices. The latter would allow one to build causal networks based on the *CorShrink* models and compare them to the GLASSO based causal networks. An example of a causal network on the samples of the Deng et al data using the three shrinkage approaches *CorShrink-ML*, Schäfer-Strimmer and GLASSO ($\rho = 1$) is provided here. This method also opens other areas of applications and extensions that we intend to pursue in future. One can combine Linear Discriminant Analysis and Multiple regression problem with the shrunk covariance matrices obtained from the *CorShrink* approaches, in the same way the Schäfer Strimmer method has been used in these domains [4] [2]. In this paper, we did not consider any additional structure on the covariance matrices. But for structured covariance matrices, one would want to pool the knowledge of the structure into the shrinkage method. For example, for a block covariance matrix, it makes more sense to apply *CorShrink* separately on each block and pool the blocks together.

The codes to fit the *CorShrink* models on data are implemented in an R package **CorShrink** which is available on Github at `https://github.com/kkdey/CorShrink`. It also contains a README demonstrating how these models could were fitted on simulated data. The Deng et al single cell data [1] is available as a R data package with instructions for downloading and loading into R at `https://github.com/kkdey/singleCellRNASeqMouseDeng2014`.

References

[1] Deng Q, Ramskold D, Reinius B, Sandberg R. 2014. Single-Cell RNA-Seq Reveals Dynamic, Random Monoallelic Gene Expression in Mammalian Cells. *Science*. 343 (6167) 193-196.

[2] Schäfer J and Strimmer K. 2005. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statist. Appl. Genet. Mol. Biol*.4.32.

[3] Schäfer J and Strimmer K. 2005. An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics*. 21: 754-764.

[4] Xu P., Brock GN, Parrish RS. 2009. Modified linear discriminant analysis approaches for classification of high-dimensional microarray data. *Computational Statistics* & *Data Analysis*. 53.5.

[5] Friedman J, Hastie T, Tibshirani R. 2008. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*. 9.3.

[6] Witten DM, Friedman JH, Simon N. 2010. New Insights and Faster Computations for the Graphical Lasso. *Journal of Computational and Graphical Statistics*, 20, 4, 892?900.

[7] Stephens M. 2016. False discovery rates: a new deal. *Biostatistics* Advance Access.

[8] Blei DM, Kucukelbir A, McAuliffe JD. 2016. Variational Inference: A Review for Statisticians. https://arxiv.org/pdf/1601.00670.

[9] Beal MJ, Ghahramani Z. The Variational Bayesian EM Algorithm for Incomplete Data: with Application to Scoring Graphical Model Structures *Bayesian Statistics*, 7.