IJMAO

International Journal of Mathematical Sciences and Optimization: Theory and Applications
Vol. 6, No. 2, pp. 800 - 813.
doi.org/10.6084/m9.figshare.13359620.

On Estimation of Sparse Factor Loadings Using Distribution-free Approach

Traore, Bintou [1]*, Adeleke, Ismaila Adedeji [2], Adekeye, Kayode Samuel [3], Nkemnole, Edesiri Bridget [4]

1*, 4. Department of Mathematics, University of Lagos, Nigeria.
2. Department of Actuarial Science and Insurance, University of Lagos, Lagos, Nigeria
3. Department of mathematics, Redeemer's University, Ede, Nigeria
*Corresponding author: bintou.traore64@yahoo.fr

Article Info

Received: 22 April 2020 Revised: 11 June 2020
Accepted: 10 August 2020 Available online: 01 September 2020

Abstract

Sparse Factor Analysis (SFA) is often used for the analysis of high dimensional data, providing simpler pattern of factor loadings by constraining insignificant loadings to be zero. However, existing SFA approaches require the assumption of normality of data since sparse factor loadings are obtained through a likelihood function with additional constraint or penalty function. This work proposes a method for obtaining sparse factor loadings without requiring any distributional assumption. In this method, the orthogonal sparse eigenvectors were computed based on Procrustes reformulation, and thereafter, an iterative procedure was provided to find sparse factor loadings corresponding to the orthogonal sparse eigenvectors. In the end, the proposed method was compared with penalized likelihood factor analysis via nonconvex penalties using simulated data. Results show that sparse factor loadings from both methods provide simpler structure of factor loadings than the structure obtained from standard Exploratory Factor Analysis. In addition, the new method out-performs the penalized likelihood factor analysis via nonconvex penalties as it provides smaller values of MSE even when the two methods have the same level of sparsity.

Keywords: Exploratory Factor Analysis, factor loadings, Sparsity, sparse eigenvectors, regularized maximum likelihood.

MSC2010:47H09, 47H10

IJMAO

International Journal of Mathematical Sciences and Optimization: Theory and Applications
Vol. 6, No. 2, pp. 800 - 813.
doi.org/10.6084/m9.figshare.13359620.

1 Introduction

Sparse Factor Analysis (SFA) is a variant of the standard Factor Analysis (FA) for analyzing correlation structures among observed variables. SFA improves interpretability of factor loadings by avoiding rotational indeterminacy associated with FA models and constraining many entries of the matrix of loadings to be zero. This technique became a very active research area for the analysis of high dimensional data in the last few years (e.g., Carvalho, *et al.*, 2008 ; Engelhardt, and Stephen, 2010 ; Adachi and Trendafilov, 2014 ; Bunte, et al., 2016 ; Zhou, *et al.*, 2016 ; Yamamoto, *et al.*, 2017). However, current literature on SFA suggests the normality assumption of data as the sparse factor loadings are obtained from likelihood function with additional constraint or penalty function. For example, Choi, Zou and Oehler (2011), Ning and Georgiou (2011), Hirose and Yamamoto (2015) and Trendafilov *et al.* (2017) discussed the penalized maximum likelihood procedure to obtain sparse factor loadings. Choi, *et al.* (2011) and Hirose and Yamamoto (2015) proposed the sparse approach of Factor Analysis based on the maximum likelihood with the least absolute shrinkage and selection operator (LASSO), Tibshirani (1996) and nonconvex penalty respectively, and developed Expectation-Maximization (EM) algorithm to compute the sparser solutions. Ning and Georgiou (2011) addressed the standard maximum likelihood (ML) EFA with L1-norm penalty, while Trendafilov, *et al.* (2017) suggested a special reparameterization with additional LASSO penalties introduced into the standard factor analysis problems. Authors have shown, through empirical examples, SFA approaches outperformed the classical FA in terms of interpretability of factor loadings. However, the use of the maximum likelihood function to obtained sparse factor loadings is more complicated since this implies normality assumption which may affect the overall fit especially when the data under investigation deviate substantially from the normal distribution (Trendafilov, *et al.* (2017)). Furthermore, regularized maximum likelihood-based method tends to produce too many zero loadings.
In this article, a distribution-free sparse principal component is suggested to provide sparse factor loadings. Indeed, sparse principal component is used to obtain sparse eigenvectors from which sparse factor loadings are calculated. This study has employed the method proposed by Benidis, Sun, Babu, and Palomar (2016) in obtaining orthogonal sparse eigenvectors of a sample correlation matrix. This method is deemed appropriate as it is very suitable for high-dimensional data and provides sparser orthogonal eigenvectors. Then, an iterative procedure is applied to provide the sparse factor loadings based on obtained orthogonal sparse eigenvectors.
The rest of this paper is organized as follows: Section 2 presents a brief description of the penalized likelihood factor analysis (via nonconvex penalties) proposed by Hirose and Yamamoto (2015). Section 3 focuses on the proposed methodology to find sparse factor loadings. In Section 4, the performance of the proposed method is compared with the regularized maximum likelihood method proposed by Hirose and Yamamoto (2015) through numerical examples using simulated data while discussion of results and conclusion are presented in Sections 5 and 6, respectively.

2 Penalized likelihood factor analysis

The penalized likelihood estimation in factor analysis is the optimization of the likelihood function of the factor model with additional penalty function which can be convex or nonconvex function. The penalized likelihood factor analysis via nonconvex penalties suggested by Hirose and Yamamoto (2015) is based on the following maximization problem:

IJMAO

International Journal of Mathematical Sciences and Optimization: Theory and Applications
Vol. 6, No. 2, pp. 800 - 813.
doi.org/10.6084/m9.figshare.13359620.

$$\max_{\Lambda} l(\Lambda, \Psi) - n \sum_{i=1}^{p} \sum_{j=1}^{k} \rho P(|\lambda_{ij}|) \tag{2.0.1}$$

where,

$l(\Lambda, \Psi) = -\frac{n}{2} p \log(2\pi) + \log|\Sigma| + tr\left(S\Sigma^{-1}\right)$, is the log-likelihood function of the factor model with $\Sigma = \Lambda\Lambda^{'} + \Psi$ and S being the sample correlation matrix; $\Lambda = [\lambda_{ij}]$, is a $p \times k$ matrix of factor loadings ; Ψ is a $p \times p$ diagonal matrix whose diagonal elements, ψ_{ii} are called unique variances; n, p and k are the number of observations, manifest variables, and factors respectively. $P(.)$ is a nonconvex penalty function and $\rho > 0$ is the parameter of regularization. As nonconvex penalty functions, the smoothly clipped absolute deviation (SCAD) and the minimax concave (MC+) introduced by Fan & Li (2001) and Zhang 2010) respectively, are the most widely used to enhance the sparsity of factor loadings. SCAD and MC+ are considered as alternative to the lasso (the L_1 - penalty) and they are defined respectively as follow:

$$P(\theta; \gamma; a) = I(\theta \leq \gamma) + \frac{(a\gamma - \theta)_+}{(a-1)\gamma} I(\theta > \gamma) \; for a > 2 \tag{2.0.2}$$

$$\gamma P(\theta; \gamma; a) = \gamma \left(|\theta| - \frac{\theta^2}{2\gamma a} \right) I(\theta < \gamma a) + \frac{\gamma^2 a}{2} I(|\theta| \geq \gamma a) \tag{2.0.3}$$

where, θ represents the parameter subject to the constraint; γ and a are the parameters of regulation and $I(.)$ is an indicator function for a given value of γ and a .

The maximization problem (1) using SCAD or MC+ can be solved by an iterative procedure which combined the EM and coordinate descent algorithms (see, Hirose and Yamamoto, 2015).

3 Methodology

The extraction or estimation of the sparse factor loadings is generally done via a likelihood function with additional penalty function introduced in the factor analysis model. However, since this approach requires normality of data, we used the sparse eigenvectors to obtain the sparse factor loadings which is independent of the distribution of the data. Thus, the approach requires finding sparse eigenvectors before the computation of the sparse factor loadings.

3.1 Computation of sparse eigenvectors from a sample correlation matrix

There are various methods of computation of sparse eigenvectors of a matrix among which the orthogonal sparse eigenvector via Procrustes reformulation proposed by Benidis, Sun, Babu, andPalomar (2016). Estimation of orthogonal sparse eigenvectors from the sample covariance matrix is done based on the following maximization problem:

$$\underset{V}{maximize} \quad Trace\left(V^T S V D\right) - \sum_{i=1}^{k} \rho_i \|v_i\|_0 \quad Subject \;\; to \;\; V^T V = I_k \tag{3.1.1}$$

IJMAO

International Journal of Mathematical Sciences and Optimization: Theory and Applications
Vol. 6, No. 2, pp. 800 - 813.
doi.org/10.6084/m9.figshare.13359620.

where,

k is the number of eigenvectors to be estimated; p is the number of random variables; V is a $p \times k$ matrix containing the k eigenvectors that correspond to the k first eigenvalues (leading eigenvectors); D is a diagonal matrix of weights to ensure that V contains the leading eigenvectors without an arbitrary rotation. D is chosen such that the diagonal of entries $V^T SVD$ are different from those of $V^T SV$ while keeping the right descending order of the eigenvectors, V at each iteration. D can be a scaled identity matrix since $k \neq p$ in factor analysis; $\|v_i\|_0$ the l_0 -norm function with $v_i \in V\,(i = 1, \ldots, k)$.

The optimization problem (4) is directly intractable due to the nonconvexity and discontinuous implied by the penalty function, $\|v_i\|_0$. To deal with this problem, the following smoothened function presented by Song *et al* (2015) is used to approximate the penalty function:

$$z_\delta^\epsilon(x) = \begin{cases} \frac{x^2}{2\epsilon(\delta+\epsilon)\log\left(1+\frac{1}{\delta}\right)}, & |x| \leq \epsilon \\ \frac{\log\left(\frac{\delta+|x|}{\delta+\epsilon}\right)+\frac{\epsilon}{2(\delta+\epsilon)}}{\log\left(1+\frac{1}{\delta}\right)}, & |x| > \epsilon \end{cases} \tag{3.1.2}$$

where,

x is the variable of interest; $\delta > 0\ and 0 < \epsilon \ll 1$ are parameters that controls the approximation of the penalty function.

Then, substituting the l_0 -norm function, $\|v_i\|_0$ by the function, $z_\delta(x)$ in problem (4) yields a nonconvex but differentiable optimization problem which can be written as:

$$\max_V \quad Trace\left(V^T SVD\right) - \sum_{i=1}^{k} \rho_i \sum_{j=1}^{p} z_\delta^\epsilon(v_{ij}), \quad Subject\ \ to \quad V^T V = I_k \tag{3.1.3}$$

Where,

z_δ^ϵ is the penalty function which approximate the l_0 -norm function; ρ_i , $\delta\ and \epsilon$ are the parameters that controls the whole optimization process with $0 < \delta \leq 1, 0 < \epsilon \ll 1$, $\rho_i > 0$ and $v_{ij} \in V\,(i = 1, \ldots, k; j = 1, \ldots, p)$.

The optimization problem (6) is a non-convex problem which can be solved using minorization-maximization (MM) based mainly on two key quantities, G and H obtained from functions that majorize $Trace\left(V^T SVD\right)$ and $\sum_{i=1}^{k} \rho_i \sum_{j=1}^{n} z_\delta^\epsilon(v_{ij})$ respectively. The quantities G and H are expressed at each iteration, l , as follow:

$$G^{(l)} = SV^{(l)}D \tag{3.1.4}$$

and

$$H^{(l)} = \left[diag\left(w^{(l)} - w_{\max}^{(l)} \bigotimes 1_p\right)\widetilde{v}^{(l)}\right]_{p\times k} \tag{3.1.5}$$

where,

IJMAO

International Journal of Mathematical Sciences and Optimization: Theory and Applications
Vol. 6, No. 2, pp. 800 - 813.
doi.org/10.6084/m9.figshare.13359620.

$w^{(l)} \in R_+^{nk}$; $\widetilde{v}^{(l)} = vec\left(V^{(l)}\right) \in R_+^{nk}$; $w_{\max}^{(l)} \in R_+^k$, with $w_{max,i}$ being the maximum weight that corresponds to the i th eigenvector $\widetilde{v}_i^{(l)}$ $(i = 1, \ldots, k)$.

Each weight, $w_i^{(l)} \in w^{(l)}$ is given by:

$$w_i^{(l)} = \begin{cases} \frac{\rho_i}{2\epsilon(\delta+\epsilon)\log\left(1+\frac{1}{\delta}\right)}, & |\widetilde{v}_i^{(l)}| \leq \epsilon \\ \frac{\rho_i}{2\log\left(1+\frac{1}{\delta}\right)|\widetilde{v}_i^{(l)}|\left(|\widetilde{v}_i^{(l)}|+\delta\right)}, & |\widetilde{v}_i^{(l)}| > \epsilon \end{cases} \tag{3.1.6}$$

The algorithm for performing the optimization problem is described as follow:

1) Initialization: Set the initial iteration, $l = 0$ and choose the starting point $V^{(0)}$.
2) Compute $G^{(l)}$ and $H^{(l)}$ as defined previously.
3) Perform the Singular Value Decomposition (SVD) of $G^{(l)} - H^{(l)}$ to obtained the eigenvectors, L_{left} and L_{right} .
4) Compute the new eigenvectors, $V^{(l+1)} = L_{left}\ (L_{right})^T$.
5) $l = l + 1$
6) Repeat steps 2-5 until convergence.

3.2 Computation of sparse factor loadings

Let Y be a $p \times p$ matrix of scores of observed variables and S , a correlation matrix obtained from Y . The principle of factoring by principal component assumes that if u_{ii} and q_i $(i = 1 \ldots, p)$ are the eigenvalues and eigenvectors of S respectively, then, S and Λ , the matrix of factor loadings, can be expressed by:

$$S = QUQ^T \tag{3.2.1}$$

and

$$\Lambda = V\,(U^*)^{\frac{1}{2}} \tag{3.2.2}$$

where,

$U = [u_{ij}]$, is a $p \times p$ diagonal matrix (diagonal entries u_{ii}); $Q = [q_{ij}]$ is a $p \times p$ matrix containing the p eigenvectors corresponding to the p eigenvalues of S . Q^T is the transpose of Q . $\Lambda = [\lambda_{ij}]$, is a $p \times k$ matrix $(-1 \leq \lambda_{ij} \leq 1)$; p and k are the number of observed variables and common factors, respectively. U^* , is a $k \times k$ diagonal matrix whose diagonal contains the first k eigenvalues, $u_{ii}\,(i = 1, \ldots, k)$ and V is the $p \times k$ matrix of eigenvectors corresponding to the first k eigenvalues of S .

IJMAO

International Journal of Mathematical Sciences and Optimization: Theory and Applications
Vol. 6, No. 2, pp. 800 - 813.
doi.org/10.6084/m9.figshare.13359620.

When q_{ij} 's are not the true eigenvectors of S , then, $Q^T SQ \neq U$ and computing factor loadings with the true eigenvalues of **S** is inappropriate because such estimation may produce out of range values leading to incorrect estimate of the sample correlation matrix. Moreover, the matrix ($Q^T SQ$) can be used to compute the sparse factor loadings that fit with these eigenvectors.

First, it should be noted that the fundamental concept of factor analysis requires that $\Lambda\Lambda^T$ must be Gramian Matrix and of rank k (Everitt, 1984; Thompson, 2004). Subsequently, this implies that the matrix product of the sparse factor loadings by its transpose should be symmetric and positive semi-definite with rank k .

Now, let $V^s \in R^{p\times k}$ be the sparse orthogonal eigenvectors generated from S ($V^s \neq Q$, true eigenvectors of **S**) and QUQ^T with Q and U , being the $p \times p$ eigenvectors matrix and the $p \times p$ diagonal matrix containing the eigenvalues of **S** respectively. The matrix product $(V^s)^T SV^s$ can be expressed as:

$$(V^s)^T SV^s = (V^s)^T \left[QUQ^T\right] V^s \tag{3.2.3}$$

$$(V^s)^T SV^s = (V^s)^T QUQ^T V^s \tag{3.2.4}$$

As V^s and Q are both orthogonal eigenvectors ($V^s not \perp Q$), we can see that,

$$(V^s)^T QQ^T V^s = I_k \tag{3.2.5}$$

Accordingly, $(V^s)^T Q \in R^{kp}$ is an orthogonal matrix and $Q^T V^s \in R^{pk}$ is its transpose. Let $T = (V^s)^T Q$, then,

$$(V^s)^T SV^s = TU^{1/2}U^{1/2}T^T \tag{3.2.6}$$

$$(V^s)^T SV^s = \left(TU^{1/2}\right)\left(TU^{1/2}\right)^T \tag{3.2.7}$$

The matrix product $\left(TU^{1/2}\right)\left(TU^{1/2}\right)^T$ is a Gramian matrix. Therefore, the diagonal entries of the matrix product $\left(TU^{1/2}\right)\left(TU^{1/2}\right)^T$ are strictly positive, but the off-diagonal elements are necessarily neither positive nor equal to zero ($(V^s)^T Q \neq I$). Moreover, these diagonal entries are not unavoidably in decreasing order, but they can be used as eigenvalues with the sparse eigenvectors, V^s to obtain the sparse factor loadings, Λ^s because:

1) The number of coefficients on the diagonal of $\left[\left(TU^{1/2}\right)\left(TU^{1/2}\right)^T\right]$ equals the number of columns of V^s ;
2) The matrix product, $V^s\left[\left(diag\left[\left(TU^{1/2}\right)\left(TU^{1/2}\right)^T\right]\right)^{\frac{1}{2}} I_k\right].\left[\left(diag\left[\left(TU^{1/2}\right)\left(TU^{1/2}\right)^T\right]\right)^{\frac{1}{2}} I_k\right](V^s)^T$, is a $p \times p$ symmetric matrix of rank k ;

IJMAO

International Journal of Mathematical Sciences and Optimization: Theory and Applications
Vol. 6, No. 2, pp. 800 - 813.
doi.org/10.6084/m9.figshare.13359620.

3) The diagonal of $\left[\left(TU^{1/2}\right)\left(TU^{1/2}\right)^{T}\right]$ has positive values which can be arranged in decreasing exactly like the eigenvalues of a positive semi-definite matrix.

Subsequently, the matrix of factor loadings, Λ^s can be expressed by:

$$\Lambda^s = V^s\left[\left(diag\left[\left(TU^{1/2}\right)\left(TU^{1/2}\right)^{T}\right]\right)^{\frac{1}{2}} I_k\right] \tag{3.2.8}$$

which is equivalent to

$$\Lambda^s = V^s\left[\left(diag\left[\left(V^s\right)^T SV^s\right]\right)^{\frac{1}{2}} .I_k\right] \tag{3.2.9}$$

where,

$V^s \in R^{p\times k}$ is the sparse eigenvectors generated from S ; Q and U are the $p\times p$ eigenvectors matrix and the $p\times p$ diagonal matrix containing the eigenvalues of S respectively; $(V^s)^T$ and Q^T are the transpose of V^s and Q respectively; $T = (V^s)^T Q$.

However, one major problem with the use of the diagonal of $\left[\left(V^s\right)^T SV^s\right]$ for computing sparse factor loadings, Λ^s , is the fact that some coefficients, λ_{ij}^s of the latter may not lie between -1 and 1 and this may lead to improper solutions. To avoid this issue, we introduce a parameter for adjusting the square root of the quantity, $\left(V^s\right)^T SV^s$ such that all coefficients of the sparse factor loadings lie between -1 and 1. Let $\tau \geq 0$ be the adjustment parameter, if a coefficient, $\lambda_{ij}^s \in \Lambda^s$ is out of the normal range, then, the sparse factor loadings, Λ^s is re-expressed as:

$$\Lambda^{s*} = V^s\left[\left(diag\left[\left(V^s\right)^T SV^s\right]\right)^{\frac{1}{2+\tau}} .I_k\right] \tag{3.2.10}$$

Having the sparse eigenvectors, V^s and the sample correlation matrix, S , the estimation of sparse factor loadings becomes very simple if $\lambda_{ij}^s \in [-1;1]$. But, since this case is often not likely, an iterative procedure can be used to find $\lambda_{ij}^s \in [-1;1]$. This iteration procedure consists in choosing a very small value of τ so that the quantity $2+\tau$ is very close to 2 and computing the sparse factor loadings, Λ^{s*} . Then, sequentially increase τ and compute again the sparse factor loadings (for each sequence) until all coefficients of the latter lie between -1 and 1. The coefficients of Λ^{s*} decrease only with respect to τ since the quantity $\frac{1}{2+\tau}$ decreases when τ increases while V^s and S remain constant.

3.3 Algorithm for Computation of Sparse Factor Loadings

Based on the details provided in Section 3.2, the algorithm for computing the sparse factor loadings Λ^{s*} is presented in the following steps:

Step 1: Compute the sparse eigenvectors, V^s from the sample correlation matrix, S using Benidis *et al.*, (2016) algorithm described previously.

Step 2: Choose τ such that $2+\tau \approx 2$ and set $l = 0$; $\tau^l = \tau$;

IJMAO

International Journal of Mathematical Sciences and Optimization: Theory and Applications
Vol. 6, No. 2, pp. 800 - 813.
doi.org/10.6084/m9.figshare.13359620.

Step 3: Compute $\Lambda^l = V^s \left[\left(diag \left[(V^s)^T S V^s \right] \right)^{\frac{1}{2+\tau^l}} I_k \right]$;

Step 4: $l = l + 1$ and $\tau^{l+1} = (l+1)\,\tau^l$.

Step 5: Repeat Steps 3 and 4 until all coefficients of Λ^l lie between -1 and 1;

Step 6: Return Λ^l .

4 Numerical application

The main purpose of this application is to see how the proposed method referred as "Method 1" perform comparatively to the method proposed by Hirose & Yamamoto (2015) referred as "Method 2" in term of sparsity level and error introduce by the sparsity. Accordingly, we generated 50 sample of $p = 20$ random variables following multivariate normal distribution with vector of mean, $\mu = 0$ and Σ lying between -1 and 1. We choose $n = 200$ and Σ such that the correlations between the random variables have four disconnected cliques of size 5.

To obtain this correlation pattern, each sample (20 random variables) is divided into 4 sets of 5 random variables. For each set, five random variables are separately generated following normal distribution with $\mu = 0$ and $\sigma = 1$. In order to have strong correlation between random variables within the same set, we first generated a 5×5 symmetric correlation matrix, $\Sigma_i \, (i = 1, \ldots, 4)$ with off-diagonal entries equal to $\pm$ 0.999 (with random signs). Then, the data matrix of each set is multiplied by the upper triangular matrix of the Cholesky decomposition of Σ_i to provide the final correlated random variables. By combining these new random variables, we finally got a 200 $\times$20 data matrix for each sample where random variables belonging to the same set are strongly correlated while correlations between different sets are weak. Figure 1 shows how the structure of the correlation matrix of each generated sample is generally presented.

For each sample, the correlation matrix, $S_j \, (j = 1, \ldots, 50)$ is computed and the number of eigenvectors, k_j , to be estimated is fixed with respect to the most important eigenvalues of each S_j . Figure 2 displays the plots of each eigenvectors with respect to its corresponding eigenvalues for each of the 50 samples. Following these plots, each curve decreases significantly from the 4^{th} eigenvectors. Accordingly, $k = 4$ for each sample because only the first four eigenvectors correspond to the largest eigenvalues while the remaining eigenvectors have eigenvalues almost equal to zero.

We choose $\epsilon = 10^{-9}$, $\tau^0 = 10^{-9}$, $\delta = 0.5$ and $d_{ii} = w_1 - (i-1)\,w$, with $i = 2, \ldots, k-1$, $w_1 = 1$, $w = \frac{1-0.5}{k-1}$ and $d_{kk} = 0.5$ $(d_{ii} \in D)$. However, for Method 1, ρ is taken at six levels for each sample notably: 0.0001, 0.01, 0.1, 0.5, 1, 10. With respect to Method 2, γ is taken at four levels namely, $1, 50, 100$, and ∞ for each of the six levels of ρ . Subsequently, for each sample, Method 1 is performed six times leading to 300 cases for all samples (6×50). While Method 2 is carried out $6 \times 50 \times 4$ times leading to 300 cases for each level of γ .

The two methods are compared in terms of sparsity level and mean squared error (MSE). The sparsity level is given by the number of zero in the matrices of loadings, Λ^s while the error is represented by the average squared difference between S and $\Lambda^s(\Lambda^s)^T$.

The summary (minimum, mean and maximum) of the number zero and MSE provided by the 50 samples taken at different levels of ρ and γ is presented in table 1 and 2 for Method 1 and 2 respectively. The comparison of MSE with respect to the number of zero provided by each case is

IJMAO

International Journal of Mathematical Sciences and Optimization: Theory and Applications
Vol. 6, No. 2, pp. 800 - 813.
doi.org/10.6084/m9.figshare.13359620.

given in figure 3.

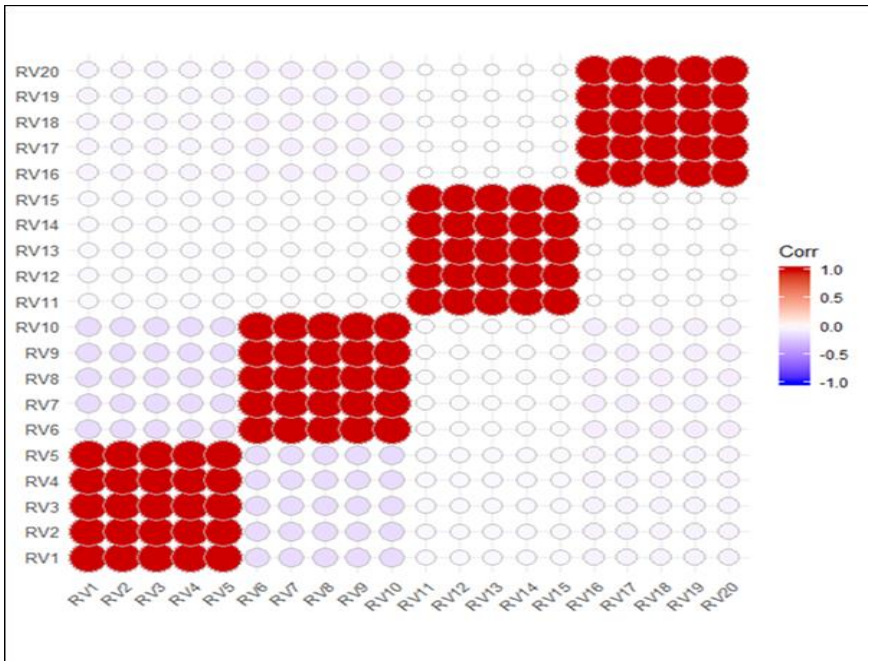


Figure 1: The structure or pattern of the sample correlation matrix.

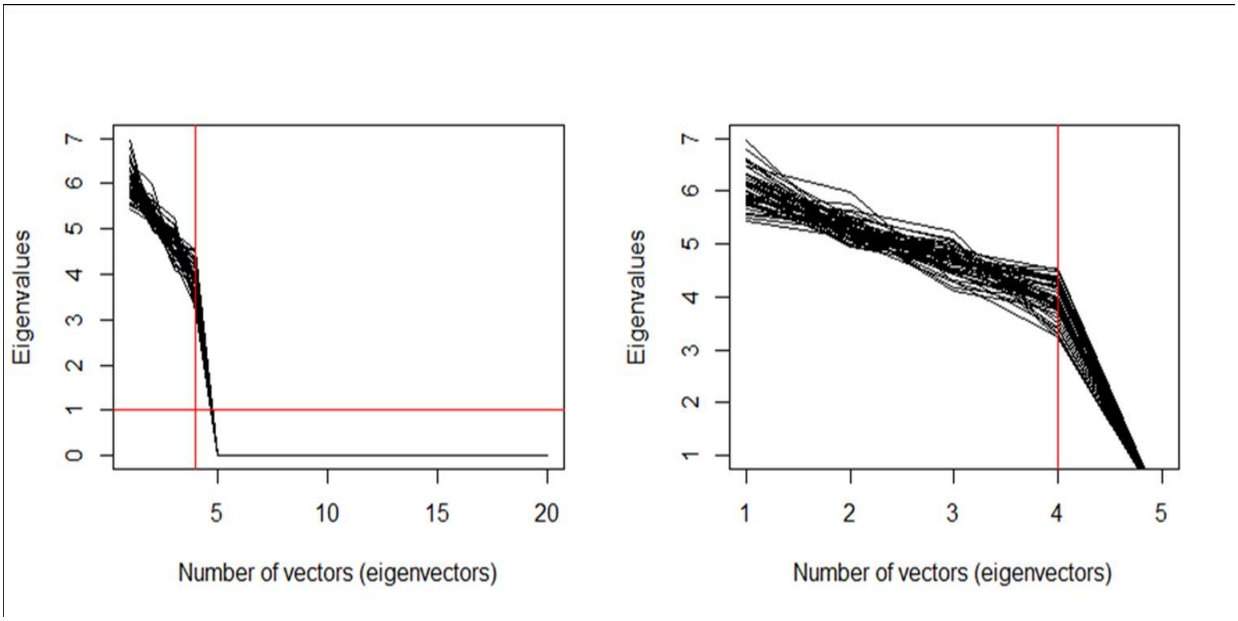


Figure 2: Plot of successive eigenvalues for the 50 sample correlation matrices.

IJMAO

International Journal of Mathematical Sciences and Optimization: Theory and Applications
Vol. 6, No. 2, pp. 800 - 813.
doi.org/10.6084/m9.figshare.13359620.

Table 1: Summary of the number of zero and MSE for Method 1.

		10^-4	**0.01**	**0.1**	**0.5**	**1**	**10**
Number	*min*	0	0	15	50	60	60
of	*mean*	0.12	15.22	45.1	59.2	60	74.28
Zeros	*max*	3	40	60	60	60	76
	min	0	0.002	0.0182	0.0283	0.0469	0.119
MSE	*mean*	0.0002	0.0133	0.0537	0.0807	0.0842	0.4701
	max	0.0011	0.0316	0.0825	0.1189	0.1274	0.5041

Table 2: Summary of the number of zero and MSE for Method 2.

gamma=1		10^-4	**0.01**	**0.1**	**0.5**	**1**	**10**
Number	*min*	51	51	51	60	60	60
of	*mean*	58	58	58	60	60	65
Zero	*max*	60	60	60	60	60	70
	min	0.027	0.027	0.027	0.0386	0.2277	0.3899
MSE	*mean*	0.078	0.078	0.078	0.0821	0.2397	0.4241
	max	0.1254	0.1254	0.1254	0.1374	0.2631	0.4588

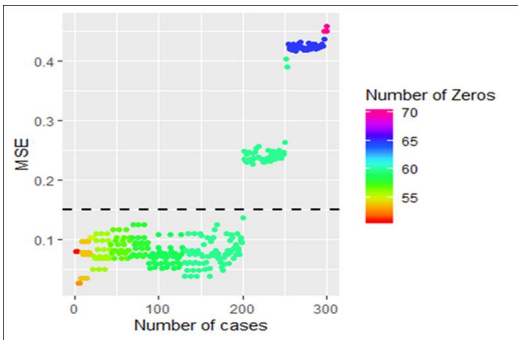
gamma=50		10^-4	**0.01**	**0.1**	**0.5**	**1**	**10**
Number	*min*	7	7	39	60	60	60
of	*mean*	13	13	47	60	66	66
Zero	*max*	19	19	54	65	70	70
	min	0.0296	0.0296	0.127	0.3007	0.3653	0.3653
MSE	*mean*	0.0405	0.0405	0.1452	0.3102	0.4103	0.4103
	max	0.0513	0.0513	0.1757	0.3641	0.4528	0.4528

gam=100		10^-4	**0.01**	**0.1**	**0.5**	**1**	**10**
Number	*min*	6	6	38	59	60	60
of	*mean*	9	10	48	60	65	65
Zero	*max*	17	20	55	60	70	70
	min	0.0147	0.0147	0.1401	0.299	0.3651	0.3651
MSE	*mean*	0.0291	0.0341	0.16	0.309	0.4083	0.4083
	max	0.0342	0.0404	0.1882	0.3244	0.4472	0.4472

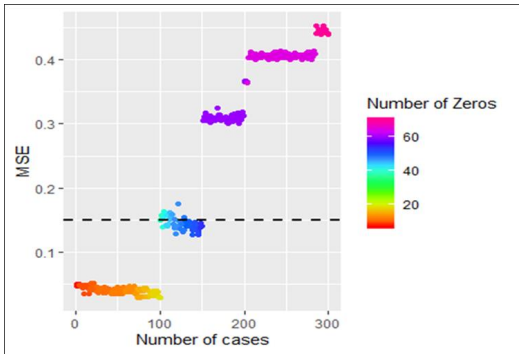
IJMAO

International Journal of Mathematical Sciences and Optimization: Theory and Applications
Vol. 6, No. 2, pp. 800 - 813.
doi.org/10.6084/m9.figshare.13359620.

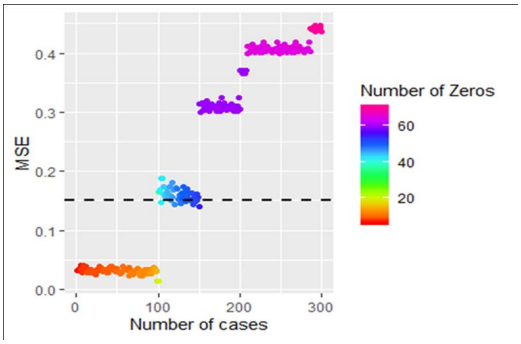
gamma=Inf		10^-4	**0.01**	**0.1**	**0.5**	**1**	**10**
Number	*min*	3	4	40	60	60	60
of	*mean*	5	9	49	60	65	65
Zero	*max*	8	15	56	65	70	70
	min	0.005	0.0404	0.1654	0.3023	0.3648	0.3648
MSE	*mean*	0.0057	0.0471	0.1831	0.3151	0.4093	0.4093
	max	0.0071	0.0583	0.2233	0.3707	0.4475	0.4475



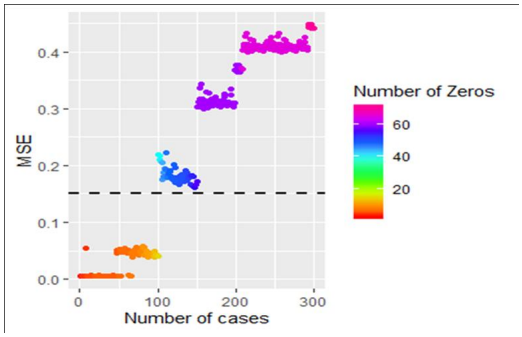
$\gamma = 1$



$\gamma = 50$



$\gamma = 100$



$\gamma = \infty$

IJMAO

International Journal of Mathematical Sciences and Optimization: Theory and Applications
Vol. 6, No. 2, pp. 800 - 813.
doi.org/10.6084/m9.figshare.13359620.

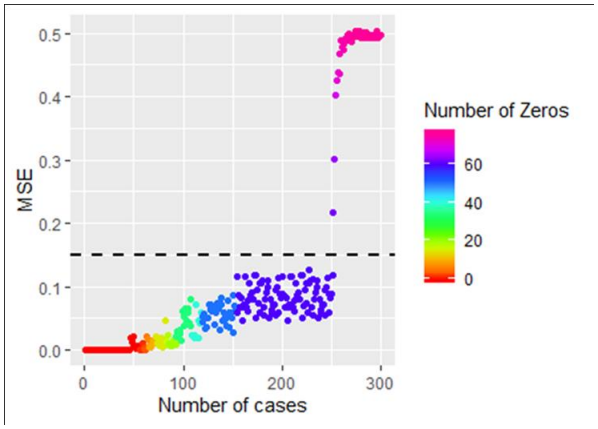


Figure 3: Plots of MSE versus number of cases for Method 1 (first 4 plots at the top) and Method 2 (last plot at the bottom).

5 Discussion of Results

Table 1 and 2 presents the summary of the number of zeros and MSE obtained by varying ρ for Method 1 and ρ and γ for Method 2 respectively. Following these tables, both methods provide sparser factor loadings according to the chosen parameters of regulation ρ and γ . For Method 1, the number of zero increases as the parameter ρ increases. While for Method 2, the level of sparsity increases when ρ increases and γ decreases simultaneously. With $\rho = 10^{-4}$, Method 1 provides very few zero loadings whereas small number of zeros are obtained in Method 2 only when $\rho = 10^{-4}$ and γ tend toward infinity. Furthermore, the number of zeros is particularly high for every level of ρ when $\gamma = 1$, but almost constant for each level of γ with $\rho \geq 0.5$. This is evidenced by the first part and last 3 columns of table 2 respectively. For some values of ρ and γ , both methods produce the same number zero.

With respect to MSE, figure 3 displays errors introduced by the sparsity of each case using Method 1 and 2. The first four plots represent values of MSE for different cases taken at the four level of γ respectively using Method 2, whereas the last plot indicates MSE's values provided by Method 1. For Method 2, except for $\gamma = 1$, most cases (almost 200 over 300 cases) have more than 20 zeros with MSE's values above 0.1 as shows by plots at the top right and middle of figure 3. But, for $\gamma = 1$, most cases (almost 250 over 300) have number of zeros between 50 and 60 with MSE's values less than 0.15 for some and above 0.2 for others (see top left plot of figure 3). In addition, good levels of sparsity with small MSE's values are provide with $\rho \leq 0.5$ for every level of γ . This is also evidenced by different values of MSE and corresponding number of zeros given in table 2. The last plot obtained from Method 1 indicates that all cases with number of zeros less or equal to 60 have MSE's values under 0.15. Compared to Method 2, Method 1 seems to be better because it gives smaller values of MSE even when both have the same level of sparsity.

IJMAO

International Journal of Mathematical Sciences and Optimization: Theory and Applications
Vol. 6, No. 2, pp. 800 - 813.
doi.org/10.6084/m9.figshare.13359620.

6 Conclusion

Based on the obtained results, we found that both methods provide sparser factor loadings and the sparsity level depends strongly on the parameters of regulation required for each method. The level of sparsity increases as ρ increases for Method 1 whereas for Method 2, it increases when ρ increases and γ decreases simultaneously. The proposed method performs better than Method 2 since it provides smaller values of MSE even when both methods have the same level of sparsity or number of zero. However, this does not prevent the loss of information which increases with the level of sparsity.

Competing Financial Interests

The authors declare no competing financial interests.

References

[1] Adachi, K., & Trendafilov, N. T. Sparse orthogonal facto analysis. In: Carpita M., Brentari E., Qannari E. (eds) Advances in Latent Variables. Studies in Theoretical and Applied Statistics. *Springer, Cham*, 227-239, (2014).

[2] Bai, J., & Liao, Y. Efficient estimation of approximate factor models via penalized maximum likelihood. *Journal of Econometrics*, **191(1)**, 1–18, (2016).

[3] Bartholomew, D.,Knott, M., & Moustaki, I. Latent Variable Models and Factor Analysis. *John Wiley & Sons, Ltd*, (2011).

[4] Basilevsky, A.T. Statistical Factor Analysis and Related Methods. *John Wiley & Sons, Inc.*, (1994).

[5] Benidis, K., Sun, Y., Babu, P. & Palomar, D. P. Orthogonal Sparse PCA and Covariance Estimation via Procrustes Reformulation. *IEEE Transactions on Signal Processing.* **64(23)**, 6211–6226, (2016).

[6] Bunte, K., Leppaaho, E., Saarinen, I., & Kaski, S. Sparse group factor analysis for biclustering of multiple data sources. *Bioinformatics* **32(16)**, 2457-2463, (2016).

[7] Edelman, A., Arias, T. A., & Smith, S. T. The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications*, **20**, 303–353, (1998).

[8] Carvalho, C. M., Chang, J., Lucas, J. E., Nevins, J. R., Wang, Q., & West, M. High-Dimensional Sparse Factor Modeling: Applications in Gene Expression Genomics. *Journal of the American Statistical Association*, **103(484)**, 1438-1456, (2008).

[9] Engelhardt, B. E., & Stephen, M. Analysis of Population Structure: A Unifying Framework and Novel Methods Based on Sparse Factor Analysis. *Plos Genetics, Public Library of Science*, **6(9)**, 1-12, (2010).

[10] Everitt, B. S. An introduction to latent variables models. *Chapman and Hall Ltd*, (1984).

[11] Fontanella, S., Trendafilov, N., & Adachi, K. Sparse exploratory factor analysis. In Proceedings of COMPSTAT 2014: 21st International Conference on Computational Statistics, 281–288, (2014).

IJMAO

International Journal of Mathematical Sciences and Optimization: Theory and Applications
Vol. 6, No. 2, pp. 800 - 813.
doi.org/10.6084/m9.figshare.13359620.

[12] Hirose, K., & Yamamoto, M. Sparse estimation via nonconcave penalized likelihood in factor analysis model. *Statistics and Computing*, **25(5)**, 863–875, (2015).

[13] Hirose, K., & Yamamoto, M. Estimation of an oblique structure via penalized likelihood factor analysis. *Computational Statistics and Data Analysis*, **79(Kaiser)**, 120–132, (2014).

[14] Jung, S., & Takane, Y. Regularized common factor analysis. *New Trends in Psychometrics*,, **1(1)**, 1–10, (2007).

[15] Ma, Z. Sparse principal component analysis and iterative thresholding. *Ann. Statist*, **41(2)**, 772–801, (2013).

[16] Song, J., Babu, P. & Palomar, D. P. Sparse generalized eigenvalue problem via smooth optimization. *IEEE Trans. Signal Process*, **63(7)**, 1627–1642, (2015).

[17] Sun, Y., Babu, P. & Palomar, D. P. Majorization-minimization algorithms in signal processing, communications, and machine learning. *IEEE Transactions on Signal Processing*, **65(3)**, 794–816, (2017).

[18] Trendafilov, N. T., & Adachi, K. Sparse Versus Simple Structure Loadings. *Psychometrika*, **80(3)**, 776–790, (2015).

[19] Trendafilov, N. T., Fontanella, S., & Adachi, K. Sparse Exploratory Factor Analysis. *Psychometrika*, **82**, 778-794, (2017).

[20] Thompson, B. Exploratory and confirmatory factor analysis: Understanding concepts and applications. *American Psychological Association*, (2004).

[21] Yamamoto, M., Hirose, K. & Nagata, H. Graphical tool of sparse factor analysis. *Behaviormetrika,* **44**, 229–250, (2017).

[22] Yuan, X.-T. & Zhang, T. Truncated power method for sparse eigen- value problems. *J. Mach. Learn. Res.*, **14 (1)**, 899–925, (2013).

[23] Zhou, Y.,Wang, P., Wang, X., Zhu, J., & Song, P. X-K. Sparse multivariate factor analysis regression models and its application to integrative genomics analysis. *International Genetic Epidemiology Society*, , 70-80, (2016).