Research Article **Open Access**
Special Issue: Proceedings of the 24th International Workshop on Matrices and Statistics

Xiangzhao Cui, Chun Li, Jine Zhao, Li Zeng, Defei Zhang, and Jianxin Pan*

Regularization for high-dimensional covariance matrix

DOI 10.1515/spma-2016-0018
Received October 9, 2015; accepted March 28, 2016

Abstract: In many applications, high-dimensional problem may occur often for various reasons, for example, when the number of variables under consideration is much bigger than the sample size, i.e., $p >> n$. For high-dimensional data, the underlying structures of certain covariance matrix estimates are usually blurred due to substantial random noises, which is an obstacle to draw statistical inferences. In this paper, we propose a method to identify the underlying covariance structure by regularizing a given/estimated covariance matrix so that the noises can be filtered. By choosing an optimal structure from a class of candidate structures for the covariance matrix, the regularization is made in terms of minimizing Frobenius-norm discrepancy. The candidate class considered here includes the structures of order-1 moving average, compound symmetry, order-1 autoregressive and order-1 autoregressive moving average. Very intensive simulation studies are conducted to assess the performance of the proposed regularization method for very high-dimensional covariance problem. The simulation studies also show that the sample covariance matrix, although performs very badly in covariance estimation for high-dimensional data, can be used to correctly identify the underlying structure of the covariance matrix. The approach is also applied to real data analysis, which shows that the proposed regularization method works well in practice.

Keywords: High-dimensional; Covariance estimation; Covariance structure; Regularization

MSC: 62H20; 62H99

1 Introduction

In many practical fields such as genomics, biomedical imaging, functional magnetic resonance imaging, tomography and finance, etc, high-dimensional data are collected routinely. Assume X is an $n \times m$ data matrix, where n is the sample size and m is the number of variables/dimensions. For high-dimensional data, we have $m >> n$, i.e., m is much bigger than n. In this case, $X^{'}X$ becomes singular and is not well-conditioned. The aim of high-dimensional data analysis is to make statistical inference by taking into account of the dependence/correlation among the variables. In general, increasing the sample size n has an effect of improving the precision and certainty of statistical inference, whereas increasing the dimension m has the opposite effect. Statistical analysis for high-dimensional data is extremely challenging and often has to call for new statistical methodologies and theories (Donoho [7]). Regularization of covariance matrix for high-dimensional data has attracted much more attentions in recent years. The existing works include the hard-thresholding

Xiangzhao Cui: School of Mathematics, University of Honghe, Yunnan, China, E-mail: cxzh1972@126.com
Chun Li: School of Mathematics, University of Honghe, Yunnan, China, E-mail: 496806817@qq.com
Jine Zhao: School of Mathematics, University of Honghe, Yunnan, China, E-mail: zhaojine0829@163.com
Li Zeng: School of Mathematics, University of Honghe, Yunnan, China, E-mail: zengli23@126.com
Defei Zhang: School of Mathematics, University of Honghe, Yunnan, China, E-mail: zhdefei@163.com
***Corresponding Author: Jianxin Pan:** School of Mathematics, University of Manchester, UK, E-mail: jianxin.pan@manchester.ac.uk

method (Bickel and Levina [2]; El Karoui [8]), soft-thresholding method with generalization (Rothman et al. [12]), and adaptive thresholding method (Cai and Liu [4]) among others. However, such threshold estimates of the covariance matrix may yield negative eigenvalues. To overcome this difficulty, Rothman [11] proposed to use a Lasso-type penalty to encourage sparsity and introduce a logarithmic barrier function to enforce positive-definiteness of the covariance estimates. Xue et al. [13] proposed an alternating direction method to ensure the positive-definiteness of Lasso penalty based covariance estimates. However, such regularization approaches require a tuning parameter, which though can be estimated by cross-validation method or AIC- or BIC-based method remains a very challenging issue in high-dimensional data analysis due to its very intensive computation.

Based on the entropy loss function, very recently Lin et al. [10] proposed a technique which avoids the tuning parameter and regularizes estimates of the covariance matrix by considering a class of certain candidate covariance structures, such as tri-diagonal Toeplitz structure, compound symmetry, order-1 autoregressive models, banded Toeplitz structures, etc. Unfortunately, their entropy loss function based method can not be applied to high-dimensional covariance regularization, because it involves the inverse of the sample covariance matrix, which becomes singular for high-dimensional data.

To conquer the difficulty, Cui et al. [5] proposed to use Frobenius-norm to replace the entropy loss function, so that the inverse of the covariance estimate like the sample covariance matrix is not required. However, Cui et al. [5] only considered the normal case with $n > m$. For the case of high-dimensional data, i.e., $n << m$, it is not clear whether or not the Frobenius-norm based loss function works and what the performance of the proposed method is. It is the aim of this paper to resolve such open issues.

The rest of this paper is organized as follows. In section 2, we give the regularization problem and transform it into an optimization problem in numerical analysis and explore some of its general properties. In section 3, we show that the problem of finding the optimal structure from a class of candidate covariance structures, including order-1 moving average, compound symmetry or uniform covariance, order-1 autoregressive model, and order-1 autoregressive and moving average model, that minimizes the Frobenius-norm reduces to computing the zeros of a nonlinear function. In section 4, we carry out simulation studies, illustrating how our techniques of computing the structured covariance matrix that minimizes the discrepancy function can be used in regularizing the underlying covariance structure for high-dimensional data. In section 5, we apply the proposed approach to a real data experiment in signal processing. Some further remarks and discussions are presented in section 6.

2 Problem of interest

Suppose A is a given $m \times m$ covariance matrix, that is, it is symmetric nonnegative definite. Let $\mathcal{S}$ be the set of all $m \times m$ positive definite covariance matrices with structure s, for example, compound symmetry or uniform covariance structure. A discrepancy between the given covariance matrix A and the set $\mathcal{S}$ is defined by

$$D(A, \mathcal{S}) = \min_{B \in \mathcal{S}} L(A, B), \tag{1}$$

where $L(A, B)$ is a measure of the distance between the two $m \times m$ matrices A and B. Assume there is a given class of k candidate covariance structures $\{s_1, s_2, \ldots, s_k\}$. Let S_i be the set of all covariance matrices with structure s_i. Denote the set of $m \times m$ covariance matrices with the likely structures by $\Omega = \cup_{i=1}^{k} S_i$. The discrepancy between a given covariance matrix A and the set Ω is then defined by

$$D(A, \Omega) = \min_{B \in \Omega} L(A, B). \tag{2}$$

The idea is that, in this set Ω, the structure with which A has the smallest discrepancy can be viewed as the most likely underlying structure of A, and the minimizer B with this particular structure is considered to be the regularized covariance matrix of A.

We propose to consider the distance between two matrices A and B, defined by the square of the Frobenius-norm, or hereafter F-norm,

$$L(A, B) = \mathrm{tr}\{(A - B)^T(A - B)\} \tag{3}$$

where tr(.) means the trace of matrix.

It is worth mentioning that the matrix A does not have to be nonsingular. For example, it can be the sample covariance matrix for high-dimensional data, where singularity occurs due to $m >> n$. But the matrix B that has certain structures is required to be nonsingular. In other words, regardless of whether or not the given covariance matrix A is nonsingular, we want to find covariance matrix B that has a certain structure and is nonsingular, such that the F-norm based loss function $L(A, B)$ achieves its minimum. Note here the matrix A does not have to be the sample covariance matrix. In fact, it can be any estimates of the population covariance matrix, such as those obtained by modified Cholesky decomposition methods [11] and thresholding principal orthogonal complements [9] among others.

Regarding the candidate structures of the covariance matrix, here we focus on the following four candidates that are commonly used in time series, and longitudinal or spatial studies. Other possible structures of covariance matrix may be studied in a similar manner, although cases have to be investigated individually.

(1) The first-order moving average structure, MA(1), has a tri-diagonal structure of covariance matrix,

$$B = \sigma^2 \begin{pmatrix} 1 & c & \cdots & 0 & 0 \\ c & 1 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & 1 & c \\ 0 & 0 & \cdots & c & 1 \end{pmatrix} \tag{4}$$

where $\sigma^2 > 0$ and $-\dfrac{1}{2\cos(\pi/(m+1))} < c < \dfrac{1}{2\cos(\pi/(m+1))}$.

(2) The covariance of compound symmetry (CS) structure assumes that the correlation coefficients of any two observations are the same. In other words, the covariance matrix has the form

$$B = \sigma^2 \begin{pmatrix} 1 & c & \cdots & c & c \\ c & 1 & \ddots & \ddots & c \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ c & \ddots & \ddots & 1 & c \\ c & c & \cdots & c & 1 \end{pmatrix} \tag{5}$$

where $\sigma^2 > 0$ and $-1/(m-1) < c < 1$.

(3) The first-order autoregressive structure, AR(1), has the property that the correlation between any pair of observations decays exponentially towards zero as the distance between two observations increases. The covariance matrix is of the form

$$B = \sigma^2 \begin{pmatrix} 1 & c & \cdots & c^{m-2} & c^{m-1} \\ c & 1 & \ddots & \ddots & c^{m-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ c^{m-2} & \ddots & \ddots & 1 & c \\ c^{m-1} & c^{m-2} & \cdots & c & 1 \end{pmatrix} \tag{6}$$

where $\sigma^2 > 0$ and $-1 < c < 1$.

(4) More generally, the first-order autoregressive moving average structure, ARMA(1,1), has one more parameter than AR(1), reflecting an additional decrease in correlation for each additional lag. The covariance

matrix has the form

$$B = \sigma^2 \begin{pmatrix} 1 & r & rc & \cdots & rc^{m-4} & rc^{m-3} & rc^{m-2} \\ r & 1 & r & \ddots & \ddots & rc^{m-4} & rc^{m-3} \\ rc & r & 1 & \ddots & \ddots & \ddots & rc^{m-4} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ rc^{m-4} & \ddots & \ddots & \ddots & 1 & r & rc \\ rc^{m-3} & rc^{m-4} & \ddots & \ddots & r & 1 & r \\ rc^{m-2} & rc^{m-3} & rc^{m-4} & \cdots & rc & r & 1 \end{pmatrix} \tag{7}$$

where $\sigma^2 > 0$, $-1 < c < 1$ and $-1 < r < 1$.

Owing to the fact $D(A, \Omega) = \min_{1 \le i \le k}\{D(A, S_i)\}$, the main task now is to calculate the discrepancy $D(A, S_i)$ for each of the candidate covariance structures listed above, where the covariance matrix A is given.

3 Solution of problems

We begin by considering the above structures (4)-(7) one by one, for which the optimization problem is reduced to computing the zeros of a nonlinear function. For more technical details, one may refer to [5].

3.1 MA(1)

The matrix in (4) can be rewritten as $B(c, \sigma) = \sigma^2(I + c\,T_1)$, where T_1 is a symmetric matrix with the first superdiagonal and subdiagonal equal to 1 and all other elements equal to 0. Note that the eigenvalues of $B(c, \sigma)$ are

$$\lambda_j = \sigma^2(1 + 2cs_j), \quad j = 1, \cdots, m,$$

where $s_j = \cos(\pi j/(m+1))$. Assuming $m \ge 2$, we have $s_1 > s_2 > \cdots \ge 0 \ge \cdots > s_m$, $s_j = -s_{m+1-j}$ and hence $B(c, \sigma)$ is positive definite if and only if

$$-\frac{1}{2s_1} < c < \frac{1}{2s_1}.$$

Given a covariance matrix A, the discrepancy function in (1.3) is now

$$f(c, \sigma) := \mathrm{tr}(A^T A) - 2\sigma^2(\mathrm{tr}(A) + \mathrm{tr}(AT_1)c) + \sigma^4(m + 2(m-1)c^2). \tag{8}$$

It can be shown that the stationary points (c, σ) of $f(c, \sigma)$ must satisfy following equations

$$\begin{cases} \sigma^2 = \dfrac{\mathrm{tr}(AT_1)}{2(m-1)c}, \\ h(c) := m\mathrm{tr}(AT_1) - 2(m-1)\mathrm{tr}(A)c = 0, \end{cases}$$

so that a unique stationary point is given by

$$\begin{cases} \sigma^2 = \dfrac{\mathrm{tr}(A)}{m} \\ c = \dfrac{m\mathrm{tr}(AT_1)}{2(m-1)\mathrm{tr}(A)}. \end{cases} \tag{9}$$

It can also be shown that the stationary point in (9) does minimize the discrepancy function (8).

3.2 Compound Symmetry

The matrix in (5) can be rewritten as $B(c,\sigma)=\sigma^2(I+c(ee^T-I))$, where $e=[1,\dots,1]^T\in R^m$. The eigenvalues of $B(c,\sigma)$ are $\sigma^2(1+(m-1)c)$ and $\sigma^2(1-c)$ of multiplicities 1 and $m-1$, respectively, so that $B(c,\sigma)$ is a positive definite matrix if and only if

$$-\frac{1}{m-1}<c<1.$$

Given a covariance matrix A, denoting $t:=\mathrm{tr}(A^T(ee^T-I))$, the discrepancy function is now given by

$$f(c,\sigma)=\mathrm{tr}(A^TA)-2\sigma^2\mathrm{tr}(A^T)-2\sigma^2ct+\sigma^4(m+m(m-1)c^2).$$

It can be shown that the stationary points (c,σ) of $f(c,\sigma)$ must satisfy following equations

$$\begin{cases}\sigma^2=\dfrac{\mathrm{tr}(A^T)+ct}{m+m(m-1)c^2},\\ h(c):=-2\sigma^2t+2\sigma^4m(m-1)c=0.\end{cases}$$

Thus a unique stationary point is

$$\begin{cases}c=\dfrac{t}{(m-1)\mathrm{tr}(A^T)}\\ \sigma^2=\dfrac{\mathrm{tr}(A^T)+ct}{m+m(m-1)c^2},\end{cases}\tag{10}$$

where $t=\mathrm{tr}(A^T(ee^T-I))$. By verifying the facts $(\nabla^2f)_{11}=2\sigma^4m(m-1)>0$ and $\det(\nabla^2f)=16\sigma^6m^2(m-1>0$ at the stationary points (c,σ), we conclude that ∇^2f is positive definite and so the stationary point does minimize the function $f(c,\sigma)$.

3.3 AR(1)

We rewrite B in (6) as

$$B(c,\sigma)=\sigma^2\sum_{i=0}^{m-1}c^iT_i,\tag{11}$$

where $T_0=I$ and T_i is a symmetric matrix with ones on the ith superdiagonal and subdiagonal and zeros elsewhere. It can be shown that the $k\times k$ leading principal minor of $B(c,\sigma)$ is $\sigma^{2k}(1-c^2)^{k-1}$ for $k=2,\cdots,m$. Therefore, $B(c,\sigma)$ is a positive definite covariance matrix if and only if $-1<c<1$.

The discrepancy function in (3) now becomes

$$f(c,\sigma):=\mathrm{tr}(A^TA)-2\sigma^2\sum_{i=0}^{m-1}c^i\mathrm{tr}(AT_i)+\sigma^4(m+2\sum_{i=1}^{m-1}(m-i)c^{2i}).$$

It can be shown that the stationary points (c,σ) of $f(c,\sigma)$ must satisfy

$$\begin{cases}-\sum_{i=1}^{m-1}ic^{i-1}\mathrm{tr}(AT_i)+\dfrac{2\sum_{i=0}^{m-1}c^i\mathrm{tr}(AT_i)\sum_{i=1}^{m-1}(m-i)ic^{2i-1}}{m+2\sum_{i=1}^{m-1}(m-i)c^{2i}}=0,\\ \sigma^2=\dfrac{\sum_{i=0}^{m-1}c^i\mathrm{tr}(AT_i)}{m+2\sum_{i=1}^{m-1}(m-i)c^{2i}}.\end{cases}\tag{12}$$

Since $m+2\sum_{i=1}^{m-1}(m-i)c^{2i}>0$, by rearranging the first equality in (12) we have

$$h(c):=-\sum_{i=1}^{m-1}ic^{i-1}\mathrm{tr}(AT_i)(m+2\sum_{i=1}^{m-1}(m-i)c^{2i})+2\sum_{i=0}^{m-1}c^i\mathrm{tr}(AT_i)\sum_{i=1}^{m-1}(m-i)ic^{2i-1}=0.$$

Numerical experiments show that there exists at least one root of $h(c)$ in $(-1, 1)$. Equivalently, the local minima of $f(c, \sigma)$ are achieved at the points (c, σ) satisfying (12).

3.4 ARMA(1,1)

Now we consider the problem for covariance matrix with structure of ARMA(1,1) in (7). Let $q(t) = 1 + 2r\sum_{k=1}^{m-1} c^{k-1}cos(kt)$, then $B(r, c, \sigma)$ is positive-definite if and only if $q(t) \geq 0$ and $q(t)/\equiv 0$ for all $t \in R$. Now the matrix B in (7) can be rewritten as

$$B(r, c, \sigma) = \sigma^2(I + r\sum_{i=1}^{m-1} c^{i-1}T_i),$$

where T_i is a symmetric matrix with ones on the ith superdiagonal and subdiagonal and zeros elsewhere.

The discrepancy function in (3) is now

$$f(r, c, \sigma) = \mathrm{tr}(A^TA) + \sigma^4(m + 2r^2\sum_{i=1}^{m-1}(m-i)c^{2(i-1)}) - 2\sigma^2(\mathrm{tr}(A) + r\sum_{i=1}^{m-1}\mathrm{tr}(AT_i)c^{i-1}).$$

It can be shown that the stationary points (r, c, σ) must satisfy

$$\begin{cases} 2\sigma^2 r\sum_{i=1}^{m-1}(m-i)c^{2(i-1)} - \sum_{i=1}^{m-1}\mathrm{tr}(AT_i)c^{i-1} = 0, \\ 2\sigma^2 r^2\sum_{i=2}^{m-1}(i-1)(m-i)c^{2(i-1)-1} - r\sum_{i=2}^{m-1}(i-1)\mathrm{tr}(AT_i)c^{i-2} = 0, \\ \sigma^2(m + 2r^2\sum_{i=1}^{m-1}(m-i)c^{2(i-1)}) - \mathrm{tr}(A) - r\sum_{i=1}^{m-1}\mathrm{tr}(AT_i)c^{i-1} = 0. \end{cases} \tag{13}$$

By rearranging equations in (13), we have

$$\begin{cases} \sigma^2 = \dfrac{\mathrm{tr}(A)}{m}, \\ 2r\mathrm{tr}(A)\sum_{i=1}^{m-1}(m-i)c^{2(i-1)} - m\sum_{i=1}^{m-1}\mathrm{tr}(AT_i)c^{i-1} = 0, \\ 2r^2\mathrm{tr}(A)\sum_{i=2}^{m-1}(i-1)(m-i)c^{2(i-1)-1} - mr\sum_{i=2}^{m-1}(i-1)\mathrm{tr}(AT_i)c^{i-2} = 0. \end{cases} \tag{14}$$

Numerical experiments show that there exists at least one root of equations (14) which ensures $B \in R_+^{m\times m}$. Equivalently, the local minima of $f(r, c, \sigma)$ are achieved at the points (r, c, σ) satisfying (14).

A remarkable feature of matrix A is that it can be chosen as any estimates of the population covariance matrix Σ, for example, it may be the ones obtained by using SQRT-Lasso (Belloni et al. [1]), matrix-logarithm transformation (Deng and Tsui [6]) and thresholding principal orthogonal complements (Fan et al. [9]), among others .

4 Simulation studies

To examine the proposed regularization method for high-dimensional covariance matrix, in this section we carry out very intensive simulation studies. All computations are performed with MATLAB R2012b. The root-finding problem in section 3 is solved with MATLAB *fzero* or *fsolve*.

Let m be the dimension of the covariance matrices to be tested. First, we generate an $m \times n$ data matrix Q with columns randomly drawn from the multivariate normal distribution $N(\mu, \Sigma)$, where $\mu = \mathbf{0} \in R^m$ is a

mean vector and Σ is a covariance matrix that have the structure of MA(1), CS, AR(1) or ARMA(1,1) as discussed in section 1. Second, we compute the sample covariance matrix A with the generated data Q. Finally, we find for each structure a covariance matrix that minimizes the discrepancy function in (3), divided by $\mathrm{tr}(A^TA)$, in other words,

$$L^*(A, B) = \mathrm{tr}\{(A - B)^T(A - B)\}/\mathrm{tr}(A^TA). \tag{15}$$

Note that the constant $\mathrm{tr}(A^TA)$ in the dominator has no effects on the solution of the matrix B. We test with the true covariance matrix Σ, where for each structure we consider several different values for m, c, r and σ^2. We fix the sample size $n = 100$ and choose the parameters $c \in \{0.25, 0.5, 0.75\}$ and $\sigma^2 \in \{0.5, 1, 2, 4\}$ for Σ having MA(1), CS, AR(1) structures. For Σ having an ARMA(1,1) structure we use the same n, c and σ^2 as above but consider $r \in \{0.1, 0.35, 0.6\}$ for $c = 0.25$, $r \in \{0.2, 0.45, 0.75\}$ for $c = 0.5$ and $r \in \{0.25, 0.5, 0.8\}$ for $c = 0.75$. To see the performance of the proposed regularization method for high-dimensional covariance matrix, we consider the ratio $R = m/n$ for $R = 1, 2, \ldots, 10$ and $20, 30, 40, 50$, respectively. For each combination of the above parameters, we run the simulation study for 100 times and report the average value of the F-norm discrepancy.

In Tables 1-6 we report the experimental results of $L^*(.,.)$ in (15) for MA(1), CS, AR(1) and ARMA(1,1) structures with $\sigma^2 = 1$ and $R = 1, 2, 5, 10, 20, 50$. For the cases of $R = 20$ and 50, we only considered $c = 0.5$ because the computation is very intensive in such cases. To save space, the simulation results for other values of σ^2 and R are not presented here but the results/conclusions are very similar, which can be obtained from the authors if contacted. The notation in Tables 1-6 is summarized as follows:

- Σ: the true covariance matrix;
- A: the sample covariance matrix (note that it is singular for high-dimensional data with $m >> n$);
- B: the computed covariance matrix that has a certain structure and minimizes the F-norm discrepancy function $L(A, B)$ in (3).
- $L_{\Sigma,A}$, $L_{A,B}$ and $L_{\Sigma,B}$: the discrepancy function $L^*(\Sigma, A)$, $L^*(A, B)$ and $L^*(\Sigma, B)$ defined by (15), respectively.

From Tables 1-6, we have the following observations.

(1) The matrix B having the minimum $L_{\Sigma,B}$ has the same structure as the true covariance matrix Σ and $L_{\Sigma,B} < L_{\Sigma,A}$ even for high-dimensional data. In other words, the regularized estimator B is much better than the sample covariance matrix A in terms of the F-norm discrepancy function. This shows that regularization of the sample covariance matrix A, is necessary not only for the convenient use of known structure but also for the accuracy of covariance estimation. In particular, for high-dimensional data the sample covariance matrix A is a rather poor estimate of the population covariance matrix Σ due to its singularity, but its regularized matrix B not only be nonsingular but also correctly identify the underlying structure of the population covariance matrix Σ.

(2) For Σ having one of the structures of MA(1), CS or AR(1), among different minimizers B, there are two structures clearly winning out in the sense of having smaller $L_{A,B}$: the one having the same structure as Σ and the ARMA(1,1), the latter always being the best. It is not surprising for the matrix B with ARMA(1,1) structure to win out because all MA(1), CS, and AR(1) are indeed special ARMA(1,1) structures. In fact, the resulting two structures are almost identical.

It is noted that it is very important to observe the discrepancy $L_{A,B}$, because in practice the true covariance Σ is usually unknown and so is $L_{\Sigma,B}$. Thus, the discrepancy $L_{A,B}$ can be used to identify the covariance structure.

(3) The observations above are common to all choices of the structure of Σ in the class we considered, the various values of m, c, σ^2 and r. Therefore, the findings are reliable in this sense.

In summary, the proposed covariance regularization approach works well for high-dimensional data, i.e., $m >> n$. The sample covariance matrix, though it is a rather poor estimate of the population covariance matrix for high-dimensional data, is helpful in finding the correct structure of the population covariance matrix for high-dimension data.

Note that the above simulation studies are made on a basis that the simulated data are generated from normal distribution. A reviewer raised the issue on the performance of the proposed method for non-normally distributed data. In order to address this issue, we add further simulation studies on non-normal distribution,

Table 1: Simulation results with $n = 100$; $m = 100$; $\sigma^2 = 1$.

		B							
$c = 0.25$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
MA(1)	0.47654	0.47627	0.00027	0.53268	0.05681	0.47929	0.00332	0.47616	0.00038
CS	0.13280	0.85779	0.78179	0.12042	0.01238	0.19593	0.09309	0.12032	0.01248
AR(1)	0.47371	0.47709	0.00403	0.53209	0.05989	0.47344	0.00026	0.47335	0.00035
ARMA-r=0.1	0.50013	0.50059	0.00086	0.50994	0.01025	0.50015	0.00047	0.49978	0.00035
ARMA-r=0.35	0.44725	0.45422	0.00720	0.55658	0.10984	0.44781	0.00124	0.44684	0.00042
ARMA-r=0.6	0.36803	0.38434	0.01716	0.63117	0.26522	0.38663	0.01890	0.36758	0.00045
		B							
$c = 0.5$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
MA(1)	0.40674	0.40651	0.00022	0.59889	0.19370	0.43547	0.02964	0.40640	0.00033
CS	0.04565	0.94297	0.95444	0.02892	0.01673	0.03840	0.02630	0.02888	0.01677
AR(1)	0.38182	0.44261	0.06091	0.61176	0.23106	0.38115	0.00067	0.38104	0.00078
ARMA-r=0.2	0.47963	0.49195	0.01247	0.52567	0.04707	0.48427	0.00507	0.47917	0.00046
ARMA-r=0.45	0.39952	0.45100	0.05193	0.59581	0.19678	0.39926	0.00106	0.39874	0.00078
ARMA-r=0.75	0.29355	0.39882	0.10481	0.68925	0.39797	0.30350	0.01230	0.29250	0.00106
		B							
$c = 0.75$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
CS	0.03280	0.96287	1.01062	0.00799	0.02480	0.00910	0.02594	0.00798	0.02482
AR(1)	0.23139	0.53672	0.30902	0.70680	0.48263	0.22940	0.00200	0.22929	0.00211
ARMA-r=0.25	0.44565	0.51089	0.06829	0.54914	0.10667	0.66682	0.22615	0.44459	0.00108
ARMA-r=0.5	0.32875	0.52382	0.19909	0.63583	0.31095	0.34159	0.01621	0.32700	0.00175
ARMA-r=0.8	0.21171	0.53803	0.32233	0.71928	0.50330	0.21031	0.00227	0.20979	0.00194

specifically, on mixture-normal distribution. We generate an $m \times n$ data matrix Q with sample size n and dimensionality m, of which the first $[\alpha n]$ columns are randomly drawn from normal distribution $N(\mu_1, \Sigma)$ with $\mu_1 = 0.5e$ where $e = [1, \ldots, 1]^T \in R^m$, and other $[(1-\alpha)n]$ columns are randomly drawn from normal distribution $N(\mu_2, \Sigma)$ with $\mu_2 = -\mu_1$, where $0 < \alpha < 1$. In other words, the columns of data matrix Q are random samples from the mixture-normal distribution

$$\alpha N(\mu_1, \Sigma) + (1-\alpha)N(\mu_2, \Sigma),$$

so that it has mean zero and covariance matrix $(\alpha^2 + (1-\alpha)^2)\Sigma$. We then test with the true covariance matrix $(\alpha^2 + (1-\alpha)^2)\Sigma$ being of the structure MA(1), CS, AR(1) and ARMA(1,1) with different values of parameters $\sigma^2 = 1$, $c = 0.45, 0.5$ and $\alpha = 0.25, 0.5, 0.75$. In the case of ARMA(1,1), we choose $r = 0.2, 0.45, 0.75$ for $c = 0.5$, and $r = 0.25, 0.45, 0.7$ for $c = 0.45$. We choose the sample size $n = 100$ and the dimensionality $m = 100, 500, 1000$. When calculating the sample covariance matrix A, we have to avoid centralizing the data matrix Q (i.e., by subtracting μ_1 from each of the first $[\alpha n]$ columns and μ_2 from each of other columns of Q) due to the characteristics of the mixture-normal distribution.

We run the simulation study for 100 times and report the average value of the F-norm discrepancy. To save space, in Table 7 we only present the simulation results for the cases with $\alpha = 0.5$, $\sigma^2 = 1$, $c = 0.5$ and $r = 0.2, 0.45, 0.75$. Simulation results in other combinations of the parameters are very similar to those in Table 7 and can be provided if requested.

From Table 7 we observe that the conclusions are very similar to those found when simulated data are generated from multivariate normal distribution. In other words, the true covariance structure can be correctly identified even for non-normally distributed data. Therefore, the proposed regularization methods are really robust and reliable even if the high-dimensional data are not normally distributed.

The reviewer also raised the issue of positive definiteness of the estimate of the covariance matrix. In the intensive simulation studies we conducted, we find that the estimate of the covariance matrix guarantees to be positive definite unless the true parameter values are very close to the boundary of the parameters c

Table 2: Simulation results with $n = 100$; $m = 200$; $\sigma^2 = 1$.

		B							
$c = 0.25$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
MA(1)	0.64363	0.64355	0.00008	0.68250	0.03912	0.64561	0.00217	0.64351	0.00012
CS	0.13812	0.92683	0.80513	0.12345	0.01467	0.15956	0.05093	0.12339	0.01473
AR(1)	0.64251	0.64499	0.00267	0.68342	0.04136	0.64243	0.00009	0.64239	0.00013
ARMA-r=0.1	0.66543	0.66576	0.00050	0.67229	0.00692	0.66549	0.00023	0.66532	0.00011
ARMA-r=0.35	0.61702	0.62184	0.00501	0.69447	0.07813	0.61754	0.00076	0.61688	0.00014
ARMA-r=0.6	0.53518	0.54770	0.01258	0.73280	0.19774	0.54864	0.01377	0.53500	0.00018
		B							
$c = 0.5$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
MA(1)	0.57580	0.57567	0.00013	0.71532	0.13977	0.59644	0.02119	0.57563	0.00018
CS	0.04781	0.97131	0.95543	0.02901	0.01880	0.03376	0.02359	0.02899	0.01882
AR(1)	0.55059	0.59497	0.04478	0.72344	0.17455	0.55032	0.00027	0.55027	0.00032
ARMA-r=0.2	0.64772	0.65598	0.00850	0.68014	0.03300	0.65090	0.00339	0.64756	0.00017
ARMA-r=0.45	0.56943	0.60672	0.03743	0.71517	0.14595	0.56947	0.00052	0.56914	0.00028
ARMA-r=0.75	0.45058	0.53193	0.08200	0.76960	0.31968	0.45916	0.00917	0.45022	0.00036
		B							
$c = 0.75$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
CS	0.02585	0.98142	0.96911	0.00766	0.01820	0.00819	0.01874	0.00765	0.01820
AR(1)	0.36758	0.62041	0.25627	0.79085	0.42716	0.36643	0.00116	0.36636	0.00122
ARMA-r=0.25	0.61215	0.66024	0.04784	0.69168	0.07978	0.82597	0.21476	0.61176	0.00039
ARMA-r=0.5	0.49150	0.64133	0.15221	0.74188	0.25379	0.50194	0.01168	0.49090	0.00061
ARMA-r=0.8	0.34652	0.61775	0.27218	0.79839	0.45393	0.34583	0.00144	0.34539	0.00112

and/or r. For example, for all the simulation scenarios presented above, the estimates of covariance matrices are positive definite except the case of MA(1) with $c = 0.5$, where about 50% estimates of the covariance matrices are positive definite in the 100 simulation runs. The reason is that, for the MA(1) with the true value $c = 0.5$ it is already very closing to the boundary of the space of c that makes the covariance matrix positive definite, that is, c must satisfy $-1/2cos(\pi/(m+1)) < c < 1/2cos(\pi/(m+1))$, see (4). In this case, the upper bound is given by $1/2cos(\pi/(m+1)) = 0.4997581$ for $m = 100$; 0.4999902 for $m = 500$; and 0.4999975 for $m = 1000$. However, if the true parameter is taken as $c = 0.45$ for MA(1), our simulations show that the estimates of the covariance matrix are 100% positive definite, due to $c = 0.45$ being bit far away from the upper bound.

In summary, the proposed regularization methods guarantee the positive definiteness of the estimate of the covariance matrix as long as the true parameter values are within the space of parameters that make the covariance matrix positive definite. Of course, in practice the true parameter values are unknown. We suggest to observe the value of the estimate of parameters, such as the solutions in (9) and (10) for MA(1) and CS, and those to (12) and (14) for AR(1) and ARMA(1,1), respectively. If the estimate is very close to the boundary value, it is likely that the estimate of the covariance matrix is not positive definite. In this case, extra efforts may be needed to deal with the estimation problem of covariance matrix, for example, by replacing its zero eigenvalues with the smallest non-zero eigenvalue, etc.

5 Real data analysis

The handwritten digit data set was created by Tactile Srl, Brescia, Italy (*http://www.tattile.it*) and donated in 1994 to Semeion Research Center of Sciences of Communication, Rome, Italy (*http://www.semeion.it*), for machine learning research, see [3]. The data set consists of 1593 handwritten digits from around 80 persons,

Table 3: Simulation results with $n = 100$; $m = 500$; $\sigma^2 = 1$.

		B							
$c = 0.25$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
MA(1)	0.81816	0.81814	0.00002	0.83826	0.02017	0.81920	0.00109	0.81814	0.00002
CS	0.14860	0.96907	0.86559	0.13088	0.01773	0.14645	0.03332	0.13086	0.01775
AR(1)	0.81712	0.81845	0.00136	0.83844	0.02137	0.81710	0.00002	0.81710	0.00003
ARMA-r=0.1	0.83207	0.83228	0.00023	0.83557	0.00349	0.83214	0.00009	0.83205	0.00002
ARMA-r=0.35	0.80067	0.80324	0.00259	0.84160	0.04103	0.80098	0.00036	0.80064	0.00003
ARMA-r=0.6	0.74156	0.74847	0.00704	0.85279	0.11181	0.74916	0.00767	0.74152	0.00004
		B							
$c = 0.5$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
MA(1)	0.77154	0.77151	0.00002	0.84730	0.07569	0.78291	0.01138	0.77150	0.00003
CS	0.05549	0.98828	1.01331	0.02990	0.02559	0.03191	0.02763	0.02989	0.02561
AR(1)	0.75273	0.77725	0.02466	0.85017	0.09762	0.75267	0.00006	0.75266	0.00007
ARMA-r=0.2	0.82060	0.82489	0.00432	0.83766	0.01707	0.82224	0.00171	0.82057	0.00004
ARMA-r=0.45	0.76728	0.78758	0.02031	0.84765	0.08041	0.76739	0.00021	0.76721	0.00007
ARMA-r=0.75	0.67074	0.71985	0.04945	0.86542	0.19582	0.67596	0.00541	0.67065	0.00009
		B							
$c = 0.75$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
CS	0.02637	0.99251	1.00749	0.00773	0.01864	0.00795	0.01887	0.00773	0.01865
AR(1)	0.58923	0.75443	0.16708	0.87589	0.28953	0.58901	0.00023	0.58899	0.00024
ARMA-r=0.25	0.79778	0.82267	0.02523	0.84097	0.04372	0.91132	0.11411	0.79772	0.00007
ARMA-r=0.5	0.70324	0.79203	0.08883	0.85648	0.15397	0.70968	0.00663	0.70307	0.00016
ARMA-r=0.8	0.56574	0.74662	0.18210	0.87943	0.31556	0.56578	0.00047	0.56550	0.00024

of which each was scanned, stretched in a rectangular box 16×16 in a gray scale of 256 values. Each person wrote on a paper all the digits from 0 to 9, twice. The commitment was to write the digit the first time in the normal way (trying to write each digit accurately) and the second time in a fast way (with no accuracy). As a result, the data set forms a data matrix with 1593 rows and 256 columns.

We consider the rows are independent and aim to understand the underlying covariance or correlation structure of the $m = 256$ attributes/columns. We normalize the handwritten digit data first and then use the proposed regularization approach in Sections 2 & 3 to analyze the data. Regarding the sample size n, for illustration purpose we consider several different scenarios: the full data ($n = 1593$), the group of data with handwritten digit "8" ($n = 158$) for the entire 80 persons, the subgroup of data with handwritten digit "8" for the first 40 persons ($n = 79$), the subgroup of data with handwritten digit "8" for the first 20 persons ($n = 39$), the subgroup of data with handwritten digit "8" for the second 20 persons ($n = 40$). The regularization results using the modified F-norm in (15) are presented in Table 8, where the first column indicates the sample size of the data analyzed, the modified F-norm $L_{A,B}$ defined in (15) represents the discrepancy between the sample covariance matrix A and the resulting matrix B that has one of the four structures considered, i.e., MA(1), CS, AR(1) and ARMA(1,1), and 'Time' means the time (in seconds) of our MATLAB codes spending for finding the optimal matrix B with a specific covariance structure. Note that for the real data set the population covariance matrix Σ is unknown and so there is no information available on $L_{\Sigma,A}$ and $L_{\Sigma,B}$.

In Table 8, it is made clearly that the discrepancy $L_{A,B}$ for ARMA(1,1) structure is smaller than other possible structures, indicating that the underlying covariance structure is likely to be ARMA(1,1). This seems to hold across all the sample sizes we considered, where all cases, except for the full data case ($n = 1593$), are those for high-dimensional data, i.e., $m >> n$. It is observed that for each covariance structure considered the F-norm discrepancy $L_{A,B}$ decreases with increasing of the sample size, a natural phenomenon in practices.

Finally, it is worth mentioning that the estimates of the covariance matrices considered above are all positive definite, because the estimates of the parameters in the candidate structures are clearly within the

Table 4: Simulation results with $n = 100$; $m = 1000$; $\sigma^2 = 1$.

		B							
$c = 0.25$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
MA(1)	0.89987	0.89987	0.00000	0.91098	0.01109	0.90045	0.00059	0.89987	0.00001
CS	0.15182	0.98428	0.88685	0.13274	0.01908	0.14803	0.04233	0.13274	0.01909
AR(1)	0.89927	0.90000	0.00074	0.91104	0.01182	0.89926	0.00001	0.89926	0.00001
ARMA-r=0.1	0.90825	0.90837	0.00012	0.91017	0.00191	0.90829	0.00005	0.90824	0.00001
ARMA-r=0.35	0.88911	0.89054	0.00144	0.91197	0.02289	0.88929	0.00020	0.88910	0.00001
ARMA-r=0.6	0.85126	0.85531	0.00404	0.91562	0.06433	0.85561	0.00439	0.85125	0.00001
		B							
$c = 0.5$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
MA(1)	0.87094	0.87094	0.00000	0.91383	0.04292	0.87735	0.00644	0.87093	0.00001
CS	0.05423	0.99410	1.02834	0.03055	0.02368	0.03157	0.02472	0.03055	0.02371
AR(1)	0.85883	0.87278	0.01414	0.91473	0.05625	0.85881	0.00002	0.85881	0.00002
ARMA-r=0.2	0.90141	0.90376	0.00238	0.91080	0.00946	0.90232	0.00093	0.90140	0.00001
ARMA-r=0.45	0.86785	0.87947	0.01153	0.91395	0.04590	0.86793	0.00010	0.86783	0.00002
ARMA-r=0.75	0.80192	0.83163	0.02962	0.91995	0.11787	0.80507	0.00321	0.80189	0.00003
		B							
$c = 0.75$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
CS	0.03365	0.99624	1.04017	0.00794	0.02571	0.00805	0.02583	0.00794	0.02572
AR(1)	0.73972	0.84482	0.10585	0.92421	0.18580	0.73964	0.00008	0.73964	0.00008
ARMA-r=0.25	0.88723	0.90128	0.01406	0.91186	0.02468	0.94113	0.05392	0.88721	0.00002
ARMA-r=0.5	0.82538	0.87764	0.05224	0.91708	0.09170	0.82915	0.00384	0.82534	0.00004
ARMA-r=0.8	0.72148	0.83784	0.11720	0.92581	0.20573	0.72156	0.00022	0.72141	0.00007

Table 5: Simulation results with $n = 100$; $m = 2000$; $\sigma^2 = 1$.

		B							
$c = 0.5$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
MA(1)	0.93113	0.93113	0.00000	0.95404	0.02296	0.93454	0.00344	0.93113	0.00000
CS	0.06380	0.99634	1.01947	0.03772	0.02609	0.03835	0.02673	0.03771	0.02609
AR(1)	0.92398	0.93160	0.00757	0.95430	0.03017	0.92397	0.00001	0.92397	0.00001
ARMA-r=0.2	0.94824	0.94948	0.00125	0.95320	0.00498	0.94872	0.00049	0.94823	0.00000
ARMA-r=0.45	0.92944	0.93560	0.00619	0.95405	0.02466	0.92948	0.00005	0.92943	0.00001
ARMA-r=0.75	0.89034	0.90676	0.01644	0.95584	0.06557	0.89211	0.00178	0.89033	0.00001

Table 6: Simulation results with $n = 100$; $m = 5000$; $\sigma^2 = 1$.

		B							
$c = 0.5$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
MA(1)	0.97185	0.97185	0.00000	0.98122	0.00937	0.97323	0.00141	0.97184	0.00000
CS	0.21290	0.99612	1.18655	0.10917	0.10373	0.10995	0.10453	0.10916	0.10374
AR(1)	0.96883	0.97194	0.00312	0.98123	0.01243	0.96883	0.00001	0.96882	0.00001
ARMA-r=0.2	0.97906	0.97957	0.00051	0.98108	0.00201	0.97926	0.00020	0.97906	0.00000
ARMA-r=0.45	0.97114	0.97366	0.00254	0.98119	0.01011	0.97116	0.00003	0.97114	0.00001
ARMA-r=0.75	0.95399	0.96089	0.00690	0.98144	0.02752	0.95471	0.00075	0.95398	0.00001

space of parameters that makes the covariance matrix positive definite. In other words, there is no boundary issue in this real data analysis.

Table 7: Simulation results for mixture-normal distribution with $n = 100$; $\sigma^2 = 1$; $\alpha = 0.5$.

		B							
$m = 100$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
MA(1)	0.55260	0.40710	0.14550	0.59929	0.14758	0.43605	0.14594	0.40700	0.14560
CS	0.26556	0.94302	0.25815	0.02971	0.23585	0.03921	0.23601	0.02966	0.23590
AR(1)	0.53514	0.44507	0.15134	0.61424	0.15345	0.38392	0.15122	0.38380	0.15135
ARMA-r=0.2	0.60856	0.49448	0.12647	0.52874	0.12693	0.48696	0.12656	0.48200	0.12644
ARMA-r=0.45	0.55083	0.45468	0.14796	0.59894	0.14873	0.40314	0.14819	0.40264	0.14808
ARMA-r=0.75	0.46829	0.39942	0.17390	0.68973	0.17775	0.30601	0.17362	0.29478	0.17352
		B							
$m = 500$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
MA(1)	0.82901	0.77355	0.05545	0.84853	0.05629	0.78483	0.05557	0.77355	0.05546
CS	0.27216	0.98840	0.25168	0.02963	0.24253	0.03157	0.24256	0.02962	0.24253
AR(1)	0.81442	0.77880	0.06038	0.85118	0.06122	0.75428	0.06014	0.75427	0.06015
ARMA-r=0.2	0.86566	0.82642	0.04357	0.83905	0.04371	0.82380	0.04356	0.82210	0.04356
ARMA-r=0.45	0.82557	0.78899	0.05693	0.84870	0.05753	0.76899	0.05675	0.76882	0.05674
ARMA-r=0.75	0.75261	0.72152	0.08058	0.86643	0.08228	0.67787	0.08010	0.67255	0.08006
		B							
$m = 1000$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
MA(1)	0.90339	0.87206	0.03134	0.91460	0.03177	0.87841	0.03141	0.87205	0.03134
CS	0.25820	0.99405	0.27069	0.03106	0.22709	0.03211	0.22715	0.03106	0.22714
AR(1)	0.89420	0.87388	0.03445	0.91554	0.03490	0.85989	0.03431	0.85989	0.03431
ARMA-r=0.2	0.92623	0.90465	0.02397	0.91165	0.02405	0.90320	0.02397	0.90226	0.02397
ARMA-r=0.45	0.90117	0.88056	0.03215	0.91469	0.03252	0.86927	0.03199	0.86918	0.03199
ARMA-r=0.75	0.85171	0.83305	0.04836	0.92065	0.04933	0.80684	0.04807	0.80367	0.04804

Table 8: Results of experiments on semeion handwritten digit data.

	MA(1)		CS		AR(1)		ARMA(1,1)	
	$L_{A,B}$	Time	$L_{A,B}$	Time	$L_{A,B}$	Time	$L_{A,B}$	Time
$n = 1593$	0.72107	0.00387	0.83346	0.00478	0.70520	0.37499	0.69789	0.48807
$n = 158$	0.77983	0.00858	0.86956	0.00434	0.76169	0.29108	0.75967	0.32060
$n = 79$	0.82762	0.00872	0.89816	0.00780	0.81256	0.29363	0.81099	0.29143
$n = 39$	0.86964	0.00864	0.92359	0.00765	0.85892	0.29719	0.85747	0.29345
$n = 40$	0.86991	0.00887	0.92229	0.00752	0.85710	0.30026	0.85638	0.29952

6 Discussion

Given a matrix A and a class of candidate covariance structures, a novel method was proposed to regularize the covariance matrix A so that its underlying structure becomes transparent. In other words, random noise can be filtered in this sense. Our simulation studies demonstrate the reliability of the proposed method, which filters not only random noise in A but also reveal characteristics of the stochastic process structuring the covariance matrix.

In the simulation studies and real data analysis, the available matrix A is taken as the sample covariance matrix. It has been demonstrated that even for high-dimensional data, where the sample covariance matrix A becomes singular, the proposed F-norm based regularization approach still works very well in identification of the underlying structure of the sample covariance matrix. This is very appealing because the sample covariance matrix estimates the population covariance matrix rather poorly, but it has no problem when aiming to identify the underlying covariance structure and find its regularized covariance matrix that is positive definite.

A further note is that the matrix A does not have to be the sample covariance matrix. In theory, the matrix A can be any available estimate of the population covariance matrix, e.g., obtained by various statistical methods. As long as A is provided, our proposed method can be used to regularize the covariance matrix A even if the distribution of the data is non-normal, the dimension of matrix A is high, or the matrix A is singular, etc.

Apart from the four likely candidate structures considered in this paper, other likely covariance structures such as AR(2), AR(3), factor analytic structure, general linear structure, ARMA(p, q), and banded or Toeplitz structures, may be investigated in a similar manner. In theory, the proposed approach can be applied to any possible structures of the covariance matrix, but the relevant optimization problem and computation may become difficult, especially when the dimension of matrix A is rather high and/or sparse. We will investigate this problem in our follow-up work.

References

[1] A. Belloni, V. Chernozhukov and L. Wang. Square-root lasso: Pivotal recovery of sparse signals via conic programming. *Biometrika*, **98**, (2012), 791-806.

[2] P. Bickel and E. Levina. Covariance regularization by thresholding. *Ann Stat*, **36**, (2008), 2577-2604.

[3] M. Buscema. MetaNet: The Theory of Independent Judges, *Substance Use* & *Misuse*, **33**, (1998), 439-461.

[4] T. Cai and W. Liu. Adaptive thresholding for sparse covariance estimation. *J Am Stat Assoc*, **106**, (2011), 672-684.

[5] X. Cui, C. Li, J. Zhao, L. Zeng, D. Zhang and J. Pan. Covariance structure regularization via Frobenius-norm discrepancy. Revised for *Linear Algebra Appl.* (2015).

[6] X. Deng and K. Tsui. Penalized covariance matrix estimation using a matrix-logarithm transformation. *J Comput Stat Graph*, **22**, (2013), 494-512.

[7] D. Donoho. Aide-Memoire. High-dimensional data analysis: The curses and blessings of dimensionality. *American Mathematical Society*. Available at http://www.stat.stanford.edu/∼ donoho/Lectures/AMS2000/AMS2000.html, (2000).

[8] N. El Karoui. Operator norm consistent estimation of large dimensional sparse covariance matrices. *Ann Stat*, **36**, (2008), 2712-2756.

[9] J. Fan, Y. Liao and M. Mincheva. Large covariance estimation by thresholding principal orthogonal complements. *J Roy Stat Soc B*, **75**, (2013), 656-658.

[10] L. Lin, N. J. Higham and J. Pan. Covariance structure regularization via entropy loss function. *Computational Statistics* & *Data Analysis*, **72**, (2014), 315-327.

[11] A. Rothman. Positive definite estimators of large covariance matrices. *Biometrika*, **99**, (2012), 733-740.

[12] A. Rothman, E. Levina and J. Zhu. Generalized thresholding of large covariance matrices. *J. Am. Stat. Assoc.*, **104**, (2009), 177-186.

[13] L. Xue, S. Ma and H. Zou. Positive definite ℓ_1 penalized estimation of large covariance matrices. *J. Am. Stat. Assoc.*, **107**, (2012), 1480-1491.